
Confección de un índice de vulnerabilidad estructural de violencia intrafamiliar en Guatemala (2008-2023)

Diego Andrés Alonzo Medinilla



UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería



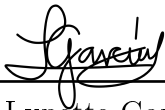
**Confección de un índice de vulnerabilidad estructural
de violencia intrafamiliar en Guatemala (2008-2023)**

Trabajo de graduación presentado por Diego Andrés Alonzo
Medinilla para optar al grado académico de Licenciado en
Ingeniería en Ciencias de la Computación y Tecnologías de la
Información

Guatemala,

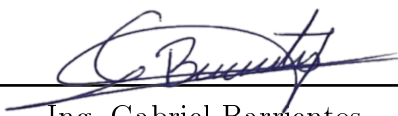
2025

Vo.Bo.:

(f) 
Ing. Lynette García

Tribunal Examinador:

(f) 
Ing. Lynette García

(f) 
Ing. Gabriel Barrientos

Fecha de aprobación: Guatemala, noviembre de 2025.

En el presente trabajo se desarrolló un índice de vulnerabilidad estructural de violencia intrafamiliar para medir la vulnerabilidad de los casos dentro del contexto guatemalteco.

De forma que se inició con un análisis exploratorio de variables sociodemográficas para poder establecer relaciones y patrones, asimismo, se prosiguió demostrando la correlación entre variables sociodemográficas y de violencia intrafamiliar.

Luego se desarrollaron reglas Apriori para identificar los patrones más frecuentes validándolos con pruebas de independencia chi-cuadrado. Esto permitiendo la selección de variables para la construcción de clases latentes.

Posteriormente, se construyeron tres tipos de modelos de clases latentes (LCA) con diferentes conjuntos de variables para poder explicar diferentes dimensiones de la vulnerabilidad estructural de violencia intrafamiliar.

A partir de ello se construyó un indicador de apoyo y presencia de infraestructura, en conjunto con mediciones de las incertidumbres e intervalos de confianza de los perfiles y del indicador para poder utilizarlos y validar que eran lo suficientemente robustos, esto utilizando bootstrapping.

Finalmente, se construyó un índice de vulnerabilidad estructural de violencia intrafamiliar a partir de los perfiles de las clases latentes y del indicador de apoyo e infraestructura, utilizando pesos iguales para cada indicador.

Estoy agradecido con todas las personas que han estado en mi travesía, aquellas que no han dejado que me rindiera, y aquellas que me retaron a no quedarme con lo inicialmente planteado. A mi familia, por su apoyo incondicional y por creer en mí aún cuando yo dudaba de mí mismo.

A mis amigos, por las conversaciones que me ayudaron a aclarar mis ideas y a mantenerme motivado. A mis profesores y mentores, por compartir su conocimiento y guiarme en este proceso, y también a las personas que ya no están porque sin ellas no sería quién soy ahora. Finalmente, agradezco a la luna que sin ella nunca hubiera podido saber de lo que era capaz.

Prefacio	III
Lista de figuras	XI
Lista de cuadros	XVIII
Resumen	XIX
Abstract	XX
1. Introducción	1
2. Antecedentes	3
3. Justificación	4
4. Objetivos	6
4.1. Objetivo general	6
4.2. Objetivos específicos	6
5. Marco teórico	7
5.1. Violencia intrafamiliar y violencia doméstica	7
5.1.1. Tipos de violencia intrafamiliar	7
5.1.2. Vulnerabilidad estructural	8
5.1.3. Grupos vulnerables	9
5.1.4. Factores que contribuyen a la vulnerabilidad estructural .	10
5.1.5. Indicadores de riesgo	12
5.2. ¿Cómo medir la vulnerabilidad?	13
5.2.1. Investigación preliminar	13
5.2.2. Ingeniería de datos	15
5.2.3. Perfilamiento	18
5.2.4. Analisis de clases latentes (LCA)	22
5.2.5. Algoritmos de agrupación	29
5.2.6. Índice	30

6. Metodología	32
6.1. Recursos utilizados	32
6.1.1. Hardware	32
6.1.2. Software	32
6.1.3. Python	33
6.2. Diseño y alcance del estudio	39
6.2.1. Diseño	39
6.2.2. Población y enfoque del estudio	40
6.2.3. Unidad de análisis	41
6.2.4. Ética y gobernanza	41
6.2.5. Constructos primarios	41
6.3. Fuentes de datos y preparación de los datos	41
6.3.1. Origen de los datos	42
6.3.2. Preparación de los datos	46
6.4. Exploración de datos	52
6.4.1. Ingeniería de características inicial	52
6.4.2. Exploración de variables	54
6.5. Establecimiento de perfiles	56
6.5.1. Pruebas chi-cuadrado	56
6.5.2. Establecimiento de asociaciones	57
6.5.3. Ingeniería de características para el desarrollo de perfiles .	57
6.5.4. Desarrollo de perfiles	57
6.6. Índice de vulnerabilidad	65
6.6.1. Indicadores del índice de vulnerabilidad	65
6.6.2. Pesaje de los indicadores	67
7. Resultados	68
7.1. Análisis exploratorio de datos	68
7.2. Apriori y pruebas chi-cuadrado de independencia entre variables .	69
7.3. Análisis de clases latentes (LCA)	70
7.3.1. Clases latentes del modelo de la víctima	70
7.3.2. Clases latentes del modelo de la situación	73
7.3.3. Clases latentes del modelo de la relación	76
7.4. Índice de vulnerabilidad estructural en violencia intrafamiliar . .	78
7.4.1. Indicador de apoyo institucional	78
7.4.2. Pesos basados en análisis de componentes principales (PCA)	79
8. Discusión	81
8.1. Hallazgos preliminares	81
8.2. Perfiles de clases latentes	82
8.2.1. Perfiles de vulnerabilidad de la víctima	82
8.2.2. Perfiles de gravedad de la situación	82
8.2.3. Perfiles de gravedad de la relación entre víctima y agresor	83
8.3. Índice de vulnerabilidad estructural(IVESVI)	83
9. Conclusiones	85
10.Recomendaciones	86
Bibliografía	87

11. Anexos	97
11.1. Accesos a información pública	98
11.2. Distribuciones y frecuencias de las variables	104
11.3. Análisis bivariados	171
11.4. Tablas de contingencia	173
11.5. Pruebas chi-cuadrada	210
11.6. Modelos entrenados	211
11.7. Código fuente para todo el trabajo	213

Lista de figuras

1.	Distribución de víctimas por grupo mayoritario	40
2.	Evolución de la violencia física según el nivel de escolaridad . . .	68
3.	Evolución de la violencia psicológica presente según el nivel de escolaridad	69
4.	Evolución de la violencia psicológica no presente según el nivel de escolaridad	69
5.	Acceso a información pública por departamento y municipio de 2017 a 2020 para el número de coordinaciones desglosadas por departamento y municipio	98
6.	Continuación figura 5	99
7.	Continuación figura 5	100
8.	Acceso a la información pública del desglose de los diagramas de trabajo	101
9.	Continuación figura 8	102
10.	Continuación figura 8	103
11.	Frecuencia de la columna AGRALFAB	104
12.	Frecuencia de la columna AGREDAD	104
13.	Frecuencia de la columna AGRESCOLARIDAD	104
14.	Frecuencia de la columna AGRESTCIV	105
15.	Frecuencia de la columna AGRGRUPET	105
16.	Frecuencia de la columna AGRNACIONAL	105
17.	Frecuencia de la columna AGRSEXO	106
18.	Frecuencia de la columna AGRTRABAJA	106
19.	Frecuencia de la columna ANOEMISION	106
20.	Frecuencia de la columna DEPTOMCPIO	107
21.	Frecuencia de la columna DIAEMISION	107
22.	Frecuencia de la columna HECANO	107
23.	Frecuencia de la columna HECAREA	108
24.	Frecuencia de la columna HECDEPTOMCPIO	108
25.	Frecuencia de la columna HECDIA	108
26.	Frecuencia de la columna HECMES	109
27.	Frecuencia de la columna HECRECURDENUN	109
28.	Frecuencia de la columna HECTIPAGRE	109
29.	Frecuencia de la columna INSTDENUNHECHO	110
30.	Frecuencia de la columna INSTDONDEDENUNCIO	110

31.	Frecuencia de la columna MESEMISION	110
32.	Frecuencia de la columna NUMHIJHOM	111
33.	Frecuencia de la columna NUMHIJMUJ	111
34.	Frecuencia de la columna OTRASVICTIMAS	111
35.	Frecuencia de la columna QUIENREPORTA	112
36.	Frecuencia de la columna TIPODISCAQ	112
37.	Frecuencia de la columna TOTALHIJOS	112
38.	Frecuencia de la columna VICALFAB	113
39.	Frecuencia de la columna VICDISC	113
40.	Frecuencia de la columna VICEDAD	113
41.	Frecuencia de la columna VICESCOLARIDAD	114
42.	Frecuencia de la columna VICESTCIV	114
43.	Frecuencia de la columna VICGRUPET	114
44.	Frecuencia de la columna VICNACIONAL	115
45.	Frecuencia de la columna VICOTRASHOM	115
46.	Frecuencia de la columna VICOTRASMUJ	115
47.	Frecuencia de la columna VICOTRASNAS	116
48.	Frecuencia de la columna VICOTRASNOS	116
49.	Frecuencia de la columna VICRELAGR	116
50.	Frecuencia de la columna VICSEXO	117
51.	Frecuencia de la columna VICTRABAJA	117
52.	Frecuencia de la columna DEPTO	117
53.	Frecuencia de la columna HECDEPTO	118
54.	Frecuencia de la columna AGROTRASNAS	118
55.	Frecuencia de la columna LEYAPLICABLE	118
56.	Frecuencia de la columna ORGANISMOREMITE	119
57.	Frecuencia de la columna ORGANISMOJURISDICCIONAL	119
58.	Frecuencia de la columna AGRESORESOTROSTOTAL	119
59.	Frecuencia de la columna CONDUCTENTE	120
60.	Frecuencia de la columna MEDIDASSEGURIDAD	120
61.	Frecuencia de la columna AGROTROSHOM	120
62.	Frecuencia de la columna TIPOMEDIDA	121
63.	Frecuencia de la columna AGROTRASMUJ	121
64.	Frecuencia de la columna AGROTROSNOS	121
65.	Frecuencia de la columna ARTICULOVIF	122
66.	Frecuencia de la columna ARTICULOCODPEN	122
67.	Frecuencia de la columna ARTICULOTRAS	122
68.	Frecuencia de la columna ARTICULOVCM	123
69.	Frecuencia de la columna ESTACIONESDEBOMBEROSPORDE- PARTAMENTO	123
70.	Frecuencia de la columna CANTIDADPNCPORDEPARTAMENTO	123
71.	Frecuencia de la columna CANTIDADORGANISMOJUDICIAL- PORDEPARTAMENTO	124
72.	Frecuencia de la columna CANTIDADIDDPORDEPARTAMEN- TO	124
73.	Frecuencia de la columna CANTIDADSALASMIXTAS	124
74.	Frecuencia de la columna CANTIDADTRIBUNALESDESENTEN- CIA	125
75.	Frecuencia de la columna CANTIDADJUZGADOSPRIMERAINS- TANCIAPENAL	125

76.	Frecuencia de la columna CANTIDADJUZGADOSPRIMERAINSTANCIA PENALNARCOACTIVIDADDELITOSAMBIENTE . .	125
77.	Frecuencia de la columna CANTIDADJUZGADOSADOLESCENTES YNINEZ	126
78.	Frecuencia de la columna CANTIDADJUZGADOSEJECUCION	126
79.	Frecuencia de la columna CANTIDADJUZGADOSPAZPENAL- COMUNITARIOS	126
80.	Frecuencia de la columna CANTIDADJUZGADOSVIOLENCIA- CONTRAMUJFEMINICIDIO	127
81.	Frecuencia de la columna CANTIDADTRIBUNALESSENTEN- CIAVIOLENCIACONTRAMUJFEMINICIDIO	127
82.	Frecuencia de la columna CANTIDADJUZGADOSTRIBUNALES- MAYORRIESGO	127
83.	Frecuencia de la columna CANTIDADFISCALIASREGIONALES	128
84.	Frecuencia de la columna CANTIDADFISCALIASDISTRITALES	128
85.	Frecuencia de la columna CANTIDADFISCALIASMUNICIPALES	128
86.	Frecuencia de la columna CANTIDADAGENCIASFISCALES . .	129
87.	Frecuencia de la columna CANTIDADCOORDINADORADEPAR- TAMENTAL	129
88.	Frecuencia de la columna CANTIDADCOORDINADORAMUNI- CIPAL	129
89.	Frecuencia de la columna CANTIDADSECCIONESESPECIALES	130
90.	Frecuencia de la columna CANTIDADCOMISARIAS	130
91.	Frecuencia de la columna CANTIDADESTACIONES	130
92.	Frecuencia de la columna CANTIDADSUBESTACIONES	131
93.	Frecuencia de la columna CANTIDADDIVISIONINVESTIGACION- CRIMINAL	131
94.	Frecuencia de la columna CANTIDADCEMENTERIO	131
95.	Frecuencia de la columna CANTIDADCENTROCOMUNITARIO	132
96.	Frecuencia de la columna CANTIDADCENTROATENCIONPER- MANENTE	132
97.	Frecuencia de la columna CANTIDADCENTROSALUD	132
98.	Frecuencia de la columna CANTIDADDIRECCION	133
99.	Frecuencia de la columna CANTIDADESCUELA	133
100.	Frecuencia de la columna CANTIDADHOSPITAL	133
101.	Frecuencia de la columna CANTIDADOTRO	134
102.	Frecuencia de la columna CANTIDADPUESTOSALUD	134
103.	Frecuencia de la columna VICOFICIO	134
104.	Frecuencia de la columna AGROFICIO	135
105.	Frecuencia de la columna VICSALARIO	135
106.	Frecuencia de la columna AGRSALARIO	135
107.	Frecuencia de la columna CANTIDADHIJOS	136
108.	Frecuencia de la columna CONHIJHOM	136
109.	Frecuencia de la columna CONHIJMUJ	136
110.	Frecuencia de la columna VICRANGOEDAD	137
111.	Frecuencia de la columna AGRRANGOEDAD	137
112.	Frecuencia de la columna VICNIVESCULARIDAD	137
113.	Frecuencia de la columna AGRNIVESCULARIDAD	138
114.	Frecuencia de la columna OTROSAGRESORES	138
115.	Frecuencia de la columna OTROSAGRESORESNAS	138
116.	Frecuencia de la columna OTROSAGRESORESNOS	139

117.	Frecuencia de la columna OTROSAGRESORESMUJ	139
118.	Frecuencia de la columna OTROSAGRESORESHOM	139
119.	Frecuencia de la columna VICOTRASVICTIMAS	140
120.	Frecuencia de la columna VICOTRASVICTIMASHOM	140
121.	Frecuencia de la columna VICOTRASVICTIMASMUJ	140
122.	Frecuencia de la columna VICOTRASVICTIMASNOS	141
123.	Frecuencia de la columna VICOTRASVICTIMASNAS	141
124.	Frecuencia de la columna VIOLENCIAFISICA	141
125.	Frecuencia de la columna VIOLENCIAPSICOLOGICA	142
126.	Frecuencia de la columna VIOLENCIASEXUAL	142
127.	Frecuencia de la columna VIOLENCIAPATRIMONIAL	142
128.	Frecuencia de la columna DIFEDADVICAGR	143
129.	Frecuencia de la columna QUIENESMAYOR	143
130.	Frecuencia de la columna DIFALFAB	143
131.	Frecuencia de la columna DIFTIEMPOREGISTROHECHO	144
132.	Frecuencia de la columna DIFSEXO	144
133.	Frecuencia de la columna DIFGRUPET	144
134.	Frecuencia de la columna VICESINDIGENA	145
135.	Frecuencia de la columna AGRESINDIGENA	145
136.	Frecuencia de la columna DIFESCOLARIDADAGRVIC	145
137.	Frecuencia de la columna DIFESCOLARIDADDETALLADA	146
138.	Frecuencia de la columna GANAMASLAVICTIMA	146
139.	Frecuencia de la columna AGRALFABENC	146
140.	Frecuencia de la columna AGRESCOLARIDADENC	147
141.	Frecuencia de la columna AGRESTCIVENC	147
142.	Frecuencia de la columna AGRGRUPETENC	147
143.	Frecuencia de la columna AGRNACIONALENC	148
144.	Frecuencia de la columna AGRSEXOENC	148
145.	Frecuencia de la columna AGRTRABAJAENC	148
146.	Frecuencia de la columna DEPTOMCPIOENC	149
147.	Frecuencia de la columna HECAREAENC	149
148.	Frecuencia de la columna HECDEPTOMCPIOENC	149
149.	Frecuencia de la columna HECMESENC	150
150.	Frecuencia de la columna HECRECURDENUNENC	150
151.	Frecuencia de la columna HECTIPAGREENC	150
152.	Frecuencia de la columna INSTDENUNHECHOENC	151
153.	Frecuencia de la columna INSTDONDEDENUNCIOENC	151
154.	Frecuencia de la columna MESEMISIONENC	151
155.	Frecuencia de la columna QUIENREPORTAENC	152
156.	Frecuencia de la columna TIPODISCAQENC	152
157.	Frecuencia de la columna VICALFABENC	152
158.	Frecuencia de la columna VICDISCENC	153
159.	Frecuencia de la columna VICESCOLARIDADENC	153
160.	Frecuencia de la columna VICESTCIVENC	153
161.	Frecuencia de la columna VICGRUPETENC	154
162.	Frecuencia de la columna VICNACIONALENC	154
163.	Frecuencia de la columna VICRELAGRENC	154
164.	Frecuencia de la columna VICSEXOENC	155
165.	Frecuencia de la columna VICTRABAJAENC	155
166.	Frecuencia de la columna DEPTOENC	155
167.	Frecuencia de la columna HECDEPTOENC	156

168. Frecuencia de la columna LEYAPLICABLEENC	156
169. Frecuencia de la columna ORGANISMOREMITEENC	156
170. Frecuencia de la columna ORGANISMOJURISDICCIONALENC	157
171. Frecuencia de la columna CONDUCTENTEENC	157
172. Frecuencia de la columna MEDIDASSEGURIDADENC	157
173. Frecuencia de la columna TIPOMEDIDAENC	158
174. Frecuencia de la columna ARTICULOVIFENC	158
175. Frecuencia de la columna ARTICULOCODPENENC	158
176. Frecuencia de la columna ARTICULOTRASENC	159
177. Frecuencia de la columna ARTICULOVCMEENC	159
178. Frecuencia de la columna VICOFICIOENC	159
179. Frecuencia de la columna AGROFICIOENC	160
180. Frecuencia de la columna CANTIDADHIJOSENC	160
181. Frecuencia de la columna CONHIJHOMENC	160
182. Frecuencia de la columna CONHIJMUJENC	161
183. Frecuencia de la columna VICRANGOEDADENC	161
184. Frecuencia de la columna AGRRANGOEDADENC	161
185. Frecuencia de la columna VICNIVESCULARIDADENC	162
186. Frecuencia de la columna AGRNIVESCULARIDADENC	162
187. Frecuencia de la columna OTROSAGRESORESENC	162
188. Frecuencia de la columna OTROSAGRESORESNASENC	163
189. Frecuencia de la columna OTROSAGRESORESNOSENC	163
190. Frecuencia de la columna OTROSAGRESORESMUJENC	163
191. Frecuencia de la columna OTROSAGRESORESHOMENC	164
192. Frecuencia de la columna VICOTRASVICTIMASENC	164
193. Frecuencia de la columna VICOTRASVICTIMASHOMENC	164
194. Frecuencia de la columna VICOTRASVICTIMASMUJENC	165
195. Frecuencia de la columna VICOTRASVICTIMASNASENC	165
196. Frecuencia de la columna VICOTRASVICTIMASNASENC	165
197. Frecuencia de la columna VIOLENCIAFISICAENC	166
198. Frecuencia de la columna VIOLENCIAPSICOLOGICAENC	166
199. Frecuencia de la columna VIOLENCIASEXUALENC	166
200. Frecuencia de la columna VIOLENCIAPATRIMONIALENC	167
201. Frecuencia de la columna QUIENESMAYORENC	167
202. Frecuencia de la columna DIFALFABENC	167
203. Frecuencia de la columna DIFTIEMPOREGISTROHECHOENC	168
204. Frecuencia de la columna DIFSEXOENC	168
205. Frecuencia de la columna DIFGRUPETENC	168
206. Frecuencia de la columna VICESINDIGENAENC	169
207. Frecuencia de la columna AGRESINDIGENAENC	169
208. Frecuencia de la columna DIFESCOLARIDADAGRVICENC	169
209. Frecuencia de la columna DIFESCOLARIDADDETALLADAENC	170
210. Frecuencia de la columna GANAMASLAVICTIMAENC	170
211. Distribución entre las edades de la víctima y el agresor	171
212. Diagrama de caja de caja y bigotes del estado civil vs salario de víctimas y agresores	171
213. Distribución porcentual de las ocupaciones según el rol (víctima o agresor)	172
214. Salario según la pertenencia indígena, según el rol (víctima o agresor)	172
215. Salario promedio según el rol (víctima o agresor)	173

Lista de cuadros

1.	Especificaciones del hardware utilizado	32
2.	Especificaciones del software utilizado	33
3.	Especificaciones de las librerías de Python utilizadas	34
4.	Especificaciones de las librerías de Python utilizadas	35
5.	Especificaciones de las librerías de Python utilizadas	36
6.	Especificaciones de las librerías de Python utilizadas	37
7.	Especificaciones de las librerías de Python utilizadas	38
8.	Especificaciones de las librerías de Python utilizadas	39
9.	Comparaciones válidas entre variables tras aplicar la prueba de independencia chi-cuadrado ($\alpha = 0,05$)	70
10.	Tabla de reglas de asociación filtradas	70
11.	Resultados del entrenamiento y evaluación de modelos para el perfil de la víctima	71
12.	Probabilidades posteriores de alfabetización de la víctima por clase latente	71
13.	Probabilidades posteriores de si es indígena o no de la víctima por clase latente	71
14.	Probabilidades posteriores del nivel de escolaridad de la víctima por clase latente	72
15.	Probabilidades posteriores de si trabaja la víctima por clase latente	72
16.	Probabilidades posteriores de la cantidad de hijos de la víctima por clase latente	72
17.	Variabilidad del modelo de clases latentes de la víctima mediante bootstrapping paramétrico	73
18.	Resultados del entrenamiento y evaluación de modelos para el perfil de la situación	73
19.	Probabilidades posteriores de alfabetización del agresor por clase latente	73
20.	Probabilidades posteriores de si es indígena o no del agresor por clase latente	73
21.	Probabilidades posteriores de la escolaridad del agresor por clase latente	74
22.	Probabilidades posteriores del sexo del agresor por clase latente	74
23.	Probabilidades posteriores de la violencia física ejercida por el agresor por clase latente	74

24.	Probabilidades posteriores de la violencia psicológica ejercida por el agresor por clase latente	75
25.	Probabilidades posteriores de la violencia sexual ejercida por el agresor por clase latente	75
26.	Probabilidades posteriores de la violencia patrimonial ejercida por el agresor por clase latente	75
27.	Variabilidad del modelo de clases latentes de la situación mediante bootstrapping paramétrico	75
28.	Resultados del entrenamiento y evaluación de modelos para el perfil de la relación entre víctima y agresor	76
29.	Probabilidades posteriores de la relación entre la víctima con el agresor por clase latente	76
30.	Probabilidades posteriores del estado civil de la víctima por clase latente	77
31.	Probabilidades posteriores del estado civil del agresor por clase latente	77
32.	Probabilidades posteriores de si la víctima trabaja por clase latente	77
33.	Probabilidades posteriores de si el agresor trabaja por clase latente	78
34.	Variabilidad del modelo de clases latentes de la relación entre el agresor y la víctima mediante bootstrapping paramétrico	78
35.	Riesgo relativo de la víctima según la clase latente	78
36.	Indicador de apoyo institucional por departamento	79
37.	Varianza explicada y razón de varianza explicada por componente principal	79
38.	Pesos de los indicadores en cada componente principal	80
39.	Tabla de contingencia entre AGR_ES_INDIGENA y AGR_ALFAB mostrando frecuencias absolutas y relativas	173
40.	Tabla de contingencia entre AGR_NIV_ESCOLARIDAD y AGR_ALFAB mostrando frecuencias absolutas y relativas	173
41.	Tabla de contingencia entre AGR_SEXO y AGR_ALFAB mostrando frecuencias absolutas y relativas	174
42.	Tabla de contingencia entre VIOLENCIA_FISICA y AGR_ALFAB mostrando frecuencias absolutas y relativas	174
43.	Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y AGR_ALFAB mostrando frecuencias absolutas y relativas	174
44.	Tabla de contingencia entre VIOLENCIA_PSICOLOGICA y AGR_ALFAB mostrando frecuencias absolutas y relativas	175
45.	Tabla de contingencia entre VIOLENCIA_SEXUAL y AGR_ALFAB mostrando frecuencias absolutas y relativas	175
46.	Tabla de contingencia entre AGR_TRABAJA y AGR_EST_CIV mostrando frecuencias absolutas y relativas	175
47.	Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y AGR_EST_CIV mostrando frecuencias absolutas y relativas	176
48.	Tabla de contingencia entre GANA_MAS_LA_VICTIMA y AGR_EST_CIV mostrando frecuencias absolutas y relativas	176
49.	Tabla de contingencia entre VIC_TRABAJA y AGR_EST_CIV mostrando frecuencias absolutas y relativas	176
50.	Tabla de contingencia entre AGR_NIV_ESCOLARIDAD y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas	177
51.	Tabla de contingencia entre AGR_SEXO y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas	177

52.	Tabla de contingencia entre VIOLENCIA_FISICA y AGR_ES_-INDIGENA mostrando frecuencias absolutas y relativas	177
53.	Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas . . .	177
54.	Tabla de contingencia entre VIOLENCIA_PSICOLOGICA y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas . . .	178
55.	Tabla de contingencia entre VIOLENCIA_SEXUAL y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas . . .	178
56.	Tabla de contingencia entre AGR_SEXO y AGR_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas	178
57.	Tabla de contingencia entre VIOLENCIA_FISICA y AGR_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas . . .	179
58.	Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y AGR_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas	179
59.	Tabla de contingencia entre VIOLENCIA_PSICOLOGICA y AGR_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas	179
60.	Tabla de contingencia entre VIOLENCIA_SEXUAL y AGR_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas	180
61.	Tabla de contingencia entre VIOLENCIA_FISICA y AGR_SEXO mostrando frecuencias absolutas y relativas	180
62.	Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y AGR_SEXO mostrando frecuencias absolutas y relativas	180
63.	Tabla de contingencia entre VIOLENCIA_PSICOLOGICA y AGR_SEXO mostrando frecuencias absolutas y relativas	180
64.	Tabla de contingencia entre VIOLENCIA_SEXUAL y AGR_SEXO mostrando frecuencias absolutas y relativas	181
65.	Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y AGR_TRABAJA mostrando frecuencias absolutas y relativas . . .	181
66.	Tabla de contingencia entre VIC_TRABAJA y AGR_TRABAJA mostrando frecuencias absolutas y relativas	181
67.	Tabla de contingencia entre AGR_TRABAJA y GANA_MAS_LA_VICTIMA mostrando frecuencias absolutas y relativas	181
68.	Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y GANA_MAS_LA_VICTIMA mostrando frecuencias absolutas y relativas	182
69.	Tabla de contingencia entre VIC_TRABAJA y GANA_MAS_LA_VICTIMA mostrando frecuencias absolutas y relativas	182
70.	Tabla de contingencia entre AGR_ES_INDIGENA y AGR_ALFAB mostrando frecuencias absolutas y relativas	182
71.	Tabla de contingencia entre AGR_NIV_ESCOLARIDAD y AGR_ALFAB mostrando frecuencias absolutas y relativas	182
72.	Tabla de contingencia entre AGR_SEXO y AGR_ALFAB mostrando frecuencias absolutas y relativas	183
73.	Tabla de contingencia entre VIOLENCIA_FISICA y AGR_ALFAB mostrando frecuencias absolutas y relativas	183
74.	Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y AGR_ALFAB mostrando frecuencias absolutas y relativas	183

75.	Tabla de contingencia entre VIOLENCIA_PSICOLOGICA y AGR_- ALFAB mostrando frecuencias absolutas y relativas	184
76.	Tabla de contingencia entre VIOLENCIA_SEXUAL y AGR_AL- FAB mostrando frecuencias absolutas y relativas	184
77.	Tabla de contingencia entre AGR_TRABAJA y AGR_EST_CIV mostrando frecuencias absolutas y relativas	184
78.	Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y AGR_EST_CIV mostrando frecuencias absolutas y relativas . .	185
79.	Tabla de contingencia entre GANA_MAS_LA_VICTIMA y AGR_- EST_CIV mostrando frecuencias absolutas y relativas	185
80.	Tabla de contingencia entre VIC_TRABAJA y AGR_EST_CIV mostrando frecuencias absolutas y relativas	185
81.	Tabla de contingencia entre AGR_NIV_ESCOLARIDAD y AGR_- ES_INDIGENA mostrando frecuencias absolutas y relativas . .	186
82.	Tabla de contingencia entre AGR_SEXO y AGR_ES_INDIGE- NA mostrando frecuencias absolutas y relativas	186
83.	Tabla de contingencia entre VIOLENCIA_FISICA y AGR_ES_- INDIGENA mostrando frecuencias absolutas y relativas	186
84.	Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y AGR_- ES_INDIGENA mostrando frecuencias absolutas y relativas . .	186
85.	Tabla de contingencia entre VIOLENCIA_PSICOLOGICA y AGR_- ES_INDIGENA mostrando frecuencias absolutas y relativas . .	187
86.	Tabla de contingencia entre VIOLENCIA_SEXUAL y AGR_- ES_INDIGENA mostrando frecuencias absolutas y relativas . .	187
87.	Tabla de contingencia entre AGR_SEXO y AGR_NIV_ESCO- LARIDAD mostrando frecuencias absolutas y relativas	187
88.	Tabla de contingencia entre VIOLENCIA_FISICA y AGR_NIV_- ESCOLARIDAD mostrando frecuencias absolutas y relativas . .	188
89.	Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y AGR_- NIV_ESCOLARIDAD mostrando frecuencias absolutas y relati- vas	188
90.	Tabla de contingencia entre VIOLENCIA_PSICOLOGICA y AGR_- NIV_ESCOLARIDAD mostrando frecuencias absolutas y relati- vas	188
91.	Tabla de contingencia entre VIOLENCIA_SEXUAL y AGR_- NIV_ESCOLARIDAD mostrando frecuencias absolutas y relati- vas	189
92.	Tabla de contingencia entre VIOLENCIA_FISICA y AGR_SEXO mostrando frecuencias absolutas y relativas	189
93.	Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y AGR_- SEXO mostrando frecuencias absolutas y relativas	189
94.	Tabla de contingencia entre VIOLENCIA_PSICOLOGICA y AGR_- SEXO mostrando frecuencias absolutas y relativas	189
95.	Tabla de contingencia entre VIOLENCIA_SEXUAL y AGR_SE- XO mostrando frecuencias absolutas y relativas	190
96.	Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y AGR_TRABAJA mostrando frecuencias absolutas y relativas .	190
97.	Tabla de contingencia entre VIC_TRABAJA y AGR_TRABAJA mostrando frecuencias absolutas y relativas	190
98.	Tabla de contingencia entre AGR_TRABAJA y GANA_MAS_- LA_VICTIMA mostrando frecuencias absolutas y relativas . . .	190

99.	Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y GANA_MAS_LA_VICTIMA mostrando frecuencias absolutas y relativas	190
100.	Tabla de contingencia entre VIC_TRABAJA y GANA_MAS_LA_VICTIMA mostrando frecuencias absolutas y relativas . . .	191
101.	Tabla de contingencia entre CANTIDAD_HIJOS y VIC_ALFAB mostrando frecuencias absolutas y relativas	191
102.	Tabla de contingencia entre VIC_ES_INDIGENA y VIC_ALFAB mostrando frecuencias absolutas y relativas	191
103.	Tabla de contingencia entre VIC_NIV_ESCOLARIDAD y VIC_ALFAB mostrando frecuencias absolutas y relativas	191
104.	Tabla de contingencia entre VIC_TRABAJA y VIC_ALFAB mostrando frecuencias absolutas y relativas	192
105.	Tabla de contingencia entre AGR_EST_CIV y VIC_EST_CIV mostrando frecuencias absolutas y relativas	192
106.	Tabla de contingencia entre AGR_TRABAJA y VIC_EST_CIV mostrando frecuencias absolutas y relativas	192
107.	Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y VIC_EST_CIV mostrando frecuencias absolutas y relativas . .	193
108.	Tabla de contingencia entre GANA_MAS_LA_VICTIMA y VIC_EST_CIV mostrando frecuencias absolutas y relativas	193
109.	Tabla de contingencia entre VIC_REL_AGR y VIC_EST_CIV mostrando frecuencias absolutas y relativas	193
110.	Tabla de contingencia entre VIC_TRABAJA y VIC_EST_CIV mostrando frecuencias absolutas y relativas	194
111.	Tabla de contingencia entre CANTIDAD_HIJOS y VIC_ES_INDIGENA mostrando frecuencias absolutas y relativas	194
112.	Tabla de contingencia entre VIC_NIV_ESCOLARIDAD y VIC_ES_INDIGENA mostrando frecuencias absolutas y relativas . .	194
113.	Tabla de contingencia entre VIC_TRABAJA y VIC_ES_INDIGENA mostrando frecuencias absolutas y relativas	194
114.	Tabla de contingencia entre CANTIDAD_HIJOS y VIC_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas . .	195
115.	Tabla de contingencia entre VIC_TRABAJA y VIC_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas	195
116.	Tabla de contingencia entre AGR_EST_CIV y VIC_REL_AGR mostrando frecuencias absolutas y relativas	195
117.	Tabla de contingencia entre AGR_TRABAJA y VIC_REL_AGR mostrando frecuencias absolutas y relativas	196
118.	Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y VIC_REL_AGR mostrando frecuencias absolutas y relativas . .	196
119.	Tabla de contingencia entre GANA_MAS_LA_VICTIMA y VIC_REL_AGR mostrando frecuencias absolutas y relativas	197
120.	Tabla de contingencia entre VIC_TRABAJA y VIC_REL_AGR mostrando frecuencias absolutas y relativas	197
121.	Tabla de contingencia entre CANTIDAD_HIJOS y VIC_TRABAJA mostrando frecuencias absolutas y relativas	198
122.	Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y VIC_TRABAJA mostrando frecuencias absolutas y relativas . .	198
123.	Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y VIOLENCIA_FISICA mostrando frecuencias absolutas y relativas .	198

124.	Tabla de contingencia entre VIOLENCIA_PSICOLOGICA y VIOLENCIA_FISICA mostrando frecuencias absolutas y relativas .	198
125.	Tabla de contingencia entre VIOLENCIA_SEXUAL y VIOLENCIA_FISICA mostrando frecuencias absolutas y relativas	198
126.	Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y VIOLENCIA_PSICOLOGICA mostrando frecuencias absolutas y relativas	199
127.	Tabla de contingencia entre VIOLENCIA_SEXUAL y VIOLENCIA_PSICOLOGICA mostrando frecuencias absolutas y relativas	199
128.	Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y VIOLENCIA_SEXUAL mostrando frecuencias absolutas y relativas	199
129.	Tabla de contingencia entre CANTIDAD_HIJOS y VIC_ALFAB mostrando frecuencias absolutas y relativas	199
130.	Tabla de contingencia entre VIC_ES_INDIGENA y VIC_ALFAB mostrando frecuencias absolutas y relativas	200
131.	Tabla de contingencia entre VIC_NIV_ESCOLARIDAD y VIC_ALFAB mostrando frecuencias absolutas y relativas	200
132.	Tabla de contingencia entre VIC_TRABAJA y VIC_ALFAB mostrando frecuencias absolutas y relativas	200
133.	Tabla de contingencia entre AGR_EST_CIV y VIC_EST_CIV mostrando frecuencias absolutas y relativas	200
134.	Tabla de contingencia entre AGR_TRABAJA y VIC_EST_CIV mostrando frecuencias absolutas y relativas	201
135.	Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y VIC_EST_CIV mostrando frecuencias absolutas y relativas . .	201
136.	Tabla de contingencia entre GANA_MAS_LA_VICTIMA y VIC_EST_CIV mostrando frecuencias absolutas y relativas	201
137.	Tabla de contingencia entre VIC_REL_AGR y VIC_EST_CIV mostrando frecuencias absolutas y relativas	202
138.	Tabla de contingencia entre VIC_TRABAJA y VIC_EST_CIV mostrando frecuencias absolutas y relativas	202
139.	Tabla de contingencia entre CANTIDAD_HIJOS y VIC_ES_INDIGENA mostrando frecuencias absolutas y relativas	202
140.	Tabla de contingencia entre VIC_NIV_ESCOLARIDAD y VIC_ES_INDIGENA mostrando frecuencias absolutas y relativas . .	202
141.	Tabla de contingencia entre VIC_TRABAJA y VIC_ES_INDIGENA mostrando frecuencias absolutas y relativas	203
142.	Tabla de contingencia entre CANTIDAD_HIJOS y VIC_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas . .	203
143.	Tabla de contingencia entre VIC_TRABAJA y VIC_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas	203
144.	Tabla de contingencia entre AGR_EST_CIV y VIC_REL_AGR mostrando frecuencias absolutas y relativas	204
145.	Tabla de contingencia entre AGR_TRABAJA y VIC_REL_AGR mostrando frecuencias absolutas y relativas	204
146.	Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y VIC_REL_AGR mostrando frecuencias absolutas y relativas . .	205
147.	Tabla de contingencia entre GANA_MAS_LA_VICTIMA y VIC_REL_AGR mostrando frecuencias absolutas y relativas	205

148.	Tabla de contingencia entre VIC_TRABAJA y VIC_REL_AGR mostrando frecuencias absolutas y relativas	206
149.	Tabla de contingencia entre CANTIDAD_HIJOS y VIC_TRABAJA mostrando frecuencias absolutas y relativas	206
150.	Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y VIC_TRABAJA mostrando frecuencias absolutas y relativas . .	206
151.	Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y VIOLENCIA_FISICA mostrando frecuencias absolutas y relativas .	207
152.	Tabla de contingencia entre VIOLENCIA_PSICOLOGICA y VIOLENCIA_FISICA mostrando frecuencias absolutas y relativas .	207
153.	Tabla de contingencia entre VIOLENCIA_SEXUAL y VIOLENCIA_FISICA mostrando frecuencias absolutas y relativas	207
154.	Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y VIOLENCIA_PSICOLOGICA mostrando frecuencias absolutas y relativas	207
155.	Tabla de contingencia entre VIOLENCIA_SEXUAL y VIOLENCIA_PSICOLOGICA mostrando frecuencias absolutas y relativas	208
156.	Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y VIOLENCIA_SEXUAL mostrando frecuencias absolutas y relativas	208
157.	Resultados de la prueba chi-cuadrado para mujeres	210
158.	Resultados del entrenamiento y evaluación de modelos para el perfil de la víctima	211
159.	Resultados del entrenamiento y evaluación de modelos para el perfil de la situación de la violencia intrafamiliar	212
160.	Resultados del entrenamiento y evaluación de modelos para la relación entre agresor y víctima	213

La violencia intrafamiliar constituye un problema social crítico en Guatemala, afectando a hombres, mujeres y niños, con un alarmante incremento de casos reportados en los últimos años. A pesar de los registros existentes, aún no se ha desarrollado una caracterización estructurada de los mismos utilizando análisis de datos.

Este proyecto tiene como objetivo construir perfiles de víctimas, agresores, y relaciones dentro del contexto de violencia intrafamiliar a partir de variables sociodemográficas y contextuales. A partir de datos públicos del Instituto Nacional de Estadística de Guatemala (2008-2023), Asociaciones de Bomberos, IGGSS.

A partir de dichos perfilamientos generar un índice que calcule la vulnerabilidad estructural de las víctimas, por lo que se plantea un enfoque metodológico que incluye la unificación y limpieza de datos, exploración estadística, establecimiento de asociaciones significativas mediante reglas de asociación y pruebas estadísticas, ingeniería de características, construcción de los diversos perfiles a partir del uso técnicas de clases latentes para utilizarlos como indicadores.

La construcción de estos perfiles permite identificar características comunes en víctimas, agresores y relaciones, mientras que el índice permite conocer el riesgo que corren las víctimas de violencia intrafamiliar, sirviendo como base para estrategias de prevención e intervención. Con ello, se busca contribuir de manera efectiva a la comprensión y mitigación de la violencia intrafamiliar en el país.

Abstract

Intrafamilial violence constitutes a critical social problem in Guatemala, affecting men, women, and children, with an alarming increase in reported cases in recent years. Despite existing records, a structured characterization of these cases using data analysis has not yet been developed.

This project aims to build profiles of victims, perpetrators, and relationships within the context of intrafamilial violence based on sociodemographic and contextual variables. It uses public data from the National Institute of Statistics of Guatemala (2008-2023), Fire Department Associations, and IGGSS.

From these profiles, the project seeks to generate an index that calculates the structural vulnerability of victims, proposing a methodological approach that includes data unification and cleaning, statistical exploration, establishment of significant associations through association rules and statistical tests, and feature engineering, construction of the various profiles based on the use of latent class techniques to use them as indicators.

The construction of these profiles allows for the identification of common characteristics in victims, perpetrators, and relationships, while the index provides insight into the risk faced by victims of domestic violence, serving as a basis for prevention and intervention strategies. In this way, the aim is to contribute effectively to the understanding and mitigation of domestic violence in the country.

La violencia intrafamiliar es un problema latente a nivel global, afectando a hombres, mujeres y niños; en Guatemala tan solo en 2021 se reportaron 36435 casos, la cifra más alta en el período de 2010 a 2021. Así mismo, se recibieron 100 denuncias por día de hechos de violencia intrafamiliar, en la mayoría de los casos siendo las víctimas mujeres [1]. Esta problemática actualmente afecta todo el ámbito de salud, con secuelas a nivel físico (trastornos gastrointestinales, cefaleas, dolores de espalda), psicológico, (insomnio, pesadillas, adicciones, paralización temporal) y social (aislamiento, pérdida del empleo, disminución del rendimiento laboral). Actualmente no es posible señalar qué tanto afecta la violencia intrafamiliar a la salud en todas sus ramas sobre las víctimas de esta[2]. Es de vital importancia evitar que sucedan casos de revictimización, definida como “La revictimización es toda acción u omisión que lesione el estado físico, mental o psíquico de la persona víctima, convirtiendo a la persona de nuevo en víctima”[3, p. 13]; sin embargo, se ha reportado diario El País que un 65 % de las víctimas reportadas en el Polo Integral de la Mujer en Córdoba corresponden a casos de reincidencia (a pesar de que sean casos de violencia de género se extrapola a violencia intrafamiliar ya que la mayoría de víctimas de este tipo de violencia son mujeres) esto denota que además de surgir nuevos casos de víctimas, en general no se resuelven los casos ya existentes [4]. Esta situación de reincidencia se ha visto reforzada también en un estudio sudafricano revelando que muchas personas bajo esta situación regresan más de una vez a la relación abusiva a pesar de haber logrado salir de ella, causando que se mantengan los casos ya existentes y se acumulen nuevos casos. Esto puede darse por no poseer suficientes recursos para poder salir de la situación de violencia en la que viven, dependiendo de la ayuda gubernamental ya que sus abusadores no pueden o no quieren proporcionar lo suficiente para hacer subsistir a familia, provocando que las víctimas pierdan su empleo y vivienda[5]. Finalmente, un estudio en Guatemala afirma que no se ha generado una estrategia para evitar la revictimización de los menores dado que se le ha dado un tratamiento meramente formal para cumplir con normativas mínimas dando condena al delincuente pero haciendo caso omiso a la víctima [6], generando a su vez victimización secundaria que son las consecuencias psicológicas, sociales, jurídicas y económicas negativas que dejan las relaciones de la víctima con el organismo competente [3]; sin embargo, si esto ocurre en el caso de abuso a menores, cuyos casos han sido confirmados, para las personas víctimas

de violencia intrafamiliar que pueden ser niños, niñas, mujeres, hombres, adultos mayores, etcétera, es altamente probable que sean mucho menos protegidas por lo que es necesario establecer estrategias y herramientas que permitan evitar la violencia secundaria y revictimización de las víctimas. En el presente trabajo de graduación se elaboran perfiles de víctimas y agresores sobre casos de violencia intrafamiliar dentro de Guatemala, esto permite poder comprender los distintos casos existentes, de forma que se le pueda brindar una ayuda adecuada a cada caso. A su vez, utilizando dichos perfiles se utilizan para confeccionar un índice de vulnerabilidad estructural, permitiendo determinar cuáles víctimas son menos capaces de sobrellevar. Asimismo, ambas herramientas permiten mejorar el tratamiento de casos dado que se le brinda una asesoría y protección más acorde a los perfiles que se tratan, a su vez, puede ser útil para brindar prioridad a aquellos casos que estén en una vulnerabilidad mayor; disminuyendo los casos de revictimización.

CAPÍTULO 2

Antecedentes

La violencia doméstica ha sido abordada desde múltiples perspectivas, incluyendo la psicológica, jurídica, sociológica y de salud pública. Sin embargo, el enfoque estructural de la vulnerabilidad —entendido como la capacidad limitada de un individuo para resistir, prevenir o recuperarse de situaciones de violencia debido a factores socioeconómicos, institucionales y territoriales— ha ganado relevancia en estudios recientes.

La vulnerabilidad estructural ha sido introducido como concepto dentro del ámbito de salud, destacando cómo las condiciones laborales, migratorias y de acceso institucional configuran el riesgo de exposición a violencia [7]. Asimismo, se ha operado este concepto en entornos clínicos, proponiendo herramientas de evaluación basadas en indicadores estructurales [8].

En el campo de la violencia doméstica, estudios han utilizado análisis de perfiles latentes (LPA, por sus siglas en inglés) para identificar patrones de victimización y agresión, demostrando que existen tipologías recurrentes que pueden ser modeladas estadísticamente [9], [10]. Estos enfoques permiten segmentar casos en función de características sociodemográficas, contextuales y conductuales, lo cual es fundamental para el desarrollo de índices compuestos.

Por otro lado, el uso de algoritmos de asociación como Apriori, han sido validados por [11] como una herramienta eficaz para detectar relaciones significativas entre variables categóricas en grandes volúmenes de datos. Esta técnica ha sido aplicada en salud pública y criminología para identificar factores co-ocurrentes en poblaciones vulnerables.

Finalmente, la construcción de índices compuestos ha sido sistematizada por organismos como la OCDE y la Comisión Europea, quienes proponen metodologías rigurosas para la selección de indicadores, normalización, ponderación y validación, de forma que se realizan revisiones críticas sobre los métodos de agregación y asignación de pesos, incluyendo enfoques participativos y basados en datos [12], [13]. Por lo cuál, estos justifican el enfoque metodológico de esta tesis, que combina minería de patrones, análisis de perfiles y construcción de un índice de vulnerabilidad estructural para casos de violencia doméstica en Guatemala.

La violencia intrafamiliar es un problema latente a nivel global, afectando a miles de personas, significando una pobre calidad de vida a nivel de salud. A pesar de los intentos por parte de diferentes organizaciones, su complejidad provoca que se deban plantear soluciones enfocadas a ser más precisas con los casos.

Por ello, surge la necesidad de identificar patrones que permitan una mejor comprensión de este fenómeno y lograr desarrollar una herramienta que permita medir la vulnerabilidad estructural de las víctimas a la violencia doméstica; una vía prometedora para lograrlo es a través de la construcción de perfiles tanto de víctimas como de agresores.

En España se construyó un perfil psicológico de agresores para la violencia de género a partir de pruebas psicológicas; primero para el agresor basándose en 4 variables principales, historial delictivo, ajuste psicosocial, historial de violencia contra la pareja y delito u agresión actual que motivan la detención del agresor [7].

Dentro de otro estudio se construyeron perfiles de violencia para casos de acoso escolar y violencia en las escuelas, esto permitió construir perfiles de las víctimas y de los agresores. Se estableció que las víctimas muchas veces tienen características como baja autoestima, estrecha relación con sus padres, impopularidad dentro de su colegio o nulas redes sociales. En el caso del perfil de los agresores, se determinó que estos son personas sin amigos reales, aunque sí con secuaces que le apoyan al no haber conseguido amigos o por temor [8].

Estos estudios permiten comprender las dinámicas de agresores y víctimas, de manera que a nivel de criminología se establecen patrones y se comprende cómo funciona, identificando grupos vulnerables y factores que aumentan el riesgo. Esto es útil ya que se puede generar una evaluación de riesgo a partir de estos factores, de forma que se puede determinar la probabilidad de que una persona se convierta en una víctima o si una víctima es más vulnerable cuando sufre estas situaciones. Además, evalúa hasta qué punto el agresor puede tratar de llegar con la víctima analizando sus acciones [9]. Esto se ha visto reflejado en herramientas como DA (Danger Assessment), cuyo propósito es determinar la probabilidad de

que ocurra un femicidio en caso de violencia de pareja[10], permitiendo que se pueda prevenir y tomar medidas para la protección de la víctima.

4.1. Objetivo general

Desarrollar un índice de vulnerabilidad a partir de perfiles de víctimas y agresores de violencia intrafamiliar construidas sobre variables sociodemográficas y contextuales.

4.2. Objetivos específicos

- Elaborar un análisis exploratorio para la identificación de variables sociodemográficas en los casos de violencia intrafamiliar a partir de un análisis exploratorio de datos sobre el período de 2008 a 2023.
- Detectar patrones y asociaciones significativas entre las variables de víctimas y agresores permitiendo la elaboración de perfiles.
- Crear un índice de vulnerabilidad para las víctimas basado en patrones en los datos y características de víctimas y agresores.

5.1. Violencia intrafamiliar y violencia doméstica

La violencia intrafamiliar o violencia doméstica es un fenómeno donde todos los miembros de una familia están implicados en una pauta de violencia que puede manifestarse de diferentes formas, pudiendo ser física, psicológica, etcétera [14].

Asimismo, según la Ley guatemalteca 97-1996, la violencia intrafamiliar consiste en la violación de derechos humanos, causando daño físico, sexual, psicológico o patrimonial de forma pública o privada de parte de un pariente, conviviente, ex conviviente, conyugue a una persona integrante de un grupo familiar [15].

También incluye violencia emocional aumentando con el tiempo independiente de la raza, orientación sexual, religión o clase social [16]. Dado que ambos conceptos son indiferentes en el contexto guatemalteco se utilizarán de forma indiscriminada (violencia intrafamiliar y violencia doméstica).

5.1.1. Tipos de violencia intrafamiliar

La violencia intrafamiliar a pesar de abarcar la violencia existente dentro de una familia, este puede manifestarse de diversas formas, por ende, cada una de estas puede implicar una diferencia en la forma en la que se aborda el problema y en su nivel de gravedad.

5.1.1.1. Violencia Física

Es el daño que se le ocasiona al cuerpo a partir de acciones ocasionando dolor, incapacidad, o muerte [17], generando daño o sufrimiento físico [18]. Asimismo, esta tiene una mayor probabilidad de aparecer en personas con una menor escolaridad [19].

5.1.1.2. Violencia psicológica o emocional

Es el hostigamiento verbal a través de insultos, críticas, humillaciones; de forma que se destruye con el gesto y la palabra [17]. Esta puede llegar a ser igual de grave que la violencia física, ya que puede llevar a la víctima a situaciones donde ponga en riesgo su integridad física, incluso el suicidio [20].

5.1.1.3. Violencia sexual

Se manifiesta en acoso o abuso sexual, de forma que nada de la interacción ha sido voluntaria [18].

5.1.1.4. Violencia patrimonial

Es toda acción que afecta la supervivencia económica de la víctima, manifestando un control del ingreso de sus percepciones económicas [17].

Cabe de mencionar que estos tipos de violencia no son excluyentes entre sí, por lo que una víctima puede sufrir de múltiples tipos de violencia al mismo tiempo (Polivictimización) y por ende genera un impacto negativo mucho más intenso que la exposición a un solo tipo de maltrato [21].

5.1.2. Vulnerabilidad estructural

Para poder definir en qué consiste la vulnerabilidad estructural es imperativo definir ciertos conceptos primero.

5.1.2.1. Peligro

El peligro es una condición que puede provocar efectos adversos, daño, ocurriendo adversidades o contratiempos [22]. Partiendo de este concepto general, el peligro que existe en la violencia intrafamiliar es el agotamiento, discriminación, abandono, anulación constante, además de violación de la integridad física y psicológica [23]. Asimismo, afecta también en el desarrollo físico y psicológico del ser humano, resultando en disfuncionalidad de la persona hasta la muerte [24]. Por lo que en el caso de que ya se esté sufriendo de la situación de peligro en cierto grado, el nuevo peligro pasaría a ser la peor consecuencia posible del peligro original, en el caso de la violencia intrafamiliar, la muerte.

5.1.2.2. Exposición

La exposición es la cantidad de contacto que se tiene con una circunstancia que provoque una condición de peligro [22]. Por lo que en contexto de violencia intrafamiliar se definiría como qué tan expuesta está una persona a un contexto que derive a una situación de violencia intrafamiliar, se debe de denotar el hecho de que estar expuesto a violencia intrafamiliar puede cooperar a repetir dicho

peligro en su propia familia, a su vez, es mayor la probabilidad de que esto ocurra en hombres que han sufrido violencia que en mujeres, otros efectos que se presenta es la delincuencia [25]. Asimismo, en caso de que ya esté sufriendo de violencia intrafamiliar qué tan expuesto está de que la situación escale a unas consecuencias peores.

5.1.2.3. Riesgo

El riesgo es la probabilidad de que ocurra una alteración o un daño cuando hay una exposición a un peligro [22]. Por lo que en el contexto de violencia intrafamiliar se define como la probabilidad de que una persona sufra de violencia intrafamiliar al estar expuesta a un contexto que derive a una situación de violencia intrafamiliar.

5.1.2.4. Capacidad de adaptación

Es la capacidad de un sistema de modificar sus características o comportamiento para adaptarse a estímulos externos, de forma que disminuya el riesgo asociado con los peligros del exterior [26].

5.1.2.5. Vulnerabilidad

La vulnerabilidad es una función que permite medir el grado en el que un sistema es susceptible o incapaz de afrontar variaciones en su entorno, de forma que permite medir su sensibilidad y su capacidad de adaptación [26]. En el caso de edificaciones la vulnerabilidad estructural se refiere al daño o afectación que sufrirá un activo determinado ante una amenaza dada [23], por lo que se asume en este caso que el daño ya está hecho y lo que se quiere conocer es qué tanto soportará la estructura.

5.1.3. Grupos vulnerables

Existen ciertos grupos que son más propensos a ser abusados o maltratados que otros [27], de forma que estos grupos son definidos como vulnerables y por ende se generan políticas especiales diseñadas a protegerlos [28]. En el caso de la violencia intrafamiliar, los grupos vulnerables son niños, mujeres, ancianos y discapacitados que debido a sus características de vulnerabilidad suelen ser los más victimizados por este fenómeno [29], a su vez, la Ley guatemalteca 97-1996 se enfoca en la protección de estos grupos de riesgo [15].

Dado que estamos en un contexto guatemalteco, se debe definir cada grupo vulnerable conforme el país. Por ende, se comprende como niño todo aquel que comprenda la edad desde su concepción hasta los trece años, y adolescente a todo aquel que comprenda desde los trece hasta los dieciocho años [30], un adulto mayor es toda persona por encima de los sesenta años [31]. Finalmente, se considera discapacidad cualquier deficiencia física, mental o sensorial, trastornos de talla y peso, genéticas, congénita o adquiridas que limite sustancialmente una o más de las actividades consideradas normales para una persona [Ley de atención a

las personas con discapacidad] y, por ende, cualquier persona que sufra de las mismas se considera alguien discapacitado.

5.1.4. Factores que contribuyen a la vulnerabilidad estructural

Es necesario determinar qué conjunto de factores contribuyen a que ciertos individuos sean más vulnerables estructuralmente a la violencia intrafamiliar, lo que incluye tanto factores demográficos como contextuales.

5.1.4.1. Fuentes de Ingreso

El factor económico limita a las víctimas de tener autonomía, por lo que terminan dependiendo económicamente de su pareja, además, el empleo puede significar una posible solución para la situación de violencia ya que le aporta independencia, seguridad, autoestima y estabilidad, sin este factor muchas mujeres terminan siendo víctimas de agresiones [32], [33].

Además, se ha determinado que el ingreso dentro del hogar es una de las variables con mayor peso en los modelos explicativos de vulnerabilidad al ser este el que determina la capacidad de resistencia y adaptación al fenómeno que azota a la familia [34].

Asimismo, se ha demostrado que la pobreza es un factor que contribuye a la victimización de las mujeres que en situaciones de mayor holgura económica, a su vez, la autonomía para trabajar se ve afectada bajo estas circunstancias ya que se ha demostrado que las mujeres que han experimentado agresiones y violencia tenían aproximadamente una quinta parte de probabilidades de trabajar a tiempo completo durante 6 meses o más en el año siguiente en comparación con aquellas que no hubieran experimentado violencia recientemente, de forma que termina siendo un ciclo [35].

5.1.4.2. Apoyo institucional

El apoyo institucional es un factor que influye en que las víctimas se mantengan o no dentro del ciclo de violencia, ya que dependen de un apoyo institucional adecuado, de forma que la red de apoyo familiar y social es un factor protector contra la violencia grave [36], [37].

Asimismo, la violencia requiere de inversión gubernamental adecuada en los servicios públicos, en caso de que sea inadecuada se promueve que grupos vulnerables como niños y adolescentes formen parte de la amplificación y reproducción de estos grupos de población [38].

Por ello, el apoyo institucional debe ser efectivo en sus intervenciones, priorizando territorios o áreas de impacto, fortaleciendo la articulación entre instituciones para garantizar una respuesta efectiva para poder dar atención a las víctimas [39].

Sin embargo, para que siquiera exista apoyo primero deben existir instituciones disponibles para brindar ayuda, por lo que esto es una condición previa para

poder implementar políticas que contribuyan a una mejor calidad de vida [40].

5.1.4.3. Discapacidad

Se ha demostrado que aquellos individuos que presentan discapacidades visuales o auditivas son más vulnerables de sufrir violencia física y psicológica que aquellos sin esas discapacidades [41], de forma que por sí misma, una discapacidad aumenta el riesgo de sufrir violencia.

A su vez, aquellas personas con discapacidades quedan a merced del agresor ya que dependen totalmente de sus cuidados para sus necesidades básicas; además, dichas víctimas se encuentran en una situación de indefensión ya que instituciones rechazan su autonomía invalidando cualquier denuncia [42].

Asimismo, en dichas circunstancias la víctima sufre de cierta violencia invisible dado que no puede responsabilizar ni identificar que el agresor está violentando sus derechos [43].

Existe un consenso a nivel mundial que los menores con discapacidad se encuentran desproporcionadamente más expuestos a violencia y maltrato [44]. Por ello, la discapacidad de una víctima es un factor que contribuye a su vulnerabilidad de sufrir o sobrellevar una situación de violencia.

5.1.4.4. Escolaridad

Se ha demostrado en estudios que existe una correlación inversamente proporcional entre la probabilidad de sufrir violencia y el nivel educativo de las mujeres [45], además, se ha demostrado que el nivel educativo abarca un rol central en el empoderamiento de las mujeres y por ende su autonomía de forma que esto se traduce en un factor de riesgo a tomar en cuenta [46], [47].

Asimismo, también se ha demostrado que en el caso de que tanto el agresor como la víctima tengan un nivel educativo bajo y que tengan un nivel educativo similar entonces la probabilidad de violencia se reduce y aumenta en cualquier otro caso (educación superior de la víctima, educación superior del agresor) [45].

Finalmente, se ha demostrado que aquellos individuos que son menos educados tienen los registros de adversidad mayores, de forma que la educación es un factor protector contra la violencia intrafamiliar [48].

5.1.4.5. Poder de decisión de la mujer y factores comunitarios

Este es un factor de riesgo dentro de la violencia intrafamiliar [47], este consiste en que las mujeres tengan capacidad de tomar decisiones dentro del hogar tales como la administración de recursos, la educación de los hijos, si pueden trabajar, si puede participar en la vida social de su comunidad, etcétera. De forma que si la mujer no tiene poder de decisión dentro del hogar es más probable que sufra de violencia intrafamiliar [47], [46].

Se ha demostrado que es mucho más frecuente la violencia interpersonal de

pareja (que está englobada dentro de la violencia intrafamiliar) en sociedades donde las situaciones de conflictos y luchas políticas son comunes [49], factor que ocurre dentro de las comunidades indígenas al haber sido discriminadas, tratadas con desigualdad y excluidas históricamente, sufriendo un impacto diferenciado de violencia [50].

Asimismo, dentro de las comunidades indígenas existe una discriminación y racismo sobre las mujeres, y el rol de género está arraigado implicando que las mujeres tengan una limitada participación en el acceso a trabajo, educación, salud o justicia [51]. Esto conlleva a que sufran barreras culturales al estigmatizarlas en un rol de madres y cuidadoras, limitando su desarrollo personal y profesional [52].

Finalmente, se ha demostrado que existe una tendencia a violentar a las mujeres indígenas por parte de una pareja o familiar cercano. Sin embargo, producto del rol que se les da dentro de la comunidad son víctimas pasivas y silenciadas, lo que incita a que ocurran casos de violencia doméstica dentro de estas comunidades [53].

5.1.4.6. Relación entre la víctima y el agresor

La relación más frecuente es la violencia entre parejas, esto es debido a que pueden convertirse en canales de descarga emocional. El agresor suele percibir a sus vínculos más cercanos como amenazas a su estabilidad psíquica lo que intensifica la violencia en relaciones íntimas. Finalmente, el vínculo afectivo, la dependencia económica o los temores de la víctima son mociones para el aislamiento de la víctima [54].

Se ha demostrado que la mayoría de casos de violencia severa ocurren en relaciones no formalizadas o en proceso de ruptura. El caso de divorciados es el más grave, mientras que los convivientes están en segundo lugar, y los casados, los menos graves [55].

5.1.5. Indicadores de riesgo

La escala de predicción del riesgo de violencia EPV-R es una escala validada y ampliamente utilizada en el ámbito forense y policial. Esta que recolecta características de la situación para poder establecer un riesgo dentro del contexto de violencia permitiendo que sea una captura de la situación en un momento específico [56].

Este mide aspectos en específico como los datos personales de los involucrados, la situación de relación de los involucrados, el tipo de relación que existe por parte de la pareja, el tipo de violencia, el perfil del agresor y la vulnerabilidad de la víctima, recopilando variables sociodemográficas de la situación de violencia [57].

Se debe de denotar que dicha escala asigna una puntuación según las preguntas contestadas de forma que a partir de un sistema de puntos se genera un nivel de riesgo basado en dicho puntaje. Dependiendo del ítem (variable) medido se le dará de peso al puntaje de forma que a aquellos ítems que implican una mayor severidad se les da un mayor peso [57].

En la misma línea se han desarrollado índices para medir la severidad en la violencia de pareja de forma que se calcularon pesos para cada tipo de violencia [57].

5.2. ¿Cómo medir la vulnerabilidad?

La vulnerabilidad al ser un concepto tan abstracto es difícil de medir, por lo que es necesario explorar sus dimensiones y factores asociados. De forma que se pueda comprender qué es lo que contribuye a la vulnerabilidad estructural en violencia intrafamiliar y cómo se puede medir.

5.2.1. Investigación preliminar

Para medir la vulnerabilidad estructural en violencia intrafamiliar es necesario recopilar información y realizar un análisis numérico, por lo que, es imperativo que se utilice una investigación cuantitativa. La investigación cuantitativa es aquella que se basa en la recolección y el análisis de datos numéricos para comprender y explicar fenómenos sociales, económicos, científicos y empresariales; usa el análisis estadístico para la transformación de datos en información significativa y confiable. A su vez, puede ser de tipo descriptivo y correlacional siendo este el tipo de investigación que se centra en describir y caracterizar fenómenos, descubriendo patrones [58].

5.2.1.1. Estadística de la investigación

Para poder llevar a cabo una investigación cuantitativa es necesario comprender ciertos conceptos estadísticos que permitan el análisis de los datos recopilados.

5.2.1.1.1. Muestra, población y confianza. Dentro de un estudio es necesario contar con una población y una muestra. La población es el conjunto completo de individuos o elementos que poseen ciertas características comunes que son de interés para el estudio [59], mientras que la muestra es un subconjunto de la población que se selecciona para participar en el estudio [59].

A su vez, al momento de estimar algún parámetro de la población (media, mediana, etcétera) es necesario establecer ciertos aspectos:

- Estimación por intervalo: también conocido como intervalo de confianza es un rango que se estima que debe de contener al parámetro poblacional que se está calculando, con un margen de error [59].
- Nivel de confianza: es la probabilidad de que el intervalo de confianza contenga el valor verdadero del parámetro poblacional [59]. Comúnmente se utiliza un nivel de confianza del 95 % o 99 % [59].
- Margen de error: es un valor que se le suma y se le resta al parámetro estimado para obtener un rango de estimación del parámetro poblacional. Este valor depende del nivel de confianza y del tamaño de la muestra [59].

5.2.1.1.2. Variables. Una variable es una característica o atributo que puede tomar diferentes valores, por otra parte una variable aleatoria es una descripción numérica del resultado de un experimento o de algo que puede ocurrir [59]. Estas pueden ser discretas o continuas, las discretas que son aquellas que tienen un número finito de valores o una sucesión infinita de valores enteros, mientras que las continuas son aquellas que pueden tomar cualquier valor numérico dentro de un intervalo o una colección de intervalos [59].

5.2.1.1.3. Distribuciones de probabilidad. Las distribuciones de probabilidad indican cómo es que los valores de una variable aleatoria tienen cierta probabilidad de ocurrir. En el caso de las variables aleatorias discretas su distribución está definida por una función de probabilidad dándole específicamente una probabilidad a cada valor dentro de la variable aleatoria. Las condiciones requeridas para una función de probabilidad discreta es que la sumatoria de las probabilidades de 1 y que las probabilidades sean mayores o iguales que 0 [59].

En el caso de variables categóricas, se utiliza la distribución multinomial que es una extensión de la distribución binomial que modela dos casos, uno de éxito y otro de fracaso, solo que en este caso con K categorías [59], [60], cuya función de probabilidad es:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{x_1!x_2!\dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

donde n es el número total de ensayos, x_i es el número de veces que ocurre la categoría i , y p_i es la probabilidad de que ocurra la categoría i .

5.2.1.1.4. Métodos de bootstrap. Para poder estimar parámetros poblacionales (un parámetro poblacional puede ser una media, una varianza, un *likelihood*, etcétera) es necesario calcular la incertidumbre de los mismos. Es por ello que existe el arranque (*bootstrapping*), de forma que a partir de una distribución F desconocida se simulan múltiples distribuciones F nuevas. Con las mismas se generan las estimaciones poblacionales obteniendo el sesgo y su varianza, por lo que se utiliza las siguientes fórmulas [61], [62]:

$$b(F) = \frac{1}{R} * \sum_{r=1}^R \hat{F}_r^* - \hat{F}$$

$$v(F) = \frac{1}{R-1} * \sum_{r=1}^R (\hat{F}_r^* - \bar{F}^*)^2$$

donde R es el número de muestras *bootstrap*, $\hat{F}_r^*$ es el estimador del parámetro de interés en la muestra *bootstrap* r , $\hat{F}$ es el estimador del parámetro de interés en la muestra original, y $\bar{F}^*$ es el promedio de los estimadores del parámetro de interés en las muestras *bootstrap*. Además se considera que R debe de ser lo suficientemente grande para que las estimaciones sean confiables y se considera grande cuando es igual o mayor a 1000 [62] [61].

5.2.2. Ingeniería de datos

Es necesario trabajar los datos, transformarlos y filtrarlos, de forma que se profundice y se comprendan para poder interpretarlos en virtud de medir la vulnerabilidad estructural en violencia intrafamiliar.

5.2.2.1. Fundamentos computacionales para la ingeniería de datos

Para el procesamiento de datos masivos y poder sintetizar conclusiones sobre los datos, es necesario disponer de una computadora capaz que procese y extraiga los datos que se requieren, además de un entorno apto para el procesamiento de los mismos.

5.2.2.1.1. Unidades de procesamiento.

5.2.2.1.1.1. Unidades de procesamiento central (CPU). La implementación de una computadora está compuesta por una organización (microarquitectura). Son los aspectos a alto nivel del diseño de una computadora como su sistema de memoria, la conexión entre sus distintos componentes y el diseño de su CPU. En el CPU es donde la aritmética, lógica, y los datos se implementan [63].

Este se define como el componente principal de una computadora, encargado de realizar diferentes operaciones computacionales, ejecutando instrucciones para ejecutar programas. Está compuesto por una unidad de control que se encarga de guiar el sistema computacional a través de pulsos eléctricos, notificando para ejecutar instrucciones a alto nivel. La unidad aritmética-lógica se encarga de manejar toda la aritmética y las operaciones para generar las acciones dentro de una computadora, y la unidad de memoria maneja el uso de memoria [64].

5.2.2.1.1.2. Unidades de procesamiento gráfico (GPU). Es un circuito diseñado para acelerar las gráficas y el video de una computadora. De forma que a partir de calculo de múltiples operaciones matemáticas reduce el tiempo para una computadora de correr múltiples programas. Posee de su propia RAM y esta se especializa en realizar procesamiento paralelo de forma que dentro de una GPU se realizan múltiples cálculos de forma simultánea haciendolos más veloces y eficientes que los CPU's. Sin embargo, las GPU's no reemplazan a las CPU's debido a que están especializadas en el procesamiento gráfico [65]. Existe un problema al momento de manejar GPU y es el hecho de que se debe manejar la calendarización del cómputo en el sistema, el procesador y el de la GPU transfiriendo los datos entre la memoria del sistema y la memoria del GPU [63]. Para atacar el problema previamente mencionado es que NVIDIA creo Arquitectura de Dispositivos Unificados de Cómputo (CUDA) [63] que consiste en un software que está compuesto por un driver de Hardware y una API que permite extender el lenguaje C para especificar el nivel de paralelismo además de especificar las operaciones a realizar específicas del dispositivo GPU [66].

5.2.2.1.2. Lenguaje y entorno de desarrollo. Para el desarrollo de todo el proceso de ingeniería de datos es necesario contar con un lenguaje de programación y un entorno de desarrollo adecuado para el procesamiento de los datos.

5.2.2.1.2.1. Python. Es un lenguaje de programación interpretado, orientado a objetos, y de alto nivel; es ideal para el desarrollo rápido de aplicaciones. Permite el uso de múltiples librerías dependiendo de la funcionalidad que se requiera además de generar ambientes específicos para proyectos [67]. Venv, precisamente es un ambiente virtual que posee su propio conjunto de paquetes y módulos instalados de forma que sean específicas del proyecto [68].

5.2.2.1.2.2. Control de versiones. Git es una herramienta que permite almacenar y gestionar cambios en archivos y proyectos de software a lo largo del tiempo. Funciona habilitando capturas de versiones de los archivos, permitiendo que se rastreen los cambios y se puedan revertir en caso se requiera. Esto implica que se puedan comparar versiones en caso de ser necesario [69].

5.2.2.2. Tratamiento de datos

El tratamiento de datos permite obtener resultados relevantes asegurando la confiabilidad de los datos. Debe analizarse si los datos están completos, si son válidos y si no tienen duplicados. Estos problemas se tratan aplicando estrategias que permiten corregir errores humanos y manejar los datos a lo largo del tiempo [70]. En el caso de escalas de tiempo es necesario estandarizar datos ya que permite que se sitúen los datos dentro de una escala de valores equivalentes, permitiendo su comparación, y que sean internamente coherentes[71].

A su vez, los datos deben limpiarse de cualquier contaminación posible como: datos con incoherencias, errores humanos, errores de formato, errores de caracteres. Por otra parte, los datos faltantes son otro tipo de contaminación de datos y es producto de la resistencia humana ya que los campos deberían capturar información valiosa[72]. En la mayoría de los casos es mejor una estimación que dejar los datos en blanco[73]. Para poder resolver este tipo de contaminación se utilizan técnicas de imputación de datos, tal como la aplicación de una estimación razonable. También es válido la eliminación del registro si la mayor parte de las variables están vacías siendo la estrategia más segura y que no requiere de ningún tipo de análisis o método estadístico para llevarla a cabo[73].

Sin embargo, para desarrollar una estimación razonable es necesario mantener la relación entre variables, evitando que se distorsione la relación existente entre las mismas mientras se imputen los datos [72]; siempre y cuando se mantenga la misma se puede inferir un valor para rellenar los datos faltantes.

Además, es necesario transformar los datos a formatos manipulables para hacer análisis, tal es el uso de la librería pandas en python que permite manipular datos en estructuras de datos llamadas DataFrames. Estas permiten manipular datos tabulares de forma eficiente [74]. Asimismo, la librería camelot permite extraer tablas de documentos PDF, de forma que se pueden transformar a estructuras tabulares manipulables [75].

También, es necesaria su limpieza a través de homogenización de los datos mediante métodos como la eliminación de espacios en blanco, conversión de mayúsculas y minúsculas, eliminación de caracteres especiales, corrección de errores tipográficos, entre otros [76]. Lo que es alcanzable mediante métodos como: `.lower()`, `.strip()`, `.replace()` en python, además de utilizar el módulo de `diffib` para verificar la homogeneidad de los datos, utilizar el módulo de `re` para manejar diferentes regex y patrones de texto [77], y en caso de no poder terminar de homogeneizar todos los datos utilizar algoritmos de semejanza de texto para poder relacionar textos similares y homogeneizarlos tal como lo permite el algoritmo de Levchenstein al contar la cantidad de operaciones necesarias para convertir un texto a otro [78].

Asimismo, para el análisis de datos es necesario realizar una codificación de los datos, de forma que se puedan transformar variables categóricas en variables numéricas para su análisis, utilizando técnicas como la codificación one-hot que crea variables binarias para cada categoría o la codificación de etiquetas que asigna un número entero a cada categoría [79]. Otra técnica de codificación es la división por rangos de variables discretas de forma que debido a sus características se puede dividir en rangos simplificando la información [80]

5.2.2.3. Reducción de datos

A partir de la reducción de los datos se puede filtrar para determinar cuáles variables son realmente necesarias para el objetivo. Para lograr la reducción se pueden utilizar técnicas como la correlación para determinar la relación entre variables. En caso de que exista demasiada correlación entre variables puede afectar la interpretación de resultados. Se pueden utilizar algoritmos como Análisis de Componentes Principales para disminuir la cantidad de variables e identificar qué tanto se relacionan las variables con una variable objetivo, reduciendo la cantidad de variables útiles [73].

5.2.2.4. Construcción de variables

Utilizando la construcción de nuevas variables se puede desarrollar discretización de datos para poder categorizar variables continuas, generando rangos que puedan brindar una mayor información o agrupar información para generar variables predictoras. A su vez, se pueden combinar variables existentes aplicando transformaciones, algoritmos o funciones para generar nuevas variables que puedan contribuir al análisis o a la predicción de ciertos factores [73].

5.2.2.5. Análisis exploratorio de datos (EDA)

Sirve para investigar conjuntos de datos, resumiendo sus características principales a partir de métodos de visualización de datos. Además, permite identificar patrones, anomalías, identificar errores, poner a prueba hipótesis o verificar suposiciones de forma que se revele más información para poder comprender de una forma más profunda las variables y sus relaciones entre sí. Asimismo, este análisis se puede realizar utilizando una o múltiples variables, de forma gráfica o no [81].

5.2.3. Perfilamiento

Para el desarrollo de índices compuestos, el análisis de las variables que componen cada perfil es clave, de forma que se pueda comprender la importancia de una variable dentro del perfil y su contribución al mismo.

5.2.3.1. Apriori

Es un algoritmo de aprendizaje de máquina no supervisado que genera reglas de asociación, de forma que identifica patrones frecuentes, conexiones y dependencias entre grupos de elementos, estableciendo relaciones entre variables categóricas.

Funciona de forma que identifica los elementos únicos en conjunto con sus frecuencias, luego combina los elementos que aparecen en conjunto y filtra aquellos que tengan una probabilidad mayor a un umbral especificado. Esto elimina aquellas asociaciones con muy poca frecuencia. Las combinaciones frecuentes se recombinan para tener conjuntos más grandes [82]. El algoritmo emplea diversas métricas de soporte, confianza y elevación para definir sus criterios operativos y mejorar la eficiencia.

Un patrón determinante frecuente es un conjunto de elementos (elementos antecedentes y sus consecuentes) que aparece con frecuencia en un conjunto de datos. Es frecuente cuándo tanto los elementos antecedentes como los consecuentes poseen un alto soporte y confianza en el conjunto de datos, para ello se utiliza una métrica llamada convicción, permitiendo que se generen reglas de asociación fuertes [83].

- Soporte: es la proporción entre el número transacciones que contienen a los elementos antecedente y a los elementos consecuentes sobre el número total de transacciones. Define la probabilidad de que ocurran transacciones que contengan ambos elementos. Es decir que es la probabilidad que se presenten ambos elementos (antecedente y consecuente) en una misma transacción [84].

$$\text{Soporte}(x, y) = P(x \wedge y) = \frac{\text{conteo}(x \cup y)}{|D|}$$

- x : es el elemento antecedente
 - y : es el elemento consecuente
 - D : es el conjunto de datos
 - $\text{conteo}(x \cup y)$: es el número de transacciones que contienen a ambos elementos
- Confianza: esta es la proporción entre el número de transacciones que contienen a los elementos antecedentes y consecuentes sobre el número de transacciones que contienen al elemento consecuente. Es decir que es la probabilidad de que dado que una transacción contiene al elemento consecuente se presente también el elemento antecedente [84].

$$\text{Confianza}(x, y) = P(x \mid y) = \frac{P(x \wedge y)}{P(y)} = \frac{\text{soporte}(x \cup y)}{\text{soporte}(y)}$$

- x : es el elemento antecedente
 - y : es el elemento consecuente
 - $\text{soporte}(x \cup y)$: es el soporte de ambos elementos
 - $\text{soporte}(y)$: es el soporte del elemento consecuente
- Elevación: es la razón entre el número de transacciones que contienen a los elementos antecedentes dado que contiene el elemento consecuentes y el número de transacciones que contienen al elemento antecedente. Es decir que es la probabilidad de que dado que una transacción contiene al elemento consecuente ocurre el elemento antecedente dividido la probabilidad que ocurra el elemento antecedente [84].

$$\text{Elevación}(x, y) = \frac{P(x | y)}{P(x)} = \frac{\text{confianza}(x, y)}{P(x)}$$

- x : es el elemento antecedente
 - y : es el elemento consecuente
 - $\text{confianza}(x, y)$: es la confianza entre ambos elementos
 - $P(x)$: es la probabilidad del elemento antecedente
- Convicción: es la probabilidad de que una transacción contiene al elemento antecedente por la probabilidad de que una transacción no contenga al elemento consecuente sobre la probabilidad de que una transacción tenga el elemento antecedente pero no el consecuente.

$$\text{Conviction}(x, y) = \frac{P(x) \cdot P(\neg y)}{P(x \wedge \neg y)}$$

- x : es el elemento antecedente
- y : es el elemento consecuente
- $P(x)$: es la probabilidad del elemento antecedente
- $P(\neg y)$: es la probabilidad de que no ocurra el elemento consecuente
- $P(x \wedge \neg y)$: es la probabilidad de que ocurra el elemento antecedente pero no el consecuente

Este algoritmo está implementado en una librería llamada Efficient-Apriori, de forma que permite descubrir estructuras ocultas en los datos categóricos. La función principal *apriori* identifica los conjuntos de ítems frecuentes que cumplan con el umbral mínimo de soporte y luego genera las reglas de asociación que satisfacen un nivel mínimo de confianza. El código está optimizado para la búsqueda mediante poda anticipada mejorando el desempeño. Finalmente, la librería permite ajustar parámetros de soporte mínimo, confianza mínima, y longitud máxima de conjunto de ítems [85].

5.2.3.2. Prueba de chi-cuadrada para la independencia

La prueba de chi-cuadrada de independencia (conocida también como estadístico de prueba para independencia) es una herramienta estadística no paramétrica que permite evaluar si existe una relación significativa entre dos variables categóricas. Su fundamento radica en comparar las frecuencias observadas en una tabla

de contingencia con las frecuencias esperadas bajo el supuesto de independencia entre las variables.

La prueba plantea una hipótesis nula H_0 que supone que no existe asociación entre las variables, mientras que la hipótesis alternativa H_1 sostiene que sí existe una relación significativa; en caso de que el valor p obtenido es menor al nivel de significación preestablecido (usualmente $\alpha = 0,05$), se rechaza la hipótesis nula y se concluye que hay evidencia estadística que demuestra una asociación entre las variables [59].

La distribución de probabilidad utilizada para estas pruebas es la distribución chi-cuadrada que es continua, la misma, surge de la suma de los cuadrados de variables aleatorias independientes que siguen una distribución normal estándar. Esta distribución es asimétrica y depende de un parámetro conocido como grados de libertad (df, por sus siglas en inglés), que está relacionado con el número de categorías o grupos en las variables analizadas. A medida que los grados de libertad aumentan, la distribución chi-cuadrada se aproxima a una distribución normal [86], [87], definiéndose como:

$$\chi^2 = \sum_i^n X_i^2$$

Cuya función de densidad de probabilidad es:

$$f(x|n) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n/2)-1} e^{-x/2} \text{ para } x > 0, n > 0$$

- n: es el número de grados de libertad
- Γ : es la función gamma
- e: es la base del logaritmo natural
- x: es la variable aleatoria

5.2.3.2.0.1. Tablas de contingencia Las tablas de contingencia son específicamente tablas para presentar las frecuencias observadas y esperadas en una prueba de independencia, por lo que realizan un conteo de frecuencias observadas según las variables y además calculan la frecuencia esperada siguiendo la fórmula: La frecuencia esperada para la celda (i, j) se calcula como:

$$E_{ij} = \frac{(\text{Total del renglón } i) \times (\text{Total de la columna } j)}{\text{Tamaño de la muestra}} = \frac{(T_i \times T_j)}{N}$$

Por lo que estas son producto de la tabulación cruzada, que es un resumen tabular entre los datos de dos o más variables, de forma que permite examinar la relación entre variables, y consiste en el conteo de la frecuencia de ocurrencia de cada combinación posible de las variables [59]. A su vez, estas tablas de contingencia son representadas en python mediante la librería pandas con la función `pandas.crosstab` generando una tabulación cruzada entre dos o más factores, de forma que genera una tabla de frecuencia (tabla de contingencia) necesaria para la prueba chi-cuadrada [74].

5.2.3.2.1. Suposiciones. Se denota que para que la aplicación de esta prueba sea válida, deben cumplirse ciertos requisitos metodológicos:

- Las variables deben ser categóricas y mutuamente excluyentes, es decir, cada observación debe pertenecer a una sola categoría por variable [88].
- Las observaciones deben ser independientes entre sí, lo que implica que la ocurrencia de una categoría no debe influir en la ocurrencia de otra [89].
- Las frecuencias esperadas deben ser 5 o más en al menos el 80 % de las celdas de la tabla de contingencia, y ninguna celda debe tener una frecuencia esperada igual a cero [59], [89].

En contextos donde estas condiciones no se cumplen—por ejemplo, en tablas 2x2 con muestras pequeñas—se recomienda utilizar el test exacto de Fisher como alternativa. Sin embargo, cuando el tamaño muestral es grande, como en estudios poblacionales con millones de registros, la prueba de chi-cuadrada es robusta y preferible por su eficiencia computacional y su capacidad para detectar asociaciones significativas [88].

La fórmula del estadístico chi-cuadrada para la independencia es:

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Donde:

- O_{ij} representa la frecuencia observada en la celda (i, j) ,
- E_{ij} representa la frecuencia esperada en la celda (i, j) ,
- y la suma se realiza sobre todas las celdas de la tabla de contingencia. El valor calculado se compara con el valor crítico de la distribución chi-cuadrada, determinado por los grados de libertad

$$df = (r - 1)(c - 1)$$

, donde r y c son el número de filas y columnas respectivamente [90]. En síntesis, la prueba de chi-cuadrada es adecuada para analizar la independencia entre variables categóricas en muestras grandes, siempre que se cumplan los supuestos de independencia, exclusividad y suficiencia de frecuencias esperadas. Su aplicación rigurosa permite identificar patrones de asociación relevantes en contextos como el análisis de vulnerabilidad social, salud pública, y estudios demográficos.

Dentro de Scipy existe una función llamada `chi2_contingency` que computa la prueba de independencia de variables utilizando una tabla de contingencia, de forma que genera el estadístico de chi-cuadrada además del valor-p para comprobar la hipótesis de independencia. Además, utilizando la función `expected_freq`, se calculan las frecuencias esperadas dentro de la tabla de contingencia que es pasada como parámetro [91].

5.2.4. Analisis de clases latentes (LCA)

Los algoritmos de agrupamiento son algoritmos no supervisados que organizan y clasifican las observaciones en conjuntos basados en patrones o datos similares [92]. Para el desarrollo de indicadores compuestos, el análisis por grupos permite clasificar grandes cantidades de datos en grupos manejables, de forma que es un método de agregación de los indicadores. Es una herramienta diagnóstica para explorar el impacto de las decisiones metodológicas durante la fase de construcción del indicador compuesto, un método para diseminar información del indicador sin perder en las dimensiones individuales del indicador. También, es un método para seleccionar grupos para la imputación de datos. La agrupación permite que se reduzcan los datos y se sinteticen simultáneamente en dirección de objetos y variables [93].

5.2.4.1. Definición

5.2.4.1.1. Finite mixture modeling (FMM). Es un modelo de aprendizaje no supervisado que consiste en agrupar utilizando una variable modelo para representar datos como una distribución mezclada, de forma que indica la probabilidad de pertenecer a diferentes grupos [94], [95]. Se denota que este método asume que existen agrupaciones finitas para una variable [96]. Este tipo de agrupación utiliza un acercamiento probabilístico a diferencia de otros tipos de agrupamiento que utilizan medidas de distancia como K-Means, agrupación jerárquica, etcétera [95].

5.2.4.1.2. Variables Latentes. Una variable latente es aquella que no posee un error medible, de forma que está libre de error. Las variables categóricas latentes son aquellas variables donde existen diferencias cualitativas entre los grupos de objetos o personas, mientras que las variables latentes continuas son aquellas donde los grupos de personas u objetos difieren cuantitativamente de forma continua [97]. Asimismo, las variables latentes son aquellas variables que no son observadas a simple vista pero se asume que existen dentro del modelo [98].

Por ende, *Latent Class Analysis* (LCA, por sus siglas en inglés) es un método (FFM por sus siglas en inglés) que permite descubrir agrupaciones dentro de los datos, las observaciones se dividen a partir de variables latentes categóricas que están compuestas por diferentes clases o grupos mutuamente excluyentes, por lo que existe una independencia local ya que cada grupo solo posee características comunes a nivel local pero no comparten características entre grupos [95], [99], [100]. A su vez, este está compuesto por dos tipos de parámetros, el primero es la probabilidad de que un caso pertenezca a una clase latente, y la segunda es la probabilidad dado que pertenece a cierta clase latente una variable tenga cierto valor [95], aunque también permite que las variables observadas puedan ser continuas, cuentas, categorías ordinales, binarias o categorías no ordenadas [101].

Es por estas propiedades que la librería Pomegranate permite el modelamiento de variables categóricas a partir de la clase `Categorical`, la cual modela variables asumiendo que son mutuamente excluyentes, siguiendo una distribución de *Bernoulli*, donde una variable puede ser exitosa (pertenecer a la clase latente) o

no (no pertenecer a la clase latente).

Esta clase utiliza varios parámetros, entre ellos: **probs**, que representa las probabilidades teóricas para cada clase latente; **n_categories**, que indica el número de categorías por variable (aunque, al utilizar Torch como base, el número de categorías debe coincidir con el número máximo de categorías entre todas las variables); **check_data**, que permite verificar las propiedades de los datos y convertirlos en tensores de **torch**; **frozen**, que congela los parámetros de la distribución para mantener los valores iniciales durante el entrenamiento; y **inertia**, que controla la proporción de actualización de los parámetros: un valor de 0 implica una actualización completa, mientras que un valor de 1 conserva los parámetros iniciales sin cambios [102].

5.2.4.1.3. Modelo matemático. La probabilidad marginal para estimar una característica dentro de una variable latente para una característica $i_j=1$ en una variable latente C con k clases latentes es:

$$P(i_j) = \sum_{k=1}^K P(C = k)P(i_j = 1|C = k)$$

- C : es la variable latente
- K : es la cantidad total de clases latentes
- k : es la clase latente
- i : es la característica (ítem) de tipo 1

Dada la suposición que es probabilidad disjunta, la probabilidad conjunta para todas las variables observadas es:

$$P(i_1, i_2 \dots i_r) = \sum_{k=1}^K P(k = j)^K P(i_1|C = k) P(i_2|C = k) \dots P(i_r|C = k)$$

- K : cantidad total de clases latentes
- k : es la clase latente
- i : es la característica (ítem)
- C : es la variable latente

Asimismo, la probabilidad condicional de que una observación i pertenezca a la clase latente k está dado por:

$$t_{ik} = \frac{p_k * f(x_i, a_k)}{\sum_{j=1}^K p_j * f(x_i, a_j)}$$

- t_{ik} : es la probabilidad posterior de que la observación i pertenezca a la clase latente k
- p_k : es la probabilidad de pertenecer a la clase latente k
- $f(x_i, a_k)$: es la función de densidad de probabilidad para la observación i en la clase latente k
- K : es el número de clases latentes
- i : es la observación, registro
- j : es la clase latente
- a_k : son los parámetros de la función de densidad de probabilidad para la clase latente k

5.2.4.1.3.0.1. Likelihood y Log-Likelihood El likelihood es un método que permite estimar qué tan probable es que una observación pertenezca a un componente K basado en los parámetros que se han pasado [103]. Esta función incrementa en cada iteración [104]. Por otra parte, el log-likelihood se define como:

$$L(K) = \sum_{i=1}^n \ln\left(\sum_{k=1}^K (p_k * f(x_i, a_k))\right)$$

- $L(K)$: es la log-likelihood del modelo con K clases latentes
- p_k : es la probabilidad de pertenecer a la clase latente k
- $f(x_i, a_k)$: es la función de densidad de probabilidad para la observación i en la clase latente k
- K : es el número de clases latentes
- n : es el número de observaciones, registros

Existe un término de clasificación del likelihood, el cuál se define como [105]:

$$C(K) = \sum_{k=1}^K \sum_{i=1}^n p(x_i|C=k) \ln p(x_i|C=k)$$

- $C(K)$: es la parte de la log-likelihood que mide el ajuste del modelo a los datos
- $p(x_i|C=k)$: es la probabilidad de que la observación i pertenezca a la clase latente k
- K : es el número de clases latentes
- n : es el número de observaciones, registros
- i : es la observación, registro
- k : es la clase latente

- x_i : es la observación i
- C : es la variable latente
- p : es la probabilidad
- $\ln$: es el logaritmo natural
- f : es la función de densidad de probabilidad
- a_k : son los parámetros de la función de densidad de probabilidad para la clase latente k
- p_k : es la probabilidad de pertenecer a la clase latente k
- x_i : es la observación i
- n : es el número de observaciones, registros

La log-likelihood se puede descomponer en dos partes: la parte que mide el ajuste del modelo a los datos y la parte que mide el traslape de los componentes mezclados [106].

$$L(K) = C(K) + E(K)$$

- $L(K)$: es la log-likelihood del modelo con K clases latentes
- K : es el número de clases latentes
- $C(K)$: es la parte de la log-likelihood que mide el ajuste del modelo a los datos
- $E(K)$: es la parte de la log-likelihood que mide la incertidumbre en la clasificación de las observaciones en las clases latentes

5.2.4.1.3.1. Algoritmo de expectación-maximización Este es un algoritmo que parte de la base de que, a partir de un conjunto de datos, existen variables escondidas que se pueden obtener a partir de las variables observables. Para extraer dichas variables latentes, se postula que existe una familia de densidades $f(x|\Phi)$ que, dependiendo de los parámetros Φ , deriva en la correspondiente familia de densidades de probabilidad $g(y|\Phi)$, siguiendo la ecuación:

$$g(y|\Phi) = \int_{\chi(y)} f(x|\Phi) dx$$

De forma que el algoritmo busca encontrar un valor Φ que permita maximizar la función $g(y|\Phi)$, dado un valor observado y , realizando un mapeo mediante la función $f(x|\Phi)$.

El algoritmo está compuesto por dos pasos iterativos: el paso de expectación (E-step) y el paso de maximización (M-step) [107]. En el E-step se calcula la esperanza condicional del logaritmo de la verosimilitud completa, dado los datos observables y el valor actual del parámetro estimado. En el M-step, se maximiza esta esperanza con respecto a los parámetros, obteniendo así una nueva estimación. Este procedimiento se repite iterativamente hasta alcanzar la convergencia.

En este contexto, la función que se maximiza es el **log-likelihood** de la iteración anterior [98].

Sin embargo, debido a que tiene la propiedad de que maximiza el **log likelihood** de forma de que $L(K)^{t+1} \geq L(K)^t$ tiende a converger a un máximo local y La convergencia es lenta. Este problema puede mitigarse si se busca inicializaciones dispersadas, por lo que es recomendado correr el algoritmo con múltiples valores iniciales [98].

5.2.4.1.3.2. Implementación de LCA Las FFM están implementadas en Python mediante la librería Pomegranate, a partir de su clase `pomegranate.gmm.GeneralMixtureModel`. Esta clase permite modelar cualquier tipo de FMM, como *Gaussian Mixture Models*, LCA, modelos de Markov, entre otros, ya que puede utilizar cualquier tipo de distribución probabilística.

Los parámetros que requiere incluyen: **distributions**, que define la lista de distribuciones probabilísticas (Gaussiana, Bernoulli, Normal, Uniforme, etc.); **priors**, que especifica las probabilidades iniciales de pertenencia a cada clase latente; **max_iter**, que indica el número máximo de iteraciones del algoritmo EM; **tol**, que define el umbral de convergencia; **check_data**, que verifica las propiedades de los datos y los convierte en tensores de Torch; **frozen**, que permite congelar los parámetros de la distribución; y **inertia**, que controla la proporción de actualización de los parámetros: un valor de 0 implica actualización completa, mientras que un valor de 1 conserva los parámetros iniciales [102].

5.2.4.1.4. Evaluación de LCA. Existen ciertos conceptos clave para poder evaluar la calidad de un modelo LCA:

5.2.4.1.4.0.1. Entropía La entropía mide el solapamiento entre las clases latentes dentro de un modelo de mezcla finita, de forma que mide la incertidumbre en la clasificación de las observaciones en las clases latentes [106]. Se define como:

$$E(K) = \sum_{k=1}^K \sum_{i=1}^n t_{ik} \ln t_{ik}$$

- K: es el número de clases latentes
- n: es el número de observaciones, registros

5.2.4.1.4.0.2. Validación cruzada Dentro del entrenamiento de un modelo si se repiten las muestras con las que se ha entrenado, siempre se tendrá una predicción perfecta pero en el caso de los datos que no han sido utilizados no se realizaría una predicción correcta. Para evitar el sobreajuste es una práctica común tener una parte de los datos como un conjunto de prueba para poder validar que realmente el modelo generaliza. Sin embargo, incluso a pesar de esta generalización se puede sobreajustar el modelo en el conjunto de pruebas. Para resolver este problema se utiliza un conjunto de validación, que es otra división del conjunto de datos [108].

Una solución a dicho problema es la validación cruzada, de forma que ya no se requiere de un conjunto de validación, en vez de ello el conjunto de entrenamiento se divide en k conjuntos (k -folds) de forma que el modelo utiliza $k-1$ conjuntos para entrenar los datos mientras que el último conjunto lo utiliza como conjunto de validación. Luego itera reiniciando el modelo pero entrenándolo con un conjunto de validación distinto de forma que se itera por k repeticiones donde todos los conjuntos terminan siendo una vez el conjunto de validación. Se extraen las métricas en cada iteración y se promedian para poder decidir si el modelo es óptimo. Finalmente, se prueba dicho modelo con el conjunto de prueba utilizando los valores inicializados del modelo, lo que es útil para filtrar cuáles son modelos útiles [109], [108].

Tal metodología es útil dentro de los FMM's ya que permite comparar distintas métricas a partir de los pliegues (*folds*), además, previene el sobreajuste y permite la replicabilidad de clases identificadas en un nuevo conjunto de datos. Esto mejora la selección del modelo, es útil para identificar clases pequeñas además de que puede ser más informativo respecto al likelihood de óptimos locales [109].

La validación cruzada está implementada en Python mediante la librería `scikit-learn` [108]. La clase `KFold` proporciona los índices necesarios para dividir el conjunto de entrenamiento en *folds* consecutivos. Los parámetros que requiere son:

- **n_splits**: número de *folds* en los que se dividirá el conjunto de entrenamiento.
- **shuffle**: indica si se deben mezclar los datos antes de la división.
- **random_state**: semilla para garantizar la reproducibilidad de la mezcla.

Por otro lado, la función `train_test_split` permite dividir arreglos o matrices en conjuntos aleatorios de entrenamiento y prueba. Sus parámetros incluyen:

- **array**: datos que serán divididos.
- **test_size**: proporción del conjunto de datos que se usará como conjunto de prueba (valor entre 0 y 1).
- **train_size**: proporción del conjunto de entrenamiento.
- **random_state**: semilla para reproducir la separación.
- **shuffle**: indica si se deben mezclar los datos antes de la separación.
- **stratify**: permite realizar una separación estratificada según la variable objetivo (solo aplicable en algoritmos supervisados).

5.2.4.1.4.1. Métricas de evaluación Finalmente, el número óptimo de clases latentes se deciden a partir de reglas probabilísticas explícitas como el *Bayesian Information Criterion* (BIC), *Akaike Information Criterion* (AIC), *Normalized Entropy Criterion* (NEC), aunque también a partir de la interpretabilidad

de los parámetros del modelo basándose en la teoría y en investigación preliminar [95], [110]. También se ha visto a partir de estudios de simulaciones Monte Carlo que el mejor indicador del número real de clases depende en de las características de los datos simulados o del modelo especificado [109].

5.2.4.1.4.1.1. Akaike Information Criterion (AIC) Esta es una métrica estadística utilizada para evaluar que tan bueno es un modelo para ajustarse, a su vez penaliza la complejidad del modelo para evitar que ocurra sobre ajuste, de esta forma identifica al modelo que mejor explique la distribución de los datos [111], [112], [104].

$$AIC(K) = -2 * L(K) + v2(K)$$

- L es la log-likelihood del modelo dado los datos.
- K es el número de parámetros en el modelo
- N es la longitud de registros de los datos, puntos de datos.
- $v2(K)$ es el número de parámetros libres en el modelo.

5.2.4.1.4.1.2. Bayesian Information Criterion (BIC) Este es una métrica estadística utilizada para evaluar que tan bueno es un modelo para ajustarse, a su vez penaliza la complejidad del modelo para evitar que ocurra sobre ajuste, de esta forma identifica al modelo que mejor explique la distribución de los datos [111].

$$BIC = -2 * L(K) + v(K) * \ln(N)$$

- L es la log-likelihood del modelo dado los datos.
- K es el número de parámetros en el modelo
- N es la longitud de registros de los datos, puntos de datos.

5.2.4.1.4.1.3. Normalized Entropy Criterion (NEC) Es una métrica basada en la entropía de probabilidades posteriores que la observación i pertenezca a la clase latente k, este modelo solo es posible de aplicar para modelos de mezclas gaussianas [106], [95]. Esta métrica es útil para comparar modelos con diferente número de clases latentes, de forma que el modelo con menor NEC es el que mejor se ajusta a los datos [95].

$$NEC(K) = \frac{E(K)}{L(K) - L(1)}$$

5.2.4.2. Asignación de factores de riesgo a las clases latentes

La vulnerabilidad se debe a la presencia o ausencia de múltiples características que actuando tanto individual como colectivamente desencadenan la presencia de un suceso. A partir de la ausencia o presencia de estos factores surge el riesgo.

Existen características dentro de un grupo que están asociadas al aumento en la probabilidad de estar expuestos a un suceso, riesgo [22]. Por ende, se le aplica un factor de riesgo, de forma que un factor de riesgo tiene como objetivo predecir la presencia de un suceso adverso, facilitando el proceso diagnóstico y previniendo la situación [113], [114].

Tal como se ha observado en Escalas de Predicción de Riesgo de Violencia Grave Contra la Pareja (EPV) de forma que dependiendo de ítems se clasifican como un riesgo bajo, medio y alto [57].

5.2.5. Algoritmos de agrupación

K-means es un algoritmo de aprendizaje no supervisado que particiona un conjunto de datos en k grupos distintos minimizando la varianza intra-cluster. El algoritmo asigna iterativamente cada punto de datos al centroide más cercano y actualiza los centroides como la media de los puntos asignados. La función objetivo es:

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

- J : es la suma total de las distancias cuadradas entre los puntos y sus centroides.
- k : es el número de clusters.
- C_i : es el conjunto de puntos en el cluster i .
- x : es un punto de datos.

K-Means es uno de los algoritmos de agrupamiento más potentes y populares, aunque sufre de limitaciones como la sensibilidad a la inicialización y la dificultad para manejar tipos de datos mixtos [115]. Se han propuesto variantes como K-Means++, K-Means restringido y K-Means en kernel para abordar estos problemas.

Asimismo, se han propuesto variantes con complejidad de tiempo lineal utilizando el desplazamiento de clusters, mejorando la escalabilidad para grandes conjuntos de datos [116]. También, se ha propuesto X-Means, que extiende K-Means al estimar el número óptimo de clusters utilizando el Criterio de Información Bayesiano (BIC) [117]. También, se desarrolló un K-Means ponderado por entropía (EWKM) para datos dispersos de alta dimensión, optimizando la selección de características dentro de los clusters [118].

Para datos categóricos se utiliza el algoritmo K-Modes, que reemplaza las medias por modas y utiliza una medida de disimilitud adecuada para atributos no numéricos. Este consiste en un algoritmo de agrupamiento basándose en atributos categóricos, de forma que se identifican los valores más frecuentes dentro de cada grupo para determinar su centroide. Se basa en la disimilitud entre los puntos de datos, de forma que entre menos disimilitud existe mayor similitud entre los puntos [119].

Este posee una inicialización aleatoria donde se utiliza cualquier punto como centroide inicial para empezar a generar las agrupaciones. Sin embargo, este método solo funciona cuándo el número de agrupaciones es pequeño y las probabilidades de que al menos una inicialización aleatoria está cercano a una solución buena [120]. Es por ello que este algoritmo puede generar un óptimo local, por lo que se deberían de generar distintos modelos con diferentes semillas para lograr obtener un modelo que sea óptimo global [121].

Precisamente, por ello existen múltiples inicializaciones tales como la inicialización basada en frecuencia donde se busca aquel punto que contenga las características más frecuentes dentro del conjunto de datos [122]. También la inicialización basada en densidad donde se busca aquel punto que sea más parecido al resto basado en sus características [120].

Finalmente, es ligeramente superior al análisis de clases latentes (LCA, según sus siglas en inglés) para casos donde se conoce la estructura de los grupos, sin embargo, *Latent Class Analysis* es preferible para la modelación compleja cuando la estructura es incierta por lo que se debe de inferir [121]

5.2.6. Índice

Un indicador es una medida cuantitativa o cualitativa derivada de las observaciones que puede revelar ciertas características. Es una variable que simplifica información relevante de forma que el fenómeno a medir sea perceptible [93], [123], [124].

Un indicador sintético es aquel que está compuesto por cualquier combinación de indicadores individuales, además, consisten en una aproximación meramente empírica al problema de la medición en nivel y ritmo. Tiene la ventaja de ser sencillo de implementar y permite realizar estimaciones rápidas [125].

Por otro lado, un índice es una medida compuesta que relaciona variables o constructos de diferente naturaleza [126]. Sin embargo, un indicador compuesto es una medida obtenida de agregación y normalización y pesaje de múltiples indicadores en una única medida, permitiendo representar fenómenos complejos e interpretables que no puedan ser medidos por un único indicador. Esto permite que se puedan sintetizar realidades complejas, una fácil interpretación, reducir la visibilidad de múltiples indicadores [93], [123]; por ello, no existe una diferencia entre índice e indicador compuesto.

5.2.6.1. Análisis de clases latentes para la construcción de índices

El Análisis de clases latentes puede ser utilizado para la construcción de análisis estadísticos posteriores (regresiones, indicadores, etcétera) dado que se puede utilizar las probabilidades posteriores para asignar individuos a ciertas clases y utilizarlas como medidas continuas de pertenencia [127]. Esto permite análisis estadísticos posteriores tal como la regresión logística, generación de indicadores, generación de índices o regresiones lineales. Sin embargo, esto solo es posible cuándo se toma en cuenta la incertidumbre en la clasificación que está presente dentro de cada uno de los análisis de clases latentes o es una exploración o que

la separación de clases sea lo suficientemente certera (prácticamente sin incertidumbre) [99].

5.2.6.2. Asignación de pesos

La mayoría de indicadores compuestos utilizan la asignación de pesos igualitarios, de forma que todas las variables son igual de importantes, esto suele distinguir que a veces hay ciertos indicadores que les falta base estadística o empírica para asignar los pesos [93].

Sin embargo, existen otros métodos para asignar pesos a los indicadores, entre ellos se encuentran: la asignación de pesos basada en la opinión de expertos, la asignación de pesos basada en encuestas a la población objetivo, la asignación de pesos basada en el análisis estadístico (Análisis de Componentes Principales, Análisis Factorial, Análisis de clases latentes, entre otros) [93].

5.2.6.2.1. Análisis de componentes principales (PCA). El análisis de componentes principales se basa en calcular el plano que mejor se ajusta basado en que es perpendicular al eje mayor del promedio de los residuos. Contiene el mínimo cuadrado residual que varía inversamente cómo la longitud del eje [128]. De esta forma captura la mayor varianza posible de las variables originales en los primeros componentes principales. Se busca hallar aquel vector que contiene la mayor varianza para sus variables aleatorias [129]

PCA es parte de un tipo de asignación de pesos que se basa en los datos. Asimismo, se ha utilizado en la construcción de indicadores compuestos para asignar pesos según la variabilidad que explique cada una de los indicadores según los componentes principales, de forma que los indicadores que expliquen mayor variabilidad dentro del primer componente principal obtienen el mayor peso [130] y precisamente la idea es utilizar los factores obtenidos dentro del primer componente principal como los pesos [131].

6.1. Recursos utilizados

6.1.1. Hardware

Se utilizó una computadora Dell G15 5525 con las siguientes especificaciones:

Cuadro 1: Especificaciones del hardware utilizado

Componente	Especificación
Modelo	Dell G15 5525
Procesador	AMD Ryzen 7 6800H
Gráficos integrados	AMD Radeon(TM) Graphics
Memoria RAM	16.0 GB
Cores	8 cores
Procesadores lógicos	16 procesadores
L1 cache	512.0 kB
L2 cache	4.0 MB
L3 cache	16.0 MB
GPU dedicada	NVIDIA GeForce RTX 3050 Ti Laptop GPU
Memoria GPU	4 GB
BIOS	1.23.0

*Configuración física del equipo utilizado para el desarrollo y entrenamiento de modelos.

6.1.2. Software

Se utilizó el siguiente software:

Cuadro 2: Especificaciones del software utilizado

Componente	Especificación
Sistema Operativo	Windows 11 Home 64-bit
Python	3.10.11
CUDA	V12.6.77
Visual Studio Code	1.105.1
Cores	8 cores
Procesadores lógicos	16 procesadores
Entorno virtual	venv
Git	2.51.0.windows.1

*Entorno de software utilizado para la ejecución de scripts, entrenamiento y análisis.

6.1.3. Python

Dado que el trabajo se trabajó meramente en python se utilizaron las siguientes librerías y paquetes:

Cuadro 3: Especificaciones de las librerías de Python utilizadas

Librería	Versión
apricot-select	0.6.1
asttokens	3.0.0
certifi	2025.10.5
charset-normalizer	3.4.3
colorama	0.4.6
comm	0.2.3
contourpy	1.3.2
cycler	0.12.1
debugpy	1.8.17
decorator	5.2.1
efficient_apriori	2.0.6
et_xmlfile	2.0.0
exceptiongroup	1.3.0
executing	2.2.1

*Entorno de software utilizado para la ejecución de scripts, entrenamiento y análisis.

Cuadro 4: Especificaciones de las librerías de Python utilizadas

Librería	Versión
filelock	3.20.0
fonttools	4.60.1
fsspec	2025.9.0
idna	3.11
intel-openmp	2021.4.0
ipykernel	6.30.1
ipython	8.37.0
jedi	0.19.2
Jinja2	3.1.6
joblib	1.5.2
jupyter_client	8.6.3
jupyter_core	5.8.1
kagglehub	0.3.13

Cuadro 5: Especificaciones de las librerías de Python utilizadas

Librería	Versión
kiwisolver	1.4.9
kmodes	0.12.2
kneed	0.8.5
llvmlite	0.45.1
MarkupSafe	3.0.3
matplotlib	3.10.7
matplotlib-inline	0.1.7
mkl	2021.4.0
mpmath	1.3.0
nest-asyncio	1.6.0
networkx	3.4.2
nose	1.3.7
numba	0.62.1
numpy	1.26.4

Cuadro 6: Especificaciones de las librerías de Python utilizadas

Librería	Versión
openpyxl	3.1.5
packaging	25.0
pandas	2.3.3
parso	0.8.5
pillow	11.3.0
platformdirs	4.5.0
pomegranate	1.1.2
prompt_toolkit	3.0.52
psutil	7.1.0
pure_eval	0.2.3
Pygments	2.19.2
pyparsing	3.2.5
python-dateutil	2.9.0.post0

*Continuación Tabla No. 8

Cuadro 7: Especificaciones de las librerías de Python utilizadas

Librería	Versión
pytz	2025.2
pywin32	311
PyYAML	6.0.3
pyzmq	27.1.0
requests	2.32.5
scikit-learn	1.7.2
scipy	1.15.3
seaborn	0.13.2
six	1.17.0
stack-data	0.6.3
sympy	1.14.0
tbb	2021.13.1
threadpoolctl	3.6.0

Cuadro 8: Especificaciones de las librerías de Python utilizadas

Librería	Versión
torch	2.3.1+cu121
torchaudio	2.3.1+cu121
torchvision	0.18.1+cu121
tornado	6.5.2
tqdm	4.67.1
traitlets	5.14.3
typing_extensions	4.15.0
tzdata	2025.2
Unidecode	1.4.0
urllib3	2.5.0
wcwidth	0.2.14

*Continuación Tabla No. 8

6.2. Diseño y alcance del estudio

El presente estudio tiene como objetivo principal desarrollar un índice de vulnerabilidad estructural para víctimas de violencia intrafamiliar, sin embargo, es imperativo definir las limitantes y el diseño de dicho estudio.

6.2.1. Diseño

Este estudio siguió un enfoque cuantitativo de tipo exploratorio y descriptivo, orientado a descubrir patrones en datos históricos de violencia intrafamiliar. Está compuesto por las siguientes fases:

- Unificación de los datos
- Llenado de datos

- Limpieza de datos
- Generación de variables nuevas
- Unificación de fuentes de datos alternas
- Exploración de datos
- Filtración de variables mediante pruebas de independencia
- Establecimiento de asociaciones
- Ingeniería de características para el desarrollo de perfiles
- Desarrollo de perfiles utilizando Análisis de clases latentes
- Construcción del índice de vulnerabilidad a partir de las clases latentes en conjunto con variables de interés
- Desarrollo del índice de vulnerabilidad calculando los pesos de cada factor del mismo

6.2.2. Población y enfoque del estudio

Esta investigación solo se enfocará en los casos de violencia intrafamiliar donde las víctimas sean mujeres adultas de mediana edad [18-60]; cubriendo casos desde 2008 hasta 2023 en Guatemala. Se tomó esta decisión debido a que dentro de los datos recopilados por el Instituto Nacional de Estadística de Guatemala se observa que la mayoría de víctimas de violencia intrafamiliar son mujeres adultas de mediana edad.

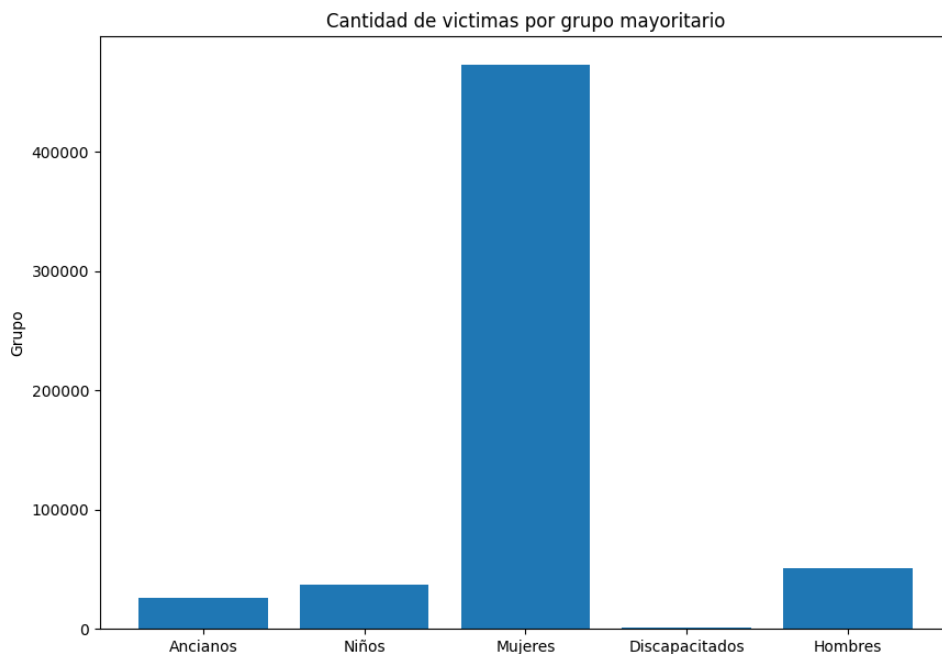


Figura 1: Distribución de víctimas por grupo mayoritario

Los datos recopilados a pesar de haber sido recopilados por un censo, no pueden ser considerados como una población y sí como una muestra debido a que no recopilan todos los casos de violencia intrafamiliar en Guatemala, sino

únicamente aquellos que han sido reportados a las autoridades correspondientes [59]. Por ende, no se deben de considerar distribuciones normales para su estudio basándose en el teorema central del límite.

6.2.3. Unidad de análisis

La unidad de análisis será a nivel de incidente reportado dentro de la base de datos, dado que no se posee un historial de los casos sino únicamente de los reportes a lo largo de los años.

6.2.4. Ética y gobernanza

Los datos utilizados poseen aprobación de uso y de ser públicos por parte de las entidades que los proveyeron (revisar anexos y bibliografía).

6.2.5. Constructos primarios

La vulnerabilidad estructural de violencia intrafamiliar es la capacidad de un individuo de sobrellevar y adaptarse a la situación de violencia intrafamiliar, es por ello que se desarrollaron indicadores para la construcción de la vulnerabilidad estructural:

1. Indicador de vulnerabilidad individual de la víctima: este indicador describe la capacidad de autonomía de la víctima, el nivel educativo de la víctima, si la víctima trabaja o no, la cantidad de hijos de la víctima y la etnia de la víctima, esto es así debido a que la autonomía y el empoderamiento de la víctima son factores clave en su capacidad para enfrentar la violencia [47].
2. Indicador de vulnerabilidad de la situación: este indicador captura características clave para medir la peligrosidad que actualmente está sobrellevando la víctima. Esto incluye el tipo de violencia que está sufriendo la víctima, la cantidad de agresores implicados en el hecho, la cantidad de víctimas adicionales implicadas en el hecho, además de características propias del agresor, es importante medir la peligrosidad de la situación ya que puede suponer un contexto de urgencia [47].
3. Indicador de vulnerabilidad de la relación: este indicador describe la relación que actualmente se está llevando entre la víctima y el agresor, incluyendo el tipo de relación que existe entre ambos, el estado civil de ambos, quién gana más dinero, si trabaja o no.
4. Indicador de vulnerabilidad contextual: este indicador describe la capacidad de acceso a servicios de salud, justicia y bomberos en el departamento donde se encuentra la víctima.

6.3. Fuentes de datos y preparación de los datos

Para poder llevar a cabo una investigación profunda sobre los datos para poder construir un índice de vulnerabilidad estructural es imperativo que se cuente con diversas fuentes de datos que permitan enriquecer el análisis y la construcción del índice, además de que los datos estén limpios y listos para utilizarse.

6.3.1. Origen de los datos

Dado que dentro del presente trabajo de graduación no se estará realizando recolección de datos, se describen estos conjuntos de datos y su procedencia.

6.3.1.1. Base de datos de violencia intrafamiliar

Se accedió a la base de datos de violencia intrafamiliar del Instituto Nacional de Estadística de Guatemala, así mismo, estas estadísticas son una aproximación a la medición del registro de denuncias y tiene como objetivo informar a la sociedad guatemalteca [132]. Esta consiste en una recopilación de variables sociodemográficas y contextuales de víctimas, agresores y hechos de violencia intrafamiliar captadas a partir de registros que se han realizado desde 2008 hasta 2023. Además, se dispone de diccionario de datos por año para comprender en qué consiste cada variable, finalmente se destaca que dicha base de datos es de dominio público: <https://www.ine.gob.gt/>

6.3.1.1.1. Variables de la víctima.

1. VIC_SEXO: sexo de la víctima
2. VIC_OCUP: ocupación de la víctima
3. VIC_TRABAJA: trabaja la víctima
4. VIC_EDAD: edad de la víctima
5. TOTAL_HIJOS: cantidad de hijos que posee la víctima
6. NUM_HIJ_HOM: cantidad de hijos varones que posee la víctima
7. NUM_HIJ_MUJ: cantidad de hijos mujeres que posee la víctima
8. VIC_GRUPET: grupo étnico de la víctima
9. VIC_ALFAB: alfabetización de la víctima
10. VIC_EST_CIV: estado civil de la víctima
11. VIC_NACIONAL: nacionalidad de la víctima
12. VIC_DEDICA: qué se dedica la víctima
13. VIC_DISC: la víctima es discapacitada
14. TIPO_DISCAQ: es el tipo de discapacidad de la víctima
15. VIC_ESCOLARIDAD: escolaridad de la víctima

6.3.1.1.2. Variables del agresor.

1. AGR_SEXO: sexo del agresor
2. AGR_OCUP: ocupación del agresor
3. AGR_TRABAJA: trabaja el agresor
4. AGR_EDAD: edad del agresor

5. AGR_EST_CI: estado civil del agresor
6. AGR_GRUPET: grupo étnico del agresor
7. AGR_NACIONAL: nacionalidad del agresor
8. AGR_DEDICA: qué se dedica el agresor
9. AGR_ALFAB: alfabetización del agresor
10. AGR_ESCOLARIDAD: escolaridad del agresor

6.3.1.1.3. Variables del hecho.

1. OTRAS_VICTIMAS: cantidad de otras víctimas implicadas
2. VIC_OTRAS_HOM: cantidad de otras víctimas hombres implicadas
3. VIC_OTRAS_MUJ: cantidad de otras víctimas mujeres implicadas
4. VIC_OTRAS_N_OS: cantidad de otras víctimas niños implicadas
5. VIC_OTRAS_N_AS: cantidad de otras víctimas niñas implicadas
6. AGRESORES_OTROS_TOTAL: cantidad de agresores extra implicados en el hecho
7. AGR_OTROS_HOM: cantidad de agresores hombres extra implicados en el hecho
8. AGR_OTRAS_MUJ: cantidad de agresores mujeres extra implicados en el hecho
9. AGR_OTROS_N_OS: cantidad de agresores niños extra implicados en el hecho
10. AGR_OTRAS_N_AS: cantidad de agresores niñas extra implicados en el hecho
11. INST_DONDE_DENUNCIO: instituto que registra el hecho
12. : Organismo Jurisdiccional que recibe la denuncia
13. LEY_APLICABLE: ley que se aplicó en el caso
14. ARTICULOVCM: artículos aplicados
15. MEDIDAS_SEGURIDAD: medidas de seguridad otorgadas por el Organismo Judicial
16. TIPO_MEDIDA: tipo de medida de seguridad aplicada
17. ORGANISMO_REMITE: Organismo Jurisdiccional a donde se remitió
18. HEC_RECUR_DENUN: si este caso ya había sido denunciado antes
19. INST_DONDE_DENUNCIO: en qué instituto fue denunciado antes
20. CONDUCTENTE: si el caso fue conductente
21. QUIEN_REPORTA: quién reporta el caso

22. HEC_TIPAGRE: tipo de agresión sufrido por la víctima
23. VIC_REL_AGR: relación entre la víctima y el agresor
24. HEC_DEPTOMCPIO: municipio en el que sucedió el hecho
25. DEPTO_MCPIO: municipio en el que se registró el hecho
26. HEC_DEPTO: departamento en el que sucedió el hecho
27. DEPTO: departamento en el que se registró el hecho
28. HEC_AREA: el área en la que sucedió el hecho
29. DIA_EMISION: día en el que se registró el hecho
30. ANO_EMISION: año en el que se registró el hecho
31. MES_EMISION: mes en el que se registró el hecho,
32. HEC_DIA: día en el que sucedió el hecho
33. HEC_ANO: año en el que se sucedió el hecho
34. HEC_MES: mes en el que se sucedió el hecho.

6.3.1.2. Clasificador Nacional de Ocupaciones (C.N.O. 2020)

Se dispone de un documento que contiene tablas que permiten agrupar las ocupaciones en diferentes grupos, subgrupos principales, subgrupos, grupo unitario y código de denominación (las ocupaciones). A su vez, utiliza las encuestas nacionales vinculadas a ocupación y empleo como los censos, y se destaca el hecho de que es un documento que unifica a nivel de país y que es de carácter público [133].

Para ello se generó un código fuente que permite extraer dichas tablas directamente del archivo pdf. Se utilizó una librería llamada Camelot para abrir el pdf y procesarlo, lo que genera un objeto que es convertible a excel, y genera una tabla por cada hoja dentro del archivo.

6.3.1.3. Encuesta Nacional de Empleo e Ingresos - E.N.E.I. - 2021

Se dispone de un tablero que recopila datos de empleo según los grupos de ocupación que se manejan. Se destaca que lo recopilado son datos del salario según el grupo de trabajo al cuál pertenece la persona. Se destaca que es un documento de carácter público [134]. Se dispone del ingreso mensual promedio según el grupo desde el primer semestre de 2014 hasta el primer semestre de 2021.

6.3.1.4. Proyecciones Departamentales 2010-2050

Se dispone de un documento por parte del Instituto Nacional de Estadística que contiene las proyecciones departamentales de población desde 2010 hasta 2050 [135].

6.3.1.5. Distribución Territorial de los Servicios de Justicia

Por parte del Instituto de Estudios Comparados en Ciencias Penales de Guatemala se tiene un programa llamado Mirador Judicial, cuyo objetivo es democratizar, transparentar el funcionamiento, incluyendo discurso crítico por parte del Sistema de Administración de Justicia. Por lo que se desarrolló un control de la distribución territorial (por departamento, dado que no se tenía un control exhaustivo por municipio en todas las fuentes de datos) de los servicios de justicia obteniendo información por parte del Centro de Información, Desarrollo y Estadística Judicial, el Instituto de la Defensa Pública Penal, Policía Nacional Civil y el Ministerio Público, realizando conteos a nivel departamental de diversas infraestructuras de justicia [136]:

1. Coordinaciones que integran la división técnica profesional.
2. Comisarías, estaciones, subestaciones y división de investigación criminal.
3. Fiscalías regionales, fiscalías distritales, fiscalías municipales y agencias fiscales.
4. Salas mixtas, tribunales de sentencia, juzgados de primera instancia penal, narcoactividad y delitos contra el ambiente, juzgados de adolescentes y niñez, juzgados de ejecución, juzgados de paz penal y comunitarios, juzgados de violencia contra la mujer y femicidio, tribunales de sentencia contra la mujer y femicidio, y juzgados y tribunales de mayor riesgo.

El conjunto de datos obtenido estaba estructurado en un archivo Excel con múltiples hojas, cada una correspondiente a instituciones de diferentes organismos gubernamentales (Organismo Judicial, Ministerio Público, Instituto de la Defensa Pública Penal y Policía Nacional Civil).

6.3.1.6. Directorio de las estaciones de bomberos por parte del Cuerpo de Bomberos Municipales

El Cuerpo de Bomberos Municipales puso a disposición el directorio de sus sedes, conteniendo la estación, su teléfono, y su ubicación, debe de denotarse que esta información solo abarca el departamento de Guatemala [137].

6.3.1.7. Directorio de la Asociación Nacional de Bomberos Departamentales de Guatemala

Por parte de la Asociación Nacional de Bomberos Departamentales de Guatemala se tiene el directorio de sus sedes, conteniendo su dirección, el departamento, el municipio, el teléfono y su correo electrónico [138].

6.3.1.8. Información pública brindada por el Ministerio de Salud Pública y Asistencia Social

Se contó con un directorio desglosado por departamento y municipio que incluye las ubicaciones exactas de cada servicio de salud (hospital, centro de salud, centro de atención permanente, vectores, puesto de salud fortalecido), publicado por el Ministerio de Salud Pública y Asistencia Social [139]. El objetivo de esta etapa fue estructurar y unificar la información de las distintas hojas del archivo provisto por el MSPAS en un único conjunto de datos normalizado y utilizable para el análisis posterior.

6.3.2. Preparación de los datos

Los datos para su correcto análisis y modelado requieren de un proceso de preparación exhaustivo que garantice su calidad, consistencia y relevancia. Este proceso incluyó varias etapas clave:

6.3.2.1. Unificación de los datos

Las variables dentro del conjunto de datos de violencia intrafamiliar proporcionado por el INE difieren a través de los años, algunas existen en todos los archivos, mientras que otras fueron agregadas a través del tiempo, además, hay unas que poseen nombres distintos según el archivo. Por ello fue necesario estandarizar los conjuntos de datos, unificando las variables, construyendo en primera instancia una tabla que muestre las variables existentes a través del tiempo, de forma que se determine cuáles variables se repiten, cuáles fueron agregadas después. Esto permite generar un único conjunto de datos que contenga todos los registros garantizando una consistencia en las variables a través de los años.

6.3.2.1.1. Excel de conteo de presencia/ausencia de variables. Se generó un excel que hace una tabla de conteo que permite visualizar si dicha variable apareció en el año en que fue generado el dataset, de forma que se sepa cuáles variables se han mantenido a lo largo del tiempo y cuáles no.

6.3.2.1.2. Procesamiento de las ocupaciones. Para las ocupaciones, se inició con el procesamiento del pdf que disponía de las mismas. Por ello, para transformar el PDF a una estructura tabular utilizable, se desarrolló un código en Jupyter Notebook que emplea la librería camelot para la extracción automatizada de tablas. Esta herramienta fue seleccionada por su capacidad para identificar y segmentar bordes de celdas en documentos PDF mediante el modo lattice, lo que facilita la detección de estructuras tabulares complejas sin necesidad de procesamiento manual [75].

6.3.2.1.2.1. Conversión de pdf a tablas El proceso de extracción se ejecutó utilizando la función `camelot.read_pdf(path, flavor='lattice', pages='all')`, que genera un conjunto de objetos convertibles a `pandas.DataFrame`, uno por cada tabla detectada en las páginas del documento. Los resultados fueron exportados a un archivo Excel, con una hoja por cada tabla detectada, utilizando las funciones nativas de `pandas` y `openpyxl`.

6.3.2.1.2.2. Limpieza de las tablas generadas Posteriormente, se aplicó un proceso de preprocesamiento y limpieza sobre las hojas del archivo generado. Para ello, se emplearon las librerías `pandas` y `openpyxl`, lo que permitió iterar sobre cada hoja del archivo Excel e identificar el encabezado correspondiente a las tablas válidas. Se eliminaron filas y columnas vacías o con contenido no tabular, y se ajustaron los encabezados de las columnas según la posición real de los nombres de campo detectados en cada hoja.

En el caso de hojas con estructuras atípicas o encabezados desalineados, se implementaron tratamientos específicos mediante código manual, ajustando los índices de inicio y fin de las tablas y completando las categorías faltantes de manera programática.

6.3.2.1.2.3. Exportación del archivo final El resultado final fue un archivo Excel estructurado con las ocupaciones clasificadas jerárquicamente y listas para su integración en las etapas posteriores de análisis y modelado.

6.3.2.1.3. Procesamiento de la información pública brindada por el Ministerio de Salud Pública y Asistencia Social. Para ello, se desarrolló un script en Python que automatiza la lectura, limpieza y consolidación de los datos utilizando las librerías `pandas`, `openpyxl`, `os` y `re`. Posteriormente, se identificaron las hojas contenidas en el archivo con `pd.ExcelFile(path).sheet_names`, lo que permitió realizar un recorrido iterativo sobre todas ellas.

Cada hoja fue cargada como un objeto `pandas.DataFrame` mediante la función `pd.read_excel()`, definiendo el parámetro `dtype=str` para evitar perder datos en caso de que sean columnas con diferentes tipos de datos. Sobre cada `DataFrame` se aplicó un proceso de limpieza estructurada, que incluyó:

- Eliminación de columnas irrelevantes, encabezados redundantes o índices no informativos.
- Renombramiento de columnas bajo un esquema estandarizado (por ejemplo: `NOMBRE DEL SERVICIO CATALOGADO COMO, DISTRITO, DEPARTAMENTO`).
- Eliminación de filas vacías mediante `dropna(how = 'all')` para mantener únicamente registros válidos.

Adicionalmente, se generaron dos columnas complementarias:

- `depto`: registraba el nombre del departamento asociado a cada hoja.
- `count_depto`: contabilizaba el número instituciones dentro de ese departamento.

Estas transformaciones facilitaron la posterior concatenación nacional de los datos mediante `pd.concat(axis=0)`, unificando los `DataFrames` departamentales en un solo conjunto. Posteriormente, se ejecutaron las funciones `drop_duplicates()` y `dropna()` para asegurar la integridad y unicidad de los registros en columnas clave.

Con el propósito de estandarizar las denominaciones geográficas, se aplicó un proceso de normalización de texto que eliminó acentos, redujo las cadenas a minúsculas y unificó los espacios. Además, se empleó la función `get_close_matches()` del módulo `difflib` para emparejar los nombres de municipios con el catálogo oficial del Instituto Nacional de Estadística (INE), permitiendo una tolerancia controlada a errores ortográficos (`n=1`, `cutoff=0.6`). Los registros sin coincidencia suficiente fueron descartados para preservar la consistencia territorial del conjunto.

Para la clasificación del tipo de servicio, se implementó una expresión regular (`regex`) aplicada sobre la columna “`NOMBRE DEL SERVICIO CATALOGADO COMO`”, procesada mediante `re.search()`. Esto permitió identificar categorías como hospital, centro de salud, puesto de salud, entre otras, generando una columna adicional con la tipología asignada a cada registro.

Con los datos ya clasificados y estandarizados, se realizaron conteos de frecuencia mediante las funciones `value_counts()` y `groupby()`, agregando los resultados por distrito y departamento. Estos conteos se consolidaron a través de la función `merge()` con los parámetros `how="left"` y `on=["servicio", "depto"]`,

de modo que se preservaran todas las combinaciones relevantes y se distinguieran las columnas mediante los sufijos `_distrito` y `_depto`.

Finalmente, se construyó una tabla dinámica (`pivot table`) que resumía la cantidad de servicios por tipo, departamento y municipio. Esta se generó con los parámetros `index=["DEPARTAMENTO_NORM", "MUNICIPIO_NORM_MATCH"]`, `columns="servicio"`, `aggfunc="size"` y `fill_value=0`.

El resultado final fue un archivo CSV estructurado, que consolida la distribución de servicios de salud por categoría y localización geográfica, sirviendo como insumo principal para las etapas subsiguientes de análisis espacial y modelado del índice de vulnerabilidad estructural.

6.3.2.1.4. Procesamiento de la información de bomberos. El objetivo de este procesamiento fue estandarizar la información de las estaciones de bomberos tanto de municipales como departamentales, unificando los formatos y asegurando la consistencia territorial.

6.3.2.1.4.1. Lectura del archivo Es por ello que se desarrolló un script en Jupyter Notebook con el cuál se leyó el archivo que contenía la distribución de las estaciones de bomberos departamentales, utilizando `pandas.read_excel` con `dtype=str`, y el encabezado desplazado se leyó el archivo.

6.3.2.1.4.2. Limpieza de columnas y textos Posteriormente se eliminaron las columnas inválidas, y se renombraron a columnas significativas, se definieron las funciones para la normalización de texto a partir de eliminación de acentos y normalización con `unidecode`, reducción a minúsculas y unificación de espacios. Esto se realizó debido a que esto permite reducir la variabilidad por diferencias ortográficas, mejorando la calidad del emparejamiento.

6.3.2.1.4.3. Normalización de localizaciones A partir de verificar una lista con departamentos oficiales se realizó un mapeo de dichos departamentos utilizando el módulo `difflib.get_close_matches` sobre la versión normalizada del texto para asignar el departamento oficial cuándo la coincidencia exacta no existe. Y solo en casos especiales se manejaron manualmente excepciones frecuentes mediante un diccionario de reemplazos.

Asimismo, se construyó un catálogo de tuplas con el municipio normalizado y el departamento normalizado para poder facilitar la coincidencia con los sitios estandarizados.

6.3.2.1.4.4. Conteo y exportación Luego se agrupó por municipio y departamento para generar los conteos, representando la presencia de estaciones de bomberos departamentales por ubicación geográfica. Y tan solo en el caso del departamento de Guatemala se sumaron 22 estaciones reportadas por la información del Cuerpo de Bomberos Municipales de Guatemala. Finalmente, se exportó el conjunto de datos consolidado como un archivo CSV.

6.3.2.1.5. Procesamiento de la información de los servicios de justicia.

6.3.2.1.5.1. Lectura de la fuente de información Para el procesamiento, se utilizó Python 3.10.11 y las librerías `pandas` y `numpy`, debido a su capacidad para manejar estructuras tabulares heterogéneas y operaciones vectorizadas de limpieza. Cada hoja fue cargada como un objeto `pd.DataFrame`,

estandarizando nombres de columnas y reemplazando valores nulos en formato texto por `np.nan`, a fin de garantizar la consistencia de tipos y permitir operaciones de limpieza automatizadas.

Durante la inspección se identificó que la estructura de las hojas incluía bifurcaciones de registros por zonas geográficas (por ejemplo, subdivisiones específicas dentro del departamento de Zacapa). Para evitar duplicaciones y preservar únicamente la información relevante, se filtraron los registros hasta el punto de separación válido.

6.3.2.1.5.2. Normalización y llenado de nombres geográficos A partir de la columna de ubicación, se extrajeron los nombres de departamentos utilizando expresiones regulares y se creó una nueva columna con el valor correspondiente. Dado que los nombres de departamentos solo aparecían en las filas iniciales de cada bloque, se aplicó la función `fillna(method="ffill")` para propagar el último valor observado hacia abajo, garantizando la correcta asociación entre instituciones y departamentos.

Posteriormente, los registros que no correspondían a departamentos válidos fueron eliminados. Los `DataFrame` generados para cada organismo se unificaron mediante la función `merge` de `pandas`, especificando `how='outer'` y `on='Ubicación'`, con el objetivo de preservar la totalidad de ubicaciones incluso cuando alguna institución no tuviera presencia en determinado municipio. Los valores faltantes resultantes se imputaron con 0, interpretándose como ausencia de instituciones del tipo correspondiente.

6.3.2.1.5.3. Conteo y exportación Finalmente, se filtraron los registros para conservar únicamente aquellos en los que la columna de departamento estuviera presente en todos los organismos, se normalizaron nuevamente los nombres de departamentos y municipios, y se exportó el resultado como un archivo CSV consolidado con el conteo total de instituciones por municipio y departamento.

6.3.2.2. Llenado y depuración de valores faltantes en el conjunto de datos

El propósito de este proceso fue maximizar la completitud del conjunto de datos nacional de violencia intrafamiliar, incrementando la utilidad analítica y facilitando posteriores etapas de limpieza y modelado. Para ello, se diseñó una metodología sistemática en Python, orientada a la identificación, análisis y tratamiento de valores nulos en las variables clave del `DataFrame` consolidado.

Utilizando librerías como: `pandas`, `numpy`, `matplotlib`, `os`, `re` y `ydata_profiling`, seleccionadas por su capacidad para manipular datos tabulares, realizar análisis exploratorio y generar reportes automáticos.

6.3.2.2.1. Identificación de valores nulos. Se implementó la función `has_nulls(df, column)`, que permite cuantificar los valores nulos por variable. Esta función fue utilizada iterativamente para priorizar el tratamiento de las columnas con mayor cantidad de datos faltantes, asegurando un enfoque eficiente y dirigido.

6.3.2.2.2. Estrategias de llenado. Para cada variable con valores nulos, se definió una estrategia de imputación específica, basada en el tipo de dato y el contexto semántico:

- **Variables categóricas:** se imputaron valores como 'Ignorado', 'Ocupación ignorada' o 'Z' (en el caso de medidas de seguridad), permitiendo distinguir entre datos realmente ausentes y aquellos ignorados por el sistema de registro.
- **Variables numéricas:** se empleó el valor 99.0 o 9999 para indicar ausencia de información, siguiendo la convención utilizada en etapas previas del procesamiento.
- **Variables territoriales:** Para los casos en que el departamento era nulo pero el municipio estaba registrado, se construyó un diccionario exhaustivo de mapeo municipio-departamento. Mediante la función `obtener_departamento()`, se rellenaron los valores faltantes de `DEPTO` y `HEC_DEPTO`, garantizando la correcta georreferenciación.
- **Variables dependientes:** En el caso de discapacidad, se imputó el tipo de discapacidad en función del valor registrado en `VIC_DISC`, siguiendo una lógica condicional que distingue entre 'Sin discapacidad', 'Ignorado' y 'Con discapacidad'.

6.3.2.2.3. Justificación de las imputaciones. La imputación de valores ignorados o convencionales responde a la necesidad de preservar la mayor cantidad de registros posible para el análisis, evitando la eliminación prematura de casos que podrían aportar información relevante. Además, la distinción explícita entre datos ignorados y ausentes permite un tratamiento diferenciado en etapas posteriores de limpieza y modelado, minimizando el sesgo y facilitando la interpretación.

6.3.2.2.4. Resultados y exportación. Tras el llenado sistemático de valores nulos, se verificó la reducción de datos faltantes mediante consultas directas (`has_nulls(df, column)`). El `DataFrame` resultante fue exportado en formato CSV, asegurando la reproducibilidad y la trazabilidad del proceso.

6.3.2.3. Transformación y enriquecimiento del conjunto de datos nacional

El objetivo de esta etapa fue depurar, transformar y enriquecer el conjunto de datos nacional de violencia intrafamiliar, preparando la base para análisis estadísticos y modelado avanzado.

6.3.2.3.1. Importación y exploración inicial. Se importaron librerías especializadas como `unidecode`, `scikit-learn`, `rapidfuzz`, `pandas`, `numpy` y `ydata_profiling`, seleccionadas por su capacidad para limpiar texto, transformar variables y realizar operaciones de codificación. El conjunto de datos fue cargado desde el archivo CSV previamente rellenado, permitiendo una inspección preliminar mediante `df.describe()` y `df.info()`.

6.3.2.3.2. Llenado y depuración de valores nulos. Se identificaron y rellenaron valores nulos en variables territoriales, categóricas y numéricas. Para los departamentos faltantes, se utilizó un diccionario exhaustivo de mapeo municipio-departamento, asegurando la correcta georreferenciación. Las variables categóricas y numéricas se imputaron con valores convencionales como 'Ignorado',

'Ocupación ignorada', 99.0 o 9999.0, siguiendo la lógica definida en etapas previas.

6.3.2.3.3. Transformación de variables. Las columnas no numéricas fueron convertidas a tipo `category`, optimizando el uso de memoria y facilitando la manipulación posterior. Se normalizaron los textos mediante la conversión a minúsculas, eliminación de acentos y espacios, utilizando la función `unidecode` y procesamiento por lotes para evitar sobrecarga de memoria.

6.3.2.3.4. Vinculación con registros externos. Se realizó la vinculación de las ocupaciones de víctimas y agresores con el registro oficial de ocupaciones ya limpio, a partir de técnicas de **fuzzy matching** (`rapidfuzz`) se mapearon descripciones libres a categorías normalizadas, corrigiendo manualmente los casos no coincidentes, asegurando la máxima cobertura y precisión en la clasificación ocupacional.

Se generaron nuevas variables derivadas (`VIC_OCUP_INE`, `AGR_OCUP_INE`, `VIC_OCUP_GRUPO`, etc.) mediante operaciones de unión con el catálogo de ocupaciones, permitiendo el análisis agrupado y la estimación de ingresos, obtenido a partir del portal oficial del INE. Los valores nulos en estas variables fueron imputados con 'Ocupación ignorada' o 'Grupo ignorado'.

6.3.2.3.5. Unificación semántica de variables. Se diseñaron funciones para unificar las variables de dedicación y ocupación (`VIC_DEDICA`, `VIC_OCUP_GRUPO`), generando la variable `VIC_OFICIO` que sintetiza la información relevante para el análisis laboral. Se aplicó la misma lógica para los agresores.

6.3.2.3.6. Normalización y limpieza adicional. Se corrigieron inconsistencias en variables como la relación víctima-agresor y el organismo jurisdiccional receptor, de forma que explícitamente se mapeo garantizando la homogeneidad en datos.

6.3.2.3.7. Estimación y vinculación de ingresos. Se incorporaron estimaciones de salario anual por grupo ocupacional y año, utilizando datos oficiales del Encuesta Nacional de Empleo e Ingresos (ENEI). Se realizó la unión correspondiente y se imputaron valores faltantes con promedios por ocupación, enriqueciendo el dataset con los salarios según las ocupaciones normalizadas.

6.3.2.3.8. Codificación de variables categóricas. Se preparó el `DataFrame` para su análisis mediante técnicas de codificación:

- **One Hot Encoding:** se generan múltiples columnas que indican la presencia o no de una categoría dentro de una variable. Permite variables categóricas con hasta 10 categorías, facilitando su uso en modelos estadísticos [108].
- **Label Encoding:** se asigna un número para cada categoría dentro de una variable, optimizando su representación mediante números [108].

6.3.2.3.9. Exportación y trazabilidad. El `DataFrame` final, limpio y enriquecido, fue exportado en formato `CSV`, asegurando la trazabilidad y reproducibilidad del proceso.

Este procedimiento permitió la depuración exhaustiva, normalización y enriquecimiento del conjunto de datos, garantizando la coherencia, comparabilidad y robustez analítica. La metodología empleada facilita el análisis multidimensional y el modelado avanzado, sentando las bases para estudios estadísticos y de machine learning sobre violencia intrafamiliar en Guatemala.

6.4. Exploración de datos

Para llevar a cabo el desarrollo de un índice de vulnerabilidad sobre violencia intrafamiliar era necesaria hacer una exploración exhaustiva del conjunto de datos, con el objetivo de identificar patrones y poder discernir que variables eran las más relevantes para el desarrollo posterior.

6.4.1. Ingeniería de características inicial

La etapa de ingeniería de características se orientó a la transformación, síntesis y generación de variables derivadas, con el objetivo de optimizar el conjunto de datos para análisis estadísticos, segmentación y modelado avanzado. El proceso se implementó en Python, empleando librerías como `pandas`, `numpy`, `scikit-learn` y `joblib`.

6.4.1.1. Depuración inicial

Se cargó el conjunto de datos y se procedió a eliminar variables redundantes o altamente correlacionadas, tales como ocupaciones y dedicaciones detalladas de víctima y agresor, para evitar redundancia y multicolinealidad.

6.4.1.2. Transformación de variables numéricas a rangos

Se diseñaron funciones para agrupar variables numéricas en categorías semánticas relevantes:

- **Cantidad de hijos:** se clasificó en rangos como sin hijos”, hijo único”, hijos medios”, muchos hijos”, demasiados hijos” e ignorado”, facilitando el análisis de la estructura familiar.
- **Edad:** se categorizaron las edades en niño”, adolescente”, adulto joven”, adulto medio”, adulto mayor” e ignorado”, permitiendo el estudio de patrones etarios.
- **Número de otros agresores y otras víctimas:** se sintetizó en sí”, no” e “ignorado”, simplificando la representación de la dinámica grupal.

6.4.1.3. Normalización de variables educativas

Se agruparon los niveles educativos en categorías principales (primaria”, básicos”, diversificado”, universidad”), utilizando expresiones regulares para identificar patrones textuales y reducir la dimensionalidad de la variable.

6.4.1.4. Construcción de variables de violencia

Se generaron variables booleanas para identificar la presencia de violencia física, psicológica, sexual y patrimonial, mediante la búsqueda de palabras clave en la descripción del tipo de agresión. Esto permitió segmentar los casos según la naturaleza del hecho.

- **Violencia física:** identificada a partir de que en la variable TIPAGRE dijera "física".
- **Violencia psicológica:** identificada a partir de que en la variable TIPAGRE dijera "psicológica".
- **Violencia sexual:** identificada a partir de que en la variable TIPAGRE dijera "sexual".
- **Violencia patrimonial:** identificada a partir de que en la variable TIPAGRE dijera "patrimonial".

6.4.1.5. Variables de diferencia y comparación

Se crearon variables que miden la diferencia entre víctima y agresor en aspectos clave:

- **Diferencia de edad:** se calculó la diferencia absoluta y se identificó quién es mayor, esto permite sintetizar la brecha generacional si es que existe.
- **Diferencia de alfabetización:** se clasificó si ambos son iguales, o si uno supera al otro, esto permite sintetizar en caso de que exista un nivel superior de conocimiento básico.
- **Diferencia de sexo y etnia:** se identificó si existe diferencia entre víctima y agresor en estos aspectos, esto permite analizar posibles patrones en cuanto a sexo y etnia.
- **Diferencia de escolaridad:** se implementaron dos enfoques, uno agrupado y otro detallado, asignando valores numéricos a cada nivel y calculando la diferencia, donde valores negativos indican mayor escolaridad de la víctima y positivos del agresor, de forma que sintetiza el nivel educativo.
- **Diferencia de ingresos:** se construyó una variable que indica si la víctima percibe mayor salario que el agresor, esto permite sintetizar la dependencia de salario por parte de la víctima.

6.4.1.6. Variables temporales

Se desarrolló una función para calcular la diferencia entre la fecha del hecho y la fecha de registro, clasificando los casos en rangos temporales (una semana", un mes", 3 meses", 1 año", 5 años", más de 10 años", etc.), permitiendo el análisis de la oportunidad en la denuncia.

6.4.1.7. Codificación de variables categóricas

Para preparar el dataset para algoritmos de clustering categórico (**kmodes**) y análisis de clases latentes (**LCA**), se identificaron todas las variables categóricas y se aplicó **Label Encoding** a cada una, almacenando los codificadores con **joblib** para asegurar la reproducibilidad y trazabilidad del proceso. Esto es importante ya que en caso de no poseer el código se puede generar nuevamente el modelo permitiendo que sea transportable [140].

6.4.1.8. Exportación del conjunto de datos enriquecido

El **DataFrame** final, con todas las variables derivadas y codificadas, fue exportado en formato **CSV**, quedando listo para análisis exploratorio.

6.4.2. Exploración de variables

Se generaron diferentes gráficos para visualizar la distribución de las variables categóricas y numéricas, de forma que se pudieran identificar patrones, tendencias, anomalías pero también para descartar variables que no aporten información relevante al análisis.

También se generaron gráficos bivariados para identificar relaciones entre variables, utilizando gráficos de dispersión, diagramas de caja y gráficos de barras apiladas. Esto permitió identificar correlaciones potenciales y dependencias entre variables clave.

Finalmente se generaron tablas de contingencia para variables categóricas, permitiendo observar la distribución conjunta y detectar asociaciones significativas preliminarmente.

6.4.2.1. Variables descartadas desde un principio

Se descartaron las siguientes variables desde un principio al no aportar información relevante al análisis o por poseer demasiados valores nulos ignorados:

- **Discapacidades del agresor:** dado que se restringieron los datos de víctimas con discapacidad, se descartó esta variable (Fig. 39).
- **Número de otros agresores y otras víctimas:** se observó que la mayoría de casos no presentan otros agresores o víctimas.
- **Organismo jurisdiccional receptor:** se observó que la mayoría de casos son atendidos por el Ministerio Público.
- **Departamento y municipio de hecho:** se observó que la mayoría de casos presentan el mismo departamento y municipio de hecho que el de residencia de la víctima. Sin embargo, se conservaron las variables en conjunto con los servicios de salud, justicia y bomberos para análisis espaciales posteriores.
- **Variables temporales:** se observó que la mayoría de casos presentan una diferencia temporal corta entre la fecha del hecho y la fecha de registro. También el objetivo es capturar patrones atemporales por lo que se descartaron estas variables.
- **Organismo que atendió el caso:** se observó que la mayoría de casos son atendidos por el Ministerio Público.
- **Medidas de seguridad:** se observó que la mayoría de casos no presentan medidas de seguridad.
- **Nacionalidad:** se observó que la mayoría de víctimas y agresores son guatemaltecos (Fig. 16, 44).
- **Otros agresores y otras víctimas:** se observó que la mayoría de casos no presentan otros agresores o víctimas por lo que se descartó (Fig. 34, 114).
- **Artículos:** se descartaron debido a que no estaban en la mayoría de años. Tampoco se aplicaban tantos por lo que estaban vacíos en la mayoría de casos.

6.4.2.2. Alfabetismo

Se generó un gráfico de barras para visualizar la distribución del alfabetismo, de víctimas y de agresores. Esto permitió identificar que la mayoría de víctimas y agresores son alfabetizados (Fig. 11, 38) aunque existe un grupo que no lo es. Existe una tendencia a que tanto la víctima como el agresor dispongan del mismo nivel de alfabetismo. Además, se observó que dependiendo del nivel de alfabetismo sería acorde el salario devengado, siendo esta una de las variables que implican autonomía por parte de la víctima [46], [47].

6.4.2.3. Edad

Se generó un histograma para visualizar la distribución de las edades, de forma que se identificó una distribución similar en víctimas y agresores (Fig. 12, 40). Sin embargo, no se encontró algo relevante en la literatura que destacara para poder seleccionar esta como una de las variables a utilizar al momento de comparar de forma bivariada entre las edades de forma que la mayoría de casos ocurrieran con edades similares o iguales entre la víctima y el agresor (Fig. 211). Esto implicaba que los datos estaban muy sesgados a este tipo de casos lo que fue representativo para no seleccionar estas variables como parte del análisis final.

6.4.2.4. Escolaridad

Se generó un gráfico de barras para visualizar la distribución de la escolaridad, de víctimas y de agresores. Esto permitió identificar que la mayoría de víctimas y agresores poseen educación primaria o básica (Fig. 13, 41). Además, se observó que dependiendo del nivel educativo sería acorde el salario devengado, siendo esta una de las variables que implican autonomía por parte de la víctima [46], [47]. También se observaron tendencias dentro de la violencia, de forma que entre mayor nivel educativo menor probabilidad de sufrir violencia física (Fig. 2) [45].

6.4.2.5. Ocupación

Esta variable se analizó desde diferentes puntos de vista, a nivel de salario, a nivel de comparativas bivariadas, sobretodo porque la ocupación por sí sola no aportaba información relevante. Se observó que en la mayoría de casos la víctima ocupa roles menos complejos que el agresor (Fig. 213), lo que explica que devengue un salario menor que el agresor (Fig. 215). Lo que genera una tendencia de que la víctima dependa económicamente del agresor [46], [47]. Se denota el hecho de que esta variable se resume con la variable de GANAMASLAVICTIMA que sintetiza el hecho de si la víctima gana más que el agresor o no, siendo esta una variable relevante para el análisis.

6.4.2.6. Estado civil y relación víctima-agresor

Esta variable se analizó para observar la distribución de los diferentes estados civiles en víctimas y agresores. Se observó que la mayoría de víctimas y agresores están casados o en unión libre (Fig. 14, 42). Se observó que existe una brecha entre el estado civil y el salario, de forma que se observa que las víctimas devengan mucho menos que sus agresores a nivel de salario (Fig. 212) lo que presenta una falta de autonomía por parte de la mayoría de víctimas [46], [47]. Además, según literatura revisada, el estado civil influye en la gravedad de violencia [55], [54]. Se observó que en la mayoría de casos la víctima y el agresor tienen una relación de

pareja (Fig. 49), lo que implica que la mayoría de violencia ocurre en relaciones cercanas, siendo esta una variable relevante para el análisis.

6.4.2.7. Cantidad de hijos

Hay diferentes variables que representan la cantidad de hijos, tanto por sexo, como totales. Es por ello que se condensó en una sola variable que representa por rangos la cantidad de hijos 107. Esta variable es relevante ya que puede implicar limitaciones en su desarrollo profesional y personal [52].

6.4.2.8. Etnia

Se realizaron distribuciones de etnia para víctimas y agresores (Fig. 15, 43, 134, 135). Existe una minoría de víctimas y agresores que pertenecen a etnias indígenas, por lo que se utilizó la variable `VIC_ES_INDIGENA` para representar si la víctima pertenece a una etnia indígena o no, generando un grupo más significativo. Asimismo, se observaron patrones de salario (Fig. 214), escolaridad, siendo esta una variable relevante para el análisis ya que puede implicar discriminación y limitaciones en su desarrollo profesional y personal [52]. Estos patrones se observan en las tablas de contingencia (Fig. 70, 130, 141).

6.4.2.9. Tipo de violencia ejercida

Se analizaron las diferentes variables que representan el tipo de violencia ejercida, tanto de forma individual como combinada. Se observó que la mayoría de casos presentan violencia física y psicológica 28. Esta variable es relevante para el análisis ya que representan la gravedad del caso [21]. Sin embargo, se decidió utilizar las variables de presencia de un cierto tipo específico de violencia (física, psicológica, sexual y patrimonial) ya que permiten un análisis más detallado y específico.

6.5. Establecimiento de perfiles

6.5.1. Pruebas chi-cuadrado

Se realizaron pruebas chi-cuadrada entre tuplas de las variables que representaran atributos de la víctima, del agresor y del caso; determinando cuáles variables categóricas estaban correlacionadas. Se generó una tabla de contingencia entre dos variables categóricas, obtenida a partir de `pd.crosstab` de la librería de `pandas` [74]. Para garantizar la validez de la prueba chi-cuadrado, se verificó que al menos el 80 % de las frecuencias esperadas fueran mayores a 5 y ninguna menor a 1; los pares de variables que no cumplieron con este criterio fueron descartados del análisis.[141].

Posteriormente, se estableció una hipótesis nula (H_0) que suponiendo que cada variable no estaba correlacionada con su par y una hipótesis alternativa (H_1) que asume que existe la correlación utilizando un 95 % de confianza; en caso de que se rechazara la H_0 se concluía que dichas variables estaban correlacionadas [59].

Dichas pruebas permitieron filtrar las variables que no poseen relación entre sí, permitiendo reducir la cantidad de variables a analizar para el establecimiento de asociaciones. Para obtener dichas pruebas se utilizó `chi2_contingency` de la librería `scipy.stats` en Python [91], obteniendo el valor p y el estadístico

chi-cuadrado adaptado a su contexto ya que las variables difieren de grados de libertad, estos se comparan con un nivel de significancia predefinido $\alpha = 0,05$. En caso de que el valor p sea menor que el nivel de significancia, se rechaza la H_0 y se concluye que existe una asociación significativa entre las variables [59].

6.5.2. Establecimiento de asociaciones

Para generar las asociaciones entre las variables primero estas variables se filtraron en base a las correlaciones válidas encontradas en las pruebas chi-cuadrada, de forma que solo se utilicen las variables que estén correlacionadas entre sí. Asimismo, el dataset utilizado para generar las asociaciones entre pares de variables se filtró para que solo estén aquellos registros que contienen ambos valores en los pares de variables a analizar, de forma que no se generen asociaciones en base a datos faltantes o ignorados [142], lo que permitió que dependiendo del par de variables a comparar variaría el número de registros disponibles, permitiendo reglas diversas.

Finalmente, se generaron las reglas de asociación utilizando el algoritmo Apriori con la clase `Rule` del paquete de `efficient-apriori` en Python [85]. Se seleccionaron las asociaciones más relevantes a partir de identificar los patrones determinantes frecuentes, por lo que se compararon las métricas de soporte, elevación y convicción, de forma que los umbrales mínimos para las reglas eran de ($\text{lift} = 1,0$, $\text{soporte} = 0,3$ y $\text{convicción} = 1,2$ [143]), y aquellas que superaron dichos umbrales fueron agregadas al conjunto de patrones frecuentes determinantes [83].

6.5.3. Ingeniería de características para el desarrollo de perfiles

Se seleccionan las variables que modelen las asociaciones encontradas y validadas, aplicando ingeniería de características se transforman dichas variables o se mantienen para poder generar el desarrollo de perfiles de víctimas y agresores.

Las variables seleccionadas pertenecen a las reglas que pasaron los umbrales mínimos ($\text{lift} = 1,0$, $\text{soporte} = 0,3$ y $\text{convicción} = 1,2$ [143]). Dichas reglas debían de presentarse al menos el 70 % del periodo temporal ya que se buscan aquellas variables atemporales

Las variables seleccionadas tuvieron un respaldo literario para poder satisfacer que son significativas para representar características.

6.5.4. Desarrollo de perfiles

6.5.4.1. Perfiles

Para poder capturar las características que mejor reflejan la vulnerabilidad y riesgo es necesario un abordaje multifactorial del mismo, por ende, se desarrollaron 3 diferentes perfiles enfocados en capturar diferentes indicadores generales del contexto.

6.5.4.1.1. Perfil de la víctima. Era imperativo capturar la autonomía de la víctima. Se seleccionaron aquellas variables que capturaran el nivel económico, social y en general la autonomía de la víctima. Por lo que se seleccionaron las siguientes variables:

1. VIC_ALFAB: es la variable que indica si una víctima es alfabeto o no lo es, se selecciona debido a que la alfabetización y el nivel de escolaridad están asociados a la autonomía y empoderamiento de las víctimas.
2. VIC_ES_INDIGENA: es la variable que indica si una víctima es de etnia indígena o no, se seleccionó debido a que la pertenencia étnica puede influir en la respuesta a la violencia ya que pueden ser discriminados y sin acceso a recursos.
3. VIC_NIV_ESCOLARIDAD: es la variable que indica el nivel de escolaridad máximo que alcanzó una víctima, se seleccionó debido a que el nivel de escolaridad influye en la autonomía y en la capacidad de la víctima para salir de su situación.
4. VIC_TRABAJA: es la variable que indica si una víctima trabaja o no, se seleccionó debido a que la situación laboral de la víctima puede influir en su autonomía económica y en su capacidad para salir de la situación de violencia siendo esta uno de los primeros factores que contribuyan a su salida.
5. CANTIDAD_HIJOS: es la variable que indica si la víctima posee ninguno, pocos, medios o muchos hijos; se seleccionó debido a que esto recae en la capacidad de la víctima para salir de la situación de violencia, mientras más hijos posea puede ser más difícil salir de la situación dado que se le asignan roles de madre.

6.5.4.1.2. Perfil de la situación. La peligrosidad del agresor y la gravedad de los hechos de violencia cometidos son elementos centrales que se seleccionaron para generar este perfilamiento.

1. AGR_ALFAB: es la variable que indica si un agresor es alfabeto o no lo es, se selecciona debido a que se ha demostrado que la alfabetización y el nivel de escolaridad están asociados a violencia
2. AGR_ES_INDIGENA: es la variable que indica si un agresor es de etnia indígena o no, se seleccionó debido a que la pertenencia étnica puede influir en la percepción y respuesta a la violencia.
3. AGR_NIV_ESCOLARIDAD: es la variable que indica el nivel de escolaridad máximo que alcanzó un agresor, se seleccionó debido a que el nivel de escolaridad puede influir en la propensión a cometer actos de violencia.
4. AGR_TRABAJA: es la variable que indica si un agresor trabaja o no, se seleccionó debido a que la situación laboral del agresor puede influir dentro del comportamiento violento.
5. AGR_SEXO: es la variable que indica el sexo de un agresor, se seleccionó debido a que el sexo del agresor puede influir en la forma en que se comete la violencia.
6. VIOLENCIA_FISICA: es la variable que indica si un agresor ha cometido violencia física, se seleccionó debido a que la violencia física es una manifestación clara de la agresión y puede tener consecuencias graves para la víctima, además, esta denota un riesgo y una gravedad alta de la situación.

7. VIOLENCIA_PSICOLOGICA: es la variable que indica si un agresor ha cometido violencia psicológica, se seleccionó debido a que la violencia psicológica puede tener efectos profundos en la salud mental y emocional de la víctima, además, esta denota un riesgo y una gravedad alta de la situación.
8. VIOLENCIA_SEXUAL: es la variable que indica si un agresor ha cometido violencia sexual, se seleccionó debido a que la violencia sexual es una forma grave de agresión que puede tener consecuencias devastadoras para la víctima, además, esta denota un riesgo y una gravedad alta de la situación.
9. VIOLENCIA_PATRIMONIAL: es la variable que indica si un agresor ha cometido violencia patrimonial, se seleccionó debido a que la violencia patrimonial puede afectar la estabilidad económica y la seguridad de la víctima, además, esta denota un riesgo y una gravedad alta de la situación.

6.5.4.1.3. Perfil de la relación. Este perfil se concentraba en la dinámica relacional, de forma que se podría visualizar la vulnerabilidad y dependencia que existiría entre las víctimas y sus agresores.

1. VIC_EST_CIV: es la variable que indica el estado civil de la víctima, se seleccionó debido a que el estado civil de la víctima puede influir en la gravedad de violencia dentro de la relación.
2. VIC_REL_AGR: es la variable que indica la relación de la víctima con el agresor, se seleccionó debido a que la relación entre víctima y agresor puede influir en el riesgo de violencia de forma que puede ser más o menos grave.
3. AGR_EST_CIV: es la variable que indica el estado civil del agresor, se seleccionó debido a que el estado civil del agresor puede influir en el riesgo que existe para la víctima.
4. GANA_MAS_LA_VICTIMA: es la variable que indica si la víctima gana más que el agresor, se seleccionó debido a que la independencia económica de la víctima puede influir en la dinámica de poder dentro de la relación.
5. AGR_TRABAJA: es la variable que indica si un agresor trabaja o no, se seleccionó debido a que dependiendo de si el agresor trabaja o no puede influir en la dinámica de poder dentro de la relación.
6. VIC_TRABAJA: es la variable que indica si la víctima trabaja o no, se seleccionó debido a que demuestra la independencia económica de la víctima
7. DIFERENCIA_NIV_ESCOLARIDAD: es la variable que indica la diferencia de escolaridad entre víctima y agresor, se seleccionó debido a que se ha demostrado que existe a ser más vulnerable cuándo la víctima posee una mayor escolaridad que su agresor, también es medianamente más alta cuándo la víctima posee una menor escolaridad que su agresor y es más baja cuándo poseen una misma escolaridad [45]. Asimismo, esta variable depende de las variables VIC_NIV_ESCOLARIDAD y AGR_NIV_ESCOLARIDAD por lo que se comprobó a partir de las pruebas χ^2 que existe una asociación significativa entre estas variables.

6.5.4.1.4. Implementación de los perfiles. A partir del análisis de clases latentes se desarrollaron diferentes modelos que representan grupos de casos de violencia intrafamiliar. Asimismo, se evaluaron los modelos generados utilizando métricas de evaluación (AIC, BIC, NEC); seleccionando el modelo que mejor represente los datos.

Estos modelos fueron realizados utilizando la librería **Pomegranate** en Python. Se usó la distribución de probabilidad para las variables categóricas **Categorical** asumiendo la independencia entre variables [95], cada clase latente es representada por estas distribuciones categóricas.

La variable que tuviese la mayor cantidad de categorías se utilizó como parámetro para definir el número máximo de categorías debido a que la librería **Pomegranate** utiliza tensores[102].

Se utilizó la clase **GeneralMixtureModel** para crear modelos generales de mezcla dado que esta clase permite crear FMM's con cualquier tipo de variable, en este caso se le proveyeron las variables generadas previamente y se le indicó el número de clases latentes (grupos) a generar.

Se estructuró el entrenamiento y las pruebas utilizando validación cruzada, específicamente utilizando **KFold** de la librería **Scikit-Learn** de forma que se dividía la clase de entrenamiento en 10 segmentos (*folds*), entrenando con 9 y validando con 1. Se entrenó el modelo en cada iteración, obteniendo las métricas (NEC, BIC, AIC, entropía, loglik) para el entrenamiento y la validación. Estas métricas se promedian para determinar el desempeño de los modelos generalizándolos [109], [108].

Se entrenaron los modelos utilizando el algoritmo Expectation-Maximization (EM) con la función **fit** de la clase **GeneralMixtureModel**. Se utilizó el método de inicialización de por defecto de K-Means y un máximo de 10000 iteraciones para la convergencia del modelo [102].

Algunos modelos se inicializaron utilizando K-Modes (huang, random y cao) para brindar una distribución primaria de probabilidades de las clases latentes, utilizando la librería **kmodes** en Python [144]. Dichos modelos se entrenaron con la misma cantidad de clases latentes que el modelo de FMM, utilizando la misma semilla con una máxima cantidad de iteraciones de 1000. Todos los modelos fueron entrenados utilizando GPU y torch dado que la librería **Pomegranate** utiliza torch como base [102].

Se generaron LCA's (con y sin inicialización de K-Modes) con clases latentes entre 2 y 9, y para diferentes semillas (42, 73) para evitar caer en óptimos locales dado que las FMM son sensibles a ello [107], [98].

Para la evaluación de los modelos se calcularon las métricas de similitud (*likelihood*) utilizando la función de **log_probability** que permite calcular el logaritmo de la verosimilitud (*log-likelihood*) de la clase **GeneralMixtureModel** [102]. Además, a partir de la función de **predict_proba** se obtuvieron las probabilidades posteriores, se calculó la entropía posterior y a partir de ello la métrica NEC, BIC, y AIC [102], [104], [95].

Se ponderaron las métricas obtenidas (AIC, BIC, NEC) para seleccionar el mejor modelo, también se verificó que todas las clases fueran relevantes. Finalmente, se analizaron los grupos generados por el mejor modelo seleccionado, generando perfiles que describan a la mayor parte de casos de violencia intrafamiliar y corroborando con literatura las características que definen a cada perfil.

6.5.4.1.5. Medición de incertidumbre. Las clases latentes permiten obtener las probabilidades posteriores de cada una para cada registro. Sin embargo,

es imprescindible medir la incertidumbre y la variabilidad de las predicciones. Es por ello que se midió la incertidumbre de las asignaciones de cada registro a una clase latente.

Siguiendo una metodología de obtener las incertidumbres a partir de bootstrapping paramétrico a partir de generar 1000 muestras con reemplazo del conjunto de datos original (mínimo recomendado para un intervalo de confianza) [145] y reestimando los modelos para cada muestra se obtuvo la incertidumbre.

Utilizando la fórmula de entropía de Shannon [146] se obtuvo la incertidumbre para cada muestra promediando las entropías obtenidas. A partir de ello se obtuvo que la medida de separación global y los puntajes individuales mostraron una baja incertidumbre y una convergencia consistente, avalando que las clases latentes sean parte del índice de vulnerabilidad.

6.5.4.1.6. Valores de riesgo. A partir de los perfiles generados, validados con literatura, y asegurados con un intervalo de confianza robusto, se procedió a corroborar con literatura el riesgo que implica cada uno de los perfiles generados. Se asignaron valores de riesgo a cada perfil en función de la severidad y vulnerabilidad asociada a las características predominantes en cada grupo, de forma que cada grupo tendría un nivel de severidad distinto dependiendo de sus características similar a la escala EPV [57].

Estos valores de riesgo fueron asignados de manera ordinal, donde perfiles con características que indican mayor vulnerabilidad y riesgo para la víctima recibieron valores más altos. La asignación se basó en una revisión exhaustiva de estudios previos que relacionan factores sociodemográficos, económicos y contextuales con la probabilidad de sufrir violencia intrafamiliar grave.

6.5.4.1.6.1. Clases latentes del perfil de la víctima

- Clase 1: representa a la víctima con las siguientes características predominantes:
 - Víctima es alfabeta: analfabetismo
 - Víctima es indígena: indígena
 - Víctima nivel de escolaridad: ninguna escolaridad
 - Víctima trabaja: no trabaja
 - Cantidad de hijos de la víctima: una cantidad media de hijos
- Clase 2: representa a la víctima con las siguientes características predominantes:
 - Víctima es alfabeta: alfabetismo
 - Víctima es indígena No indígena
 - Víctima nivel de escolaridad: Primaria
 - Víctima trabaja: no trabaja
 - Cantidad de hijos de la víctima: una cantidad alta de hijos
- Clase 3: representa a la víctima con las siguientes características predominantes:
 - Víctima es alfabeta: alfabetismo
 - Víctima es indígena No indígena

- Víctima nivel de escolaridad: Primaria
 - Víctima trabaja: no trabaja
 - Cantidad de hijos de la víctima: una cantidad media de hijos
- Clase 4: representa a la víctima con las siguientes características predominantes:
- Víctima es alfabeto: alfabetismo
 - Víctima es indígena: No indígena
 - Víctima nivel de escolaridad: diversificado
 - Víctima trabaja: Trabaja
 - Cantidad de hijos de la víctima: una cantidad media de hijos

La clase con el riesgo más alto es la clase 1 (riesgo 4), debido a su analfabetismo, nivel de escolaridad nulo, condición indígena, desempleo y carga familiar media, son factores que aumentan significativamente su vulnerabilidad tanto de vincularse al trabajo, ser excluidos socialmente, discriminados, y tener limitadas oportunidades de desarrollo y por ende una menor capacidad de sobrellevar la situación de violencia intrafamiliar [147], [51], [50] y [49].

Siguiendo la misma lógica, la clase 2 posee características para que tenga un valor de riesgo alto, gracias a que posee alfabetismo un nivel de escolaridad algo superior (primaria), no ser indígena, sin embargo, al no poseer trabajo y tener una carga familiar alta, se le asigna un valor de riesgo alto de 3. A su vez, la clase 3 se le asigna un valor de riesgo moderado, ya que posee alfabetismo, un nivel de escolaridad de primaria, no es indígena, no posee trabajo y tiene una carga familiar baja, por lo que se le asignó un valor de riesgo moderado de 2.

Finalmente la clase con el riesgo más bajo es la clase 4 ya que a diferencia de las demás clases posee un trabajo lo que le brinda una mayor autonomía económica [32], [33], [34], además son alfabetas y tienen un nivel de escolaridad diversificado lo que refuerza este concepto de autonomía y capacidad de anteponerse a la situación. Por lo tanto se le asignó el valor más bajo de 1 en riesgo.

6.5.4.1.6.2. Clases latentes del perfil de la situación

- Clase 1: representa la situación con las siguientes características predominantes:
- Agresor es alfabeto:Alfabeto
 - Agresor es indígena:No indígena
 - Nivel de escolaridad del agresor: Primaria
 - Sexo del agresor: Masculino
 - Presencia de Violencia Física: Presente
 - Presencia de Violencia Psicológica: Presente
 - Presencia de Violencia Sexual: ausente
 - Presencia de Violencia Patrimonial: ausente
- Clase 2: representa la situación con las siguientes características predominantes:
- Agresor es alfabeto:Alfabeto

- Agresor es indígena:No indígena
 - Nivel de escolaridad del agresor: Primaria
 - Sexo del agresor: Masculino
 - Presencia de Violencia Física: Presente
 - Presencia de Violencia Psicológica Ausente
 - Presencia de Violencia Sexual Ausente
 - Presencia de Violencia Patrimonial Ausente
- Clase 3: representa la situación con las siguientes características predominantes:
- Agresor es alfabeto:Alfabeto
 - Agresor es indígena:No indígena
 - Nivel de escolaridad del agresor: diversificado
 - Sexo del agresor: Masculino
 - Presencia de Violencia Física: ausente
 - Presencia de Violencia Psicológica Presente
 - Presencia de Violencia Sexual Ausente
 - Presencia de Violencia Patrimonial Ausente
- Clase 4: representa la situación con las siguientes características predominantes:
- Agresor es alfabeto:Alfabeto
 - Agresor es indígena:Indígena
 - Nivel de escolaridad del agresor: Primaria
 - Sexo del agresor: Masculino
 - Presencia de Violencia Física: Presente
 - Presencia de Violencia Psicológica Presente
 - Presencia de Violencia Sexual Ausente
 - Presencia de Violencia Patrimonial Ausente
- Clase 5: representa la situación con las siguientes características predominantes:
- Agresor es alfabeto:Analfabeto
 - Agresor es indígena:Indígena
 - Nivel de escolaridad del agresor: ninguna escolaridad
 - Sexo del agresor: Masculino
 - Presencia de Violencia Física: Presente
 - Presencia de Violencia Psicológica Presente
 - Presencia de Violencia Sexual Ausente
 - Presencia de Violencia Patrimonial Ausente

Para poder generar las ponderaciones de riesgo dentro de esta dimensión se tuvieron que tomar en cuenta las características demográficas del agresor en conjunto con las características del hecho. En primer lugar, la menor escolaridad puede implicar una mayor probabilidad de violencia física [19]. La violencia psicológica puede ser igual de mala que la violencia física dado que puede llevar a situaciones de suicidio [20]. La combinación de ambas puede ser significar el deterioro físico y mental para la víctima por lo que la presencia de múltiples violencias incrementa significativamente el riesgo de daño sobre la víctima [21]. Basado en estos factores se le otorgó un valor de riesgo de 3 a la clase latente 1, valor de riesgo 2 a la clase latente 2, valor de riesgo 1 a la clase latente 3, valor de riesgo 4 a la clase latente 4 y valor de riesgo 5 a la clase latente 5.

6.5.4.1.6.3. Clases latentes del perfil de la relación

- Clase 1: representa a la relación con las siguientes características predominantes:
 - Relación de la víctima con el agresor: Esposos
 - Estado civil de la víctima: casados
 - Estado civil del agresor: casados
 - La víctima gana más dinero que el agresor: no
 - La víctima trabaja: no trabaja
 - El agresor trabaja: sí trabaja
- Clase 2: representa a la relación con las siguientes características predominantes:
 - Relación de la víctima con el agresor: Esposos
 - Estado civil de la víctima: casados
 - Estado civil del agresor: casados
 - La víctima gana más dinero que el agresor: no
 - La víctima trabaja: sí trabaja
 - El agresor trabaja: sí trabaja
- Clase 3: representa a la relación con las siguientes características predominantes:
 - Relación de la víctima con el agresor: Ex conyuges
 - Estado civil de la víctima: soltero
 - Estado civil del agresor: soltero
 - La víctima gana más dinero que el agresor: no
 - La víctima trabaja: no trabaja
 - El agresor trabaja: sí trabaja
- Clase 4: representa a la relación con las siguientes características predominantes:
 - Relación de la víctima con el agresor: conviviente
 - Estado civil de la víctima: unidos
 - Estado civil del agresor: unidos

- La víctima gana más dinero que el agresor: no
- La víctima trabaja: no trabaja
- El agresor trabaja: sí trabaja

Para poder asignar los valores de riesgo a cada clase latente se tuvo en cuenta la dinámica relacional, y sobretodo la literatura que existe sobre la misma. En el caso de las relaciones, aquellas incompletas (unidos, convivientes, ex conyuges) poseen un mayor riesgo de niveles de violencia [55]. La dependencia económica de la víctima incrementa su riesgo al no poseer autonomía económica y una adaptación capaz [34], [32], [33]. Por ende, se le asignó un valor de riesgo de 1 a la clase 1, un valor de riesgo de 4 a la clase 2, un riesgo de 2 a la clase 3 y un riesgo de 3 a la clase 4.

6.6. Índice de vulnerabilidad

El índice de vulnerabilidad se desarrolló con el objetivo de medir la severidad de los casos de violencia intrafamiliar en función de características relevantes de las víctimas, agresores y del contexto del hecho. Este índice busca sintetizar múltiples dimensiones de vulnerabilidad en una única métrica cuantitativa que permita clasificar y priorizar los casos según su gravedad.

Se procedió a calcular el índice de vulnerabilidad como una combinación ponderada de cuatro indicadores principales compuesto por los perfiles generados para cada aspecto de la situación de violencia intrafamiliar, multiplicado por pesos y los valores de riesgo intrínsecos de cada perfil. De forma que la fórmula final del índice de vulnerabilidad es la siguiente:

$$IV = w_1 * Riesgo_v * I_v + w_2 * Riesgo_a * I_a + w_3 * Riesgo_r * I_r + w_4 * Riesgo_i * I_i$$

donde IV es el índice de vulnerabilidad, k es la dimensión de vulnerabilidad a evaluar a través de los indicadores, w_k son los pesos asignados a cada indicador, $Riesgo_k$ son los valores de riesgo asignados a cada clase y I_k son los indicadores. Asimismo, se denota que v corresponde al perfil de la víctima, a al perfil de la agresividad, r al perfil de la relación e i al indicador de presencia de apoyo institucional.

6.6.1. Indicadores del índice de vulnerabilidad

El índice de vulnerabilidad se compone de cuatro indicadores principales que capturan diferentes dimensiones de riesgo y vulnerabilidad en los casos de violencia intrafamiliar:

6.6.1.0.1. Indicador de riesgo de la víctima. Este indicador evalúa las características individuales de la víctima que pueden aumentar su vulnerabilidad ante la violencia intrafamiliar. De forma que a partir del .

6.6.1.0.2. Indicador de riesgo del agresor. Este indicador mide las características del agresor que pueden influir en la peligrosidad y recurrencia de la violencia cometida. Se consideran factores como el historial de violencia, el nivel de escolaridad, el empleo y otros aspectos relevantes.

6.6.1.0.3. Indicador de riesgo de la relación. Este indicador analiza la dinámica relacional entre la víctima y el agresor, incluyendo aspectos como la dependencia económica, el estado civil, la relación familiar y otros factores que pueden influir en la vulnerabilidad de la víctima dentro del contexto relacional.

6.6.1.0.4. Indicador de presencia de apoyo institucional. Este indicador evalúa la existencia y efectividad de redes de apoyo institucional que pueden ayudar a la víctima en situaciones de violencia intrafamiliar. Se consideran factores como la disponibilidad de servicios de atención y la accesibilidad a recursos legales y de protección.

6.6.1.0.4.1. Variables seleccionadas En primer lugar para evaluar el factor de presencia policial se seleccionaron las siguientes variables:

1. CANTIDAD_COMISARIAS
2. CANTIDAD_ESTACIONES
3. CANTIDAD_SUBESTACIONES

Asimismo se seleccionaron las siguientes variables para recopilar el apoyo de sistema de justicia:

1. CANTIDAD_JUZGADOS_VIOLENCIA_CONTRA_MUJ_FEMINICIDIO
2. CANTIDAD_TRIBUNALES_SENTENCIA_VIOLENCIA_CONTRA_MUJ_FEMINICIDIO

También se seleccionaron las siguientes variables para extraer el apoyo del ministerio público:

1. CANTIDAD_FISCALIAS_MUNICIPALES
2. CANTIDAD_AGENCIAS_FISCALES

Finalmente se seleccionaron las siguientes variables para evaluar el apoyo de la salud pública y de asistencia social:

1. CANTIDAD_CENTRO_ATENCION_PERMANENTE
2. CANTIDAD_CENTRO_SALUD
3. CANTIDAD_PUESTO_SALUD
4. CANTIDAD_HOSPITAL

6.6.1.0.5. Estandarización por tasa. Cada variable fue estandarizada por tasa por cada proyección estimada del año 2021 [135] ya que los datos de infraestructura extraídos son de ese año.

6.6.1.0.6. Normalización. Cada variable fue normalizada utilizando la normalización Min-Max para escalar los valores entre 0 y 1, facilitando la comparación entre diferentes indicadores. La fórmula utilizada fue:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

donde X es el valor original, X_{min} y X_{max} son los valores mínimo y máximo de la variable respectivamente.

6.6.1.0.7. Agregación. Cada indicador fue calculado como el promedio simple de scores normalizados de todos los subgrupos, generando el indicador de presencia de apoyo institucional. La fórmula utilizada fue:

$$I_{apoyo} = \frac{1}{K} \sum_{k=1}^K N_k$$

donde I_{apoyo} es el indicador de apoyo institucional, N_k son los puntajes normalizados de cada variable y K es el número total de variables consideradas.

6.6.1.0.8. Cálculo de la incertidumbre. Se utilizó el muestreo bootstrap para estimar la incertidumbre del indicador de presencia de apoyo institucional. Se generaron 1000 muestras con reemplazo del conjunto de datos original y se recalculó el indicador para cada muestra. La desviación estándar de los indicadores obtenidos se utilizó como medida de incertidumbre. Asimismo, se utilizó bootstrap dado que son proyecciones de poblaciones y los conteos de infraestructura pueden variar.

6.6.2. Pesaje de los indicadores

Para el pesaje de los indicadores que componen el índice de vulnerabilidad se utilizó Análisis de Componentes Principales (PCA) con el objetivo de determinar la contribución relativa de cada indicador en la variabilidad total del índice. El PCA al permitir identificar las dimensiones subyacentes que explican la mayor parte de la varianza en los datos, permitió una asignación de pesos basados en la importancia estadística de cada indicador.

Se estandarizaron los indicadores utilizando la normalización Z-score para asegurar que todos los indicadores tuvieran la misma escala y evitar que alguno dominara el análisis debido a su magnitud. Posteriormente, se aplicó PCA utilizando la librería `scikit-learn` en Python [108], obteniendo los componentes principales y sus respectivas varianzas explicadas. Los pesos de cada indicador se derivaron de las cargas de los componentes principales, asignando mayor peso a aquellos indicadores que contribuyen más significativamente a la varianza total.

Esto aplicaría siempre y cuando uno de los componentes principales explique una proporción significativa de la varianza total (por ejemplo, más del 70 %). Si no es así, se consideraría utilizar un peso igualitario para cada una de las variables.

7.1. Análisis exploratorio de datos

El análisis exploratorio de datos (EDA, por sus siglas en inglés) se realizó con el objetivo de comprender mejor las características y patrones presentes en el conjunto de datos de violencia intrafamiliar. Se llevaron a cabo múltiples tablas de contingencia que posteriormente serían utilizadas para identificar aquellas relaciones válidas a partir de las pruebas χ^2 de independencia, sin embargo, al ser demasiadas se colocó en la sección de anexos (Cuadros 70 hasta 156).

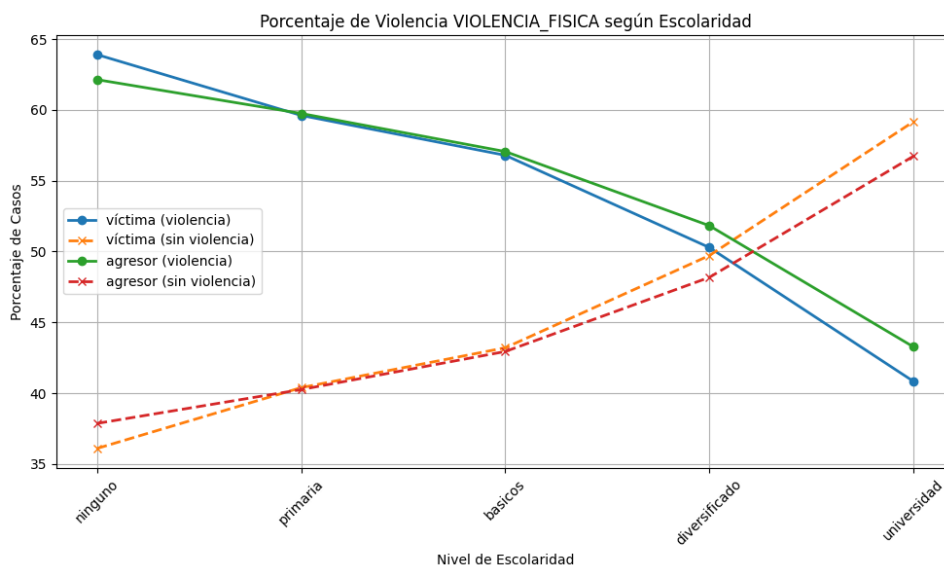


Figura 2: Evolución de la violencia física según el nivel de escolaridad
Fuente: Elaboración propia.

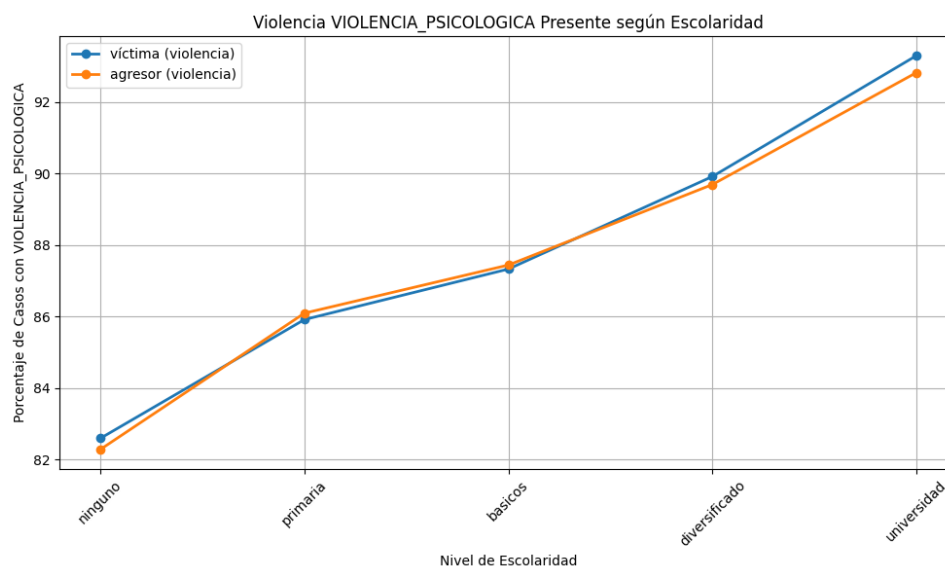


Figura 3: Evolución de la violencia psicológica presente según el nivel de escolaridad

Fuente: Elaboración propia.

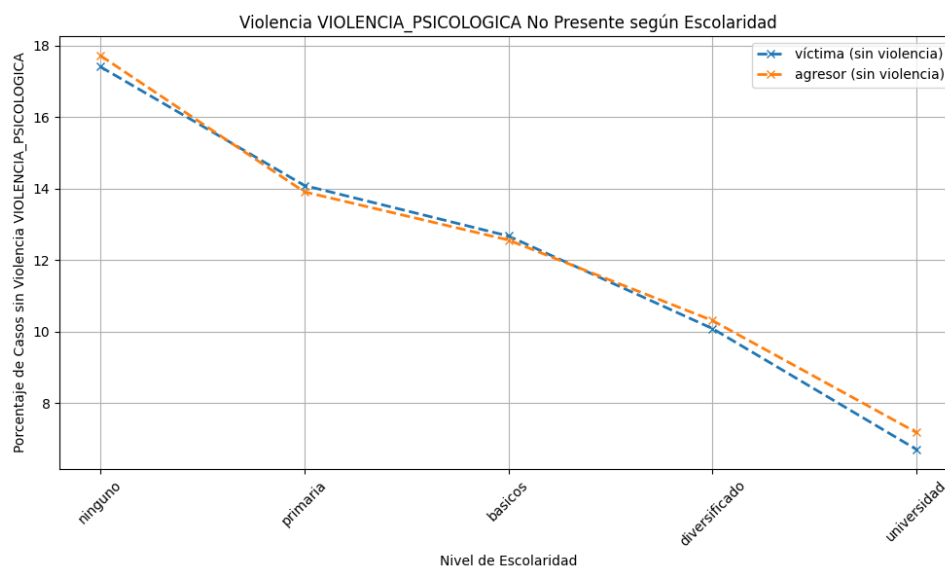


Figura 4: Evolución de la violencia psicológica no presente según el nivel de escolaridad

Fuente: Elaboración propia.

7.2. Apriori y pruebas chi-cuadrado de independencia entre variables

Se obtuvo la cantidad de comparaciones válidas entre variables tras aplicar la prueba de independencia de chi-cuadrado con un nivel de significación de $\alpha = 0,05$. En el Cuadro 9 se muestran los resultados obtenidos, donde se consideran válidas aquellas tablas de contingencia que cumplan al menos con un 80 % de las frecuencias esperadas mayores que cinco y ninguna menor que uno.

Cuadro 9: Comparaciones válidas entre variables tras aplicar la prueba de independencia chi-cuadrado ($\alpha = 0,05$)

Concepto	Cantidad	Porcentaje
Comparaciones evaluadas	1,406	100 %
Comparaciones válidas	170	12,09 %

* Se observa que del total de comparaciones únicamente el 12.09 % son válidas, lo que implica una disminución en la cantidad de variables a evaluar para las pruebas apriori.

Por consiguiente, al obtener las reglas significativas de asociaciones entre variables mediante el algoritmo apriori, se filtraron aquellas reglas que cumplieran con las comparaciones válidas obtenidas previamente, los umbrales mínimos, que tuvieran las variables seleccionadas para el análisis, y que estuvieran enfocados en los casos donde la víctima eran mujeres. En el Cuadro 10 se muestran las principales reglas de asociación, así como sus respectivas métricas asociadas.

Cuadro 10: Tabla de reglas de asociación filtradas

Variable Antecedente	Variable Consecuente	Cantidad de registros	χ^2	valor-p	Mínimo esperado	Soporte	Confianza	Elevación	Convicción	Repeticiones en años
AGR_ALFAB = alfabeto	VIC_ALFAB = alfabeto	19712.125000	32.10.6535.17	0.000000	509.400469	0.763443	0.875463	1.074464	1.487322	16
AGR_ES_INDIGENA = no	AGR_ALFAB = alfabeto	17349.187500	710.42.7884	0.000000	817.885661	0.593259	0.922576	1.067308	1.667783	16
AGR_ES_INDIGENA = no	GANA_MAS_LA_VICTIMA = si	17591.000000	138.732.198	0.000000	817.885661	0.593259	0.922576	1.067308	1.667783	16
AGR_ES_INDIGENA = no	VIC_ALFAB = alfabeto	17569.375000	3366.164641	0.000000	12176.19512	0.74313	0.894725	1.099935	1.800289	16
AGR_ES_INDIGENA = no	VIC_ES_INDIGENA = no	17274.562500	12855.723176	0.000000	2209.906116	0.612580	0.951478	1.478299	8.004850	16
AGR_ES_INDIGENA = si	VIC_ES_INDIGENA = si	17098.928571	12715.716287	0.000000	2261.447201	0.328744	0.919051	2.529946	7.905984	14
AGR_NIV_ESCOLARIDAD = primaria	AGR_ALFAB = alfabeto	19184.562500	12501.451725	0.000000	74.496808	0.444720	1.000000	1.153978	13269.0020.196715	16
AGR_NIV_ESCOLARIDAD = primaria	GANA_MAS_LA_VICTIMA = no	18364.083333	35.12.6724	0.000000	79.948921	0.384634	0.898006	1.035950	1.309608	12
AGR_NIV_ESCOLARIDAD = primaria	VIC_TRABAJA = no	18192.600000	1126.691901	0.000000	178.961432	0.327983	0.752829	1.007301	1.272433	10
AGR_SEXO = hombres	AGR_TRABAJA = si	19554.812500	3906.393481	0.000000	148.067772	0.857699	0.909543	1.043828	1.407210	16
AGR_TRABAJA = si	AGR_SEXO = hombres	19554.812500	3906.393481	0.000000	148.067772	0.857699	0.909543	1.043828	3.483109	16
AGR_TRABAJA = si	GANA_MAS_LA_VICTIMA = no	18809.857143	1132.924942	0.000000	329.627941	0.782066	0.902726	1.037712	1.335440	14
GANA_MAS_LA_VICTIMA = no	AGR_TRABAJA = si	18809.857143	1132.924942	0.000000	329.627941	0.782066	0.909718	1.037712	1.322725	14
GANA_MAS_LA_VICTIMA = no	VIC_TRABAJA = no	20136.312500	3711.632976	0.000000	762.327085	0.667030	0.759134	1.399031	1.308743	16
VIC_ALFAB = alfabeto	AGR_ALFAB = alfabeto	19712.125000	32.10.6535.17	0.000000	509.400469	0.763443	0.875463	1.074464	1.487322	16
VIC_ALFAB = alfabeto	VIC_ES_INDIGENA = no	18486.818182	1744.906247	0.000000	1377653998	0.573078	0.719160	1.115599	1.264585	16
VIC_ES_INDIGENA = no	AGR_ALFAB = alfabeto	17453.187500	734.464701	0.000000	803.182691	0.599234	0.922796	1.067129	1.668880	16
VIC_ES_INDIGENA = no	AGR_ES_INDIGENA = no	17274.562500	12855.723176	0.000000	2209.906116	0.612580	0.951478	1.478299	8.004850	16
VIC_ES_INDIGENA = si	AGR_ES_INDIGENA = si	17098.928571	12715.716287	0.000000	2261.447201	0.328744	0.919051	2.529946	7.905984	14
VIC_ES_INDIGENA = si	GANA_MAS_LA_VICTIMA = no	17712.230769	152.018047	0.000000	72.957607	0.329733	0.919122	1.045960	1.500867	13
VIC_ES_INDIGENA = no	VIC_ALFAB = alfabeto	17702.062500	1389.332416	0.000000	1194.155338	0.383206	0.960127	1.054535	1.892122	16
VIC_NIV_ESCOLARIDAD = primaria	AGR_ALFAB = alfabeto	19907.714286	3224.676735	0.000000	82.537874	0.373590	0.908107	1.047133	1.431285	14
VIC_NIV_ESCOLARIDAD = primaria	GANA_MAS_LA_VICTIMA = no	19710.937500	1315.292425	0.000000	83.296406	0.375129	0.921830	1.049941	1.558846	16
VIC_NIV_ESCOLARIDAD = primaria	VIC_ALFAB = alfabeto	19710.937500	16866.317709	0.000000	121534684	0.407187	1.000000	1.238231	190048474.424213	16
VIC_NIV_ESCOLARIDAD = primaria	VIC_TRABAJA = no	18676.200000	2350.343219	0.000000	209.550589	0.313774	0.780886	1.135437	1.433358	10
VIC_TRABAJA = no	GANA_MAS_LA_VICTIMA = no	20136.312500	3711.632976	0.000000	762.327085	0.667030	0.759134	1.399031	5.404283	16
VIOLENCIA_FISICA = presente	AGR_SEXO = hombres	20175.071429	114.13.7952	0.000000	407.498444	0.591243	0.960693	1.013382	1.345131	14
VIOLENCIA_FISICA = no presente	VIOLENCIA_PSICOLOGICA = presente	19776.230769	1397.097437	0.000000	1053.327059	0.418199	0.970876	1.115838	4.686342	13
VIOLENCIA_FISICA = no presente	VIOLENCIA_SEXUAL = no presente	20143.000000	38.793067	0.000813	177.731907	0.415736	0.985432	1.089053	1.521597	14
VIOLENCIA_PATRIMONIAL = no presente	VIOLENCIA_SEXUAL = no presente	20331.181818	135.240414	0.000000	78.210477	0.841654	0.979232	1.005404	1.247856	11

*Las reglas asociadas demuestran tendencias atemporales entre las variables analizadas al presentarse al menos 10 veces dentro de los años. Además, las mismas demuestran correlaciones fuertes al presentar métricas de confianza mayores al 0.7 y demostrar su valía aplicando estadísticos χ^2 y valor-p demostrando su validez.

7.3. Análisis de clases latentes (LCA)

A partir de las pruebas de χ^2 de independencia, las reglas apriori obtenidas y la revisión de la literatura, se seleccionaron las variables categóricas para cada uno de los tres modelos de análisis de clases latentes (LCA, según sus siglas en inglés) realizados: modelo de la víctima, modelo del agresor y modelo de la relación entre víctima y agresor.

7.3.1. Clases latentes del modelo de la víctima

A partir de la selección de las variables más relevantes para modelar el perfil de la víctima, se procedió a entrenar y evaluar modelos *Finite Mixture Models* para modelar análisis de clases latentes (LCA por sus siglas en inglés). Los resultados obtenidos se presentan en el Cuadro 158 del cuál se seleccionó el mejor modelo con el que se calcularon las diversas probabilidades posteriores promedio para cada

una de las variables categóricas utilizadas en el modelo del perfil de la víctima (Cuadros 12 a 16) además de calcular la variabilidad del modelo (Cuadro 34).

Cuadro 11: Resultados del entrenamiento y evaluación de modelos para el perfil de la víctima

Conjunto	Cantidad de Clases	Semilla	Inicialización	Log-Likelihood	Entropía	NEC	AIC	BIC	Probabilidades Posteriores
entrenamiento	4	73	no kmodes	-945908.687500	0.012445	2.085100	1891857.375000	1892065.000000	0.993206
evaluación	4	73	no kmodes	-104695.093800	0.012147	1.067300	809430.187500	809620.812500	0.993399

*Resultados del modelo seleccionado con mejor balance de métricas entre entrenamiento y evaluación.

A partir del modelo usado a partir del Cuadro 11, se presentan las probabilidades posteriores promedio para cada una de las variables categóricas utilizadas en el modelo del perfilamiento de la víctima, las cuales se pueden observar en los Cuadros 12 a 16.

Cuadro 12: Probabilidades posteriores de alfabetización de la víctima por clase latente

Condición	Clase 1	Clase 2	Clase 3	Clase 4
Alfabeta	0.037	1.000	1.000	1.000
Analfabeto	0.963	0.000	0.000	0.000

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable de la alfabetización de la víctima según la clase latente a la cuál pertenece. De forma que para la clase 1 es analfabeto, y en las demás clases es alfabeto.

Cuadro 13: Probabilidades posteriores de si es indígena o no de la víctima por clase latente

Condición Étnica	Clase 1	Clase 2	Clase 3	Clase 4
No indígena	0.345	0.614	0.632	0.800
Indígena	0.655	0.386	0.368	0.200

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable de la condición étnica de la víctima según la clase latente a la cuál pertenece. De forma que para la clase 1 es indígena, y para la clase 2, 3 y 4 es no indígena.

Cuadro 14: Probabilidades posteriores del nivel de escolaridad de la víctima por clase latente

Nivel Escolaridad	Clase 1	Clase 2	Clase 3	Clase 4
Básicos	0.000	0.000	0.000	0.393
Diversificado	0.000	0.000	0.000	0.515
Ninguno	1.000	0.033	0.030	0.000
Primaria	0.000	0.967	0.970	0.000
Universidad	0.000	0.000	0.000	0.092

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable del nivel de escolaridad de la víctima según la clase latente a la cuál pertenece. De forma que para la clase 1 no posee escolaridad, para la clase 2 y 3 posee primaria, y para la clase 4 posee básicos o diversificado.

Cuadro 15: Probabilidades posteriores de si trabaja la víctima por clase latente

Condición Laboral	Clase 1	Clase 2	Clase 3	Clase 4
No trabaja	0.818	0.737	0.757	0.509
Sí trabaja	0.182	0.262	0.243	0.491

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable del nivel de escolaridad de la víctima según la clase latente a la cuál pertenece. De forma que para la clase 1 no posee escolaridad, para la clase 2 y 3 posee primaria, y para la clase 4 posee básicos o diversificado.

Cuadro 16: Probabilidades posteriores de la cantidad de hijos de la víctima por clase latente

Cantidad de Hijos	Clase 1	Clase 2	Clase 3	Clase 4
Demasiados hijos	0.081	0.000	0.034	0.003
Hijo único	0.161	0.000	0.325	0.339
Hijos medios	0.371	0.000	0.641	0.474
Muchos hijos	0.268	0.606	0.000	0.067
Sin hijos	0.120	0.394	0.000	0.117

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable de la cantidad de hijos de la víctima según la clase latente a la cuál pertenece. De forma que para la clase 1 posee entre una cantidad media de hijos y muchos hijos, la clase 2 posee muchos hijos, la clase 3 posee entre un hijo único y una cantidad media de hijos, y la clase 4 posee entre un hijo único y una cantidad media de hijos.

En el Cuadro 17 se presenta la variabilidad del modelo de clases latentes del perfil de la víctima, obtenida mediante bootstrapping paramétrico con 1000 réplicas.

Cuadro 17: Variabilidad del modelo de clases latentes de la víctima mediante bootstrapping paramétrico

Promedio	Sesgo	Error estándar	Intervalo de confianza	Número de réplicas válidas
0.9910	$1,03 \times 10^{-5}$	$1,66 \times 10^{-4}$	[0.9907, 0.9913]	1000

*Presenta la variabilidad del modelo de clases latentes de la víctima obtenida mediante bootstrapping paramétrico, mostrando el promedio, sesgo, error estándar, intervalo de confianza y número de réplicas válidas.

7.3.2. Clases latentes del modelo de la situación

A partir de la selección de las variables más relevantes para modelar el perfil de la situación, se procedió a entrenar y evaluar modelos *Finite Mixture Models* para modelar análisis de clases latentes (LCA por sus siglas en inglés). Los resultados obtenidos se presentan en el Cuadro 159 del cuál se seleccionó el mejor modelo con el que se calcularon las diversas probabilidades posteriores promedio para cada una de las variables categóricas utilizadas en el modelo del perfil de la víctima (Cuadros 19 a 26) además de calcular la variabilidad del modelo (Cuadro 27).

Cuadro 18: Resultados del entrenamiento y evaluación de modelos para el perfil de la situación

Conjunto	Cantidad de Clases	Semilla	Inicialización	Log-Likelihood	Entropía	NEC	AIC	BIC	Probabilidades Posteriores
entrenamiento	5	73	cao	-786292.142400	0.109616	1.301000	1572624.284800	1572834.913300	0.953335
evaluación	5	73	cao	-336228.408300	0.107941	8.350300	672496.816600	672690.499200	0.954205

*Resultados del modelo seleccionado con mejor balance de métricas entre entrenamiento y evaluación.

Cuadro 19: Probabilidades posteriores de alfabetización del agresor por clase latente

Condición	Clase 1	Clase 2	Clase 3	Clase 4	Clase 5
alfabeta	1.000	1.000	1.000	1.000	0.000
analfabeta	0.000	0.000	0.000	0.000	1.000

Cuadro 20: Probabilidades posteriores de si es indígena o no del agresor por clase latente

Condición	Clase 1	Clase 2	Clase 3	Clase 4	Clase 5
No indígena	1.000	0.615	0.784	0.000	0.391
Indígena	0.000	0.385	0.216	1.000	0.609

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable de si el agresor es indígena según la clase latente a la cuál pertenece. De forma que la clase 4 y 5 el agresor es indígena, mientras que en las clases 1, 2, y 3 el agresor es no indígena.

Cuadro 21: Probabilidades posteriores de la escolaridad del agresor por clase latente

Condición	Clase 1	Clase 2	Clase 3	Clase 4	Clase 5
basicos	0.292	0.198	0.000	0.205	0.000
diversificado	0.000	0.178	1.000	0.000	0.000
ninguno	0.031	0.043	0.000	0.048	1.000
primaria	0.614	0.561	0.000	0.728	0.000
universidad	0.063	0.021	0.000	0.020	0.000

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable de escolaridad del agresor según la clase latente a la cuál pertenece. De forma que la clase 1, 2 y 4 el agresor posee una escolaridad de primaria, mientras que para la clase 3 posee una escolaridad de diversificado, y para la clase 4 no posee ninguna escolaridad.

Cuadro 22: Probabilidades posteriores del sexo del agresor por clase latente

Condición	Clase 1	Clase 2	Clase 3	Clase 4	Clase 5
hombres	0.944	0.968	0.958	0.956	0.891
mujeres	0.056	0.032	0.042	0.044	0.109

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable del sexo del agresor según la clase latente a la cuál pertenece. De forma que dentro de todas las clases se observa la tendencia a que predominen los hombres como agresores.

Cuadro 23: Probabilidades posteriores de la violencia física ejercida por el agresor por clase latente

Condición	Clase 1	Clase 2	Clase 3	Clase 4	Clase 5
no presente	0.467	0.109	0.500	0.383	0.338
presente	0.533	0.891	0.500	0.617	0.662

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable de presencia o no de violencia física ejercida por el agresor según la clase latente a la cuál pertenece. De forma que dentro de todas las clases se observa la tendencia a que esté presente la violencia física.

Cuadro 24: Probabilidades posteriores de la violencia psicológica ejercida por el agresor por clase latente

Condición	Clase 1	Clase 2	Clase 3	Clase 4	Clase 5
no presente	0.000	1.000	0.000	0.000	0.176
presente	1.000	0.000	1.000	1.000	0.824

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable de presencia o no de violencia psicológica ejercida por el agresor según la clase latente a la cuál pertenece. De forma que exceptuando la clase 2, en las demás clases se observa la tendencia a que esté presente la violencia psicológica.

Cuadro 25: Probabilidades posteriores de la violencia sexual ejercida por el agresor por clase latente

Condición	Clase 1	Clase 2	Clase 3	Clase 4	Clase 5
no presente	0.974	0.972	0.975	0.977	0.978
presente	0.026	0.028	0.025	0.023	0.022

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable de presencia o no de violencia sexual ejercida por el agresor según la clase latente a la cuál pertenece. De forma no hay tendencia a que se presente la violencia sexual en alguna clase latente.

Cuadro 26: Probabilidades posteriores de la violencia patrimonial ejercida por el agresor por clase latente

Condición	Clase 1	Clase 2	Clase 3	Clase 4	Clase 5
no presente	0.852	0.878	0.846	0.851	0.865
presente	0.148	0.122	0.154	0.149	0.135

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable de presencia o no de violencia patrimonial ejercida por el agresor según la clase latente a la cuál pertenece. De forma no hay tendencia a que se presente la violencia patrimonial en alguna clase latente.

En el Cuadro 27 se presenta la variabilidad del modelo de clases latentes del perfil de la situación, obtenida mediante bootstrapping paramétrico con 1000 réplicas.

Cuadro 27: Variabilidad del modelo de clases latentes de la situación mediante bootstrapping paramétrico

Promedio	Sesgo	Error estándar	Intervalo de confianza	Número de réplicas válidas
0.9999	$4,081 \times 10^{-2}$	$3,324 \times 10^{-8}$	[0.9999, 0.9999]	1000

*Presenta la variabilidad del modelo de clases latentes de la víctima obtenida mediante bootstrapping paramétrico, mostrando el promedio, sesgo, error estándar, intervalo de confianza y número de réplicas válidas.

7.3.3. Clases latentes del modelo de la relación

A partir de la selección de las variables más relevantes para modelar la relación entre el agresor y la víctima, se procedió a entrenar y evaluar modelos *Finite Mixture Models* para modelar análisis de clases latentes (LCA por sus siglas en inglés). Los resultados obtenidos se presentan en el Cuadro 160 del cuál se seleccionó el mejor modelo con el que se calcularon las diversas probabilidades posteriores promedio para cada una de las variables categóricas utilizadas en el modelo de la relación entre víctima y agresor (Cuadros 29 a 33) además de calcular la variabilidad del modelo (Cuadro 34).

Cuadro 28: Resultados del entrenamiento y evaluación de modelos para el perfil de la relación entre víctima y agresor

Conjunto	Cantidad de Clases	Semilla	Inicialización	Log-Likelihood	Entropía	NEC	AIC	BIC	Probabilidades Posteriores
entrenamiento	4	42	cao	-961539.242062	0.033869	0.000000	1923118.484124	1923328.568816	0.980269
evaluación	4	42	cao	-412122.798400	0.033664	0.000000	824285.506799	824478.735384	0.980271

*Resultados del modelo seleccionado con mejor balance de métricas entre entrenamiento y evaluación.

A partir del modelo seleccionado usando de el Cuadro 28, se presentan las probabilidades posteriores promedio para cada una de las variables categóricas utilizadas en el modelo de la relación entre víctima y agresor, las cuales se pueden observar en las Cuadros 29 a 33.

Cuadro 29: Probabilidades posteriores de la relación entre la víctima con el agresor por clase latente

Condición	Clase 1	Clase 2	Clase 3	Clase 4
conviviente	0.016	0.238	0.015	0.852
esposos(as)	0.906	0.015	0.909	0.021
ex-conyuges	0.016	0.250	0.033	0.086
hermanos(as)	0.014	0.130	0.011	0.010
hijastros(as)	0.000	0.005	0.000	0.001
hijos(as)	0.002	0.041	0.002	0.002
nietos(as)	0.000	0.002	0.000	0.000
otro pariente	0.036	0.217	0.025	0.021
padres/madres	0.004	0.082	0.002	0.002
suegros(as)	0.005	0.018	0.002	0.003

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable de la relación entre la víctima y el agresor según la clase latente a la cuál pertenece. De forma que para la clase 1 la mayoría de agresores son esposos(as), para la clase 2 también son esposos(as), para la clase 3 son en su mayoría ex-conyuges y para la clase 4 también son en su mayoría convivientes.

Cuadro 30: Probabilidades posteriores del estado civil de la víctima por clase latente

Condición	Clase 1	Clase 2	Clase 3	Clase 4
casados(as)	0.994	0.191	0.995	0.006
otro	0.000	0.083	0.000	0.000
solteros(as)	0.003	0.610	0.003	0.000
unidos(as)	0.003	0.099	0.002	0.994
viudos(as)	0.000	0.017	0.000	0.000

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable del estado civil de la víctima según la clase latente a la cuál pertenece. De forma que para la clase 1 son casados, para la clase 2 son solteros, para la clase 3 son casados y para la clase 4 están unidos.

Cuadro 31: Probabilidades posteriores del estado civil del agresor por clase latente

Condición	Clase 1	Clase 2	Clase 3	Clase 4
casados(as)	1.000	0.176	1.000	0.018
otro	0.000	0.087	0.000	0.000
solteros(as)	0.000	0.601	0.000	0.000
unidos(as)	0.000	0.128	0.000	0.982
viudos(as)	0.000	0.009	0.000	0.000

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable del estado civil del agresor según la clase latente a la cuál pertenece. De forma que para la clase 1 son casados, para la clase 2 son solteros, para la clase 3 son casados y para la clase 4 están unidos.

Cuadro 32: Probabilidades posteriores de si la víctima trabaja por clase latente

Condición	Clase 1	Clase 2	Clase 3	Clase 4
no	1.000	0.676	0.000	0.749
si	0.000	0.324	1.000	0.251

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable de si la víctima trabaja según la clase latente a la cuál pertenece. De forma que para la clase 1 no trabaja, para la clase 2 no trabaja, para la clase 3 no trabaja y para la clase 4 no trabaja.

Cuadro 33: Probabilidades posteriores de si el agresor trabaja por clase latente

Condición	Clase 1	Clase 2	Clase 3	Clase 4
no	0.073	0.234	0.094	0.083
si	0.927	0.766	0.906	0.917

*Presenta las probabilidades posteriores promedio respecto al modelo *Finite Mixture Model*, sobre la variable de si el agresor trabaja según la clase latente a la cuál pertenece. De forma que para la clase 1 trabaja, para la clase 2 trabaja, para la clase 3 trabaja y para la clase 4 trabaja.

En la tabla 34 se presenta la variabilidad del modelo de clases latentes de la relación entre víctima y agresor, obtenida mediante bootstrapping paramétrico con 1000 réplicas.

Cuadro 34: Variabilidad del modelo de clases latentes de la relación entre el agresor y la víctima mediante bootstrapping paramétrico

Promedio	Sesgo	Error estándar	Intervalo de confianza	Número de réplicas válidas
0.9644	$3,282 \times 10^{-3}$	$3,08 \times 10^{-4}$	[0.9638, 0.9650]	1000

*La tabla presenta métricas de variabilidad del modelo, con poca variabilidad observada entre las réplicas generadas. Asimismo, el sesgo es pequeño, al igual que el error estándar y un intervalo de confianza estrecho.

Cuadro 35: Riesgo relativo de la víctima según la clase latente

Clase	Riesgo Relativo Perfil de la Víctima	Riesgo Relativo Perfil de la Situación	Riesgo Relativo Perfil de la Relación
Clase 1	4	3	2
Clase 2	3	1	4
Clase 3	2	2	1
Clase 4	1	4	3
Clase 5	-	5	-

*Presenta el riesgo relativo asignado a cada clase latente del modelo de la víctima, siendo 1 el riesgo más bajo y 5 el riesgo más alto. Asimismo, este siendo justificado bajo la literatura revisada utilizando las características predominantes de cada clase latente.

7.4. Índice de vulnerabilidad estructural en violencia intrafamiliar

La construcción del índice de vulnerabilidad estructural en violencia intrafamiliar se realizó utilizando un análisis de componentes principales (PCA) para determinar los pesos de cada una de las variables incluidas en el índice. Además, se realizó la construcción de un indicador de apoyo institucional para determinar el apoyo que tiene a su disposición cada caso permitiendo evaluar otra dimensión de la vulnerabilidad.

7.4.1. Indicador de apoyo institucional

A continuación, en el Cuadro 36 se presentan los resultados del indicador de apoyo institucional por departamento, en conjunto con sus respectivos intervalos

de confianza e incertidumbre, obtenidos mediante bootstrapping paramétrico con 1000 réplicas, esto a través de los datos obtenidos a partir del Instituto Nacional de Estadística (I.N.E.) y el Ministerio de Salud Pública y Asistencia Social (M.S.P.A.S.).

Cuadro 36: Indicador de apoyo institucional por departamento

Departamento	Indicador de apoyo medio	I.C. Inferior	I.C. Superior	Amplitud de Incertidumbre
guatemala	0.2425	0.2096	0.3363	0.1266
quetzaltenango	0.4248	0.3683	0.5484	0.1801
huehuetenango	0.5120	0.4513	0.6361	0.1848
jalapa	0.0843	0.0638	0.1373	0.0735
escuintla	0.2720	0.2280	0.3505	0.1226
el progreso	0.5419	0.5115	0.6035	0.0920
sacatepequez	0.2276	0.1941	0.3328	0.1387
chimaltenango	0.3063	0.2629	0.3822	0.1193
santa rosa	0.3493	0.2973	0.4916	0.1943
solola	0.6115	0.5544	0.7947	0.2402
totonicapan	0.1080	0.0745	0.1568	0.0823
suchitepequez	0.4846	0.4258	0.6335	0.2077
retalhuleu	0.2651	0.2187	0.3751	0.1564
san marcos	0.4868	0.4107	0.6221	0.2114
alta verapaz	0.2119	0.1727	0.2868	0.1141
quiche	0.3641	0.3015	0.4661	0.1646
baja verapaz	0.3008	0.2574	0.3934	0.1360
peten	0.2018	0.1701	0.2627	0.0926
izabal	0.2239	0.1780	0.3156	0.1376
zacapa	0.6834	0.6335	0.7453	0.1118
chiquimula	0.2755	0.2347	0.3609	0.1263
jutiapa	0.1550	0.1286	0.2027	0.0741

*Se presentan los valores del indicador de apoyo institucional para cada departamento al haber aplicado la estandarización por tasa y normalizado utilizando Min-Max, junto con sus respectivos intervalos de confianza e incertidumbre. De forma que existe una mayor variabilidad dependiendo del departamento, siendo Zacapa el departamento con mayor apoyo institucional y Jalapa el departamento con menor apoyo institucional al realizar la estandarización poblacional.

Finalmente, en el Cuadro 37 y 38 se presentan los resultados del análisis de componentes principales (PCA) realizado para obtener los pesos del índice de vulnerabilidad estructural en violencia intrafamiliar.

7.4.2. Pesos basados en análisis de componentes principales (PCA)

Cuadro 37: Varianza explicada y razón de varianza explicada por componente principal

Componente Principal	Varianza Explicada	Razón de Varianza Explicada
PC1	1.3412	0.3353
PC2	1.0049	0.2512
PC3	0.9674	0.2418
PC4	0.6865	0.1716

*Se demuestra la varianza explicada y la razón de varianza explicada por cada componente principal del análisis de componentes principales (PCA) realizado para obtener los pesos del índice de vulnerabilidad estructural.

Cuadro 38: Pesos de los indicadores en cada componente principal

Variable	PC1	PC2	PC3	PC4
Indicador de situación	-0.6738	-0.1122	-0.2091	0.6998
Indicador de víctima	0.6811	-0.0196	0.1853	0.7080
Indicador de relación	0.0114	0.9600	-0.2664	0.0853
Indicador de apoyo institucional	-0.2862	0.2558	0.9225	0.0410

*Se presentan los pesos de cada indicador en los componentes principales obtenidos del análisis de componentes principales (PCA) realizado para obtener los pesos del índice de vulnerabilidad estructural.

El presente estudio tuvo como objetivo principal el desarrollo de un índice de vulnerabilidad estructural de violencia intrafamiliar (IVESVI) a partir de perfiles de víctimas, situaciones de violencia, relaciones entre agresores y víctimas, y contextos de apoyo de infraestructura. De forma que para lograrlo se identificarían variables sociodemográficas a partir de un análisis exploratorio, se detectarían patrones y asociaciones significativas elaborando perfiles, y finalmente se crearía el índice utilizando estos patrones y características de las víctimas y agresores. La discusión a continuación interpreta los principales hallazgos del análisis exploratorio, las reglas de asociación, y los modelos de clases latentes (LCA), buscando la coherencia en la construcción del índice en base a literatura.

8.1. Hallazgos preliminares

Los hallazgos de las pruebas de independencia, los gráficos del análisis exploratorio y las reglas de asociación, establecen una base para la construcción de perfiles de vulnerabilidad. Esto debido a que confirman interdependencia entre variables sociodemográficas de la víctima, el agresor y el caso.

En primer lugar, se descartan variables que no son relevantes o que no cumplen con la calidad de datos mínimo para la construcción del índice como variables de tiempo, artículos aplicados al caso, entre otros. También se observan variables que poseen una alta correlación entre y que son atemporales como etnia, sexo, edad, escolaridad, alfabetización, y trabajo. Estas variables son consistentes con la literatura revisada que indica que son características sociodemográficas relevantes como factores que contribuyen a la vulnerabilidad estructural [34], [46], [47].

Las reglas más repetibles y significativas indican una alta dependencia étnica y de alfabetización entre los miembros de la situación 10. La etnia tiene una alta correlación entre las víctimas al poseer elevaciones (*lift*) altos y medios de forma que las víctimas de etnia indígena tienden a relacionarse con agresores de la misma etnia y viceversa (Tabla 10). Esto es consistente con estudios previos ya que suele ocurrir la transmisión generacional de violencia en comunidades por lo que la violencia como tal no surge de forma aislada sino dentro de estructuras sociales [148].

La dependencia económica entre la víctima y el agresor demuestra que la víctima no es la principal proveedora económica lo que genera una dependencia

financiera que puede limitar la capacidad de la víctima de desarrollar una carrera profesional que le permita salir de la situación de violencia. Las mujeres que sufren de violencia tienden a tener una menor probabilidad de trabajar [35], y aquellas que trabajan tienden a ser minimizadas y a tener un menor rendimiento dentro de un entorno laboral [149].

Finalmente la violencia física tiene un componente inversamente proporcional a la presencia de la violencia psicológica tal como se observa en las reglas de asociación (Tabla 10). Esto sugiere una disociación de tipos de violencia basado en el nivel de escolaridad tal como se observa en la 2, donde se observa que a mayor nivel educativo, la violencia física y sexual tiende a disminuir [19].

8.2. Perfiles de clases latentes

Los modelos LCA permitieron la reducción de variables en perfiles distintivos y robustos (lo que es demostrado por la alta entropía y la baja incertidumbre de los modelos, Tablas 158 y 17, 159 y 27, 160 y 34), permitiendo asignar un riesgo relativo justificado a cada clase.

8.2.1. Perfiles de vulnerabilidad de la víctima

El modelo de clases latentes (LCA) identificó cuatro clases latentes distintas de víctimas con un gradiente de vulnerabilidad (Tablas 12, 13, 14, 15, 16).

La primera clase tiene un riesgo máximo dado que la víctima tiene una probabilidad alta de ser analfabeta, de no poseer escolaridad, de no trabajar y de ser indígena. Esto implica que la víctima posee una máxima dependencia económica y baja autonomía social, educativa, que es acentuada por barreras étnicas por parte de su comunidad [51], [32].

La segunda clase tiene un riesgo medio-alto, ya que este tipo de víctimas son alfabetas, tienen una escolaridad primaria, no trabajan y en su mayoría tienen muchos hijos. Esto implica que las víctimas tengan un riesgo por dependencia económica, y una carga de cuidado. A pesar de que sean más instruidas la alta carga de cuidados limitan su autonomía [32].

La tercera clase tiene un riesgo medio-bajo, ya que este tipo de víctimas son alfabetas, su escolaridad es primaria, tampoco trabajan, no son indígenas pero no tienen de muchos hijos. Por lo que a diferencia del perfil anterior, la menor carga de cuidado les permite una mayor autonomía aunque sigan dependiendo económicamente [32].

Finalmente, la cuarta clase tiene menos riesgo, debido a que poseen escolaridad de diversificado, son alfabetas, la mayoría trabajan y no son indígenas. Esto implica que tienen una mayor autonomía económica, disponen de una mayor escolaridad lo que actúa como un factor protector de violencia física, y finalmente no son oprimidas por barreras étnicas [53], [48], [33].

8.2.2. Perfiles de gravedad de la situación

El modelo de clases latentes (LCA) se enfocó en la agresividad del hecho y el perfil del agresor, identificando cinco clases latentes distintas (Tablas 19, 20, 21, 22, 23, 24, 26, 25).

La quinta clase posee un riesgo máximo, ya que se agrupan las características demográficas del agresor con la agresividad del hecho. De forma que el agresor tiene una nula escolaridad y es analfabeto lo que implica una mayor probabilidad de violencia física [19], además, el hecho es grave ya que la víctima sufre de

violencia psicológica también, y finalmente dado que el agresor es indígena lo más probable dado las reglas Apriori es que la víctima también lo sea, por lo que está limitada su autonomía social y económica [51].

La cuarta clase posee un riesgo medio alto, dado que es un alfabeto con estudios primarios de forma que la educación primaria permite tener cohesión de grupo disminuyendo la violencia desde la educación [150]. Sin embargo, el hecho es grave ya que la víctima sufre de violencia física y psicológica y se ha demostrado que la combinación de categorías de abuso causan el máximo deterioro físico y mental, incluso puede tender a suicidios [20], de forma que la violencia psicológica es al menos igual de mala que la violencia física.

La primera clase posee un riesgo medio al sufrir de violencia física y psicológica, y un nivel de educación bajo.

La tercera clase posee un riesgo medio bajo, ya que el agresor tiene una escolaridad de diversificado, lo que implica una disminución en la probabilidad de sufrir violencia física [45]. No son indígenas por lo que no existen barreras étnicas. Sin embargo, el perfil posee probabilidades altas de que la víctima esté expuesta a violencia compuesta (física y psicológica) lo que implica un riesgo considerable [20].

Finalmente, la segunda clase posee un riesgo bajo ya que a pesar de que el agresor es alfabeto y posee un nivel de escolaridad bajo (primaria), el hecho no llega a ser de igual magnitud dado que la violencia es únicamente física, lo que implica un riesgo menor [20].

8.2.3. Perfiles de gravedad de la relación entre víctima y agresor

El modelo de clases latentes (LCA) evaluó la estabilidad y la dinámica de poder por parte de los dos individuos. Cabe destacar que ningún perfil posee una probabilidad alta de una diferente escolaridad, lo que denota que en este caso los casos no suelen ser tan graves dada la misma condición escolar [45]. Sin embargo, tampoco hubo perfiles donde la víctima fuese la principal proveedora económica lo que indica una clara dependencia económica [34], [32], [33].

La segunda clase demuestra una relación extremadamente inestable al estar cómo esposos en la relación pero como solteros a nivel de estado civil. Esto denota una relación en proceso de ruptura, lo que implica un riesgo alto [55]. La dependencia económica genera una mayor vulnerabilidad al no trabajar las víctimas [34].

La cuarta clase posee un riesgo medio-alto dado que la relación es incompleta al ser unidos [55]. Asimismo, poseen una dependencia económica por parte de la víctima al no disponer de un trabajo.

La primera clase posee un riesgo medio-bajo dado que la relación es formalizada al estar casados [55]. En este caso también se denota una dependencia económica de las víctimas al no disponer de un trabajo

Finalmente, la tercera clase posee un riesgo bajo ya que la relación es estable al estar casados [55], además, la víctima depende menos económicamente al poseer un trabajo [34], de forma que se posee autonomía mutua lo que reduce coerción.

8.3. Índice de vulnerabilidad estructural(IVESVI)

El índice de vulnerabilidad estructural de violencia intrafamiliar (IVESVI) se construyó a partir de las probabilidades posteriores de los perfiles obtenidos en los modelos LCA. Estos son aptos al tener poca variabilidad (Tablas 17, 27

y 34) en los resultados permitiendo que se utilicen como insumos para análisis estadísticos posteriores [127], [99]. A partir de estos modelos se permiten integrar los patrones empíricos encontrados dentro del índice.

La inclusión del apoyo institucional a pesar de que este fue el indicador con mayor incertidumbre (Tabla 17) es imperativo al ser clave para la conceptualización de la vulnerabilidad estructural. La disparidad departamental en el acceso a recursos es un factor que excede las características individuales y refleja la capacidad del estado de proteger o no a sus ciudadanos [151]. Además, se le resta a 1 debido a que una mayor capacidad institucional disminuye la vulnerabilidad estructural.

Se destaca el rechazo de la asignación de pesos según el análisis de componentes principales (PCA, según sus siglas en inglés) debido a la falta de robustez en los resultados. Al faltar una varianza mínima explicada por el primer componente, los pesos asignados no eran representativos (Tabla 37). Este resultado sugiere colinealidad y superposición entre las distintas dimensiones de vulnerabilidad estructural, lo que es consistente con la literatura que indica que los factores de vulnerabilidad están interrelacionados y no actúan de forma aislada [152].

Por lo que se le asignaron pesos iguales a cada componente del índice, y esto es válido dado que asume la equipotencia teórica de las cuatro dimensiones de la vulnerabilidad estructural. Generando el siguiente índice de vulnerabilidad estructural de violencia intrafamiliar (IVESVI) final:

$$\text{IVESVI} = \frac{(I_V \pm \delta_{V_i}) \cdot R_v + (I_S \pm \delta_{S_i}) \cdot R_s + (I_r \pm \delta_{R_i}) \cdot R_r + (1 - C_I \pm \delta_{I_i})}{4}$$

Donde:

- i : número de la observación.
- V_i : indicador de vulnerabilidad de la víctima i
- δ_{V_i} : incertidumbre de la vulnerabilidad de la víctima i
- R_v : factor de riesgo según la clase latente de la víctima i
- I_S : indicador de gravedad de la situación
- δ_{S_i} : incertidumbre de la gravedad de la situación i
- R_s : factor de riesgo según la clase latente de la situación i
- I_R : indicador de gravedad de la relación
- δ_{R_i} : incertidumbre del perfil de la relación i
- R_r : factor de riesgo según la clase latente de la relación i
- C_I : capacidad institucional
- δ_{I_i} : incertidumbre de la capacidad institucional i

Los resultados confirman que la vulnerabilidad de violencia intrafamiliar es un fenómeno multidimensional, este es afectado por el nivel de escolaridad, la autonomía tanto económica como social, en conjunto con las barreras tanto sociales, étnicas, y de infraestructura del Estado. Asimismo, los modelos de LCA han permitido clasificar perfiles de casos en base a sus características, permitiendo la asignación de riesgos de forma empírica y justificar el diseño de un índice de vulnerabilidad estructural (IVESVI) que integra las distintas dimensiones de vulnerabilidad. Este permite la evaluación multidimensional de la vulnerabilidad al considerar el perfil de la víctima, la dinámica situacional incluyendo las características del agresor, la dinámica relacional entre la víctima y el agresor y el contexto institucional.

Asimismo, se observa una coherencia entre los hallazgos empíricos y la literatura revisada, provocando que los perfiles generados reflejen patrones conocidos y estudiados de situaciones de violencia; tal como la influencia del nivel educativo, la autonomía económica y social, en conjunto con las barreras sociales, económicas y estatales existentes. Y precisamente estos factores son los que configuran la vulnerabilidad estructural de la violencia intrafamiliar permitiendo medirla.

Finalmente, los objetivos planteados al inicio del estudio fueron cumplidos. Se elaboró un análisis exploratorio que permitió cuantificar, visualizar, interpretar e identificar variables sociodemográficas relevantes. También, se detectaron patrones y asociaciones significativas entre las variables que posibilitaron la elaboración de perfiles de víctimas, situaciones y relaciones. A partir de estos perfiles se creó un índice de vulnerabilidad estructural de violencia intrafamiliar (IVESVI) que integra las distintas dimensiones de vulnerabilidad identificadas en el análisis, desarrollando el índice IVESVI a partir de perfiles de víctimas y agresores contruidos sobre variables sociodemográficas y contextuales.

Por ende, el IVESVI puede ser una herramienta valiosa para priorizar intervenciones y recursos hacia los casos más vulnerables, permitiendo una asignación más eficiente de los esfuerzos de prevención y apoyo. Además, el enfoque metodológico utilizado en este estudio puede ser replicado en otros contextos para evaluar la vulnerabilidad estructural frente a diferentes tipos de violencia o riesgos sociales.

CAPÍTULO 10

Recomendaciones

Se recomienda la validación del índice IVESVI a través de estudios de campo y la comparación con indicadores existentes de vulnerabilidad y violencia intrafamiliar. Esto permitirá evaluar la precisión y utilidad práctica del índice en contextos reales. Además, se sugiere la actualización periódica del índice para reflejar cambios en las dinámicas sociales, económicas y políticas que puedan afectar la vulnerabilidad estructural. También, se recomienda generar IVESVI específico para casos con otros grupos relevantes como niños, discapacitados, y ancianos utilizando otro grupo de datos más amplio.

Bibliografía

- [1] I. N. de Estadística, “Datos de violencia intrafamiliar 2021,” tech. rep., Instituto Nacional de Estadística, 2021.
- [2] S. M. Walton and C. A. S. Pérez, “La violencia intrafamiliar. un problema de salud actual,” *Gaceta Médica Espirituana*, vol. 21, pp. 96–105, 2019.
- [3] A. P. L. E. D. L. Prostitución, “Revictimización que es y como prevenirla,” tech. rep., Asociación para la Eliminación de la Prostitución, Pornografía, Turismo, Tráfico Sexual de Niñas, Niños y Adolescentes - ECPAT/Guatemala., 2010.
- [4] A. Gallón, “Salvar vidas centralizando la atención de la violencia intrafamiliar en un solo lugar,” 2025.
- [5] I. Slabbert, “Domestic violence and poverty: Some women’s experiences,” *Research on Social Work Practice*, vol. 27, pp. 223–230, 3 2017.
- [6] S. J. G. Castañeda, “Revictimización de los niños, niñas y adolescentes en el proceso penal,” *Revista Ciencia Multidisciplinaria CUNORI*, vol. 8, pp. 1–14, 5 2024.
- [7] J. Quesada, L. Hart, and P. Bourgois, “Structural vulnerability and health: Latino migrant laborers in the united states,” *Medical Anthropology*, vol. 30, no. 4, pp. 339–362, 2011.
- [8] P. Bourgois, S. M. Holmes, K. Sue, and J. Quesada, “Structural vulnerability: Operationalizing the concept to address health inequities in clinical care,” *Academic Medicine*, vol. 92, no. 3, pp. 299–307, 2017.
- [9] N. H. Weiss, R. E. Johnson, A. A. Contractor, and C. Peasant, “A latent profile analysis of intimate partner victimization and aggression,” *Journal of Interpersonal Violence*, 2019.
- [10] A. Brassard, Y. Lussier, and S. Bergeron, “A latent profile analysis of ipv perpetration among men in treatment,” *Journal of Interpersonal Violence*, 2022.
- [11] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” in *Proceedings of the 20th International Conference on Very Large Data Bases*, pp. 487–499, 1994.

- [12] F. Booyesen, “An overview and evaluation of composite indices of development,” *Social Indicators Research*, vol. 59, no. 2, pp. 115–151, 2002.
- [13] L. Cherchye, W. Moesen, N. Rogge, and T. Van Puyenbroeck, “An introduction to ‘benefit-of-the-doubt’ composite indicators,” *Social Indicators Research*, vol. 82, no. 1, pp. 111–145, 2007.
- [14] R. F. Brino and M. A. de Oliveira Souza, “Concepções sobre violência intrafamiliar na Área educacional,” *Educação & Realidade*, vol. 41, pp. 1251–1273, 8 2016.
- [15] E. C. de la República de Guatemala, “Decreto numero 97-1996 ley para prevenir, sancionar y erradicar la violencia intrafamiliar,” 11 1996.
- [16] U. Nations, “¿qué es el maltrato en el hogar?,” 8 2025.
- [17] C. P. Nájera, *ESTUDIO CRIMINOLÓGICO DE LA VIOLENCIA INTRAFAMILIAR EJERCIDA SOBRE LOS ANCIANOS EN CUBA*. PhD thesis, Universidad Marta Abreu de las Villas, 7 2008.
- [18] L. A. de la ONU para los Refugiados Guatemala, “Tipos de violencia,” 2025.
- [19] B. Sanz-Barbero, N. Barón, and C. Vives-Cases, “Prevalence, associated factors and health impact of intimate partner violence against women in different life stages,” *PLoS ONE*, vol. 14, 10 2019.
- [20] L. C. Potter, R. Morris, K. Hegarty, C. García-Moreno, and G. Feder, “Categories and health impacts of intimate partner violence in the world health organization multi-country study on women’s health and domestic violence,” *International Journal of Epidemiology*, vol. 50, pp. 652–662, 4 2021.
- [21] C. Guerra, N. Pereda, G. Guilera, and J. Abad, “Internalizing symptoms and polyvictimization in a clinical sample of adolescents: The roles of social support and non-productive coping strategies,” *Child Abuse & Neglect*, vol. 54, pp. 57–65, 4 2016.
- [22] ILSI, “Evaluación de riesgo conceptos riesgo vs. peligro,” tech. rep., ILSI Argentina, 2020.
- [23] C. E. de Riesgos Naturales América Latina, “Tomo i metodologÍa de modelación probabilista de riesgos naturales vulnerabilidad de edificaciones e infraestructura,” tech. rep., Consorcio Evaluación de Riesgos Naturales América Latina, 2009.
- [24] K. J. C. Cárdenas, A. M. V. Rojas, A. S. V. Cevallos, and M. D. C. R. Jaramillo, “La violencia intrafamiliar. definición y consecuencias,” *RECI-MUNDO*, vol. 6, pp. 384–397, 9 2022.
- [25] L. S. Millett, P. L. Kohl, M. Jonson-Reid, B. Drake, and M. Petra, “Child maltreatment victimization and subsequent perpetration of young adult intimate partner violence: An exploration of mediating factors,” *Child Maltreatment*, vol. 18, pp. 71–84, 5 2013.
- [26] N. Brooks, “Vulnerability, risk and adaptation: A conceptual framework,” tech. rep., Tyndall Centre for Climate Change Research and Centre for Social and Economic Research on the Global Environment (CSERGE), 11 2003.

- [27] C. Levine, R. Faden, C. Grady, D. Hammerschmidt, L. Eckenwiler, and J. Sugarman, "The limitations of "vulnerability" as a protection for human research participants," *American Journal of Bioethics*, vol. 4, pp. 44–49, 2004.
- [28] M. Brazier and M. Lobjoit, *Protecting the Vulnerable: Autonomy and Consent in Health Care*, vol. 1. Routledge, 1 ed., 1992.
- [29] J. R. Febles, A. Z. Hernández, and M. C. Fernández, "La violencia intra-familiar y los grupos de riesgo victimal," *Derecho y Cambio Social*, vol. 1, pp. 184–203, 10 2020.
- [30] C. de la República de Guatemala, "Decreto numero 27-2003 el congreso de la republica de guatemala," 2003.
- [31] C. de la República, "Ley de protección para las personas de la tercera edad : Decreto número 80-96 del congreso de la república y sus reformas : Decretos número 2-97 y 51-98.," 4 1999.
- [32] D. P. S. Poma, A. de Jesús Quezada Coronel, M. J. R. Quezada, Y. de Jesús Luna Moreno, and J. A. A. Muñoz, "Dependencia económica en las mujeres víctimas de violencia física que acuden al ministerio de la mujer-derechos humanos de santa cruz," *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, vol. 6, pp. 768–778, 1 2025.
- [33] F. ADECCO, "Un empleo contra la violencia," tech. rep., Fundación ADECCO, 11 2019.
- [34] G. E. Clark, S. C. Moser, S. J. Ratick, K. Dow, W. B. Meyer, S. Emani, W. Jin, J. X. Kasperson, R. E. Kasperson, and H. E. Schwarz, "Assessing the vulnerability of coastal communities to extreme storms: The case of reverse, ma., usa," *Mitigation and Adaptation Strategies for Global Change*, vol. 3, pp. 59–82, 1998.
- [35] A. Browne, A. Salomon, and S. S. Bassuk, "The impact of recent partner violence on poor women's capacity to maintain work," *VIOLENCE AGAINST WOMEN*, vol. 5, pp. 393–426, 4 1999.
- [36] B. Ola, "Why battered women do not leave, part 2," *Trauma, Violence & Abuse*, vol. 2, pp. 3–35, 1 2001.
- [37] E. Echeburúa, J. Fernández-Montalvo, and P. de Corral, "¿hay diferencias entre la violencia grave y la violencia menos grave contra la pareja?: un análisis comparativo," *International Journal of Clinical and Health Psychology*, vol. 8, pp. 355–382, 2008.
- [38] R. Cuellar, "Vulnerabilidad ante la violencia: El caso de la niñez y adolescencia guatemalteca," tech. rep., Centro Internacional para Investigaciones en Derechos Humanos, 6 2024.
- [39] M. de Gobernación, "Estrategia nacional prevencion de la violecia y el delito 2017-2027," tech. rep., Ministerio de Gobernación, 2017.
- [40] N. Unidas, "Instituciones resilientes para una recuperación transformadora pospandemia en américa latina y el caribe," tech. rep., Naciones Unidas, 10 2021.

- [41] N. Olofsson, K. Lindqvist, and I. Danielsson, "Higher risk of violence exposure in men and women with physical or sensory disabilities: Results from a public health survey," *Journal of Interpersonal Violence*, vol. 30, pp. 1671–1686, 6 2015.
- [42] E. Padilla, "Sin poder huir: violencia contra las personas con discapacidad," 8 2020.
- [43] B. F. Jiménez and F. A. Muñoz, "Violencia estructural," 2004.
- [44] F. de las Naciones Unidas para la Infancia, "Voces de adolescentes con discapacidad sobre la violencia y discriminación en la familia, escuela y comunidad," tech. rep., UNICEF, 2022.
- [45] A. Mosiño and S. L. C. Torres, "Violencia contra las mujeres y su asociación con los niveles de educación: un análisis de la encuesta nacional sobre la dinámica de las relaciones de los hogares, México 2021," *Acta Universitaria*, vol. 34, pp. 1–21, 3 2024.
- [46] I. Casique, "Índices de empoderamiento de las mujeres," in *Violencia de género en las parejas mexicanas. Análisis de resultados de la Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2006* (R. Castro and I. Casique, eds.), pp. 45–68, México: Instituto Nacional de las Mujeres, 2008.
- [47] I. Casique, "Vulnerabilidad a la violencia doméstica. una propuesta de indicadores para su medición," *Revista Internacional de Estadística y Geografía*, vol. 3, 5 2012.
- [48] D. Rapp, B. Zoch, M. M. H. Khan, T. Pollmann, and A. Krämer, "Association between gap in spousal education and domestic violence in India and Bangladesh," *BMC Public Health*, vol. 12, 2012.
- [49] R. Jewkes, "Intimate partner violence: causes and prevention," *The Lancet*, vol. 359, pp. 1423–1429, 4 2002.
- [50] C. I. de Derechos Humanos, "Situación de derechos humanos en Guatemala," tech. rep., Comisión Interamericana de Derechos Humanos, 12 2015.
- [51] G. N. Miranda, "Mujeres indígenas, política pública y el reto de la descolonización del pensamiento y la acción," *Revista de Análisis de Realidad Nacional*, vol. 1, pp. 136–157, 3 2023.
- [52] R. Soto, L. López, and E. Díaz, "Análisis rápido de género," tech. rep., Cooperative for Assistance and Relief Everywhere, Inc., 2022.
- [53] Sleeper-Smith, J. Ostler, and J. L. Reid, "Review: Past and present colonialist practices: Violence in indigenous communities violence in indigenous communities on JSTOR," *Journal of Ethnic and Cultural Studies*, vol. 8, pp. 342–346, 11 2021.
- [54] J. A. Andrade, "El maltrato familiar y el escenario mental del agresor," *Revista Psicología Científica.com*, vol. 12, 2 2010.
- [55] E. Khosravipour, P. Mohammadkhani, B. Dolatshahi, A. Pourshahbaz, O. A. Sahraei, and M. Yousefnejad, "Risk factors of marital violence of married men and women in different levels of severity," vol. 30, pp. 1221–1229, 2011.

- [56] E. Echeburúa, J. Fernández-Montalvo, P. D. Corral, and J. J. López-Goñi, “Assessing risk markers in intimate partner femicide and severe violence: A new assessment instrument,” *Journal of Interpersonal Violence*, vol. 24, pp. 925–939, 6 2009.
- [57] E. Echeburúa, P. J. Amor, and I. Loinaz, “Escala de predicción del riesgo de violencia grave contra la pareja-revisada-(epv-r),” *Psicothema*, vol. 22, pp. 1054–1060, 2010.
- [58] C. D. Jesús, “La investigación cuantitativa,” tech. rep., Corporación Universitaria de Asturias, 2024.
- [59] D. R. Anderson, D. J. Sweeney, and T. A. Williams, *ESTADÍSTICA PARA ADMINISTRACIÓN Y ECONOMÍA*. CENGAGE Learning Editores S.A., 10 ed., 2008.
- [60] D. D. Wackerly, W. Mendenhall, and R. L. Scheaffer, *MATHEMATICAL STATISTICS WITH APPLICATIONS*, vol. 1. CENGAGE Learning, 7 ed., 2008.
- [61] D. Kuonen, “An introduction to the bootstrap with applications in r,” *Statistical Computing & Statistical Graphics Newsletter*, vol. 13, pp. 6–11, 2003.
- [62] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 1st ed., 1994.
- [63] J. L. Hennessy and D. A. Patterson, *In Praise of Computer Architecture: A Quantitative Approach Fifth Edition*, vol. 1. Elsevier, 5 ed., 2012.
- [64] P. Powell and I. Smalley, “What is a central processing unit (cpu)?,” 2025.
- [65] M. Flinders, S. Susnjara, and I. Smalley, “What is a gpu?,” 2025.
- [66] NVIDIA, “What is cuda?,” 9 2021.
- [67] P. Organization, “What is python? executive summary | python.org,” 2025.
- [68] P. S. Foundation, “venv — creation of virtual environments — python 3.14.0 documentation,” 10 2025.
- [69] Git, “Git - what is git?,” 2025.
- [70] IBM, “¿qué es la confiabilidad de los datos?,” 2025.
- [71] IBM, “Coherencia de los datos mediante la normalización,” 12 2024.
- [72] D. Pyle, D. D. Cerra, and M. Kaufmann, *Data Preparation for Data Mining*. Morgan Kaufmann Publishers, Inc., 1 ed., 1999.
- [73] R. Nisbet, G. Miner, K. Yale, J. IV, and A. Peterson, *HANDBOOK OF STATISTICAL ANALYSIS AND DATA MINING APPLICATIONS*, vol. 1. Academic Press, 2nd ed., 2018.
- [74] The pandas development team, “pandas-dev/pandas: Pandas.”
- [75] V. Mehta, “Camelot: Pdf table extraction for humans.” <https://github.com/camelot-dev/camelot>, 2018. Accessed: 2025-11-01.

- [76] N. V. Otten, “How to use text normalization techniques (nlp) [9 ways python],” 1 2023.
- [77] P. S. Foundation, “Python programming language.” <https://www.python.org>, 2025. Accessed: 2025-11-01.
- [78] M. Bachmann, “Rapidfuzz,” 2021.
- [79] M. All, “Handling machine learning categorical data with python tutorial | datacamp,” 2 2023.
- [80] C. E. E. Freire, “Las variables y su operacionalización en la investigación educativa. parte 1,” *Revista Conrado*, vol. 65, pp. 39–49, 2018.
- [81] IBM, “¿qué es el análisis exploratorio de datos?,” 2025.
- [82] J. Noble, “¿qué es el algoritmo apriori?,” 2025.
- [83] K. Prasanna, M. Seetha, and A. P. S. Kumar, “Capriori: Conviction based apriori algorithm for discovering frequent determinant patterns from high dimensional datasets,” in *2014 International Conference on Science Engineering and Management Research (ICSEMR)*, pp. 1–6, 2014.
- [84] H. Xie, “Research and case analysis of apriori algorithm based on mining frequent item-sets,” *Open Journal of Social Sciences*, vol. 09, pp. 458–468, 2021.
- [85] T. Odland, “Efficient-apriori: An efficient python implementation of the apriori algorithm.” <https://github.com/tommyod/Efficient-Apriori>, n.d. Accessed: 2025-09-13.
- [86] M. L. McHugh, “The chi-square test of independence,” *Biochemia Medica*, vol. 23, no. 2, pp. 143–149, 2013.
- [87] K. Krishnamoorthy, *Handbook of statistical distributions with applications*. Chapman & Hall/CRC, 2006.
- [88] R. Singhal and R. K. Rana, “Chi-square test and its application in hypothesis testing,” *Journal of the Practice of Cardiovascular Sciences*, vol. 1, no. 1, pp. 69–71, 2015.
- [89] S. P. Díaz and S. P. Fernández, “Asociación de variables cualitativas: Test de chi-cuadrado,” 2004.
- [90] M. G. Juárez, “Prueba chi-cuadrado,” 2025.
- [91] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nature Methods*, vol. 17, pp. 261–272, 2020.
- [92] IBM, “¿qué es la agrupación en clústeres?,” 2 2024.

- [93] O. F. E. Co-Operation and Development, *Handbook on Constructing Composite Indicators*, vol. 1. OECD Publishing, 1 ed., 2008.
- [94] F. Madadizadeh and S. Bahariniya, “Tutorial on statistical data reduction methods for exploring dietary patterns,” *Clinical Nutrition ESPEN*, vol. 58, pp. 228–234, 12 2023.
- [95] L. Naldi and S. Cazzaniga, “Research techniques made simple: Latent class analysis,” *Journal of Investigative Dermatology*, vol. 140, pp. 1676–1680.e1, 9 2020.
- [96] S. J. Gershman and D. M. Blei, “A tutorial on bayesian nonparametric models,” *Journal of Mathematical Psychology*, vol. 56, pp. 1–12, 2 2012.
- [97] J. Ruscio and A. M. Ruscio, “Categories and dimensions: Advancing psychological science through the study of latent structure,” *Current Directions in Psychological Science*, vol. 17, pp. 203–207, 6 2008.
- [98] X.-L. Meng, “Em algorithm,” *Wiley StatsRef: Statistics Reference Online*, 4 2014.
- [99] L. M. Collins and S. T. Lanza, *Latent Class and Latent Transition Analysis with Applications in the Social, Behavioral, and Health Sciences*, vol. 1. John Wiley & Sons, 2010.
- [100] J. D. Banfield and A. E. Raftery, “Model-based gaussian and non-gaussian clustering,” *Biometrics*, vol. 49, pp. 803–821, 9 1993.
- [101] L. K. Muthén and B. O. Muthén, *Statistical Analysis With Latent Variables User’s Guide*, vol. 1. Muthén & Muthén, 8 ed., 1998.
- [102] J. Schreiber, “Pomegranate: fast and flexible probabilistic modeling in python,” *Journal of Machine Learning Research*, vol. 18, no. 164, pp. 1–6, 2018.
- [103] J. H. Wolfe, “Pattern clustering by multivariate mixture analysis,” *Multivariate Behavioral Research*, vol. 5, pp. 329–350, 1970.
- [104] R. J. Hathaway, “Another interpretation of the em algorithm for mixture distributions,” tech. rep., 1986.
- [105] C. Biernacki, G. Celeux, and G. E. Govaert, “An improvement of the nec criterion for assessing the number of clusters in a mixture model,” *Pattern Recognition Letters*, vol. 20, pp. 267–272, 1999.
- [106] G. Celeux and G. Soromenho, “An entropy criterion for assessing the number of clusters in a mixture model,” *Journal of Classification*, vol. 13, pp. 195–212, 1996.
- [107] A. P. Dempster, . N. M. Laird, and . D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, pp. 1–38, 1977.
- [108] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

- [109] K. J. Grimm, G. L. Mazza, and P. Davoudzadeh, "Model selection in finite mixture models: A k-fold cross-validation approach," *Structural Equation Modeling*, vol. 24, pp. 246–256, 3 2017.
- [110] D. Mindrila, "Latent class analysis," *International Journal for Cross-Disciplinary Subjects in Education (IJCDSE)*, vol. 11, pp. 4323–4328, 2020.
- [111] GeeksforGeeks, "Bayesian information criterion (bic)," 7 2025.
- [112] H. Ai, "Ac-19," tech. rep., 1974.
- [113] L. D. D. L. Salud and M. V. Aguirre, "Factores determinantes de la salud: Importancia de la prevención," *Acta Médica Peruana*, vol. 28, pp. 237–241, 2011.
- [114] Centro para la Salud y Desarrollo de la Comunidad, "Capítulo 19, sección 2: Comprender los factores de riesgo y de protección: Su uso en la selección de objetivos potenciales y estrategias prometedoras para las intervenciones," capítulo en la caja de herramientas comunitarias, Universidad de Kansas, Centro para la Salud y Desarrollo de la Comunidad, 2014. Obtenido el 02 de enero de 2014.
- [115] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," 8 2020.
- [116] M. K. Pakhira, "A linear time-complexity k-means algorithm using cluster shifting," in *Proceedings - 2014 6th International Conference on Computational Intelligence and Communication Networks, CICN 2014*, pp. 1047–1051, Institute of Electrical and Electronics Engineers Inc., 3 2014.
- [117] D. Pelleg and A. W. Moore, "X-means: Extending k-means with efficient estimation of the number of clusters," in *Proceedings of the Seventeenth International Conference on Machine Learning, ICML '00*, (San Francisco, CA, USA), p. 727–734, Morgan Kaufmann Publishers Inc., 2000.
- [118] L. Jing, M. K. Ng, and J. Z. Huang, "An entropy weighting k-means algorithm for subspace clustering of high-dimensional sparse data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, pp. 1026–1041, 8 2007.
- [119] Harika, "Kmodes clustering algorithm for categorical data," 4 2025.
- [120] F. Cao, J. Liang, and L. Bai, "A new initialization method for categorical data clustering," *Expert Systems with Applications*, vol. 36, pp. 10223–10228, 9 2009.
- [121] A. Chaturvedi, K. Foods, P. E. Green, and J. D. Carroll, "K-modes clustering," *Journal of Classification*, vol. 18, pp. 35–55, 2001.
- [122] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," tech. rep., 1998.
- [123] O. F. E. Co-Operation and Development, "Oecd glossary of statistical terms," 2008.
- [124] D. of Economic and S. Affairs, "Indicators of sustainable development: Framework and methodologies," tech. rep., Department of Economic and Social Affairs, 4 2001.

- [125] E. C. Mexiquense and A. M. Mondéjar-Jiménez, “Economía, sociedad y territorio,” *Ecnomía, Sociedad y Territorio*, vol. 3, pp. 565–585, 2008.
- [126] M. Gómez, *Elementos de Estadística Descriptiva*, vol. 1. EUNED, 4 ed., 2012.
- [127] J. Magidson and J. K. Vermunt, *Latent class models*, pp. 175–198. SAGE Publications Inc., 1 ed., 2004.
- [128] K. Pearson, “Liii. on lines and planes of closest fit to systems of points in space,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, pp. 559–572, 11 1901.
- [129] I. T. Jolliffe, *Principal Component Analysis, Second Edition*, vol. 1. Springer Verlag, 2 ed., 2002.
- [130] G. Nicoletti, S. Scarpetta, and O. Boylaud, “Summary indicators of product market regulation with an extension to employment protection legislation,” *OECD Economics Department Working Papers*, vol. 226, 4 2000.
- [131] T. Greyling and F. Tregenna, “Construction and analysis of a composite quality of life index for a region of south africa,” *Social Indicators Research*, vol. 131, pp. 887–930, Apr. 2017. Publisher Copyright: © 2016, Springer Science+Business Media Dordrecht.
- [132] I. N. de Estadística, “Violencia intrafamiliar,” 2024.
- [133] D. Glejberman, “Clasificador nacional de ocupaciones-cno-2020,” tech. rep., Instituto Nacional de Estadística, 2023.
- [134] I. N. de Estadística, “Encuesta nacional de empleo e ingresos - (enei) - 2021,” 2021.
- [135] I. N. de Estadística, “Proyecciones departamentales 2010-2050,” 2020.
- [136] I. de Estudios Comparados en Ciencias Penales de Guatemala, “1. distribución territorial de las instalaciones de servicios de justicia - hojas de cálculo de google,” 2025.
- [137] C. de Bomberos Municipales de la Ciudad de Guatemala, “Direccion y telefonos de la entidad y todas sus dependencias,” 2022.
- [138] A. N. D. B. M. D. de Guatemala, “2. dirección y teléfonos – asociación nacional de bomberos municipales departamentales de guatemala,” 7 2025.
- [139] M. de Salud Pública y Asistencia Social, “Direcciones y telefonos mspas 1er semestre 2025,” 7 2025.
- [140] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015. Software available from tensorflow.org.

- [141] S. Brin, R. Motwanit, and C. Silversteint, “Beyond market baskets: Generalizing association rules to correlations,” *ACM SIGMOID Record*, vol. 26, pp. 265–276, 1997.
- [142] B. D. Gerardo, J. Lee, J. Lee, M. Park, and M. Lee, “The association rule algorithm with missing data in data mining,” in *Computational Science and Its Applications – ICCSA 2004* (A. Laganá, M. L. Gavrilova, V. Kumar, Y. Mun, C. J. K. Tan, and O. Gervasi, eds.), (Berlin, Heidelberg), pp. 97–105, Springer Berlin Heidelberg, 2004.
- [143] K. Prasanna, M. Seetha, and A. P. S. Kumar, “Capriori: Conviction based apriori algorithm for discovering frequent determinant patterns from high dimensional datasets,” *International Conference on Science, Engineering and Management Research*, 2014.
- [144] N. J. de Vos, “kmodes categorical clustering library.” <https://github.com/nicodv/kmodes>, 2015–2024.
- [145] B. Efron, “Better bootstrap confidence intervals,” *Source: Journal of the American Statistical Association*, vol. 82, pp. 171–185, 1987.
- [146] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, pp. 623–656, 10 1948.
- [147] J. B. C. García and R. M. P. Trujillo, “Exclusión de las personas analfabetas en la vinculación laboral,” *Revista Mexicana de Sociología*, vol. 4, pp. 767–799, 12 2018.
- [148] M. M. I. Zhicay, J. I. T. Segarra, and E. T. F. Lazo, “Factores socioculturales que influyen en mujeres víctimas de violencia intrafamiliar,” *Killkana Social*, vol. 2, pp. 187–196, 10 2018.
- [149] B. D. G. Guamán, G. O. Reyes, A. B. Pico, A. B. Chávez, and M. G. Zambrano, “violencia intrafamiliar y su incidencia en el desarrollo profesional de la mujer,” *Killkana Social*, vol. 9, pp. 67–86, 5 2025.
- [150] I. Alonso-Rodríguez, D. Pérez-Jorge, I. Pérez-Pérez, and E. Olmos-Raya, “Restorative practices in reducing school violence: a systematic review of positive impacts on emotional wellbeing,” 2025.
- [151] A. M. Velásquez, “La desigualdad social en guatemala evolución y respuesta institucional documentos de proyectos,” tech. rep., CEPAL, 2022.
- [152] L. V. Perez, “Principal component analysis to address multicollinearity,” tech. rep., 5 2017.

CAPÍTULO 11

Anexos

11.1. Accesos a información pública



Instituto de la Defensa Pública Penal
Por una defensa oportuna, gratuita y eficaz.

DIVISION DE COORDINACIONES TECNICO PROFESIONALES

Guatemala, 07 de diciembre de 2020
Oficio. 416-2020/DCTP-RRRDL/nera

Licenciado:
Ramón Augusto Zetina Méndez
Encargado de la Unidad de Información Pública
Instituto de la Defensa Pública Penal



Licenciado Zetina:

De manera muy atenta me dirijo a usted para saludarlo y desearle éxitos en sus actividades, tengo el agrado de dar respuesta a la solicitud de información pública número 79-2020, requerida por la usuaria LOURDES AMANDA PÉREZ CHAJÓN.

Solicitud: "Número de coordinaciones que integran la división técnica profesional desglosadas por departamento y municipio durante el periodo comprendido del año 2017 al 2020."

Respuesta: Número de Coordinaciones: desglosado por departamento y municipio, el cual consta de dos (02) folios.

Sin otro particular, me suscribo de usted,

Atentamente,

C.c. Archivo

Lic. Raúl René Kobas de León
Director
División de Coordinaciones
Técnico Profesionales
INSTITUTO DE LA DEFENSA PÚBLICA PENAL

Número de Coordinaciones	Departamento	Municipio
2	Alta Verapaz	
		1 Cobán
		1 San Pedro Carchá
1	Baja Verapaz	
		1 Salamá
1	Chimaltenango	
		1 Chimaltenango
1	Chiquimula	
		1 Chiquimula
1	El Progreso	
		1 Guastatoya
2	Escuintla	
		1 Escuintla
		1 Santa Lucía Cotzumalguapa
16	Guatemala	
		12 Guatemala
		1 Mixco
		1 Villa Nueva
		1 Amatitlán
6	Huehuetenango	
		1 Huehuetenango
		1 La Democracia
		1 Santa Eulalia
		1 Barillas
		1 Cuilco
		1 San Mateo Ixtatán
1	Izabal	
		1 Puerto Barrios
1	Jalapa	
		1 Jalapa
1	Jutiapa	
		1 Jutiapa
3	Petén	
		1 San Benito
		1 Poptún
		1 La Libertad
5	Quetzaltenango	
		1 Quetzaltenango
		1 Coatepeque

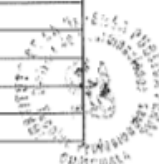


Figura 6: Continuación figura 5

	1	Cantel
	1	Palestina de los Altos
	1	San Juan Ostuncalco
3	Quiché	
	1	Santa Cruz
	1	Ixcán
	1	Nebaj
2	Retalhuleu	
	1	Retalhuleu
	1	San Sebastian
1	Sacatepéquez	
	1	Antigua Guatemala
5	San Marcos	
	1	San Marcos
	1	Malacatán
	1	Ixchiguan
	1	Comitancillo
	1	Concepción Tutuapa
1	Santa Rosa	
	1	Cuilapa
2	Sololá	
	1	Sololá
	1	Santiago Atitlán
1	Suchitepéquez	
	1	Mazatenango
2	Totonicapán	
	1	Totonicapán
	1	San Cristobal
1	Zacapa	
	1	Zacapa



NOTA IMPORTANTE: En el mes de septiembre 2020 la *Coordinación Nacional de Asistencia Legal Gratuita a la Víctima y sus Familiares* finalizó con el traslado de sus funciones al Instituto para la Asistencia y Atención a la Víctima del Delito.

Figura 7: Continuación figura 5



UNIDAD DE INFORMACIÓN PÚBLICA, MINISTERIO DE TRABAJO Y PREVISIÓN SOCIAL "MINTRAB" Guatemala
doce de mayo del año dos mil veinticinco.

Se tiene a la vista para resolver la solicitud identificada con el número 1885-2025, presentada el siete de mayo de dos mil veinticinco por **Diego Andrés Alonzo Medinilla**, quien solicita: "información de rangos de salarios para el siguiente pdf: <https://sen.inc.gob.gt/wp-content/uploads/2024/01/DIAGRAMA2.0.pdf> Estoy desarrollando una tesis en la Universidad del Valle de Guatemala focalizada en violencia intrafamiliar y necesito conocer los rangos de trabajo para las víctimas y agresores".

CONSIDERANDO

Que el Ministerio de Trabajo y Previsión Social es sujeto obligado a proporcionar la información pública que se le solicite atendiendo al principio de máxima publicidad y transparencia, de conformidad con lo establecido por la Ley de Acceso a la Información Pública Decreto número 57-2008 del Congreso de la República. Asimismo, que en su artículo 1 garantiza a toda persona interesada, sin discriminación alguna, el derecho a solicitar y a tener acceso a la información pública en posesión de autoridades y sujetos obligados.

CONSIDERANDO

Que según el artículo 2 y 9 numeral 6 de la Ley de Acceso a la Información Pública, constituye información pública la información o actos de la administración pública que se encuentren en los archivos, fichas, registros, bases de datos que se encuentren en los organismos de Estado o bien cualquier registro que documente el ejercicio de las facultades o la actividad de los sujetos obligados y sus servidores públicos, sin importar su fuente, fecha de elaboración o medio siempre que no sea confidencial o esté clasificado como reservada; y que, de acuerdo al Artículo 45 de la Ley de Acceso a la Información Pública, a toda solicitud deberá recaer una resolución por escrito, la cual podrá emitirse entregando la información solicitada según numeral 1 del Artículo 42 de la ley antes mencionada.

CONSIDERANDO

Que, para dar respuesta a la solicitud presentada por **Diego Andrés Alonzo Medinilla**, la Dirección de Estadísticas Laborales del Ministerio de Trabajo y Previsión Social, remitió a esta unidad oficio DEL No. 615/2025 JCHR/hg.

POR TANTO

Esta Unidad de Información Pública de conformidad con sus funciones y atribuciones, los considerandos y los artículos 4, 6, 9, 15, 16, 18, 21, 22, 38, 40, 41, 42 y 45 de la Ley de Acceso a la Información Pública.

RESUELVE

I. Entregar la información solicitada por **Diego Andrés Alonzo Medinilla**. II. Remitir al correo electrónico señalado por el solicitante para recibir notificaciones oficio DEL No. 615/2025 JCHR/hg, que consta de (1) folio, que contienen la información solicitada, el cual fue proporcionado bajo responsabilidad de la (s) unidad (es) responsable (s), del otorgamiento de esta. III. Los interesados tendrán responsabilidad penal y civil, por el uso, manejo y difusión de la información a la que tengan acceso, de conformidad con la Ley de Acceso a la Información Pública y demás leyes aplicables. IV. Notifíquese

Juanita Josefina Salguero Samalá
Jefe de Unidad de Acceso a la Información Pública
Ministerio de Trabajo y Previsión Social



c.c. Archivo.



Figura 8: Acceso a la información pública del desglose de los diagramas de trabajo



Ministerio de
**Trabajo y
Previsión Social**

Solicitud de información pública número:

1885-2025

Fecha y hora de creación:

2025-05-07 16:51:28

Nombre completo

Diego Andrés Alonzo Medinilla

Email

alo20172@uvg.edu.gt

Teléfono

50673581

Departamento

GUATEMALA

Municipio

MIXCO

Medio respuesta

Correo Electronico

Usuario creador

Solicitante

Nacionalidad

guatemalteco/a

Descripción clara y precisa de la información que solicita:

Solicito información de rangos de salarios para el siguiente pdf: <https://sen.ine.gob.gt/wp-content/uploads/2024/02/DIAGRAMA-2.0.pdf> Estoy desarrollando una tesis en la Universidad del Valle de Guatemala focalizada en violencia intrafamiliar y necesito conocer los rangos de trabajo para las víctimas y agresores.

Figura 9: Continuación figura 8



Ministerio de
Trabajo y
Previsión Social

77



Guatemala, 09 de mayo de 2025
Oficio DEL No. 615/2025 JCHR/hg.

Licenciada
Lesvia Josefina Salguero Sanabria
Jefa de Comunicación Social
Unidad de Acceso a la Información Pública
Ministerio de Trabajo y Previsión Social
Presente



Estimada Licenciada Salguero:

HR: L. Salguero HORA: 15:32

Le saludo cordialmente deseándole éxitos en sus labores cotidianas. Por este medio se da respuesta a la solicitud No. 1885-2025, de fecha 08 de mayo del presente año.

Respecto a la petición del señor DIEGO ANDRÉS ALONZO MEDINILLA, donde requiere lo siguiente: "Solicito información de rangos de salarios para el siguiente pdf: <https://sen.ine.gob.gt/wp-content/uploads/2024/01/DIAGRAMA2.0.pdf> Estoy desarrollando una tesis en la Universidad del Valle de Guatemala focalizada en violencia intrafamiliar y necesito conocer los rangos de trabajo para las víctimas y agresores."

Se hace de su conocimiento que la información solicitada la puede encontrar en el tablero dinámico del Informe del Empleador, el cual se ubica en la página del Ministerio de Trabajo y Previsión Social, en el apartado de Estadísticas del Mercado de Trabajo (<https://www.mintrabajo.gob.gt/statistics>) o ingresando por medio del código QR que se comparte en este oficio. Posteriormente a ingresar a la página web debe seguir los siguientes pasos: ingresar en el apartado "Informe del Empleador", dirigirse al apartado "Estadísticas sobre el Informe del Empleador", ingresar a la sección "Informe Dinámico", en la opción de "salarios de los empleados" aparece la gráfica de salario mensual promedio por tipo de ocupación, la cual utiliza el resumen a gran grupo de la Clasificación Internacional Uniforme de Ocupaciones en Guatemala (CIUO-08).



Me despido de usted, dando mis muestras de consideración y estima.

Atentamente,


Julio César Hernández Rodríguez
Director de Estadísticas Laborales
Ministerio de Trabajo y Previsión Social
Guatemala, C.A.

c.c. Archivo DEL

7a. Avenida 3-33, zona 9, Edificio Torre Empresarial | PBX: 2422-2500/01/02/03 | www.mintrabajo.gob.gt

Síguenos como Ministerio de Trabajo



Figura 10: Continuación figura 8

11.2. Distribuciones y frecuencias de las variables

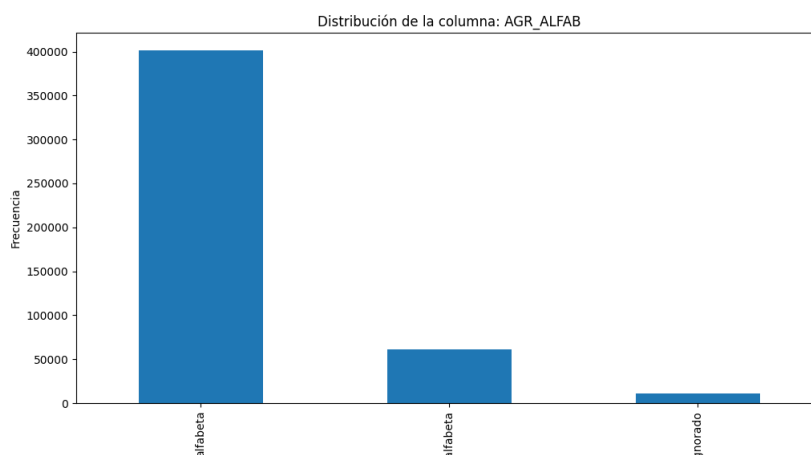


Figura 11: Frecuencia de la columna AGRALFAB

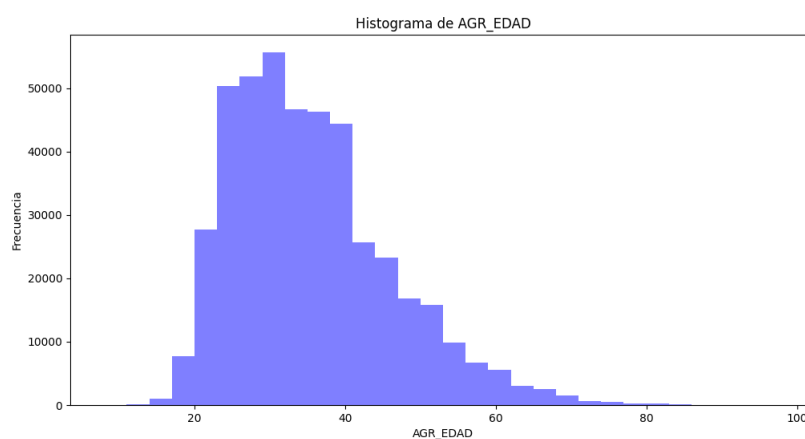


Figura 12: Frecuencia de la columna AGREDAD

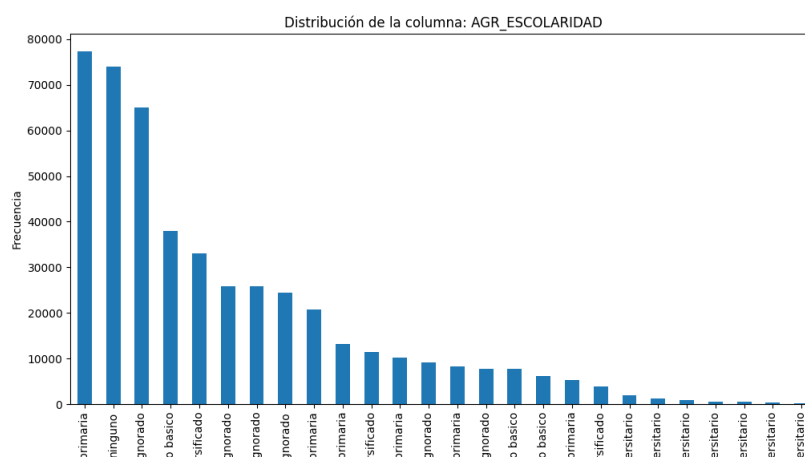


Figura 13: Frecuencia de la columna AGRESCOLARIDAD

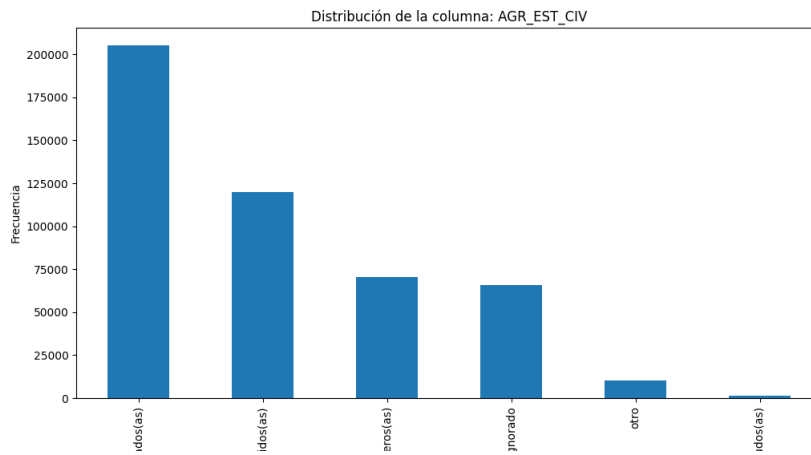


Figura 14: Frecuencia de la columna AGRESTCIV

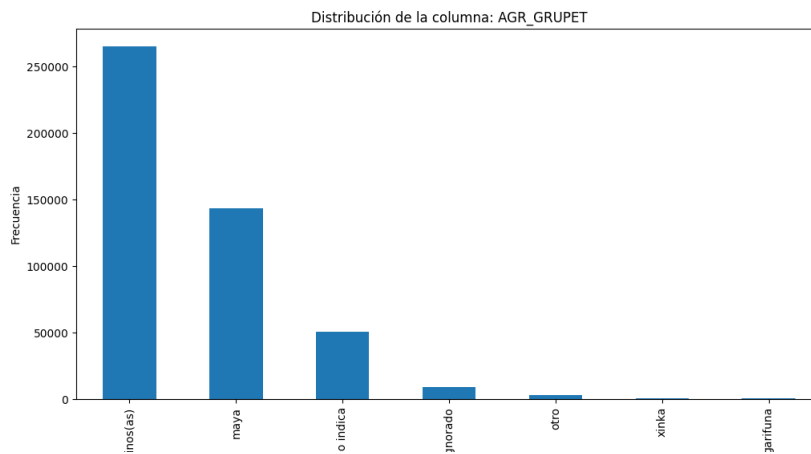


Figura 15: Frecuencia de la columna AGRGRUPET

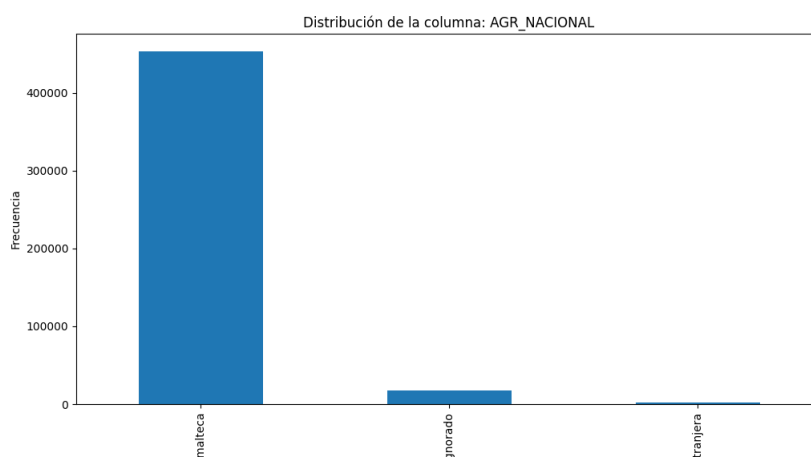


Figura 16: Frecuencia de la columna AGRNACIONAL

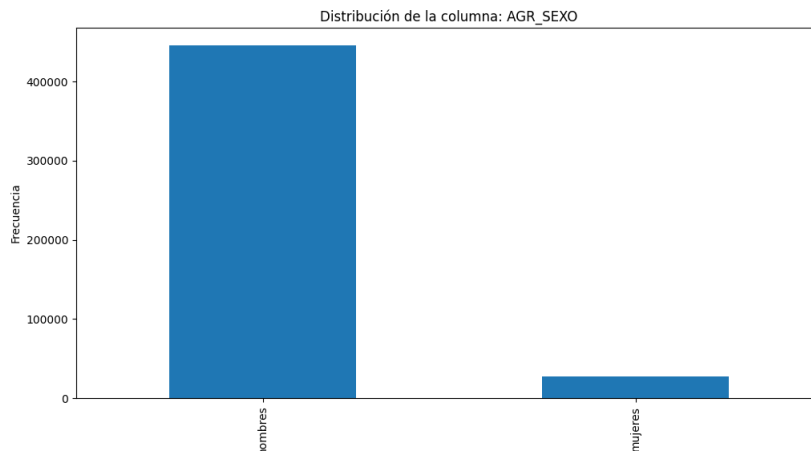


Figura 17: Frecuencia de la columna AGRSEXO

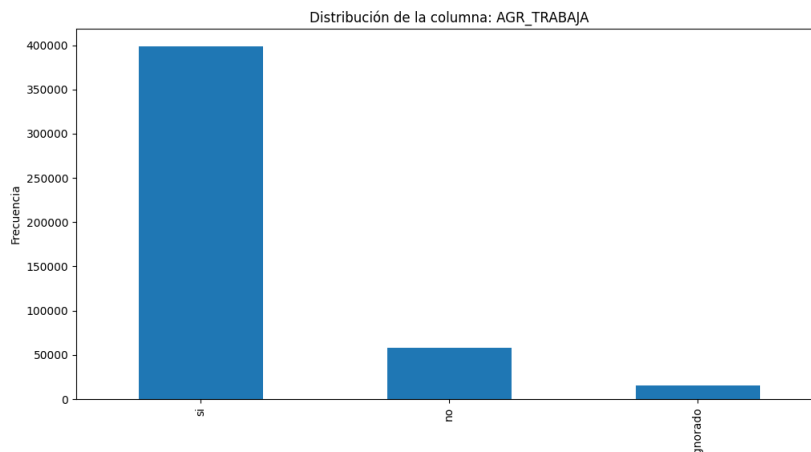


Figura 18: Frecuencia de la columna AGRTRABAJA

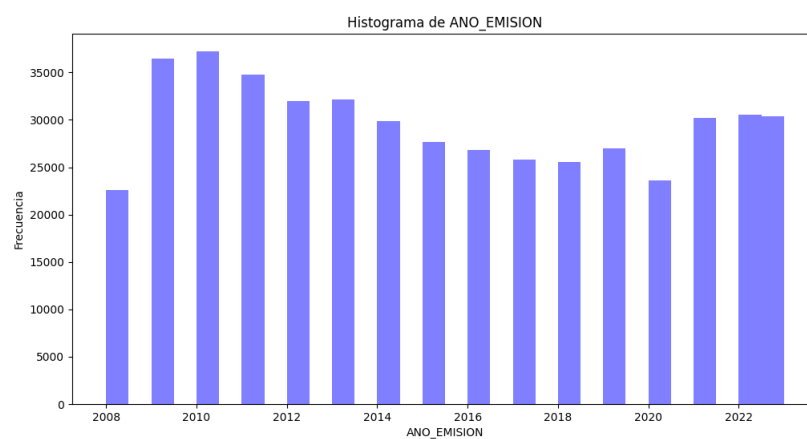


Figura 19: Frecuencia de la columna ANOEMISION

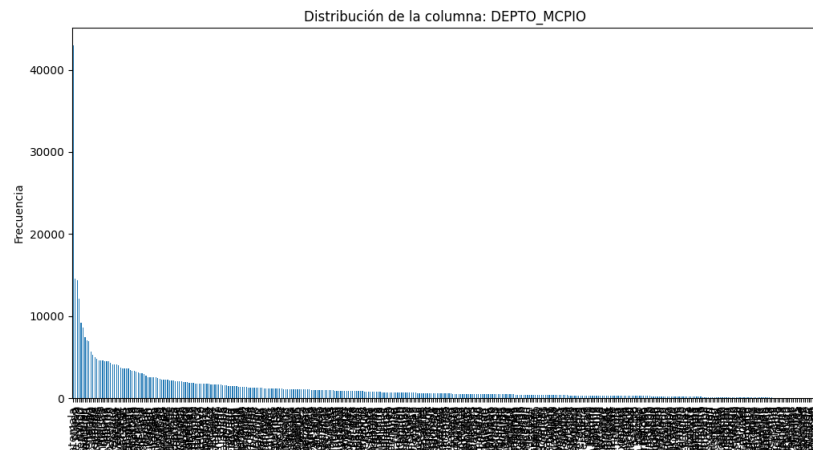


Figura 20: Frecuencia de la columna DEPTOMCPIO

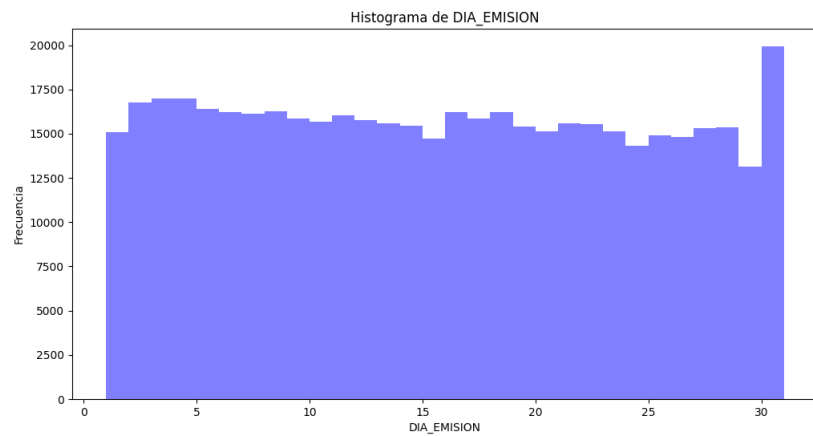


Figura 21: Frecuencia de la columna DIAEMISION

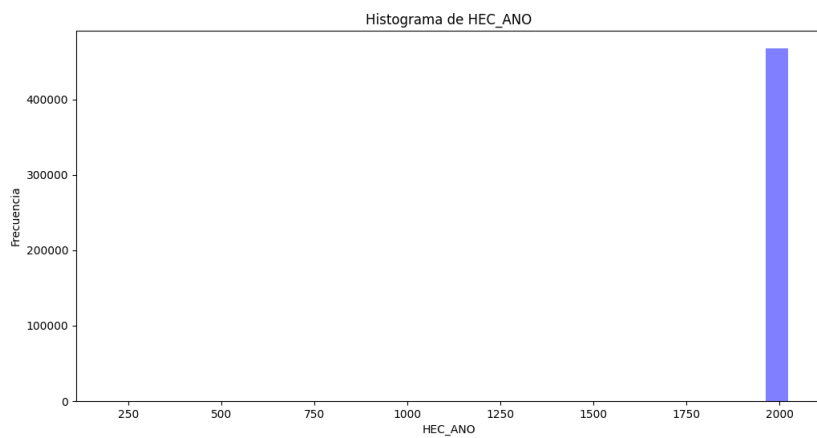


Figura 22: Frecuencia de la columna HECANO

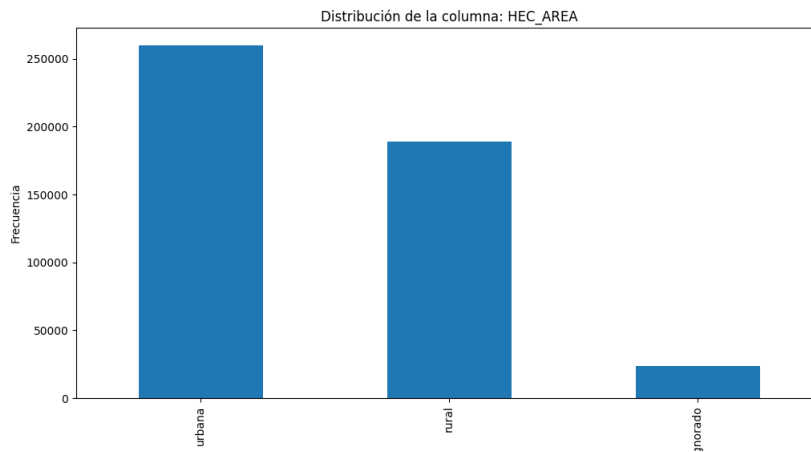


Figura 23: Frecuencia de la columna HECAREA

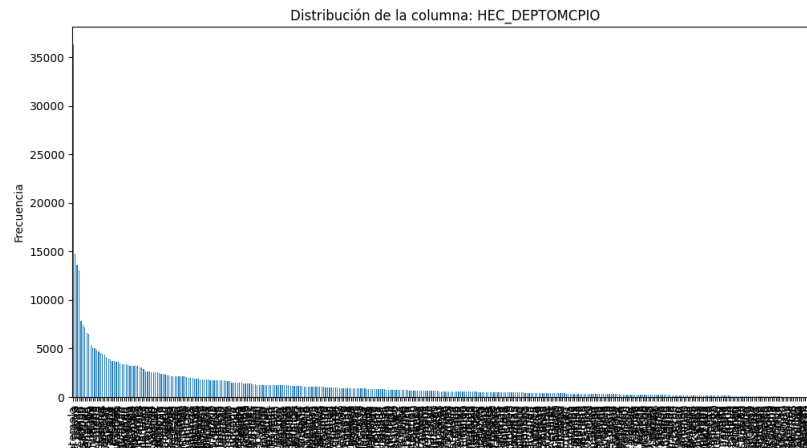


Figura 24: Frecuencia de la columna HECDEPTOMCPIO

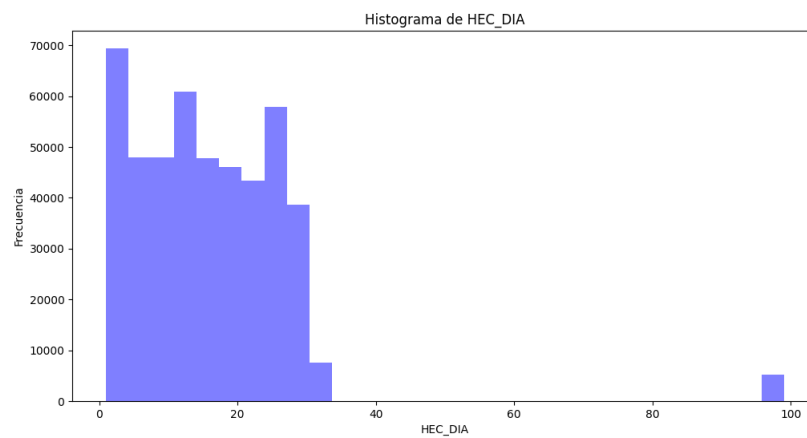


Figura 25: Frecuencia de la columna HECDIA

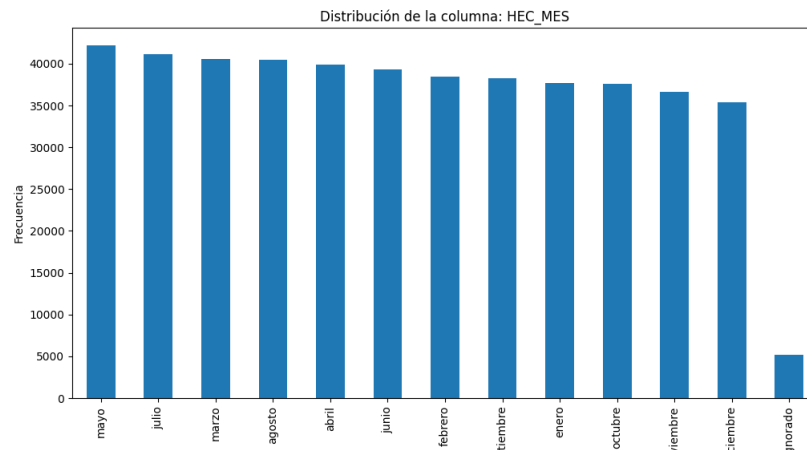


Figura 26: Frecuencia de la columna HECMES

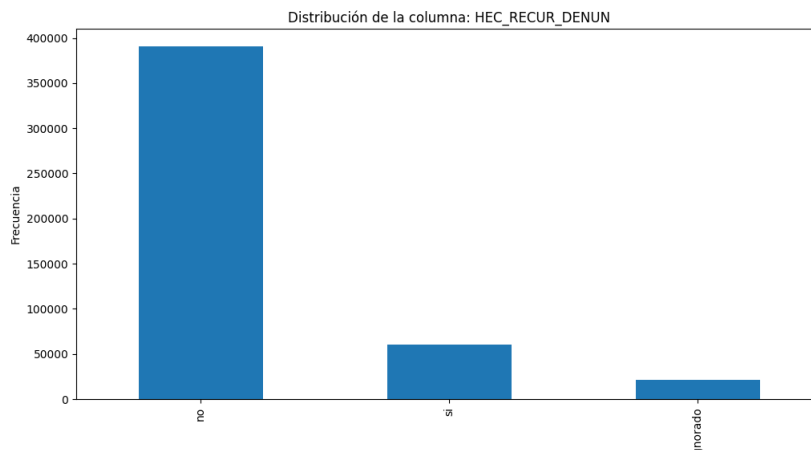


Figura 27: Frecuencia de la columna HECRECURDENUN

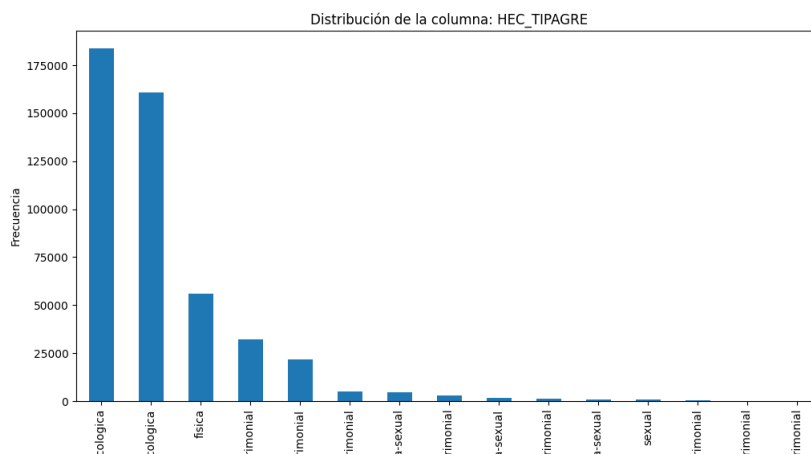


Figura 28: Frecuencia de la columna HECTIPAGRE

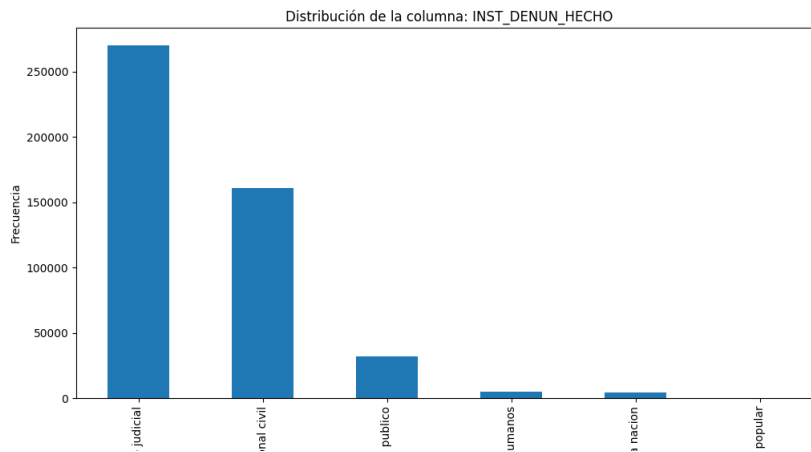


Figura 29: Frecuencia de la columna INSTDENUNHECHO

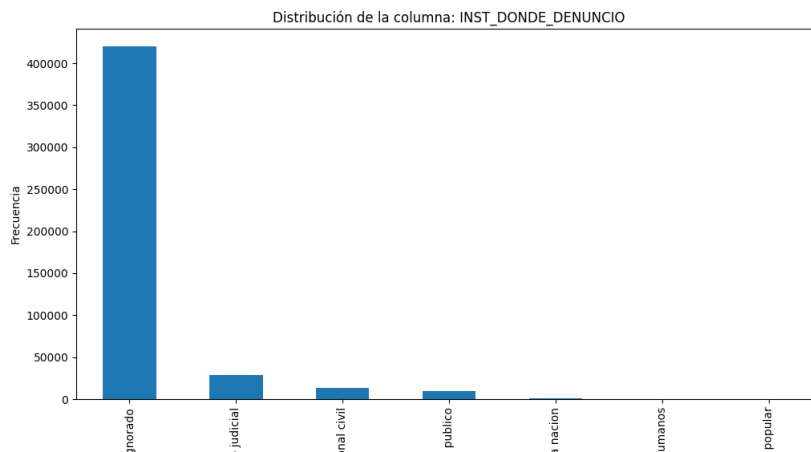


Figura 30: Frecuencia de la columna INSTDONDEDENUNCIO

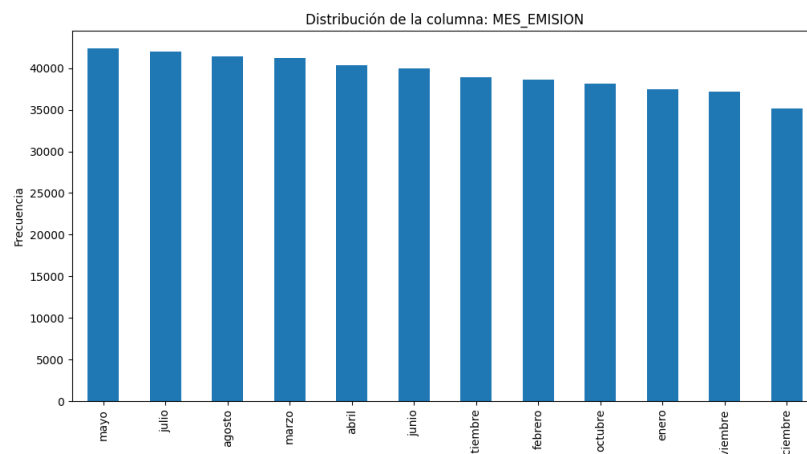


Figura 31: Frecuencia de la columna MESEMISION

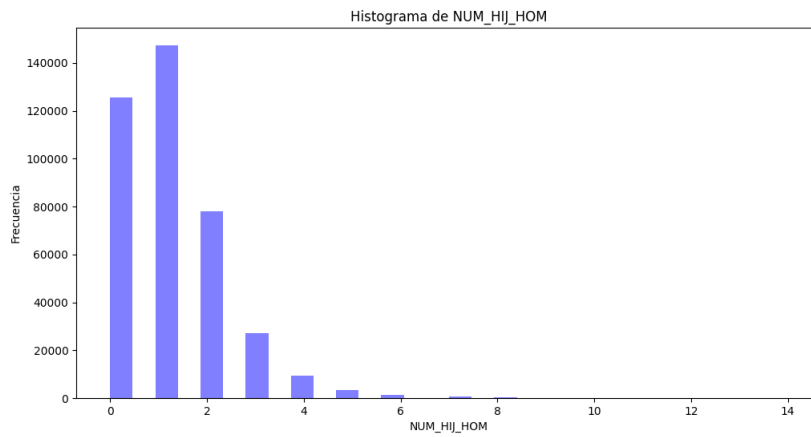


Figura 32: Frecuencia de la columna NUMHIJHOM

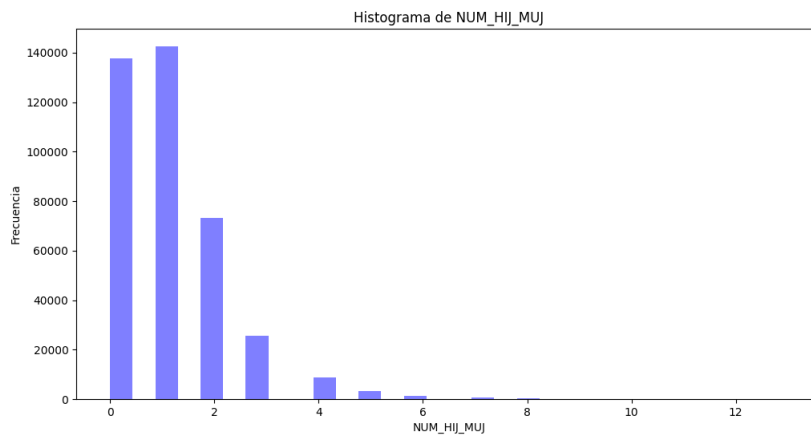


Figura 33: Frecuencia de la columna NUMHIJMUJ

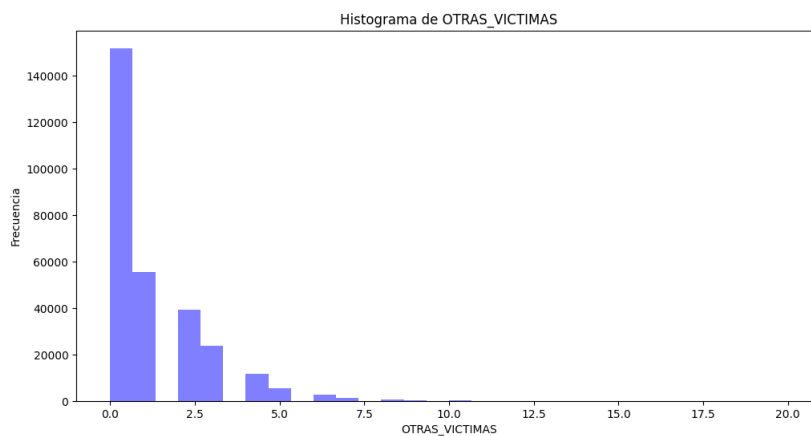


Figura 34: Frecuencia de la columna OTRASVICTIMAS

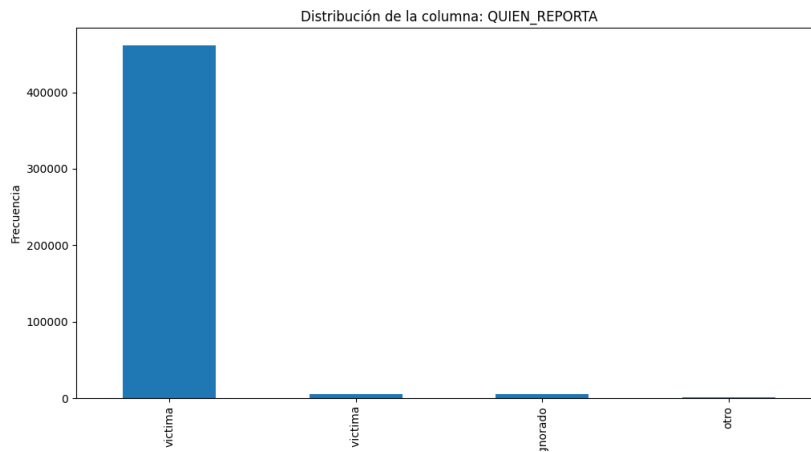


Figura 35: Frecuencia de la columna QUIENREPORTA

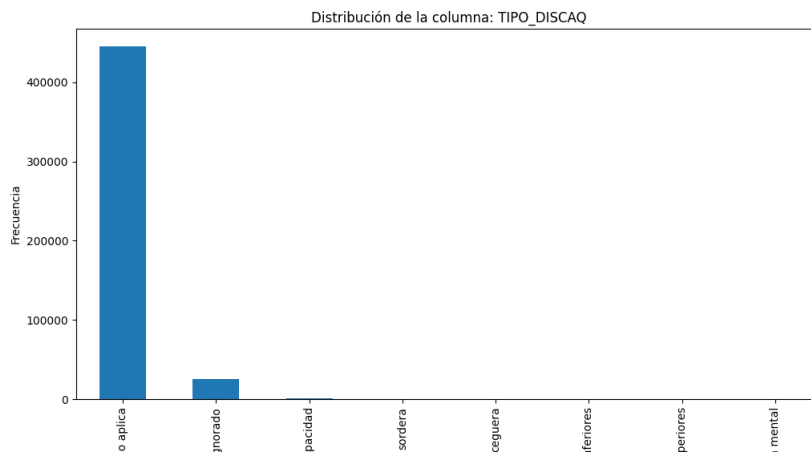


Figura 36: Frecuencia de la columna TIPODISCAQ

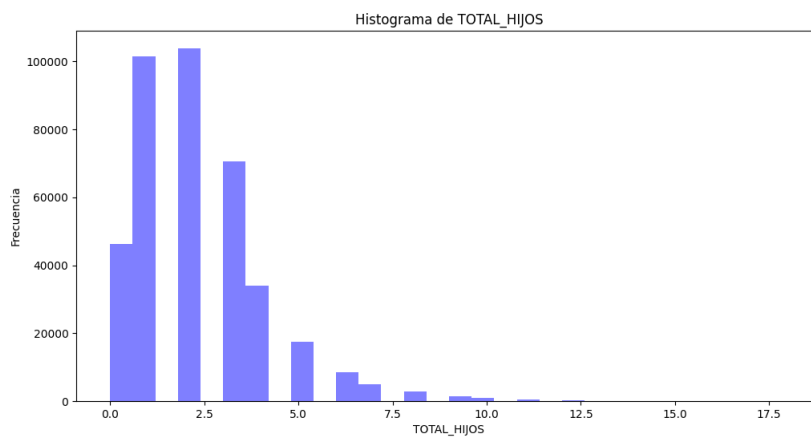


Figura 37: Frecuencia de la columna TOTALHIJOS

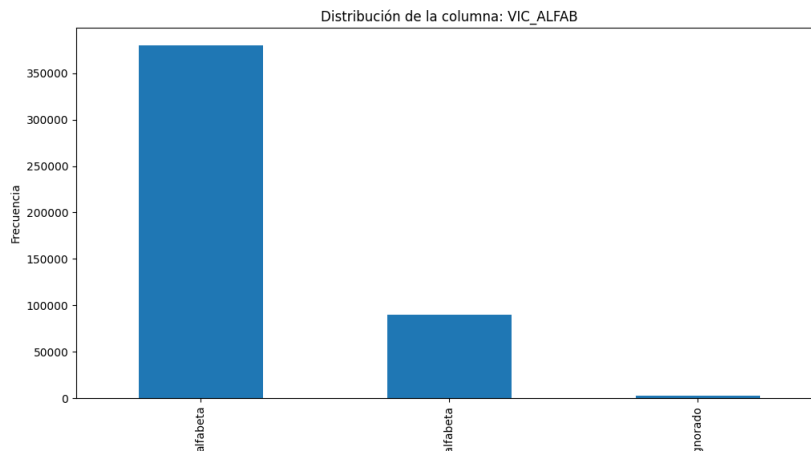


Figura 38: Frecuencia de la columna VICALFAB

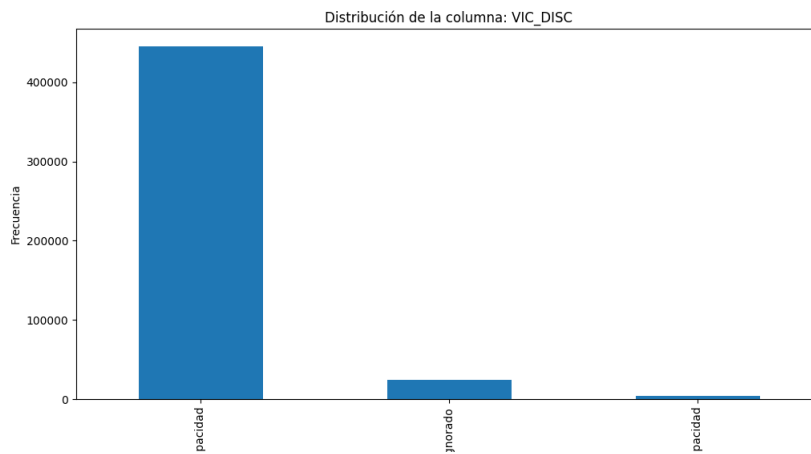


Figura 39: Frecuencia de la columna VICDISC

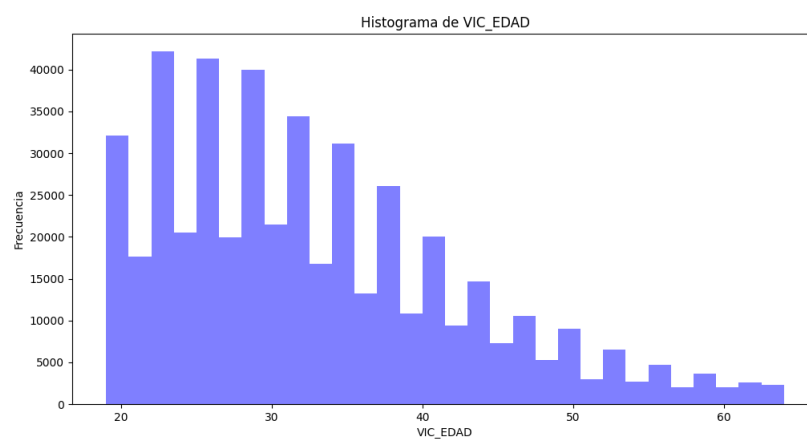


Figura 40: Frecuencia de la columna VICEDAD

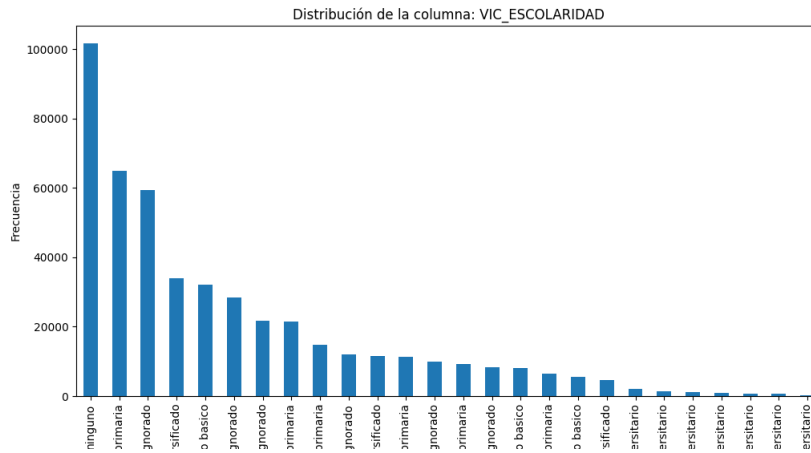


Figura 41: Frecuencia de la columna VICESCOLARIDAD

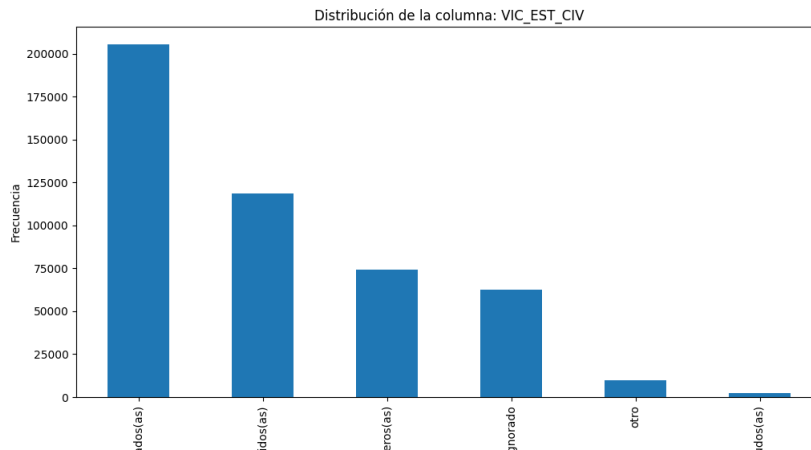


Figura 42: Frecuencia de la columna VICESTCIV

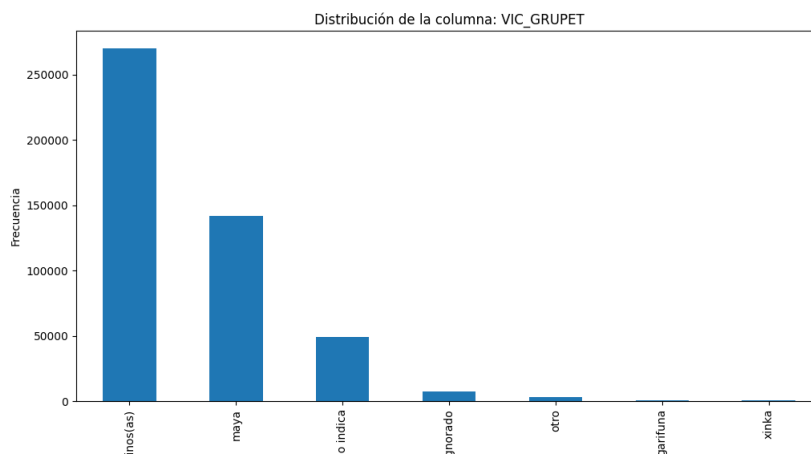


Figura 43: Frecuencia de la columna VICGRUPET

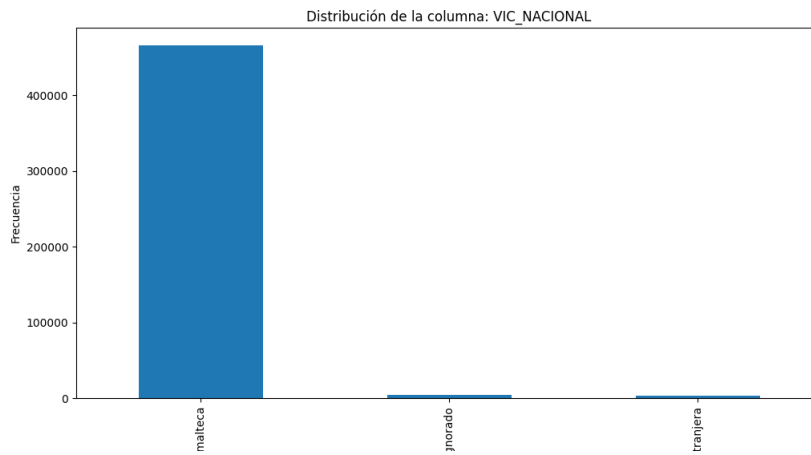


Figura 44: Frecuencia de la columna VICNACIONAL

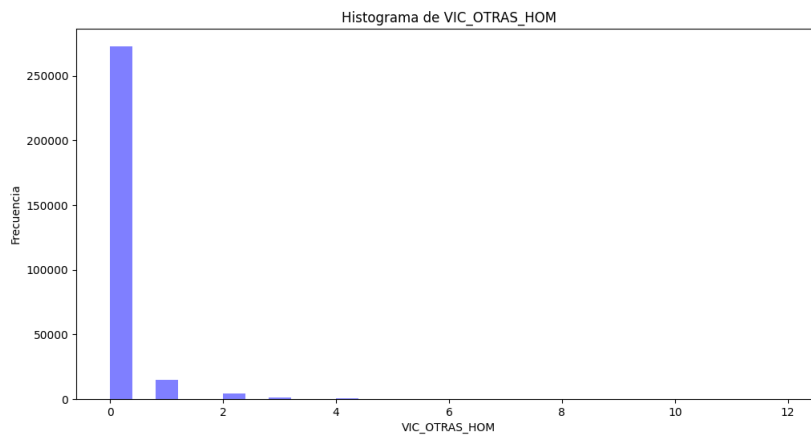


Figura 45: Frecuencia de la columna VICOTRASHOM

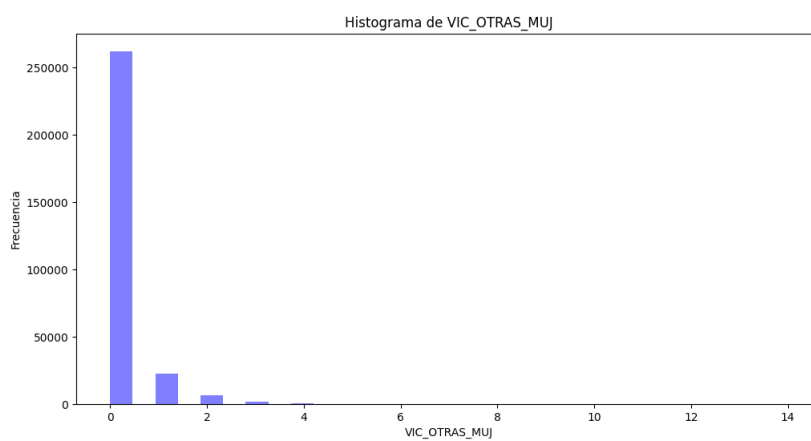


Figura 46: Frecuencia de la columna VICOTRASMUJ

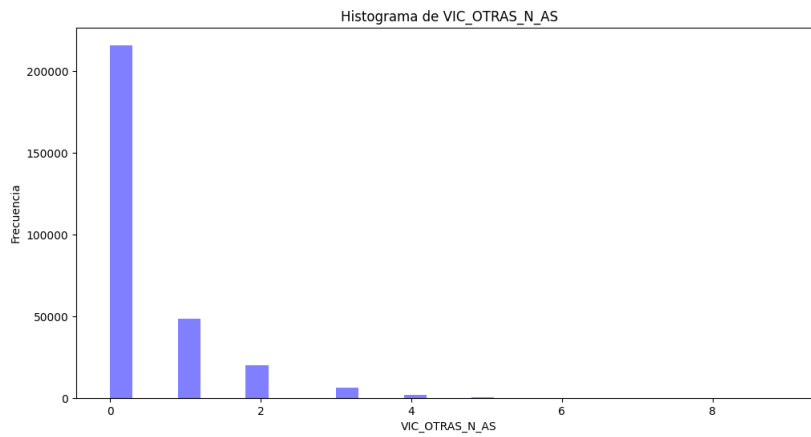


Figura 47: Frecuencia de la columna VICOTRASNAS

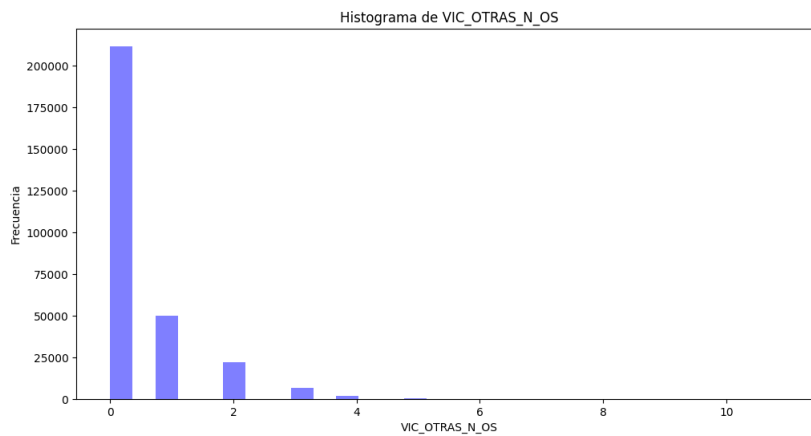


Figura 48: Frecuencia de la columna VICOTRASNOS

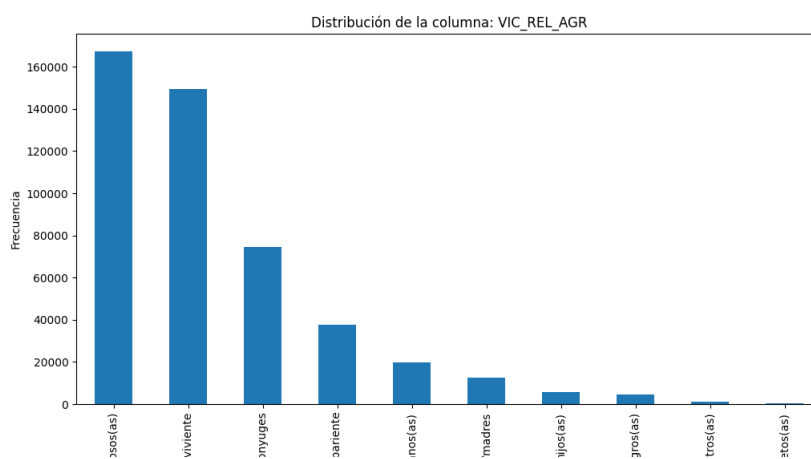


Figura 49: Frecuencia de la columna VICRELAGR

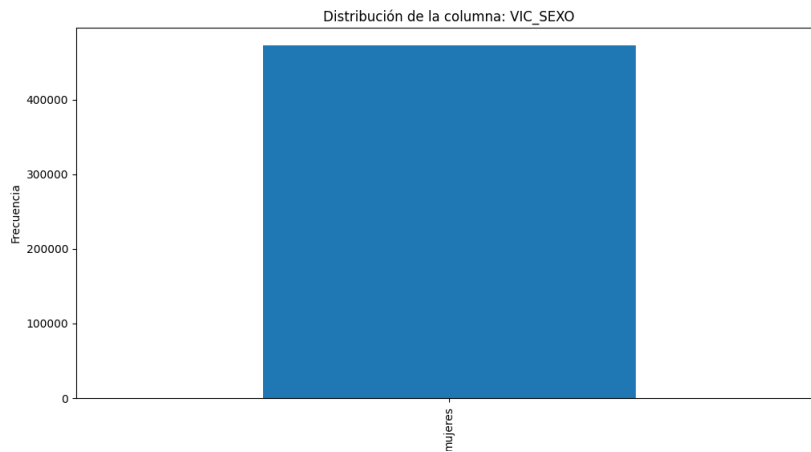


Figura 50: Frecuencia de la columna VICSEXO

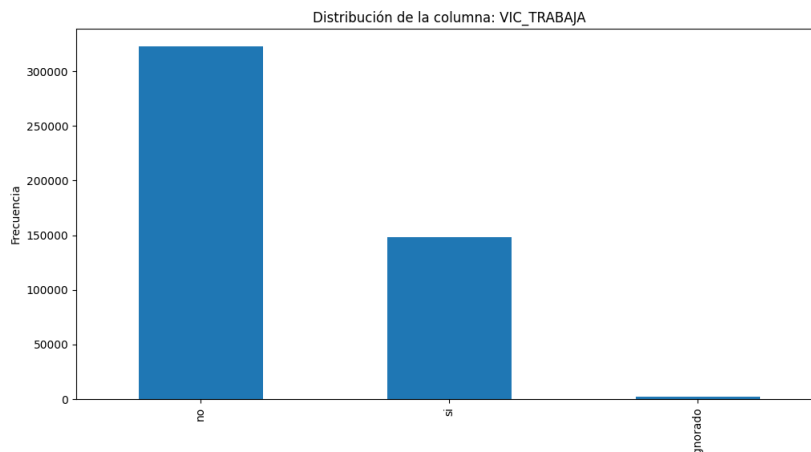


Figura 51: Frecuencia de la columna VICTRABAJA

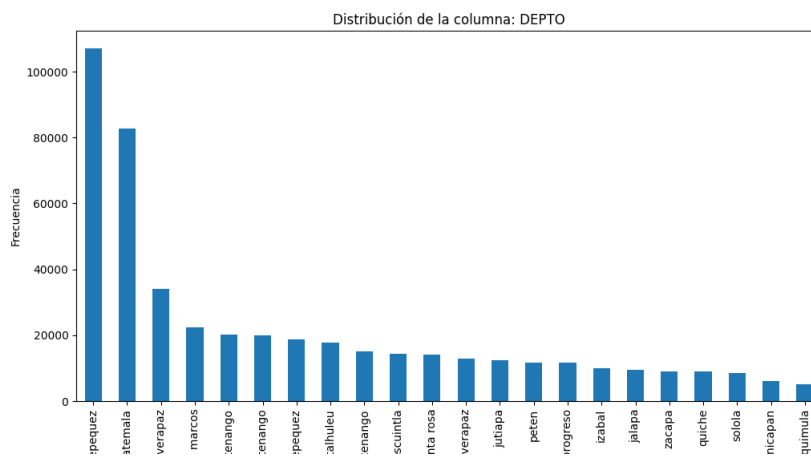


Figura 52: Frecuencia de la columna DEPTO

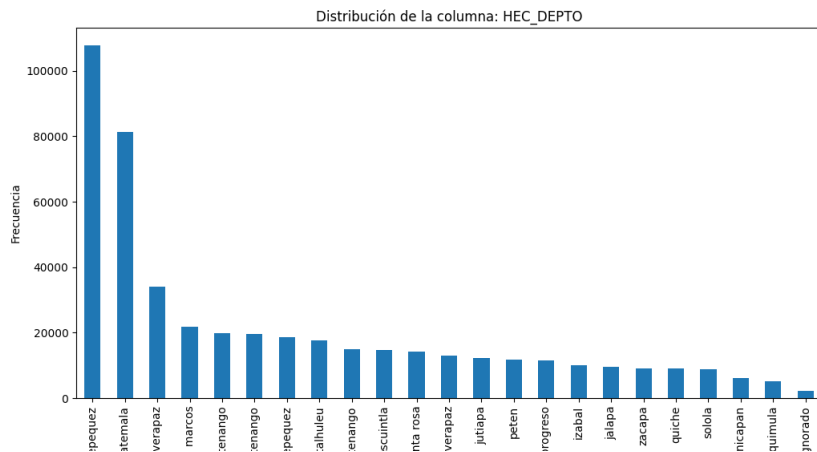


Figura 53: Frecuencia de la columna HECDEPTO

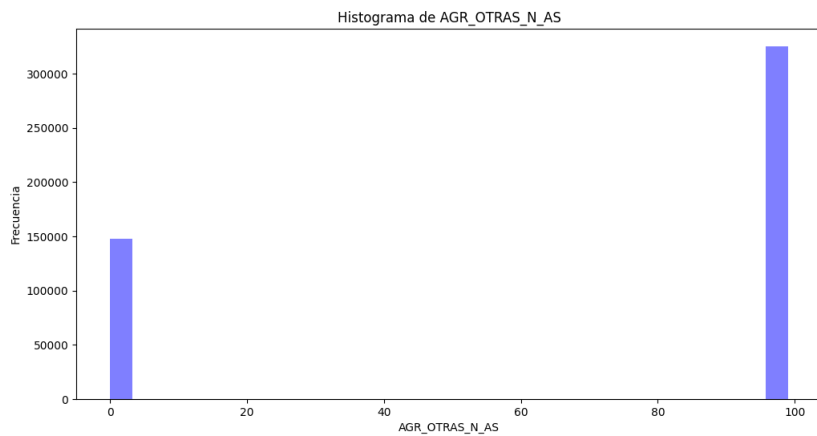


Figura 54: Frecuencia de la columna AGROTRASNAS

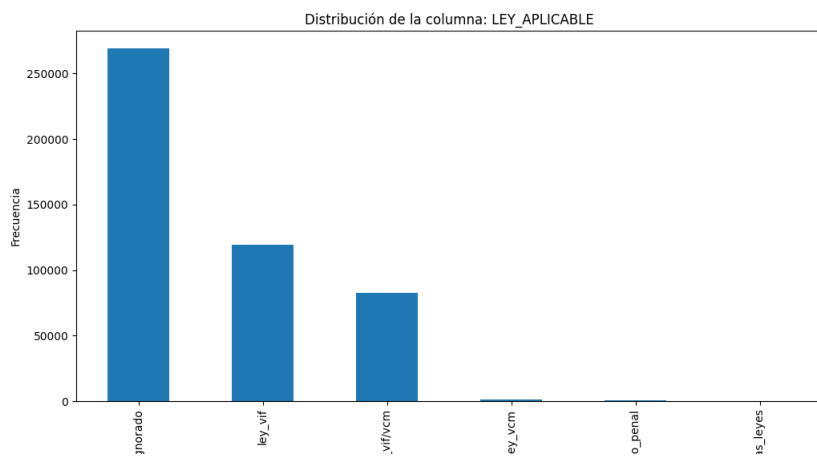


Figura 55: Frecuencia de la columna LEYAPLICABLE

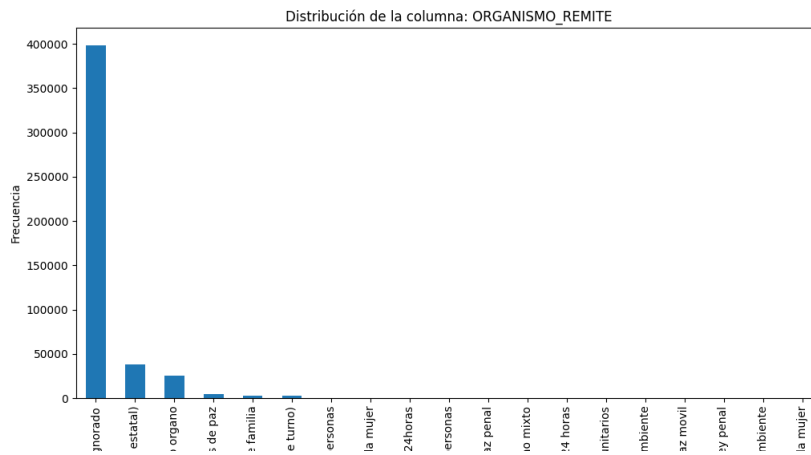


Figura 56: Frecuencia de la columna ORGANISMOREMITE

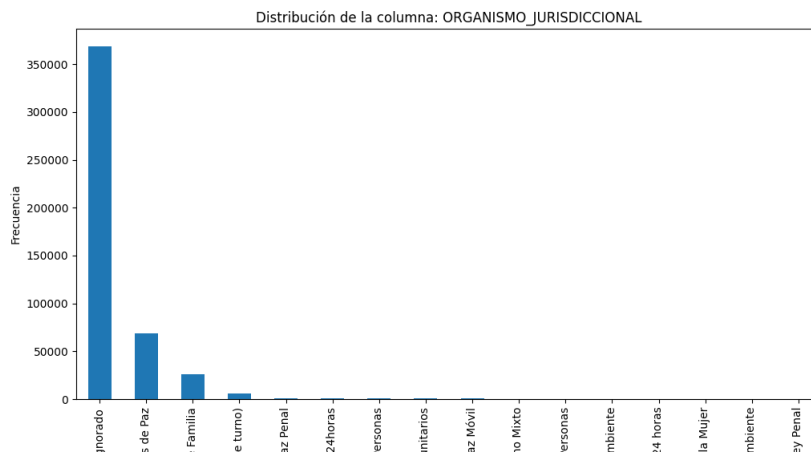


Figura 57: Frecuencia de la columna ORGANISMOJURISDICCIONAL

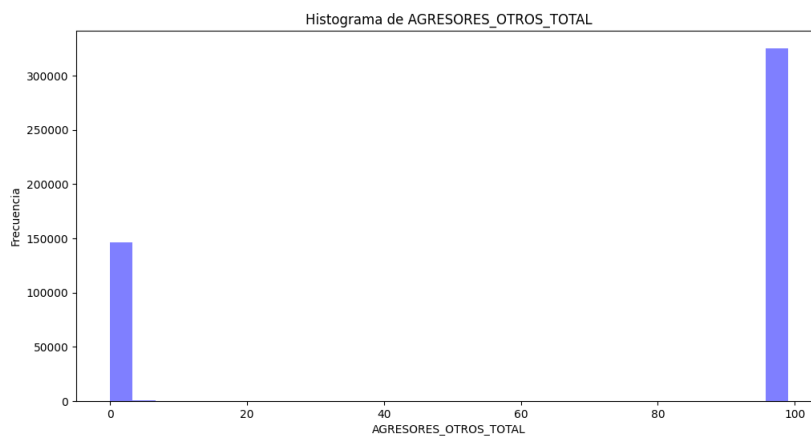


Figura 58: Frecuencia de la columna AGRESORESOTROSTOTAL

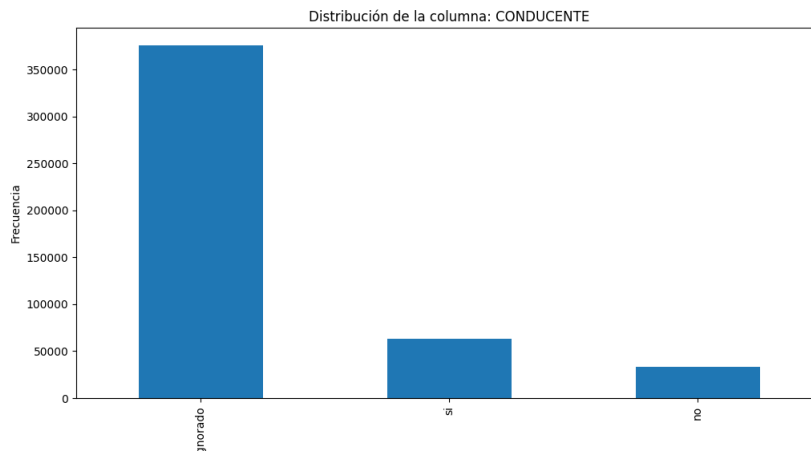


Figura 59: Frecuencia de la columna CONDUCTENTE

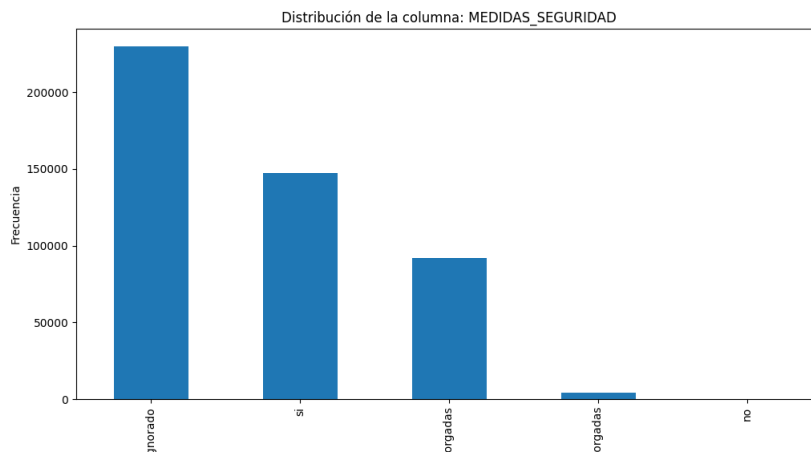


Figura 60: Frecuencia de la columna MEDIDASSEGURIDAD

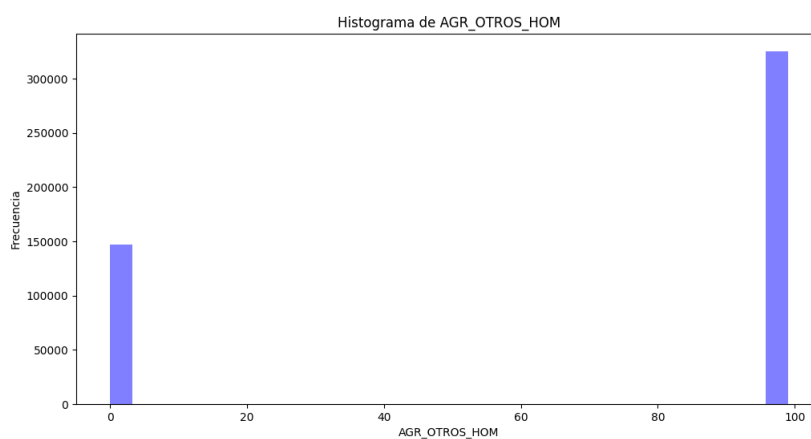


Figura 61: Frecuencia de la columna AGROTROSHOM

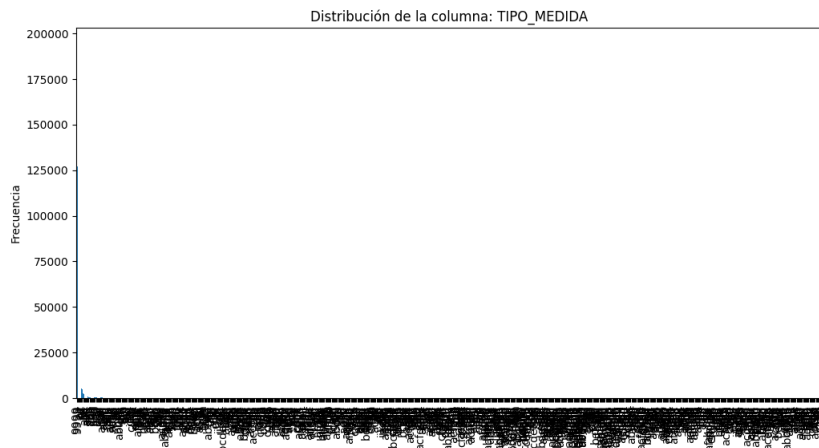


Figura 62: Frecuencia de la columna TIPOMEDIDA

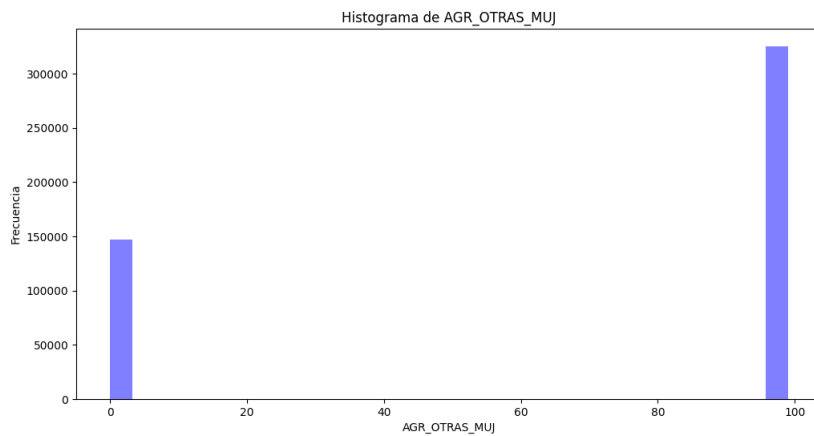


Figura 63: Frecuencia de la columna AGROTRASMUJ

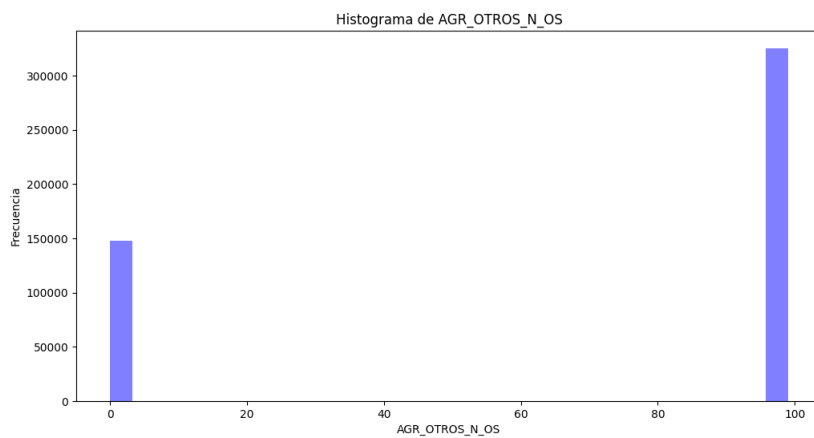


Figura 64: Frecuencia de la columna AGROTROSNOS

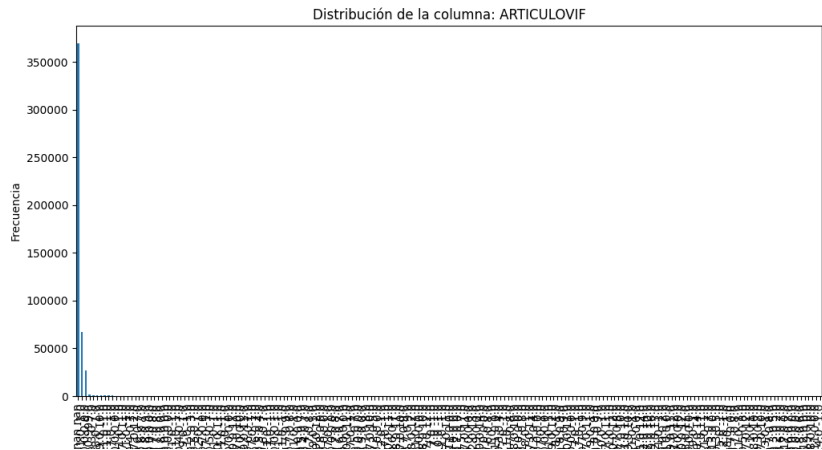


Figura 65: Frecuencia de la columna ARTICULOVIF

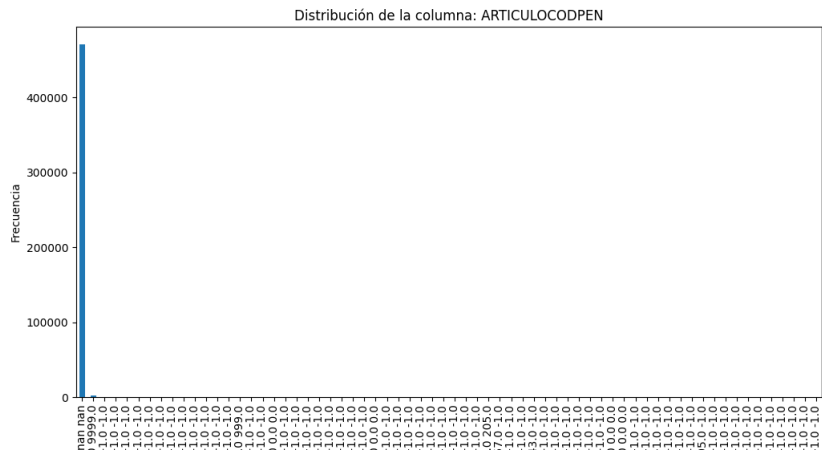


Figura 66: Frecuencia de la columna ARTICULOCODPEN

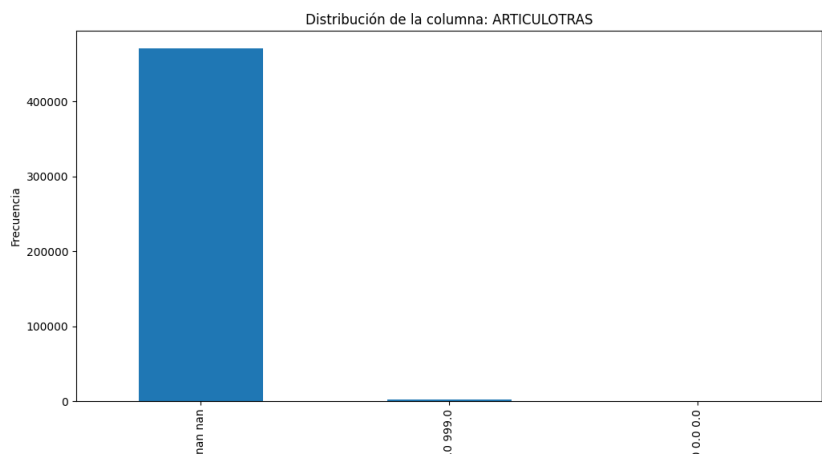


Figura 67: Frecuencia de la columna ARTICULOTRAS

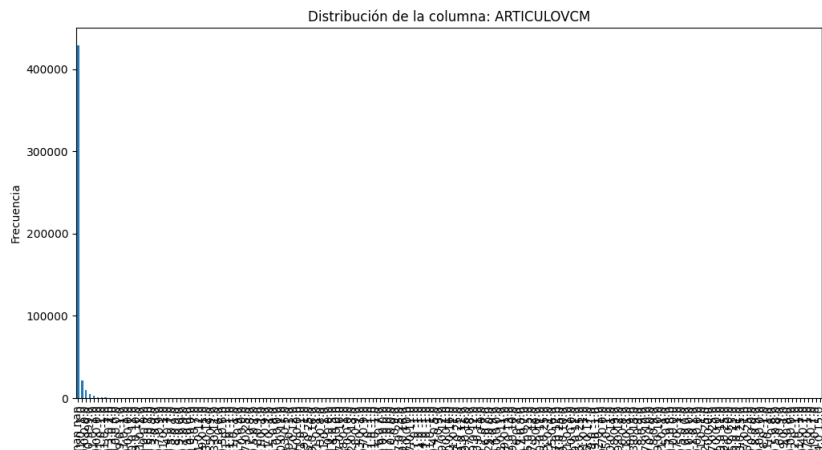


Figura 68: Frecuencia de la columna ARTICULOVCM

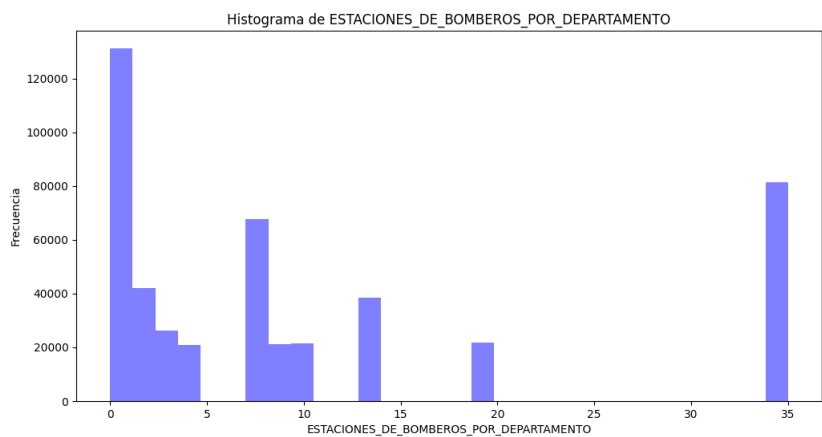


Figura 69: Frecuencia de la columna ESTACIONESDEBOMBEROSPOR-DEPARTAMENTO

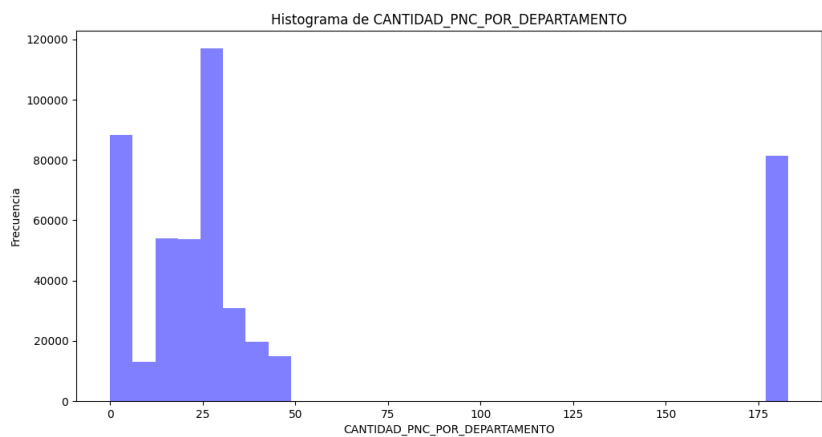


Figura 70: Frecuencia de la columna CANTIDADPNCPORDEPARTAMENTO

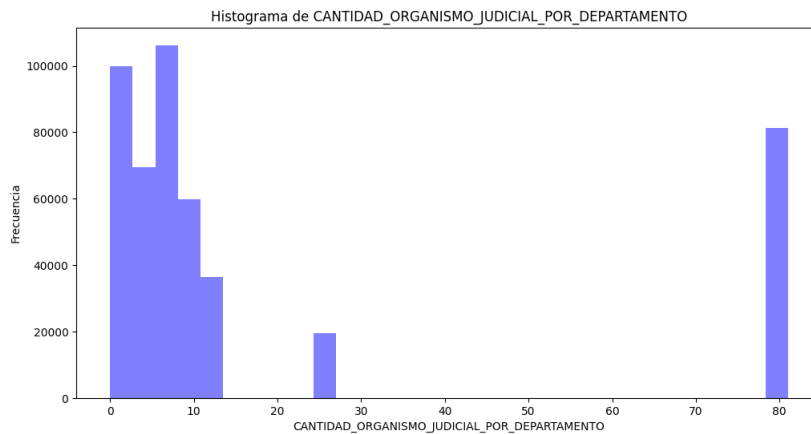


Figura 71: Frecuencia de la columna CANTIDADORGANISMOJUDICIAL-PORDEPARTAMENTO

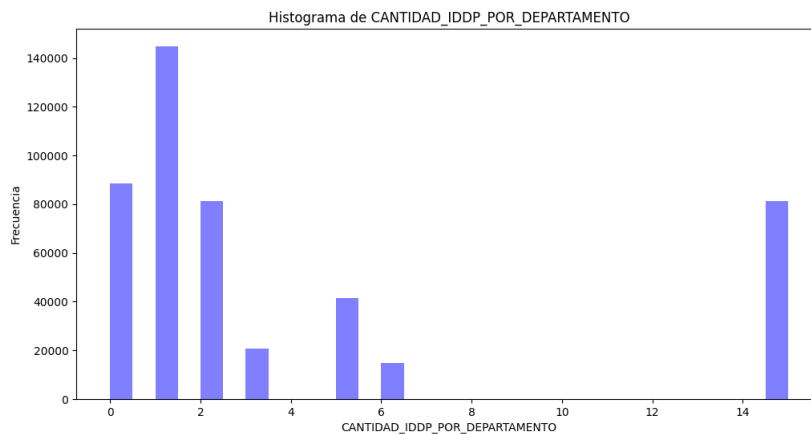


Figura 72: Frecuencia de la columna CANTIDADIDDP_PORDEPARTAMENTO

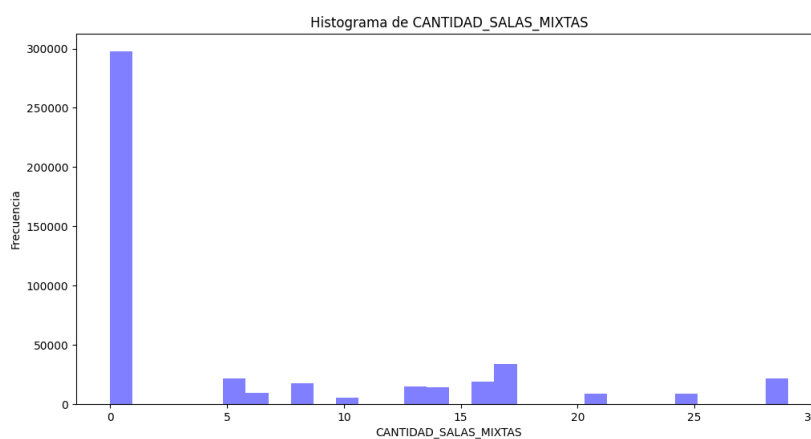


Figura 73: Frecuencia de la columna CANTIDAD_SALAS_MIXTAS

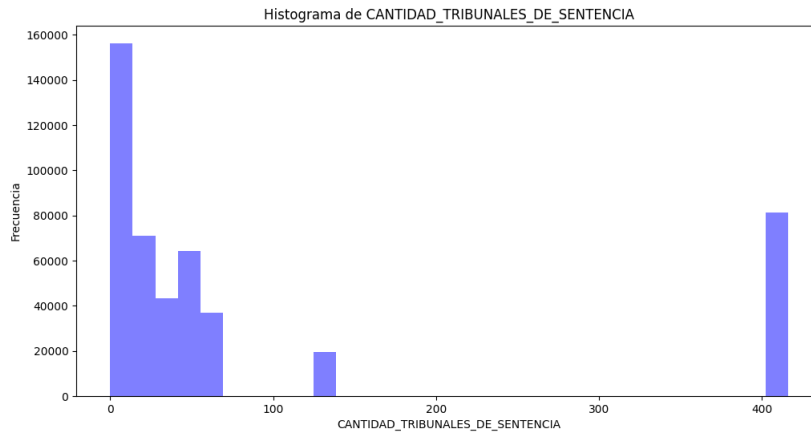


Figura 74: Frecuencia de la columna CANTIDADTRIBUNALESDESENTENCIA

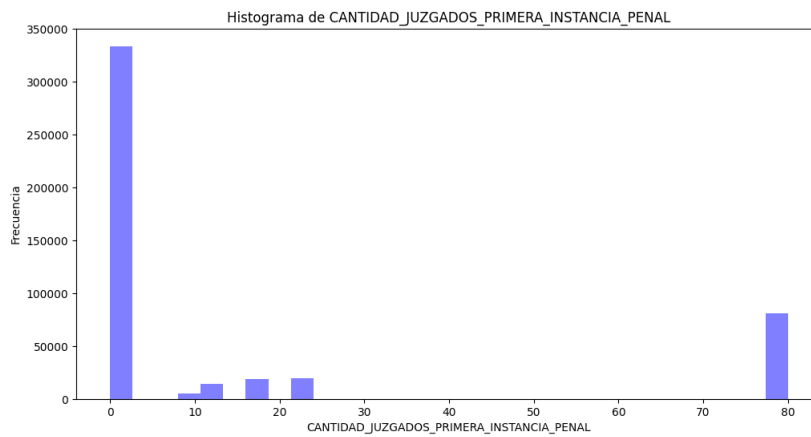


Figura 75: Frecuencia de la columna CANTIDADJUZGADOSPRIMERAINSTANCIAPENAL

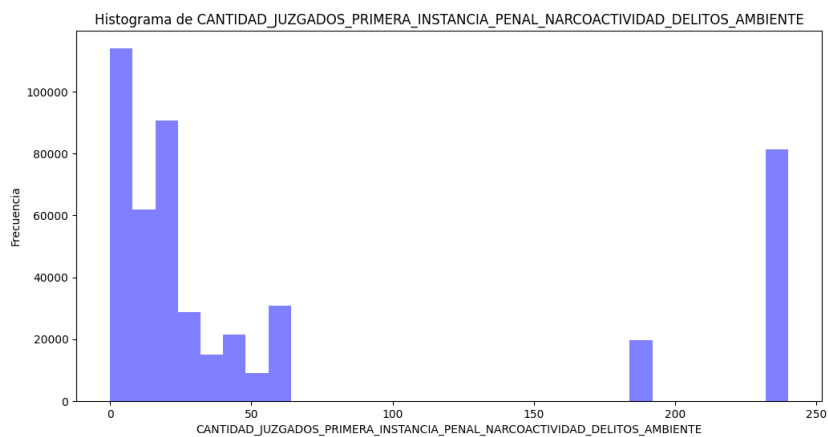


Figura 76: Frecuencia de la columna CANTIDADJUZGADOSPRIMERAINSTANCIAPENALNARCOACTIVIDADDELITOSAMBIENTE

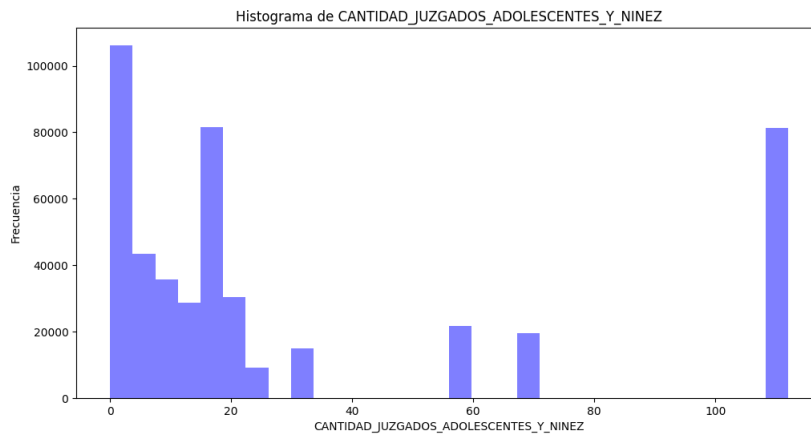


Figura 77: Frecuencia de la columna CANTIDADJUZGADOSADOLESCENTESYNINEZ

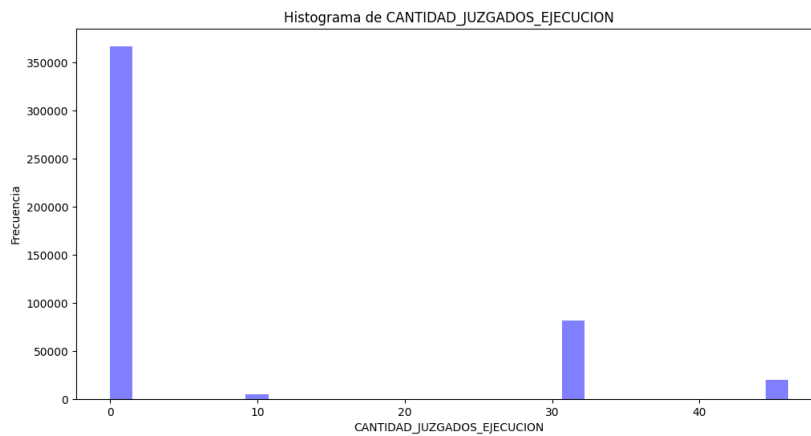


Figura 78: Frecuencia de la columna CANTIDADJUZGADOSEJECUCION

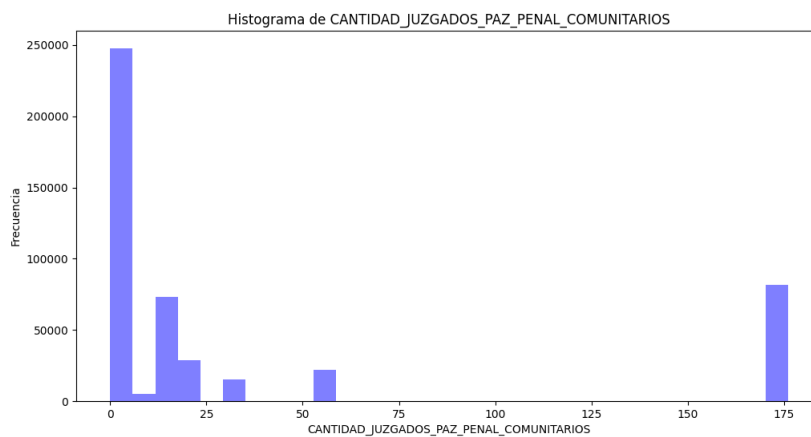


Figura 79: Frecuencia de la columna CANTIDADJUZGADOSPAZPENALCOMUNITARIOS

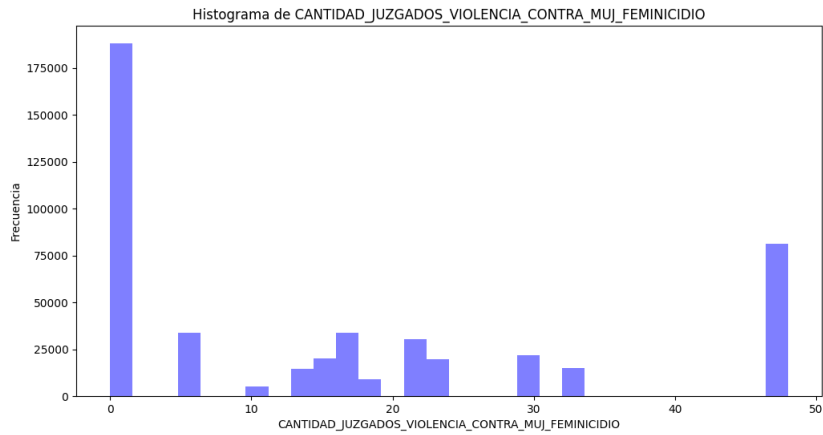


Figura 80: Frecuencia de la columna CANTIDADJUZGADOSVIOLENCIA-CONTRAMUJFEMINICIDIO

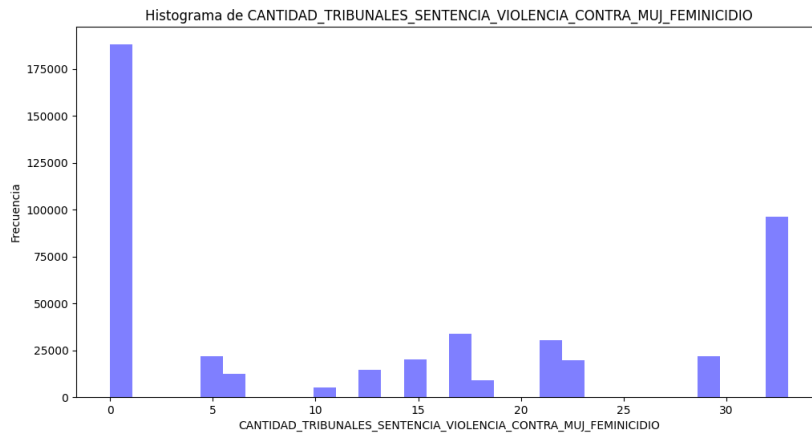


Figura 81: Frecuencia de la columna CANTIDADTRIBUNALESSENTENCIAVIOLENCIACONTRAMUJFEMINICIDIO

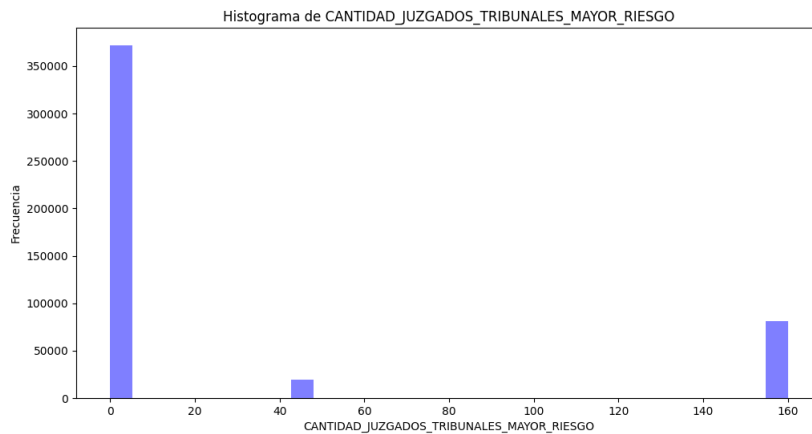


Figura 82: Frecuencia de la columna CANTIDADJUZGADOSTRIBUNALESMAJORRIESGO

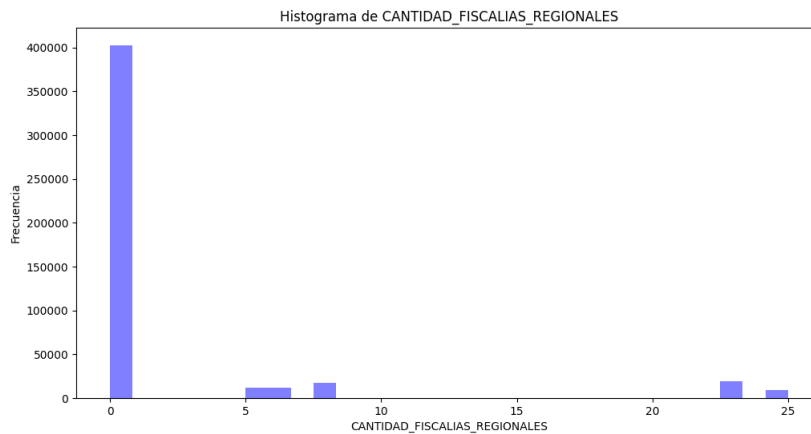


Figura 83: Frecuencia de la columna CANTIDADFISCALIASREGIONALES

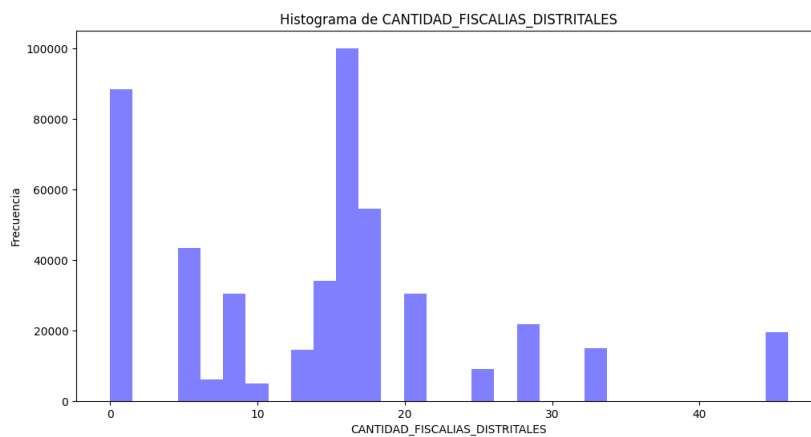


Figura 84: Frecuencia de la columna CANTIDADFISCALIASDISTRIALES

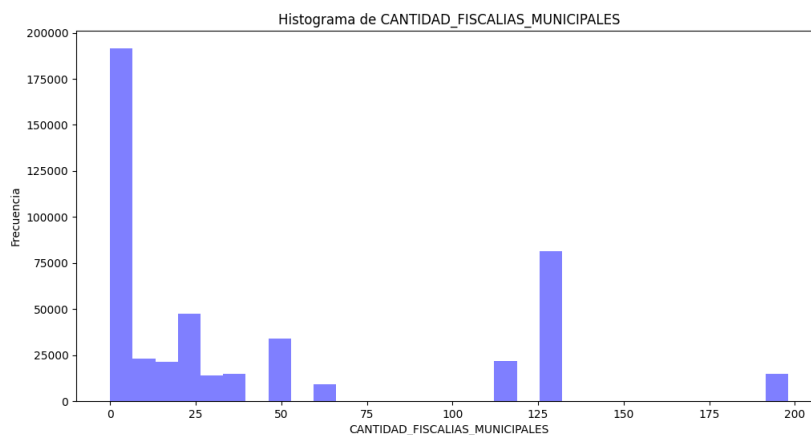


Figura 85: Frecuencia de la columna CANTIDADFISCALIASMUNICIPALES

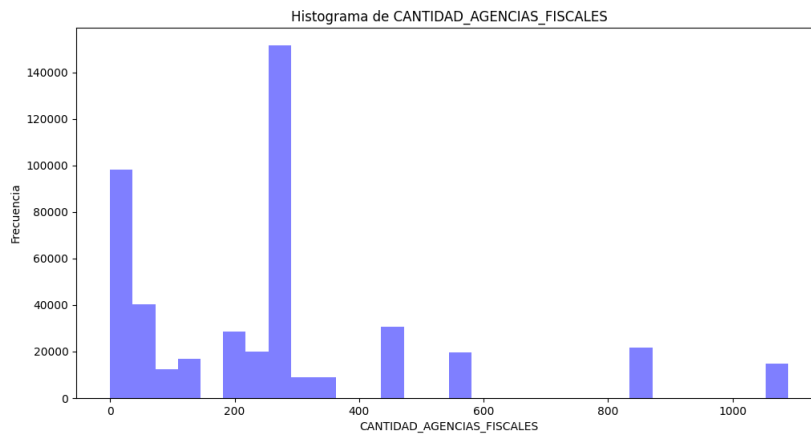


Figura 86: Frecuencia de la columna CANTIDADAGENCIASFISCALES

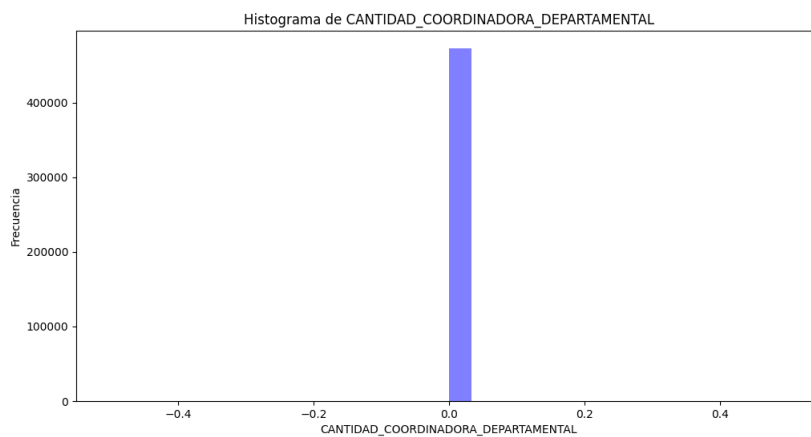


Figura 87: Frecuencia de la columna CANTIDADCOORDINADORADEPARTAMENTAL

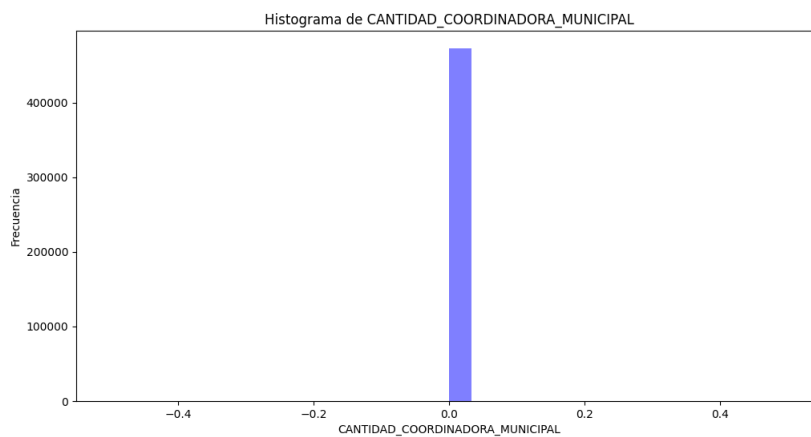


Figura 88: Frecuencia de la columna CANTIDADCOORDINADORAMUNICIPAL

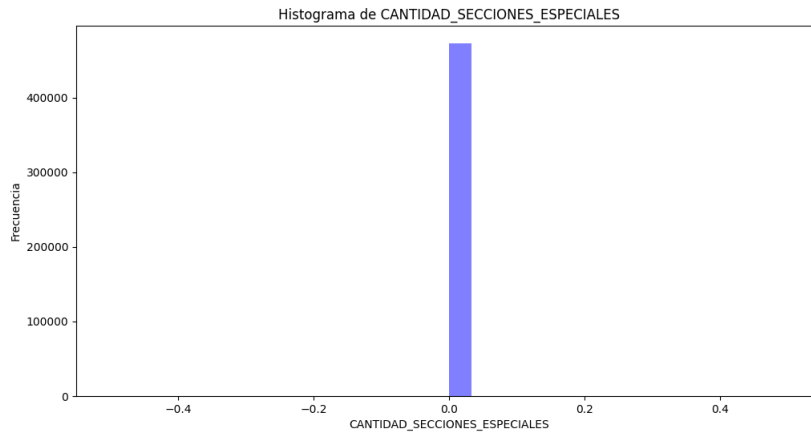


Figura 89: Frecuencia de la columna CANTIDADSECCIONESESPECIALES

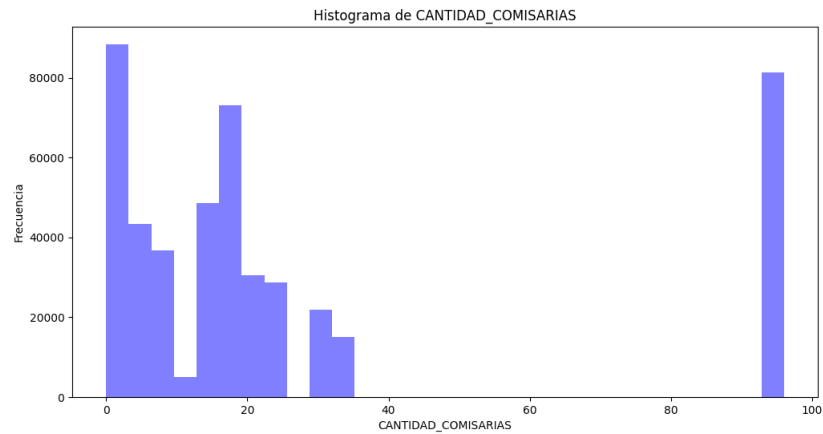


Figura 90: Frecuencia de la columna CANTIDADCOMISARIAS

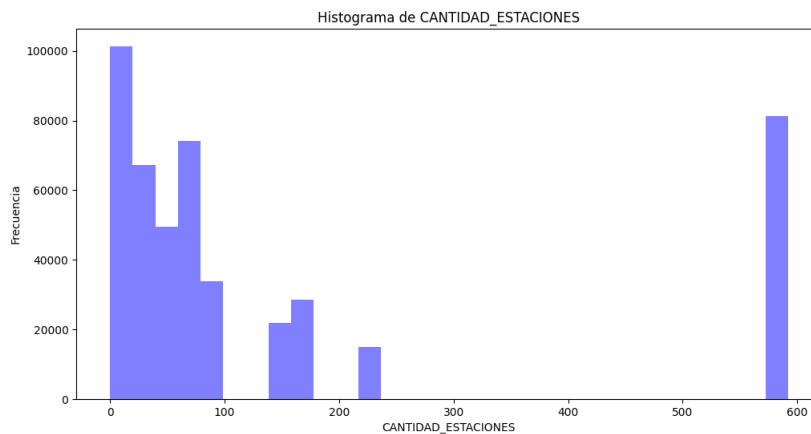


Figura 91: Frecuencia de la columna CANTIDADESTACIONES

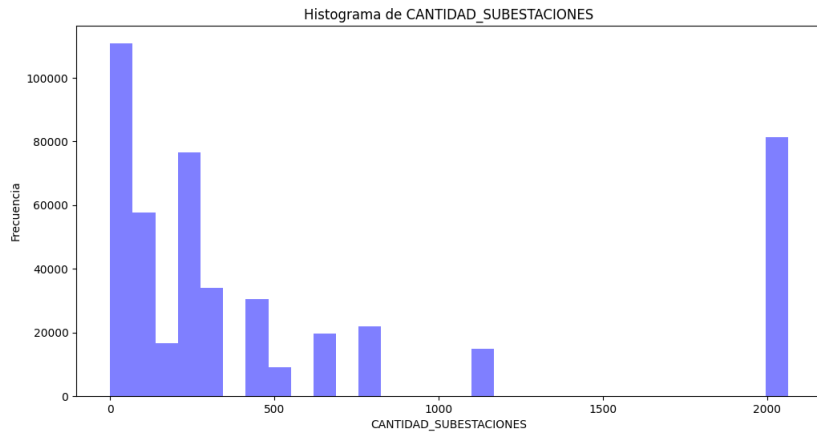


Figura 92: Frecuencia de la columna CANTIDADSUBESTACIONES

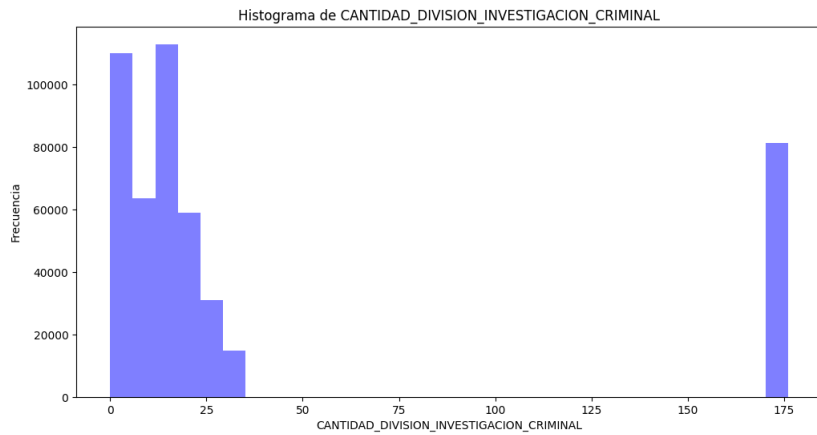


Figura 93: Frecuencia de la columna CANTIDADDIVISIONINVESTIGACIONCRIMINAL

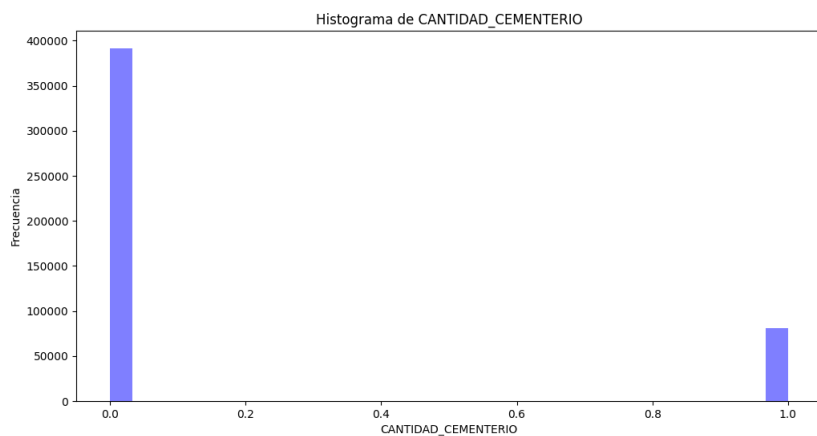


Figura 94: Frecuencia de la columna CANTIDADCEMENTERIO

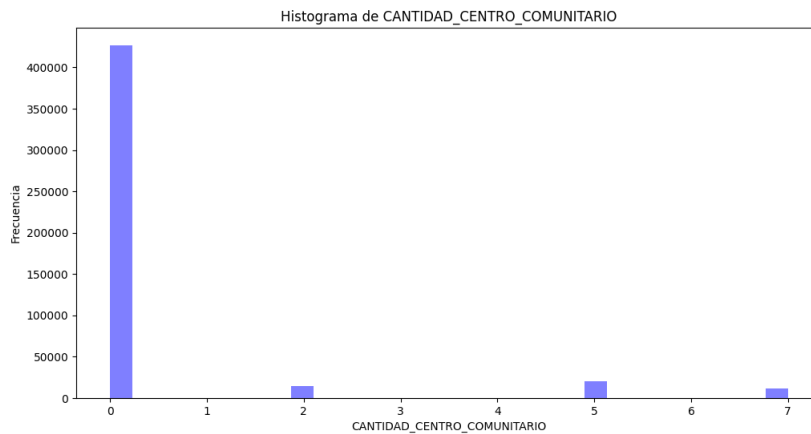


Figura 95: Frecuencia de la columna CANTIDADCENTROCOMUNITARIO

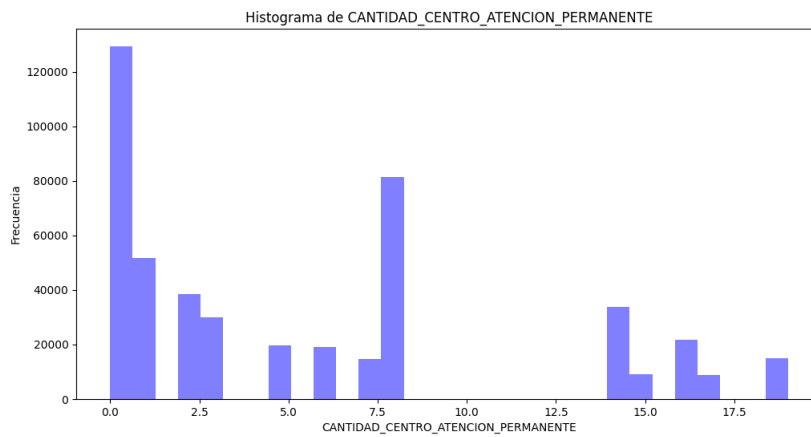


Figura 96: Frecuencia de la columna CANTIDADCENTROATENCION-PERMANENTE

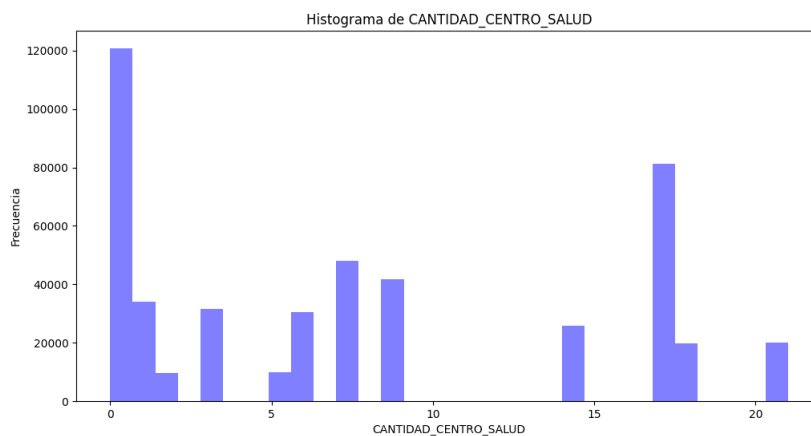


Figura 97: Frecuencia de la columna CANTIDADCENTROSALUD

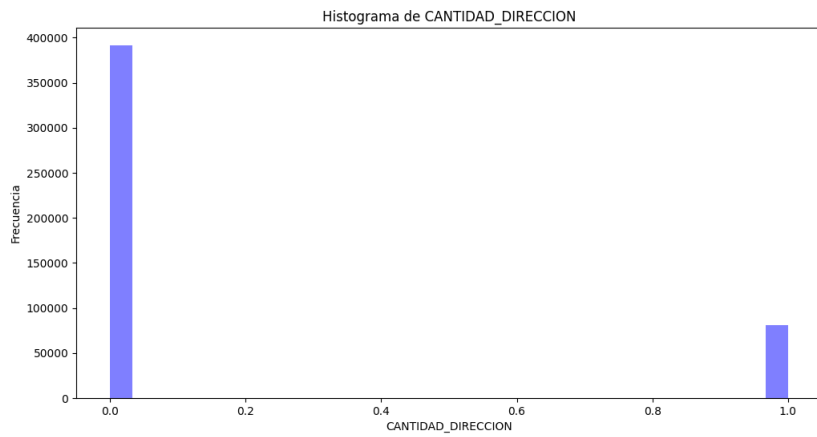


Figura 98: Frecuencia de la columna CANTIDADDIRECCION

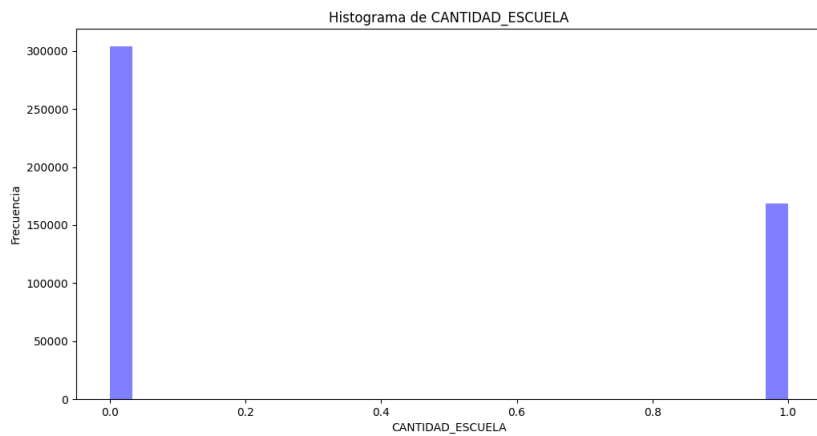


Figura 99: Frecuencia de la columna CANTIDADESCUELA

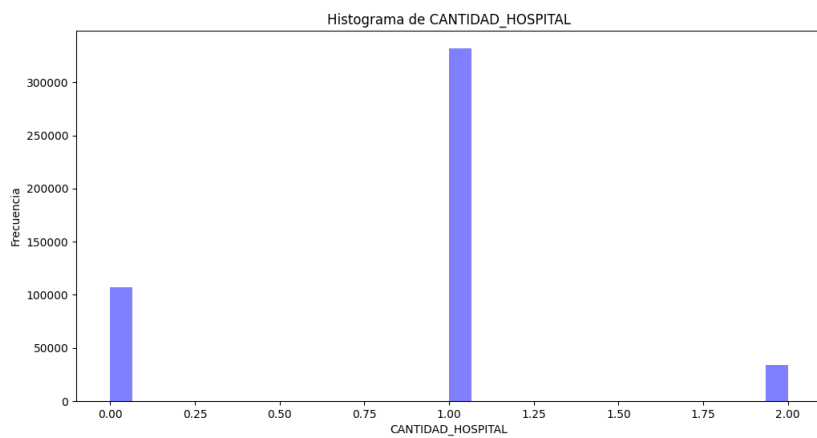


Figura 100: Frecuencia de la columna CANTIDADHOSPITAL

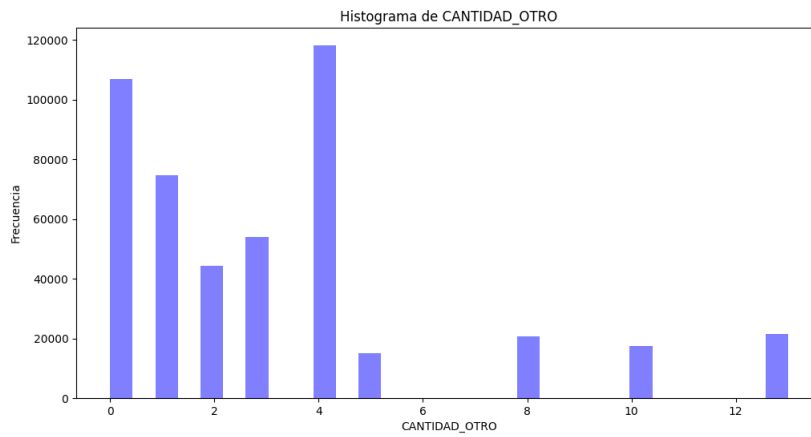


Figura 101: Frecuencia de la columna CANTIDADOTRO

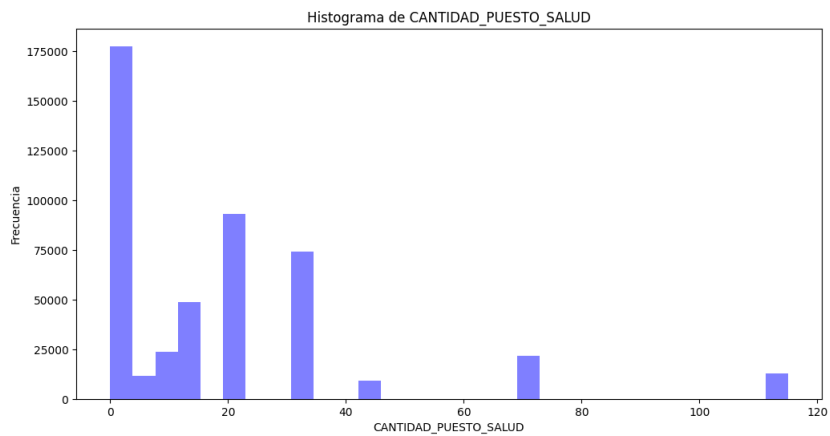


Figura 102: Frecuencia de la columna CANTIDADPUESTOSALUD

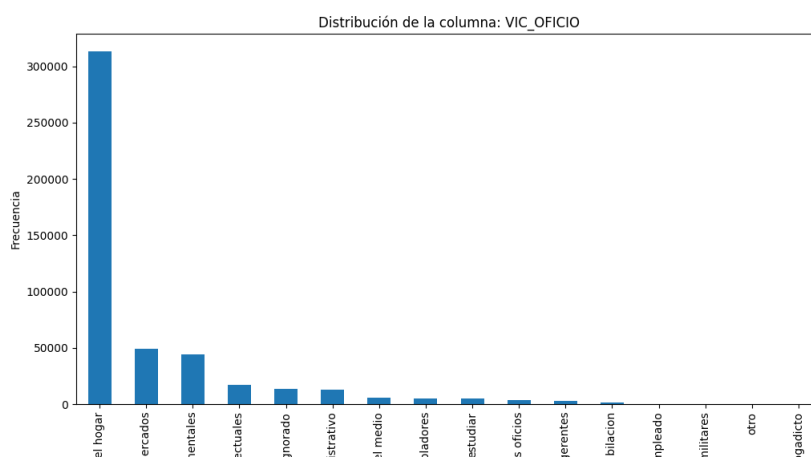


Figura 103: Frecuencia de la columna VICOFICIO

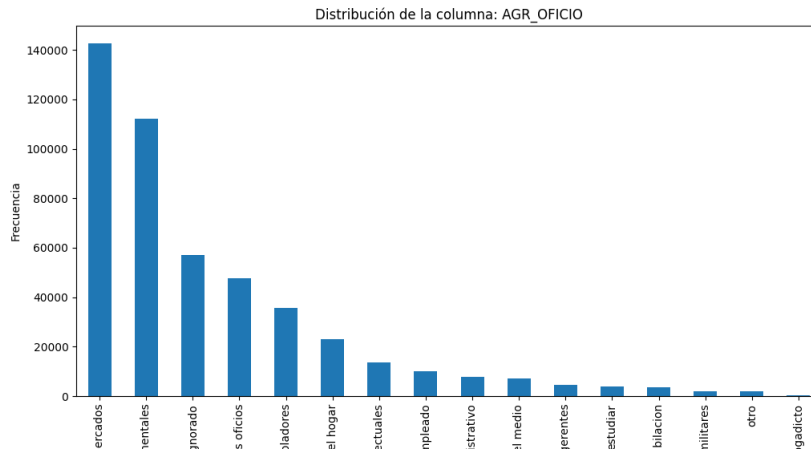


Figura 104: Frecuencia de la columna AGROFICIO

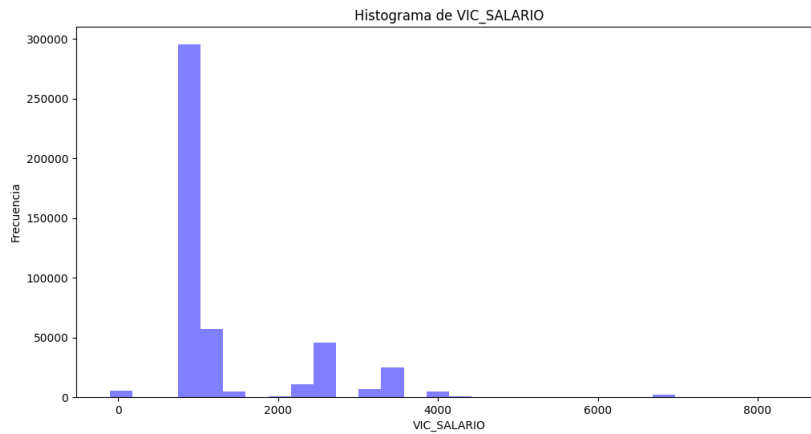


Figura 105: Frecuencia de la columna VICSALARIO

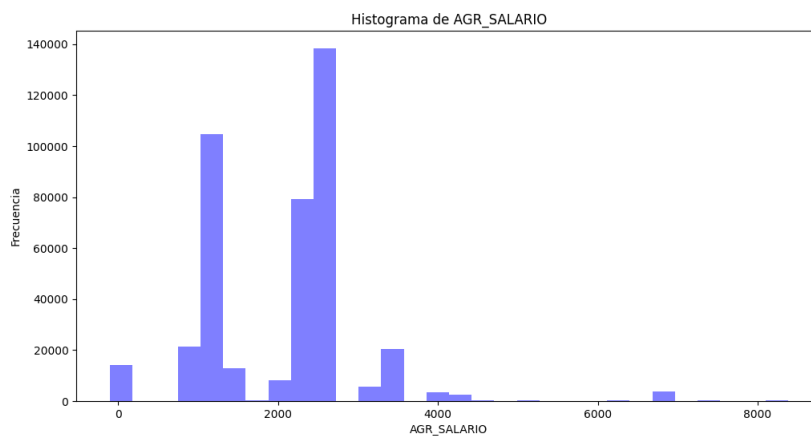


Figura 106: Frecuencia de la columna AGRSALARIO

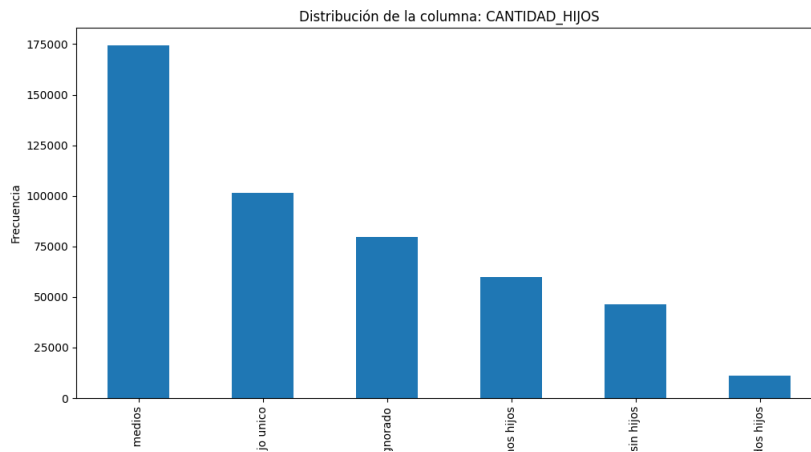


Figura 107: Frecuencia de la columna CANTIDADHIJOS

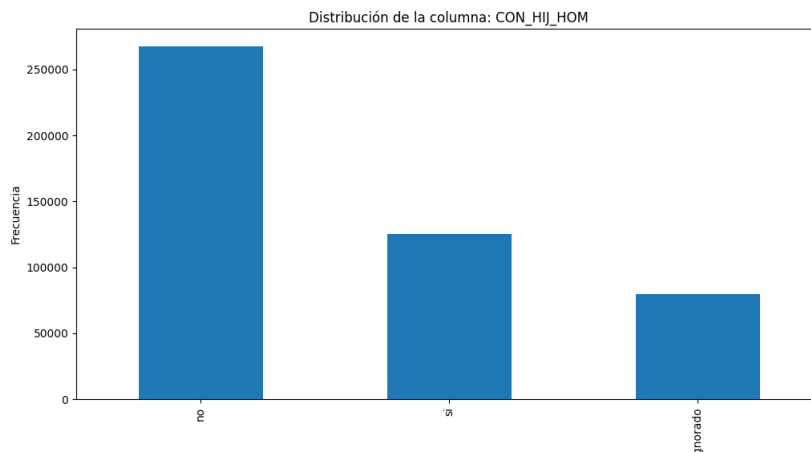


Figura 108: Frecuencia de la columna CONHIJHOM

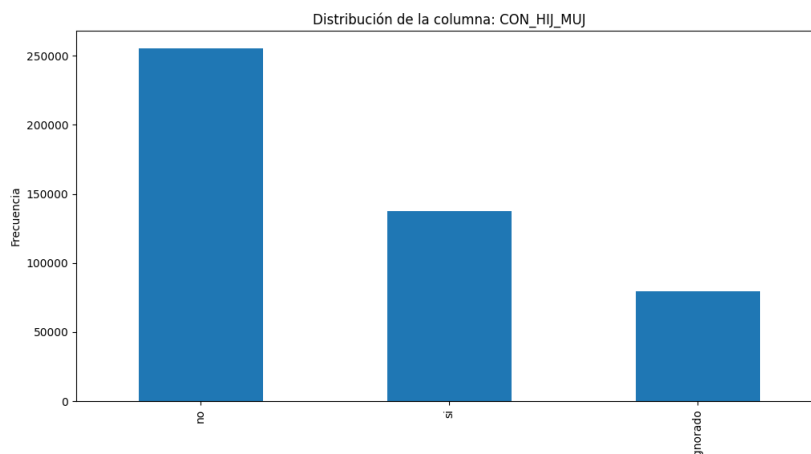


Figura 109: Frecuencia de la columna CONHIJMUIJ

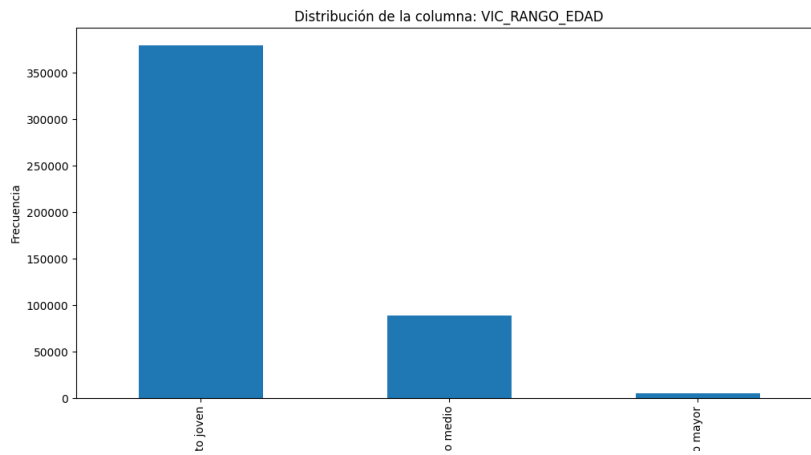


Figura 110: Frecuencia de la columna VICRANGOEDAD

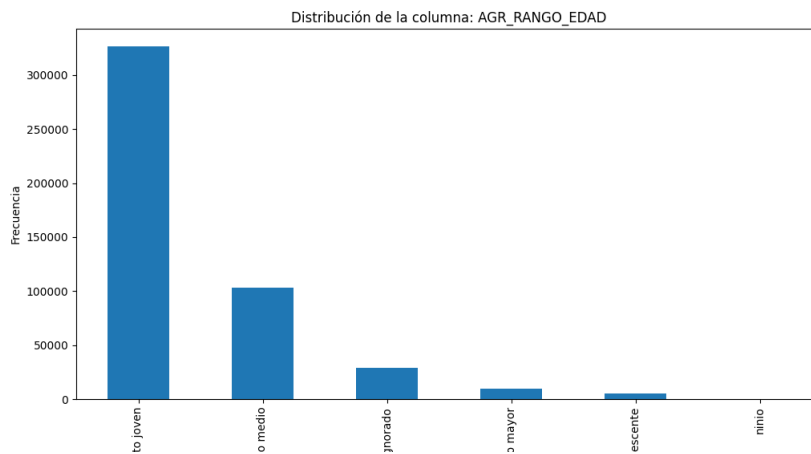


Figura 111: Frecuencia de la columna AGRRANGOEDAD

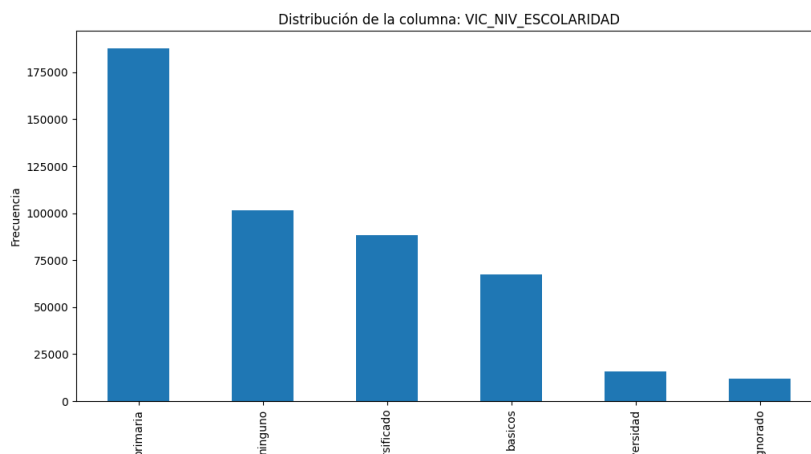


Figura 112: Frecuencia de la columna VICNIVESCOULARIDAD

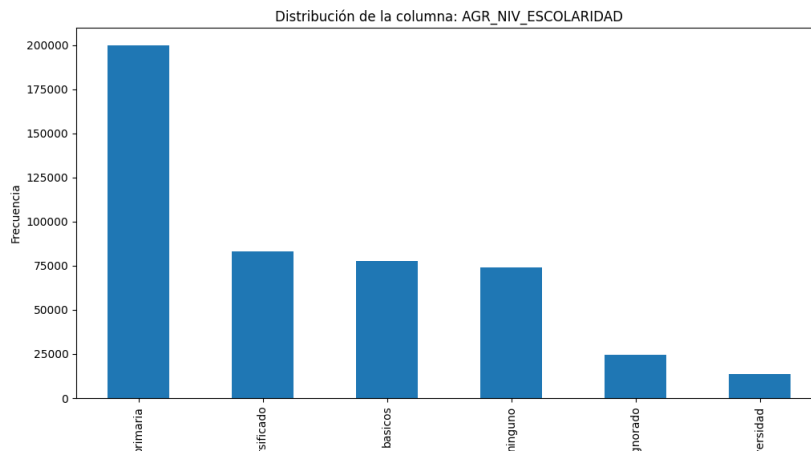


Figura 113: Frecuencia de la columna AGRNIVESCULARIDAD

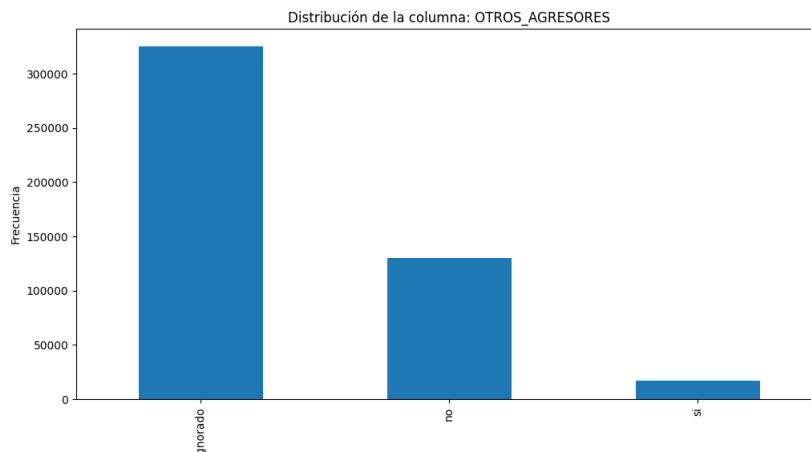


Figura 114: Frecuencia de la columna OTROSAGRESORES

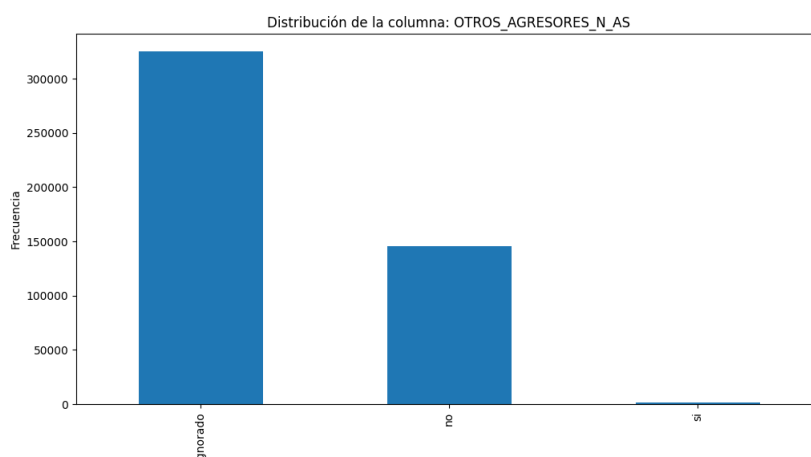


Figura 115: Frecuencia de la columna OTROSAGRESORESNAS

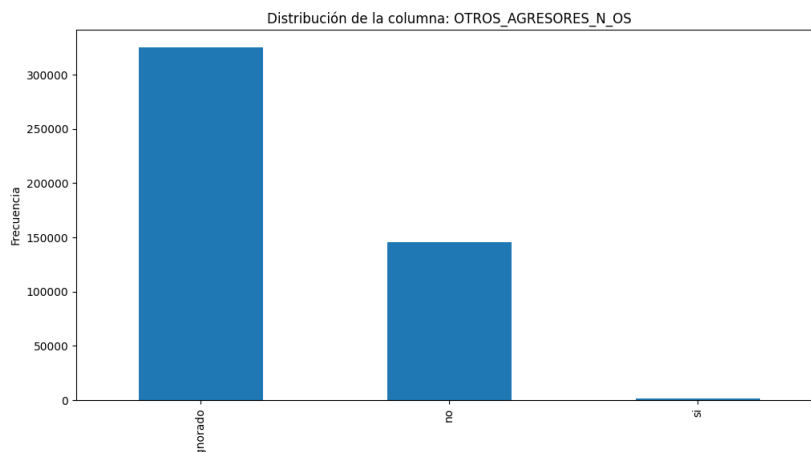


Figura 116: Frecuencia de la columna OTROSAGRESORESNOS

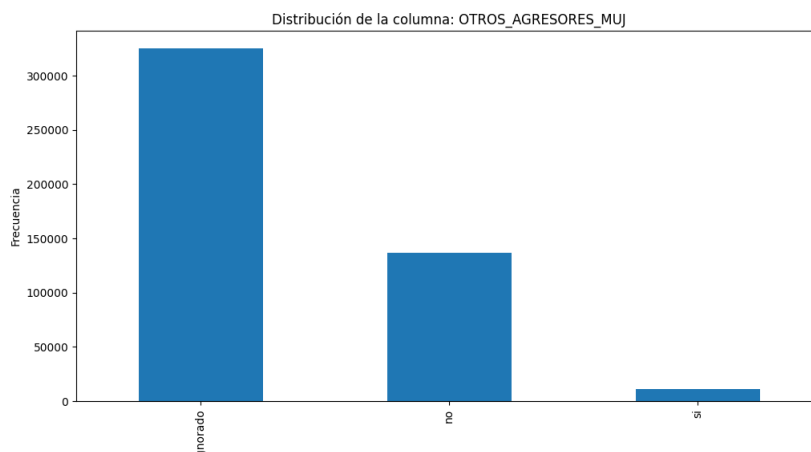


Figura 117: Frecuencia de la columna OTROSAGRESORESMUJ

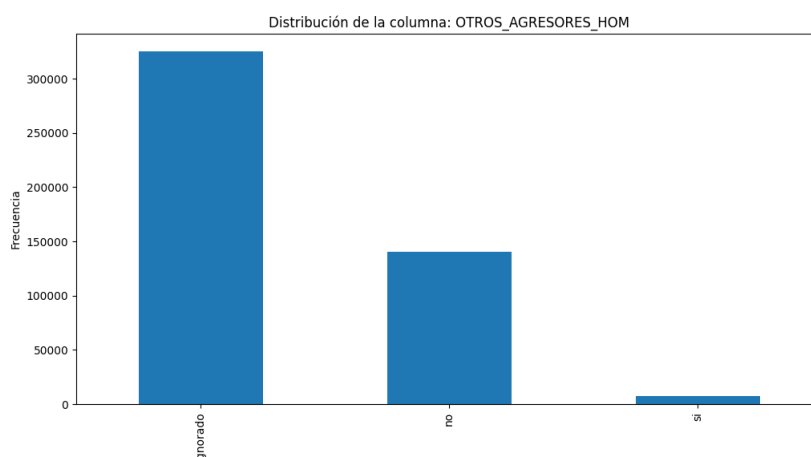


Figura 118: Frecuencia de la columna OTROSAGRESORESHOM

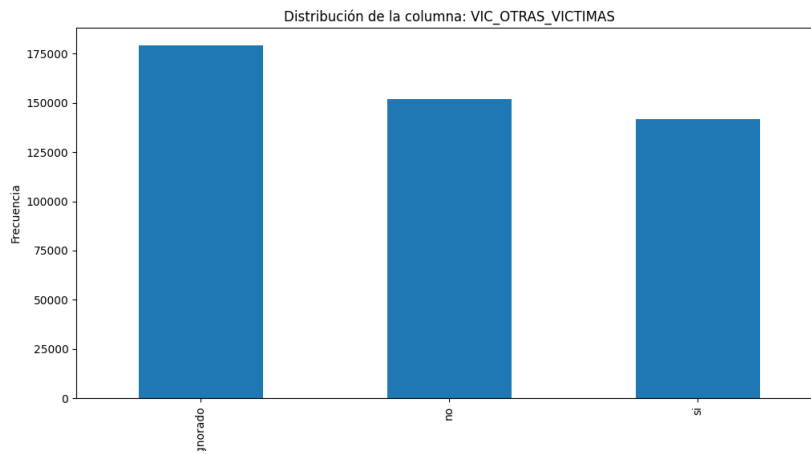


Figura 119: Frecuencia de la columna VICOTRASVICTIMAS

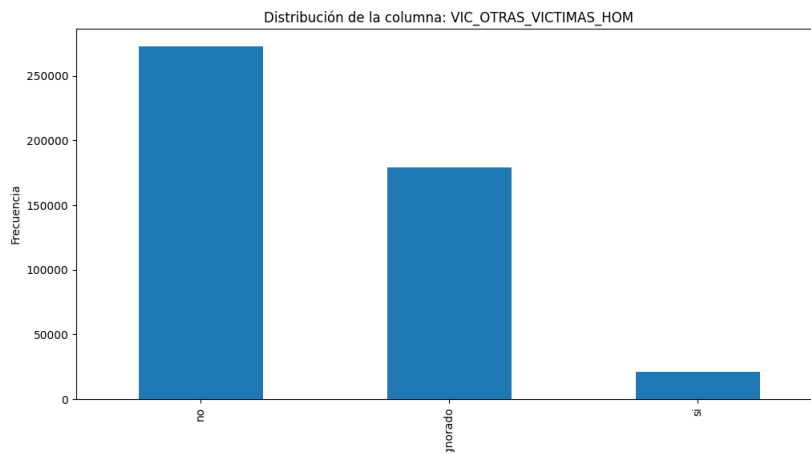


Figura 120: Frecuencia de la columna VICOTRASVICTIMASHOM

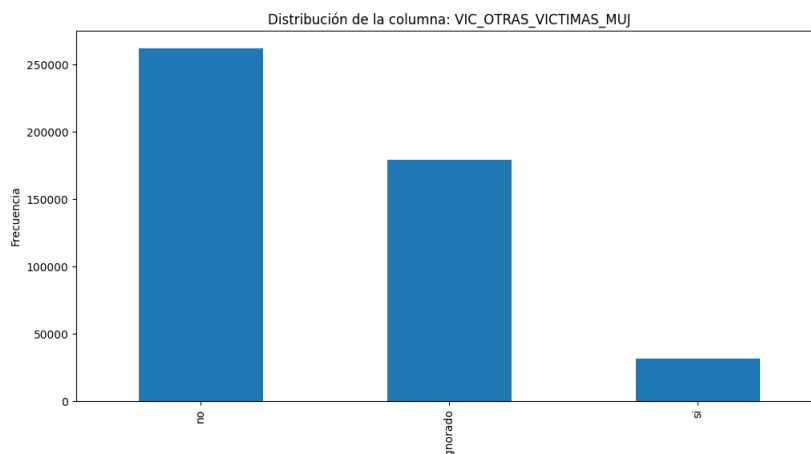


Figura 121: Frecuencia de la columna VICOTRASVICTIMASMUJ

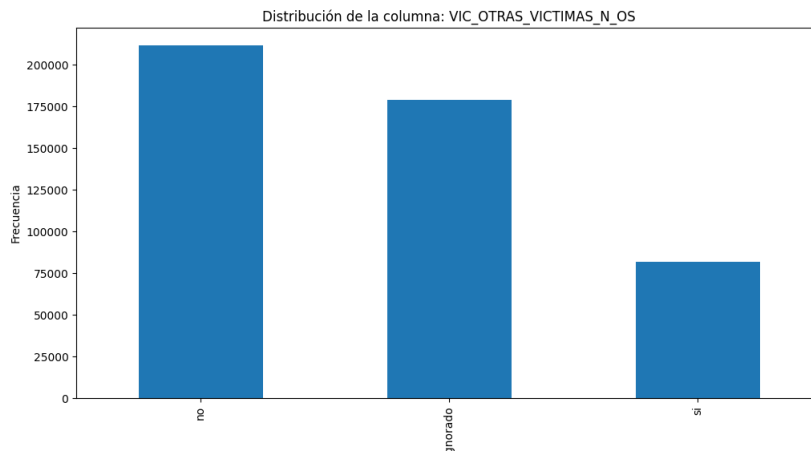


Figura 122: Frecuencia de la columna VICOTRASVICTIMASNOS

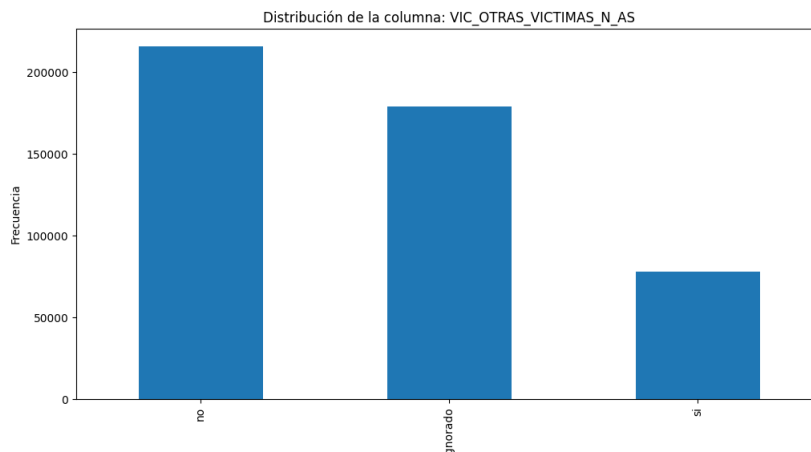


Figura 123: Frecuencia de la columna VICOTRASVICTIMASNAS

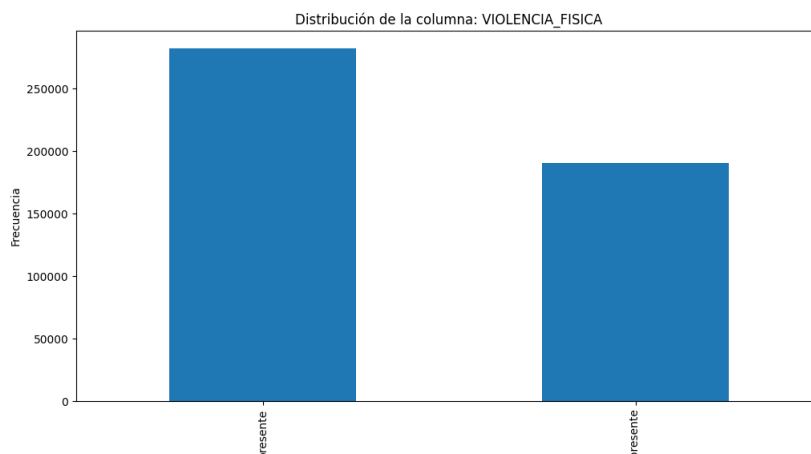


Figura 124: Frecuencia de la columna VIOLENCIAFISICA

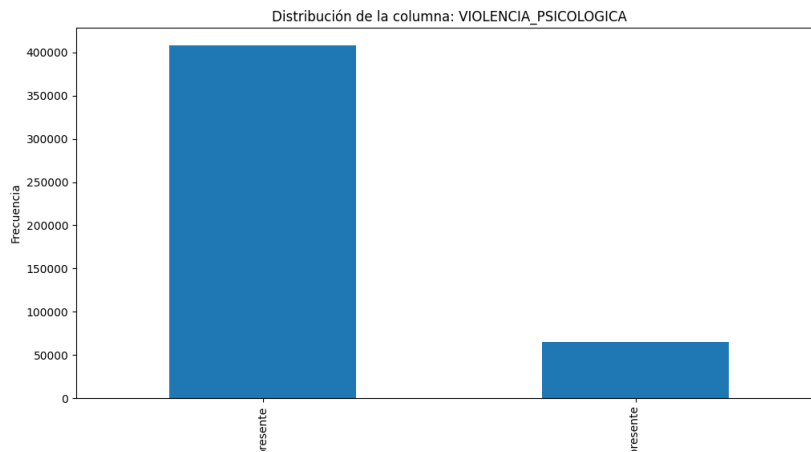


Figura 125: Frecuencia de la columna VIOLENCIAPSICOLOGICA

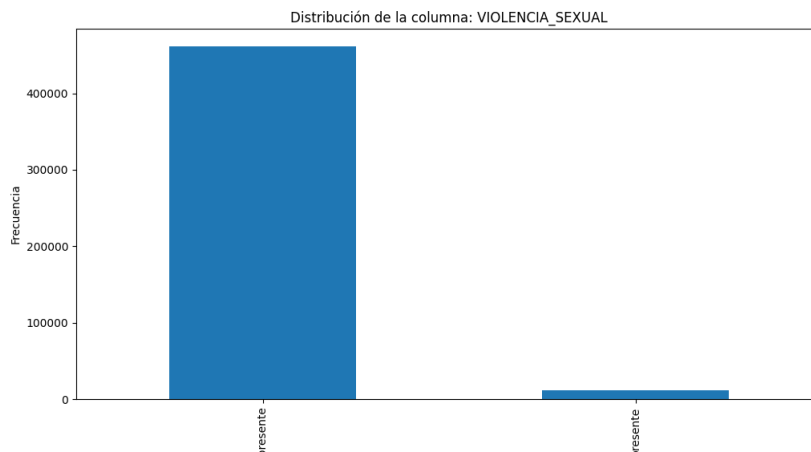


Figura 126: Frecuencia de la columna VIOLENCIASEXUAL

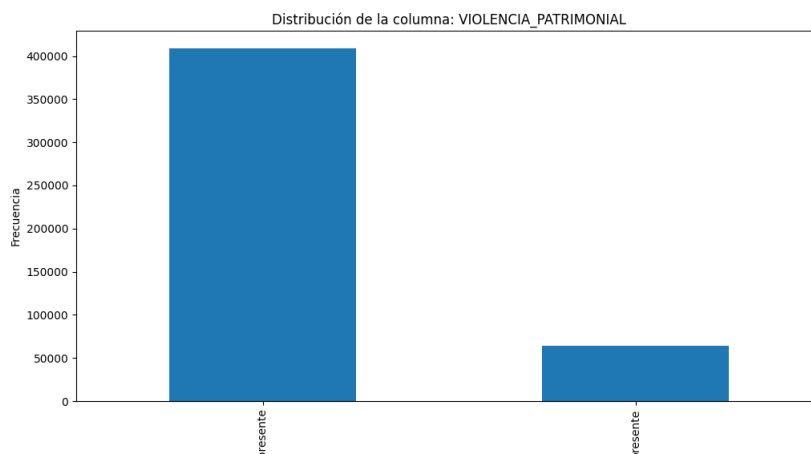


Figura 127: Frecuencia de la columna VIOLENCIAPATRIMONIAL

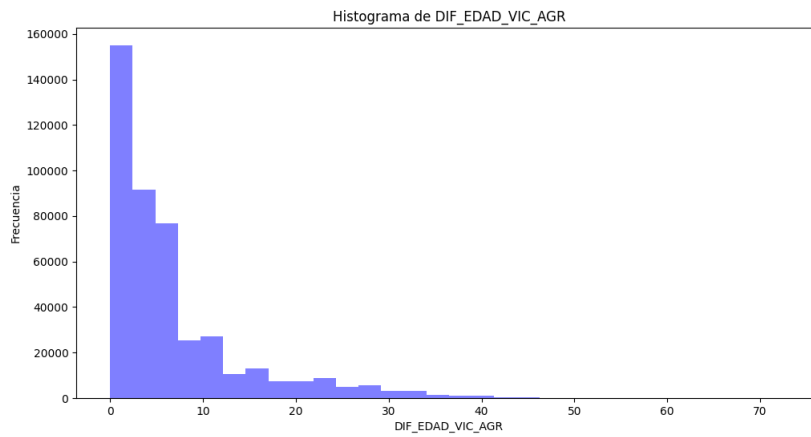


Figura 128: Frecuencia de la columna DIFEDADVICAGR

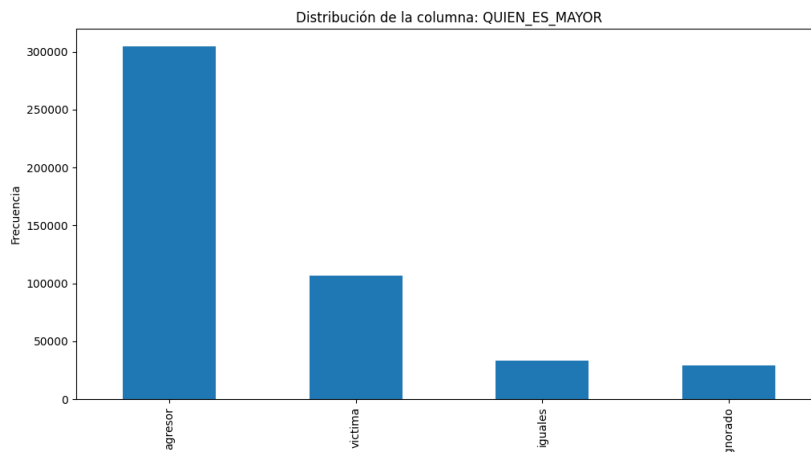


Figura 129: Frecuencia de la columna QUIENESMAYOR

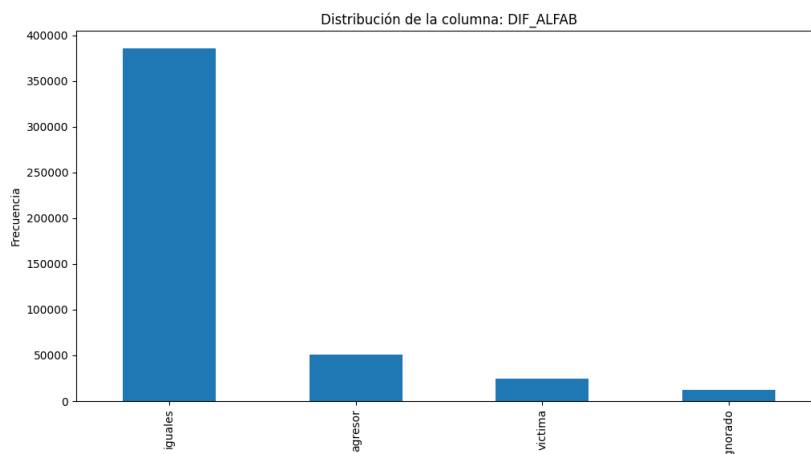


Figura 130: Frecuencia de la columna DIFALFAB

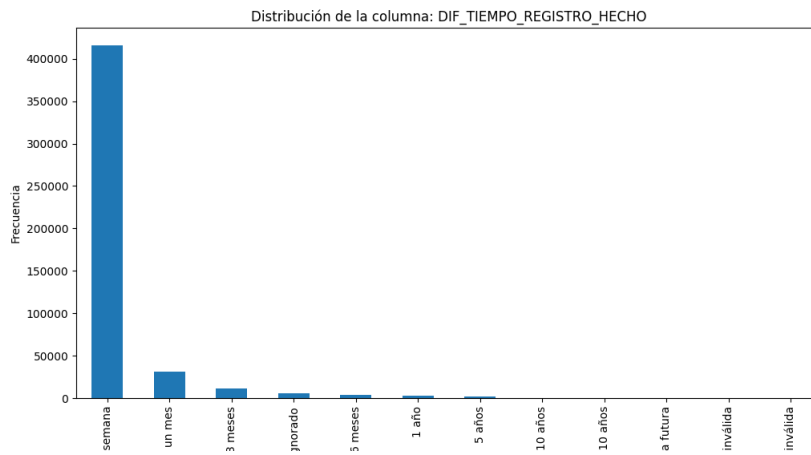


Figura 131: Frecuencia de la columna DIFTIEMPOREGISTROHECHO

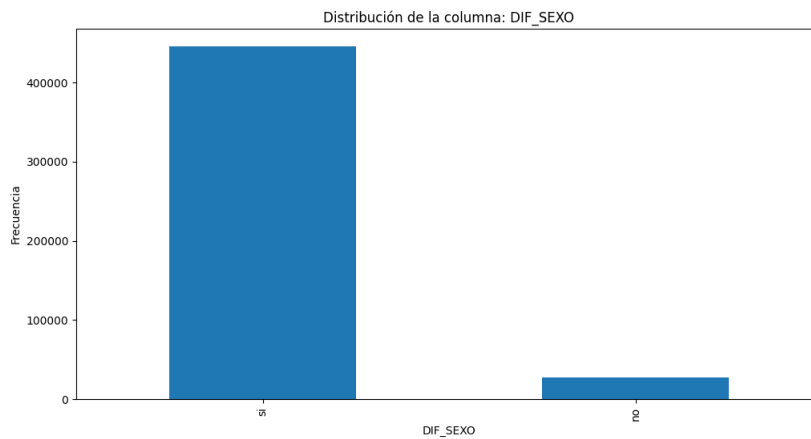


Figura 132: Frecuencia de la columna DIFSEXO

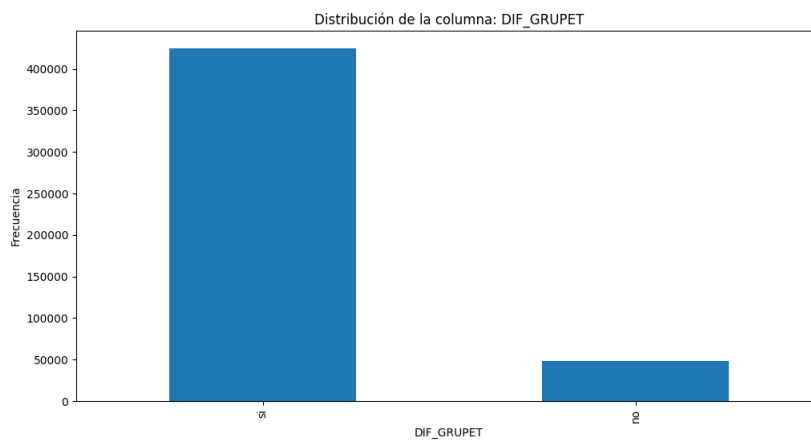


Figura 133: Frecuencia de la columna DIFGRUPET

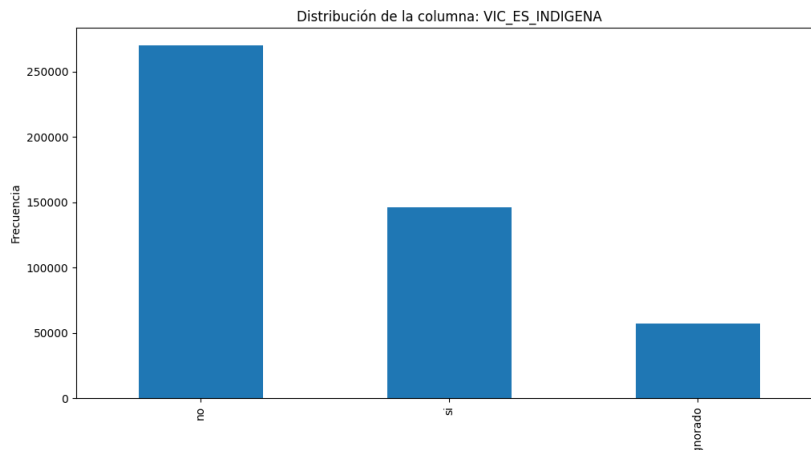


Figura 134: Frecuencia de la columna VICESINDIGENA

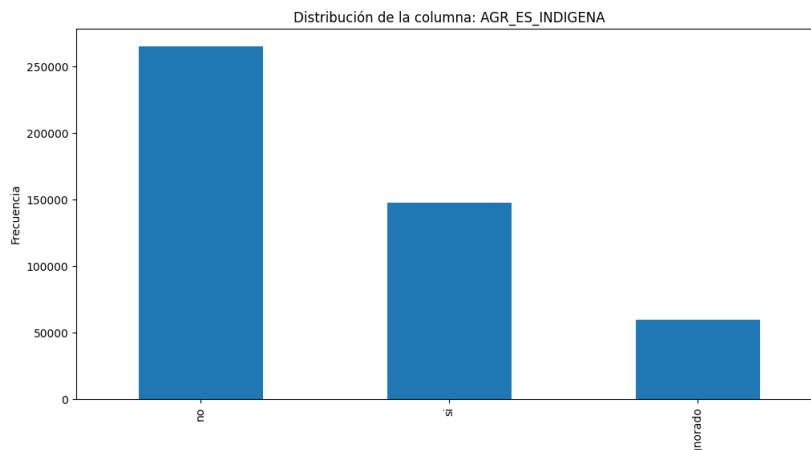


Figura 135: Frecuencia de la columna AGRESINDIGENA

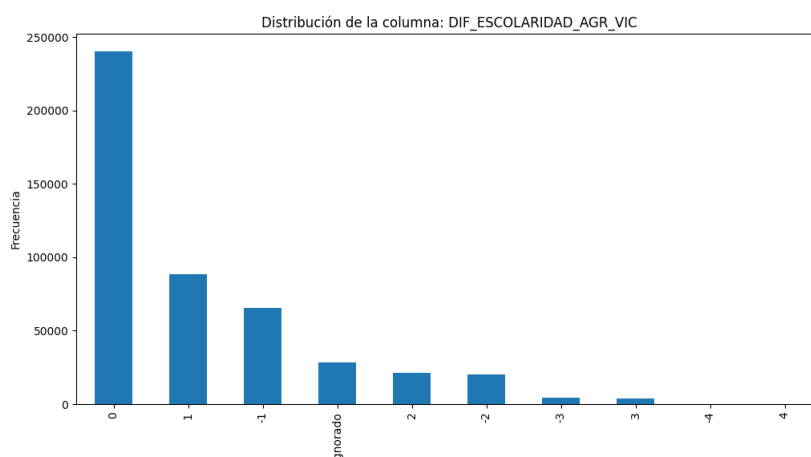


Figura 136: Frecuencia de la columna DIFESCOLARIDADAGRVIC

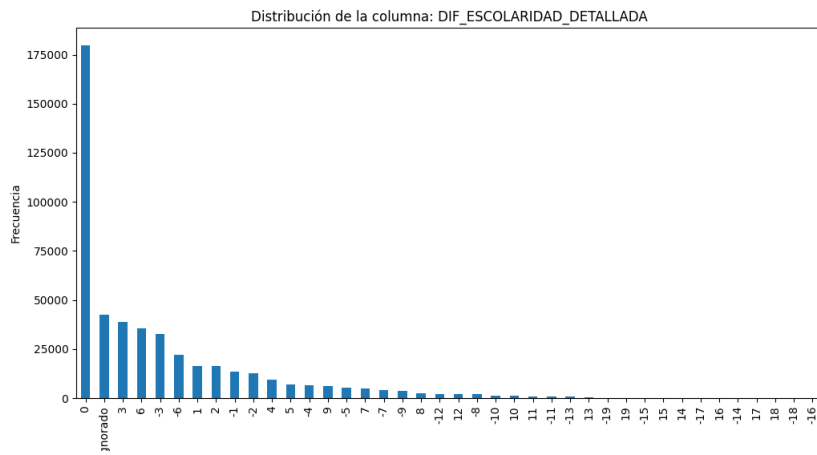


Figura 137: Frecuencia de la columna DIFESCOLARIDADDETALLADA

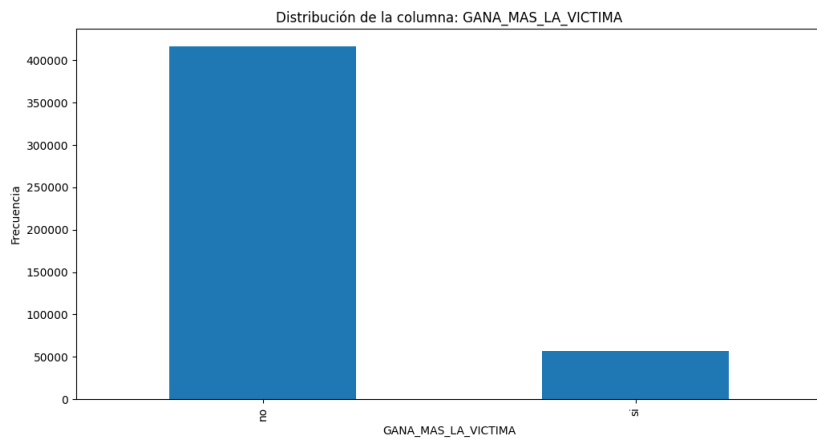


Figura 138: Frecuencia de la columna GANAMASLAVICTIMA

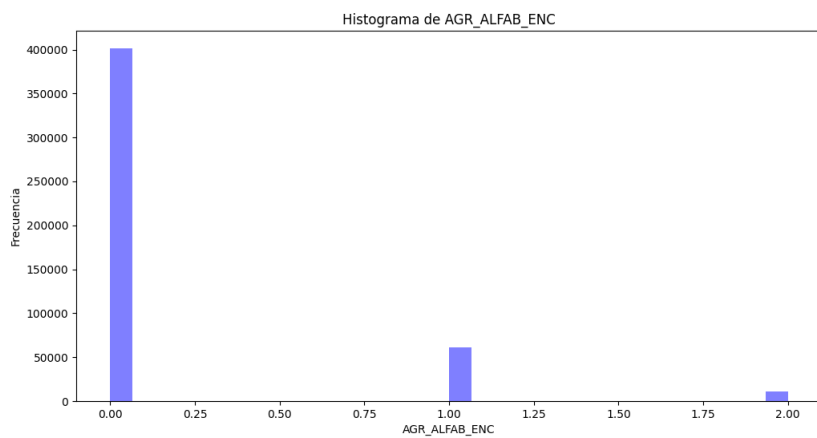


Figura 139: Frecuencia de la columna AGRALFABENC

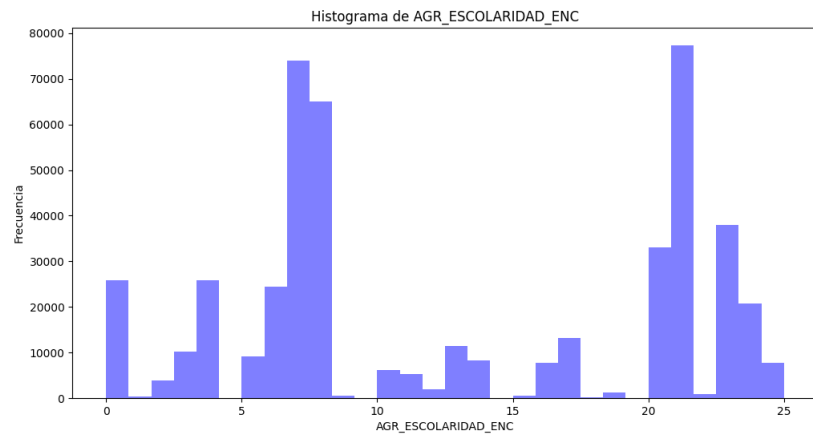


Figura 140: Frecuencia de la columna AGRESCOLARIDADENC

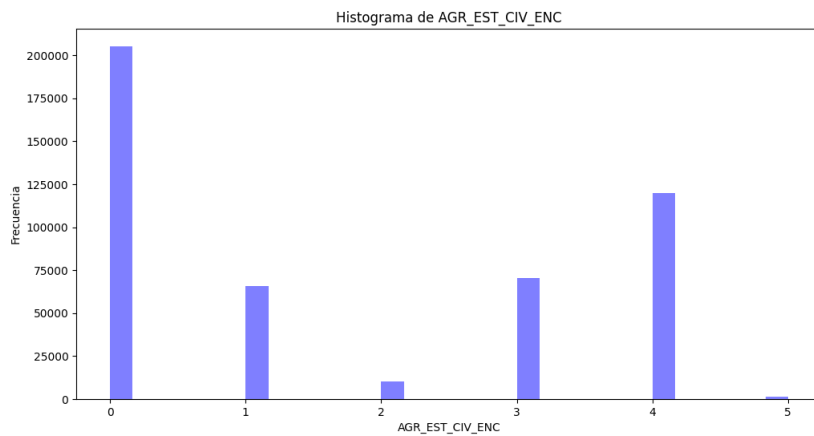


Figura 141: Frecuencia de la columna AGRESTCIVENC

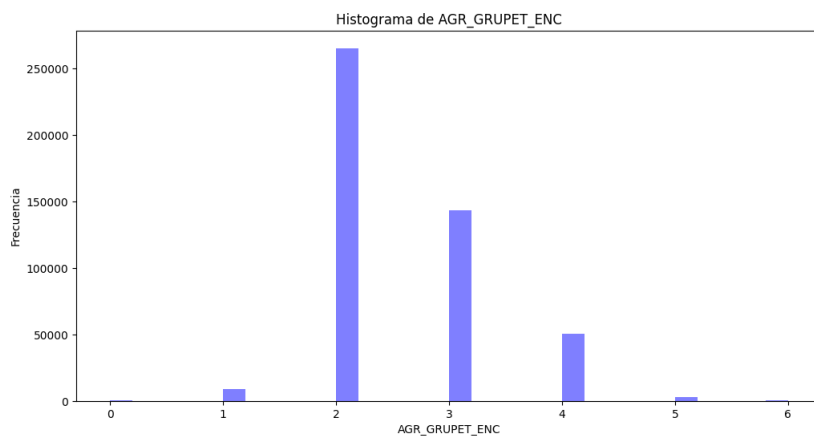


Figura 142: Frecuencia de la columna AGRGRUPETENC

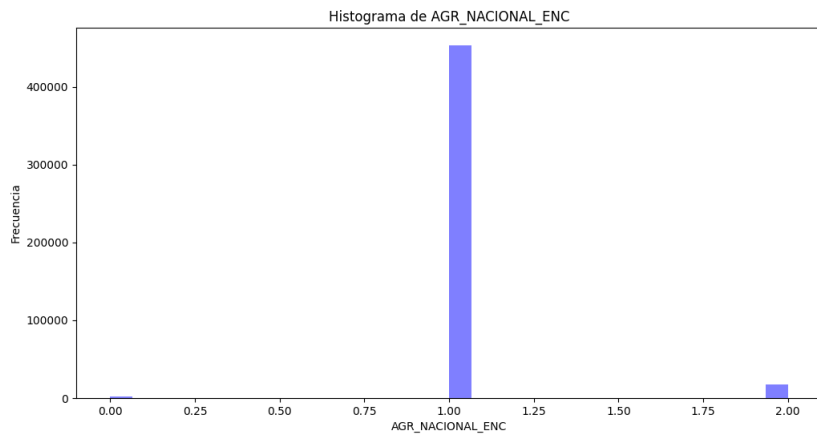


Figura 143: Frecuencia de la columna AGRNACIONALENC

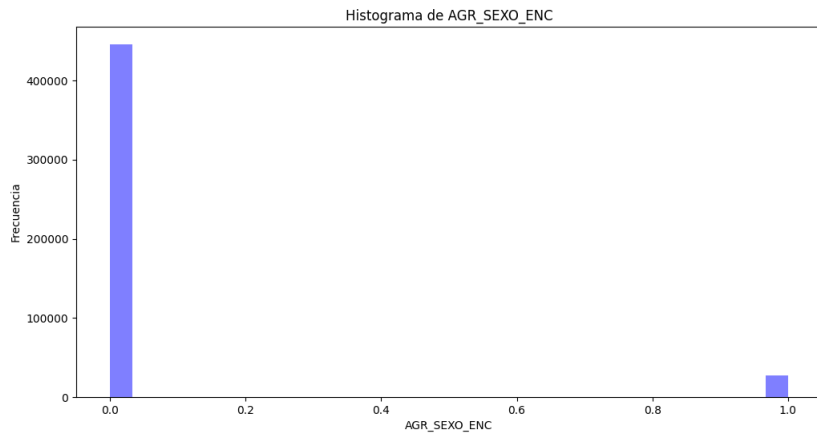


Figura 144: Frecuencia de la columna AGRSEXOENC

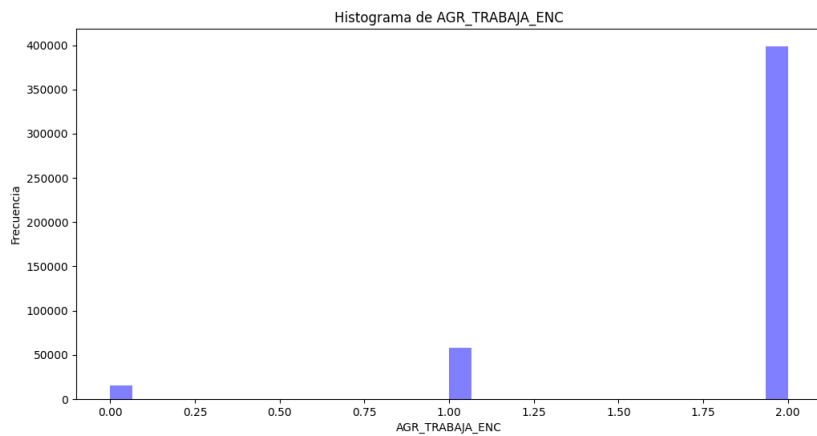


Figura 145: Frecuencia de la columna AGRTRABAJAENC

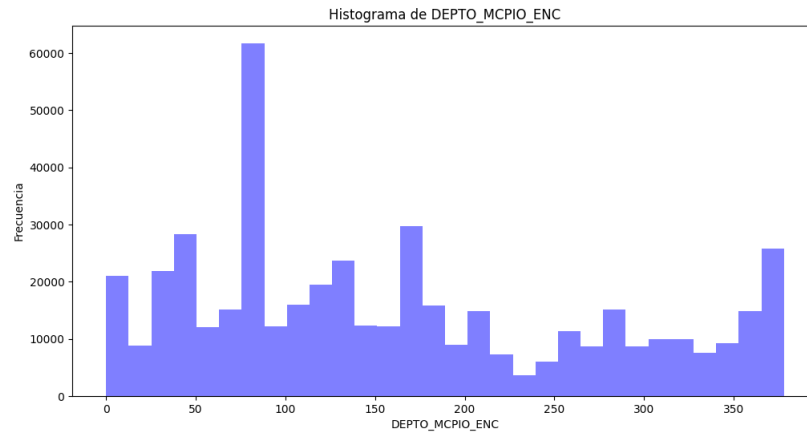


Figura 146: Frecuencia de la columna DEPTOMCPIOENC

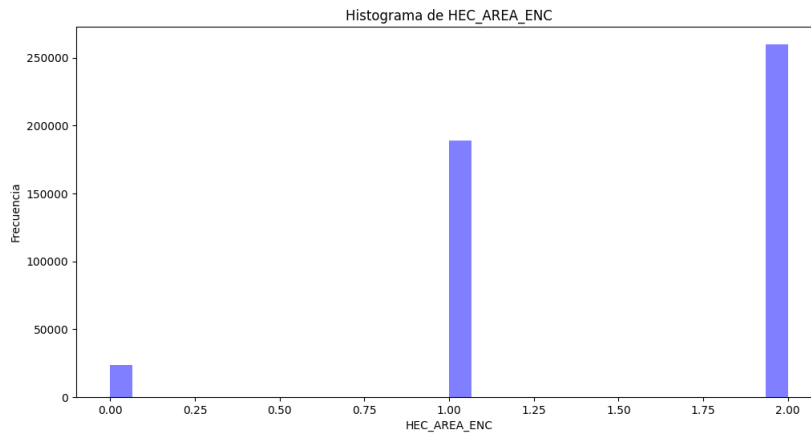


Figura 147: Frecuencia de la columna HECAREAENC

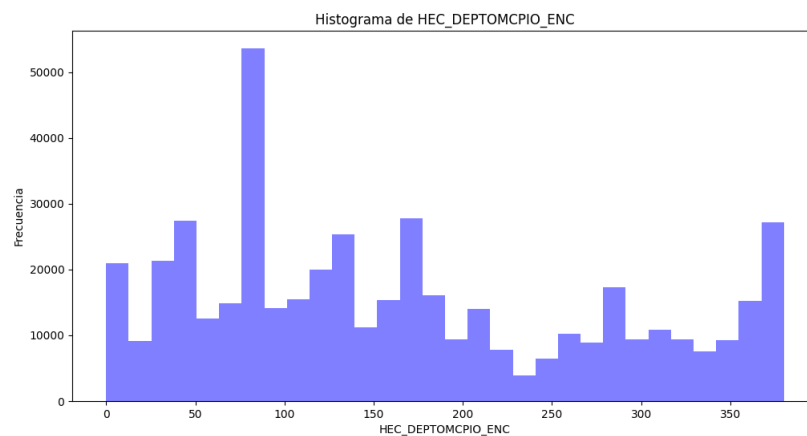


Figura 148: Frecuencia de la columna HECDEPTOMCPIOENC

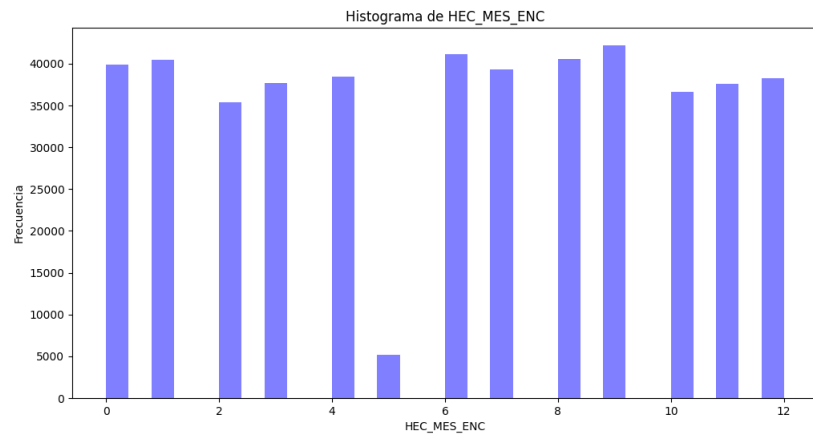


Figura 149: Frecuencia de la columna HECMESENC

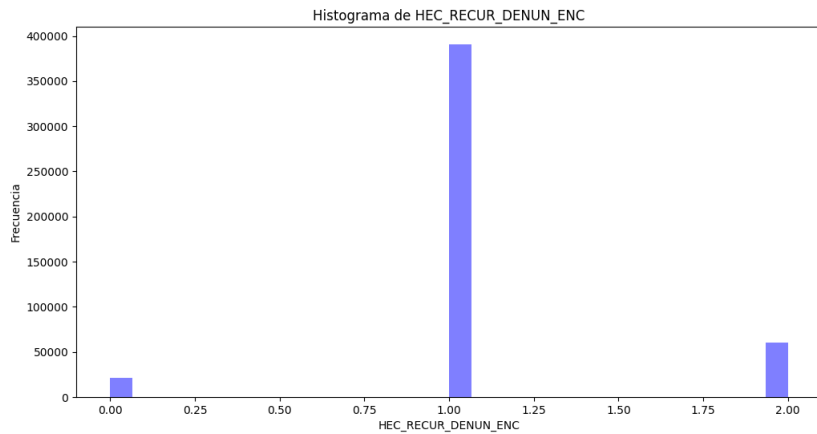


Figura 150: Frecuencia de la columna HECRECURDENUNENC

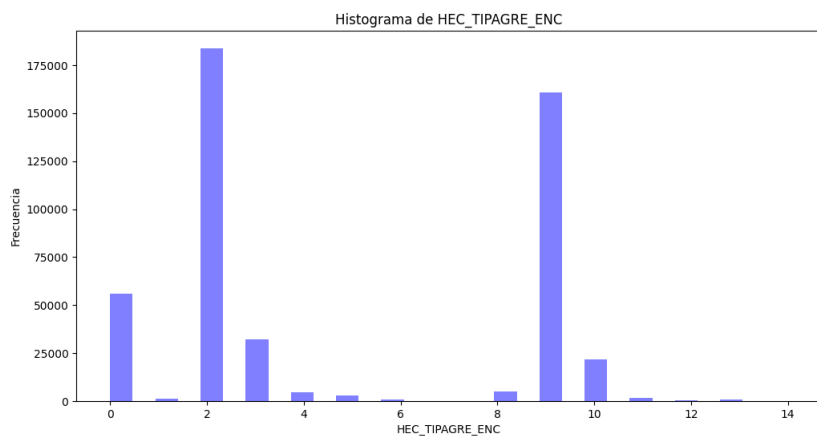


Figura 151: Frecuencia de la columna HECTIPAGREENC

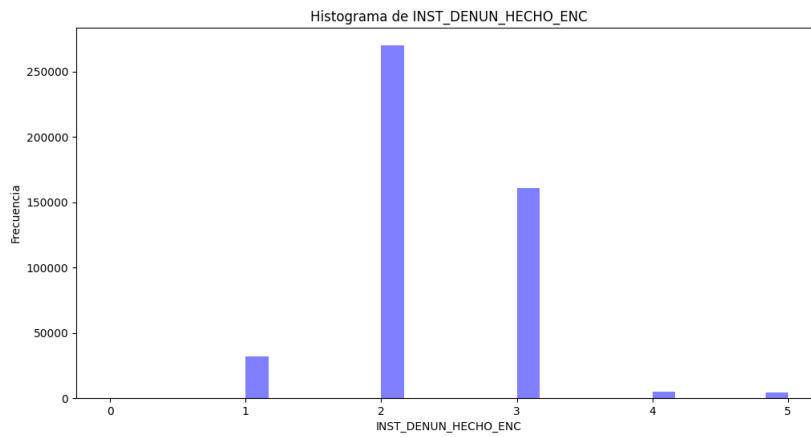


Figura 152: Frecuencia de la columna INSTDENUNHECHOENC

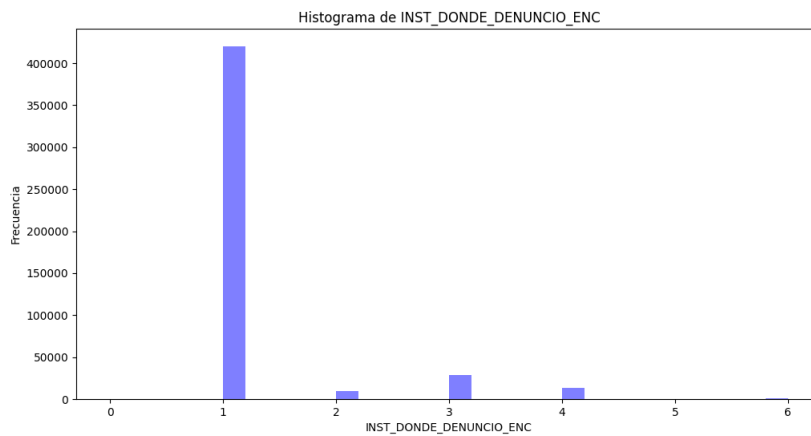


Figura 153: Frecuencia de la columna INSTDONDEDEDENUNCIOENC

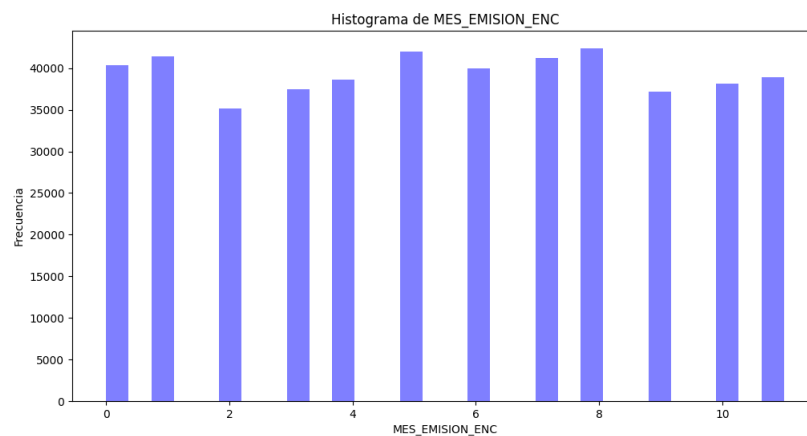


Figura 154: Frecuencia de la columna MESEMISIONENC

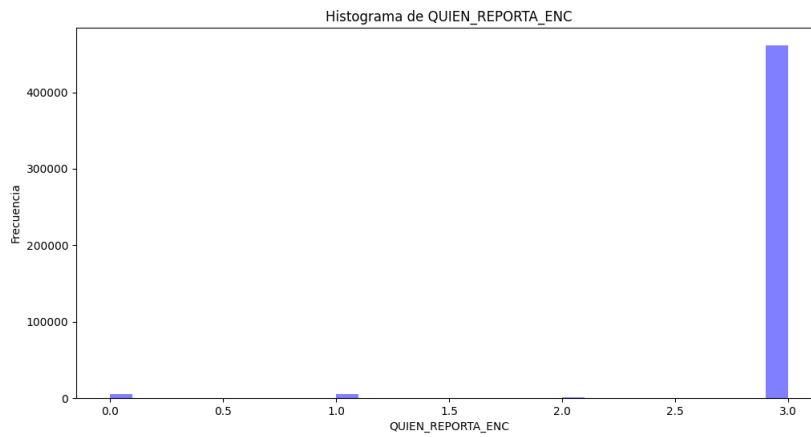


Figura 155: Frecuencia de la columna QUIENREPORTAENC

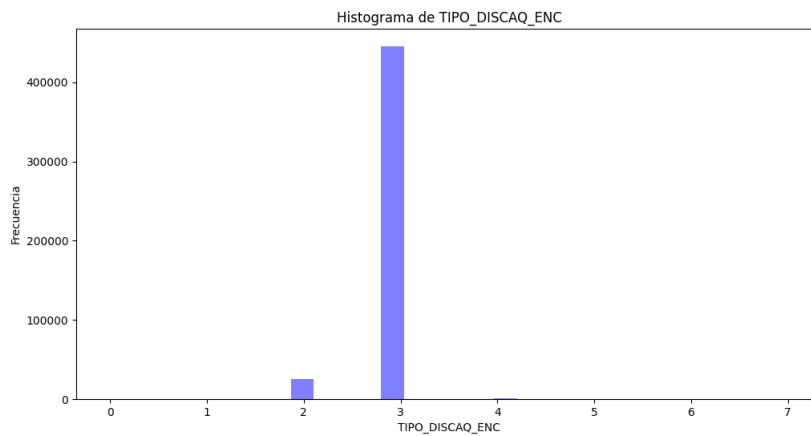


Figura 156: Frecuencia de la columna TIPODISCAQENC

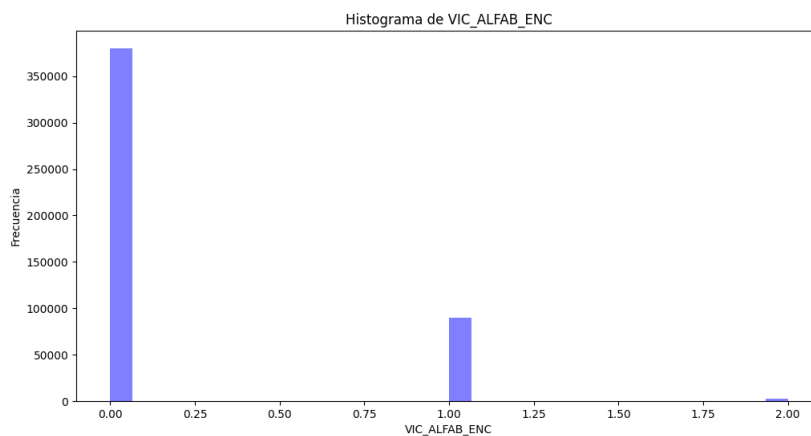


Figura 157: Frecuencia de la columna VICALFABENC

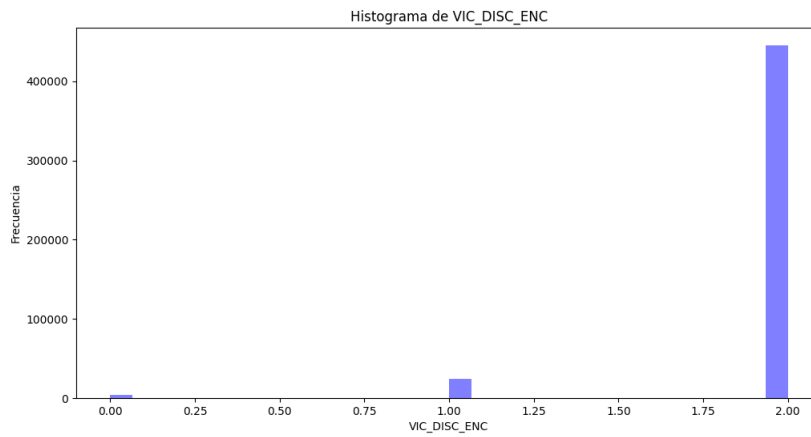


Figura 158: Frecuencia de la columna VICDISCENC

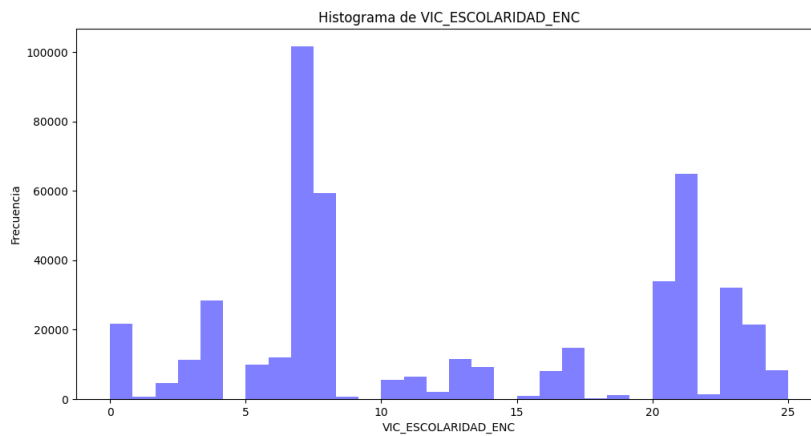


Figura 159: Frecuencia de la columna VICESCOLARIDADENC

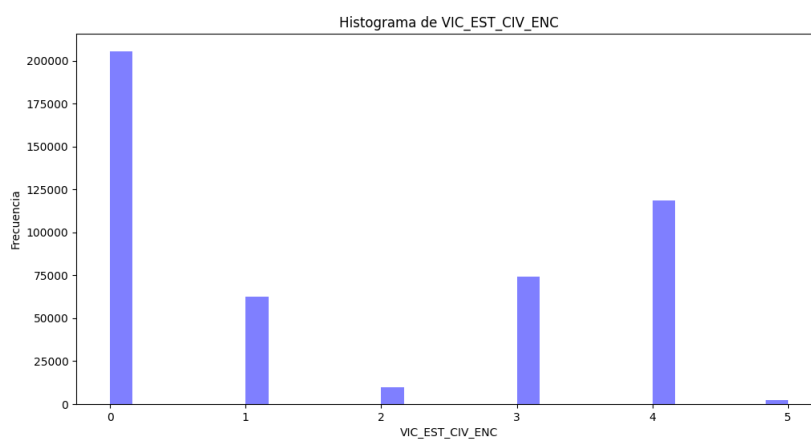


Figura 160: Frecuencia de la columna VICESTCIVENC

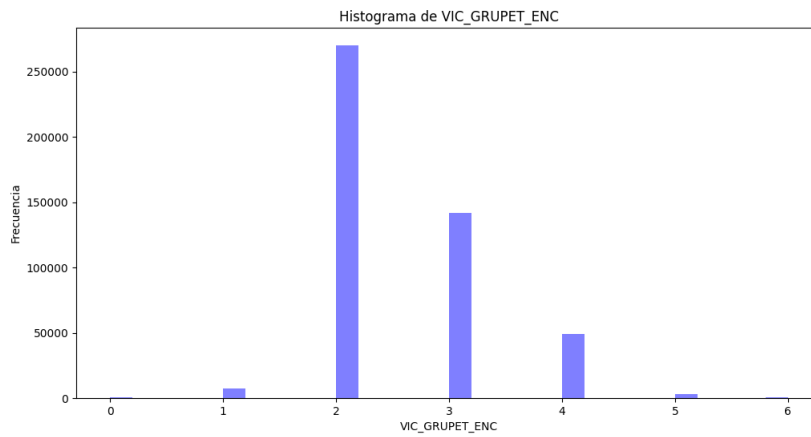


Figura 161: Frecuencia de la columna VICGRUPETENC

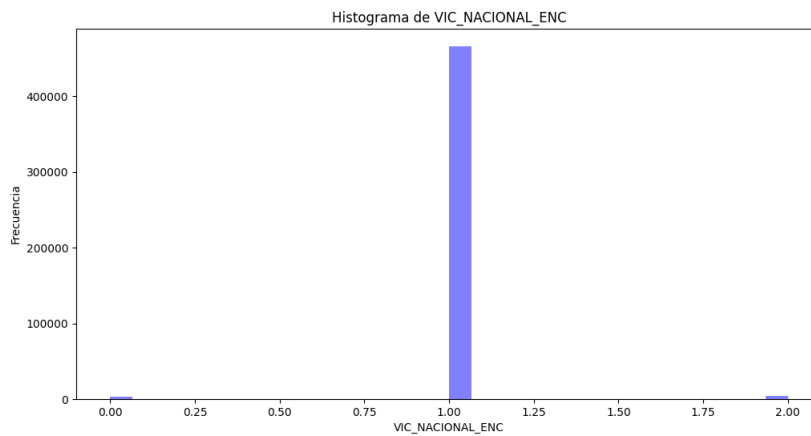


Figura 162: Frecuencia de la columna VICNACIONALENC

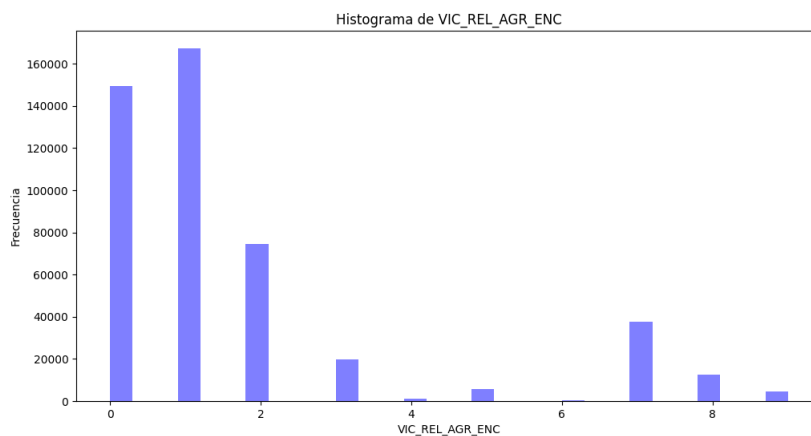


Figura 163: Frecuencia de la columna VICRELAGRENC

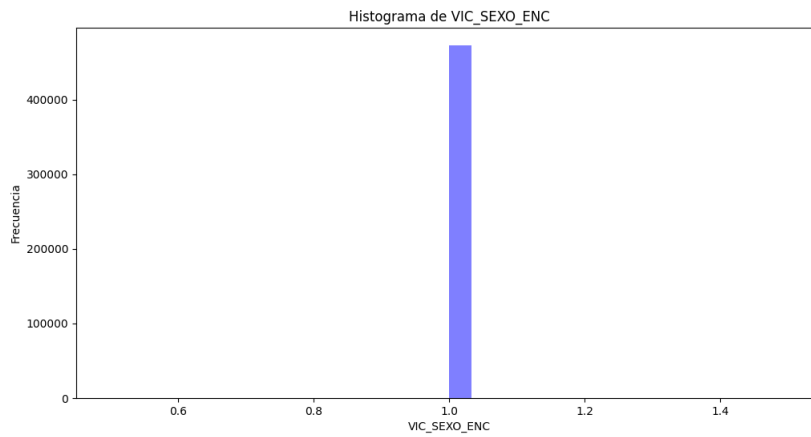


Figura 164: Frecuencia de la columna VICSEXOENC

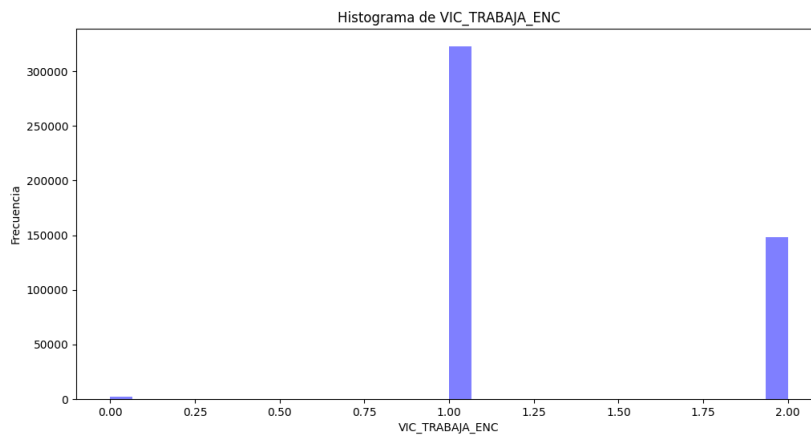


Figura 165: Frecuencia de la columna VICTRABAJAENC

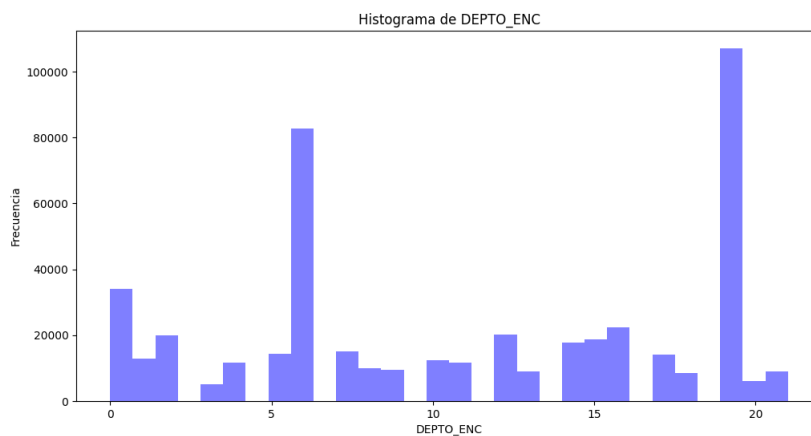


Figura 166: Frecuencia de la columna DEPTOENC

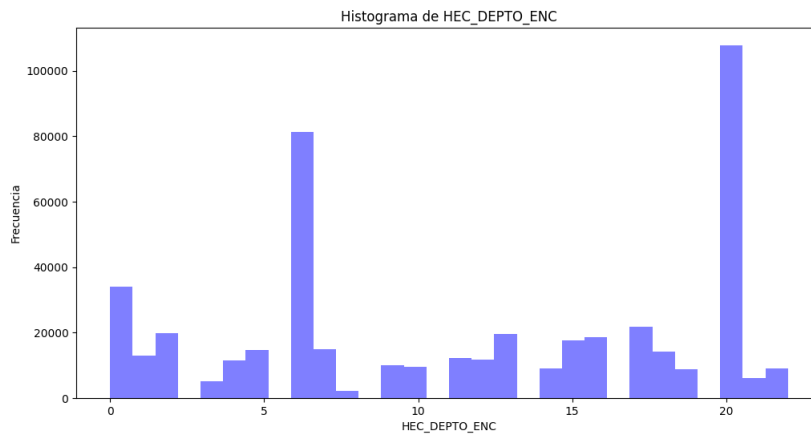


Figura 167: Frecuencia de la columna HECDEPTOENC

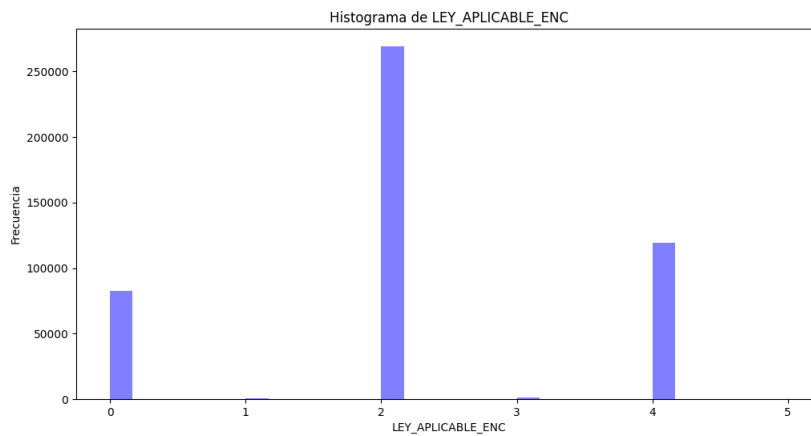


Figura 168: Frecuencia de la columna LEYAPLICABLEENC

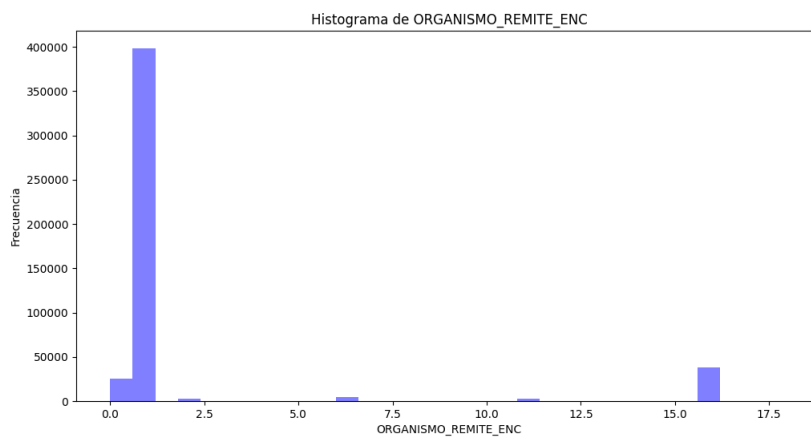


Figura 169: Frecuencia de la columna ORGANISMOREMITEENC

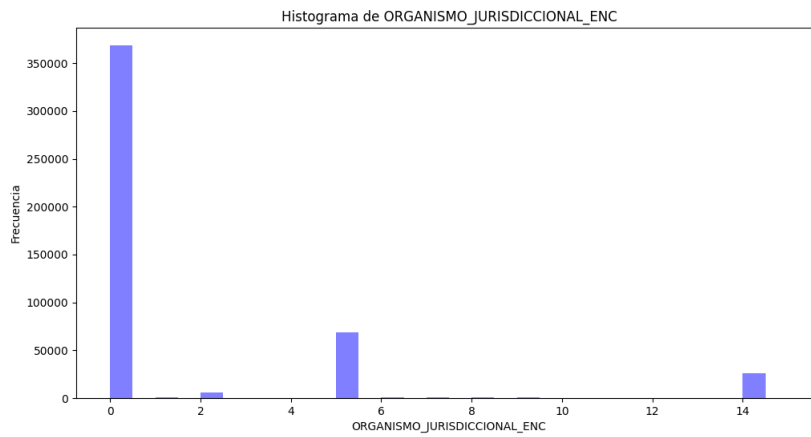


Figura 170: Frecuencia de la columna ORGANISMOJURISDICCIONALENC

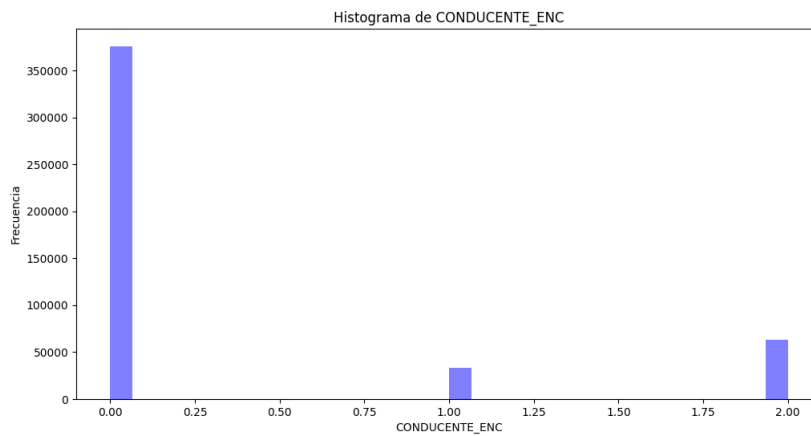


Figura 171: Frecuencia de la columna CONDUCTENTEENC

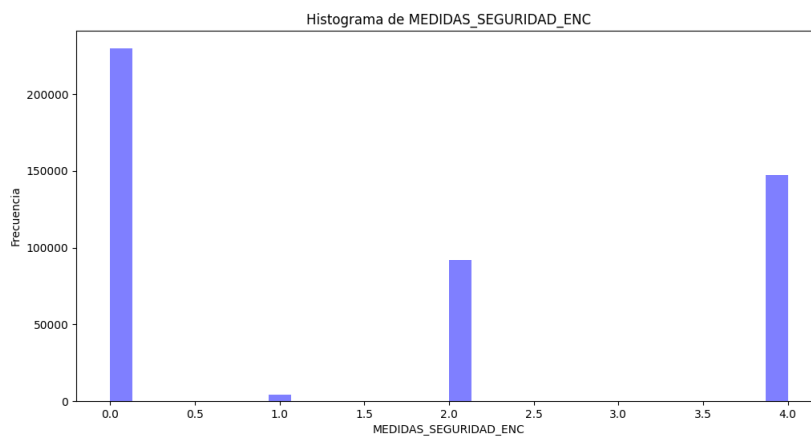


Figura 172: Frecuencia de la columna MEDIDASSEGURIDADENC

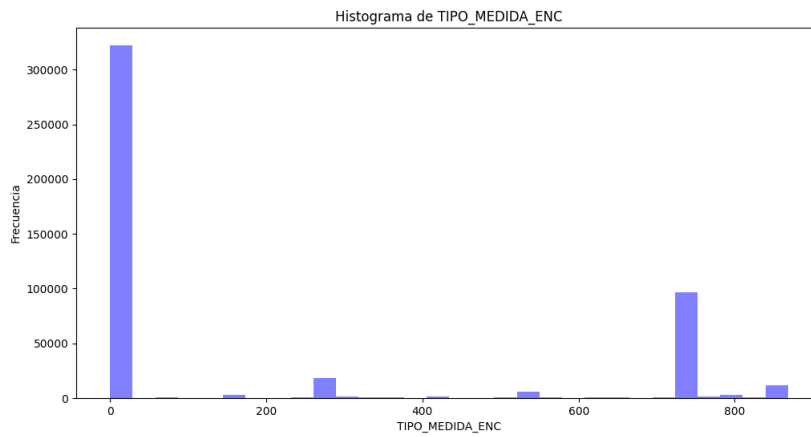


Figura 173: Frecuencia de la columna TIPOMEDIDAENC

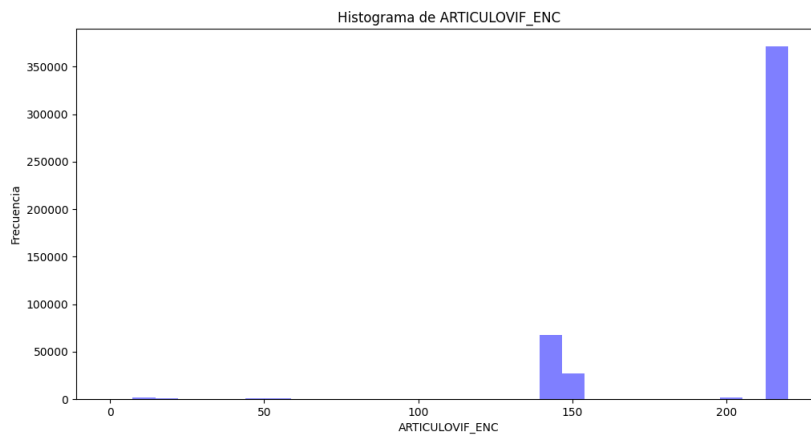


Figura 174: Frecuencia de la columna ARTICULOVIFENC

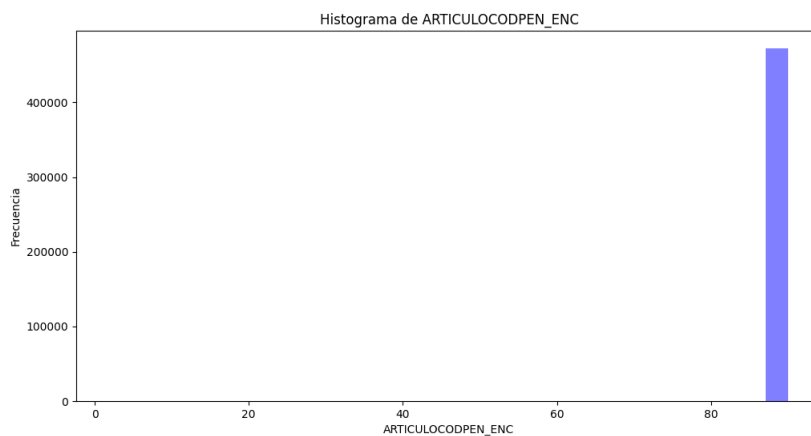


Figura 175: Frecuencia de la columna ARTICULOCODPENENC

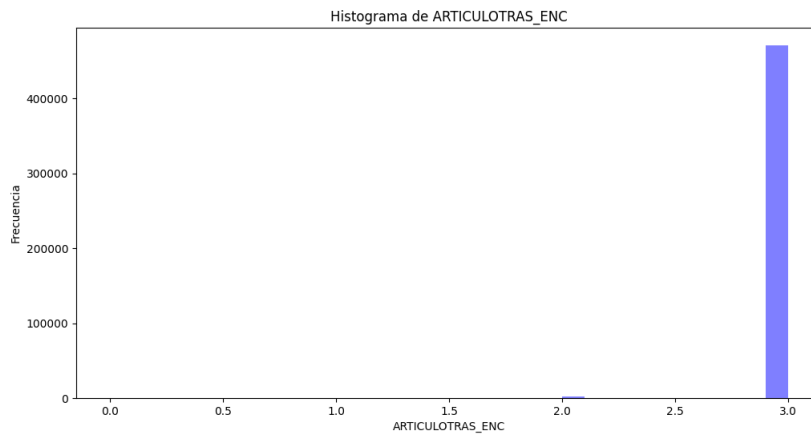


Figura 176: Frecuencia de la columna ARTICULOTRAS_ENC

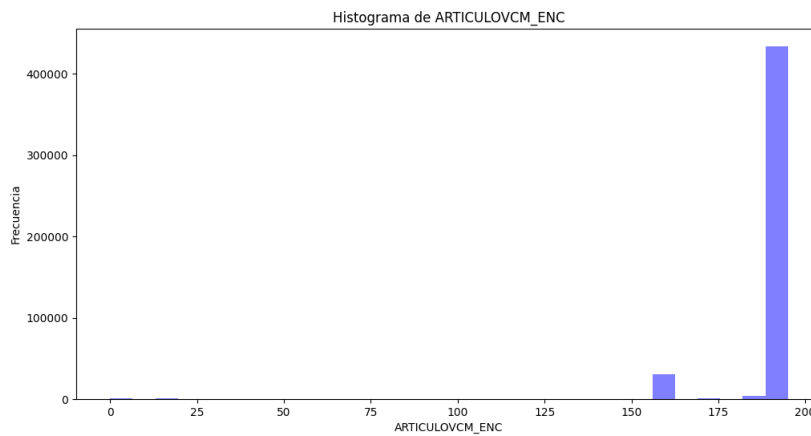


Figura 177: Frecuencia de la columna ARTICULOVCME_ENC

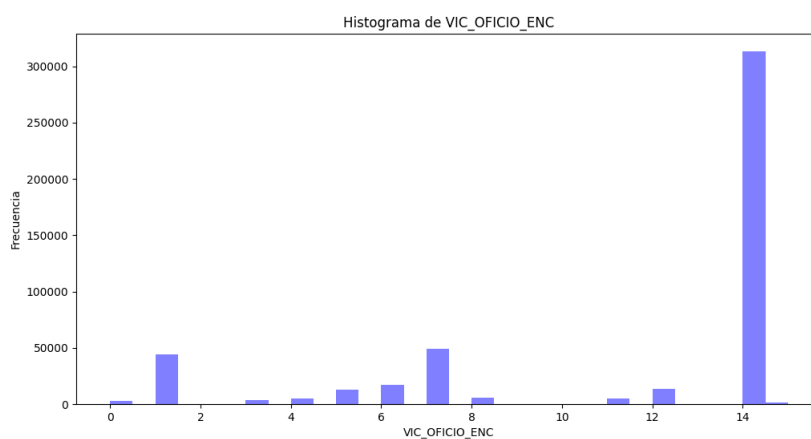


Figura 178: Frecuencia de la columna VICOFICIOENC

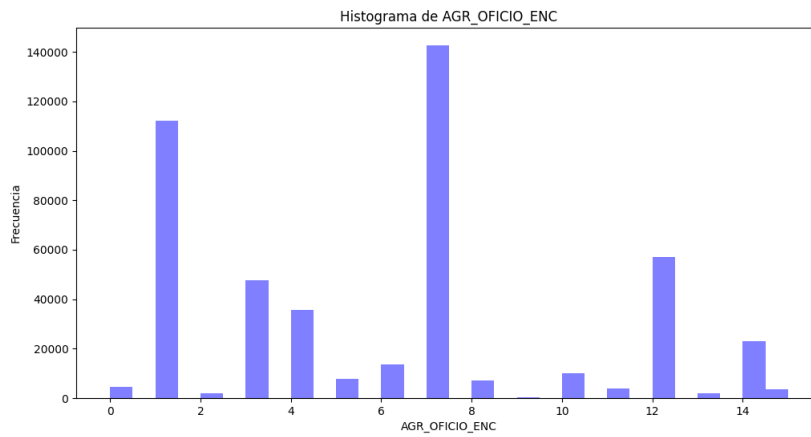


Figura 179: Frecuencia de la columna AGROFICIOENC

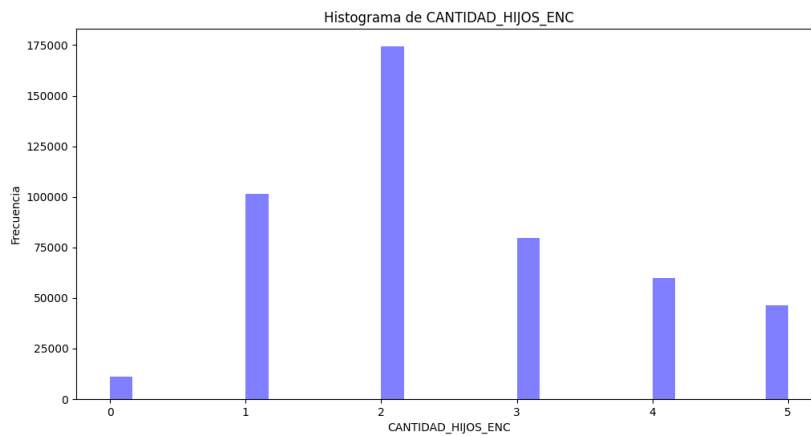


Figura 180: Frecuencia de la columna CANTIDADHIJOSENC

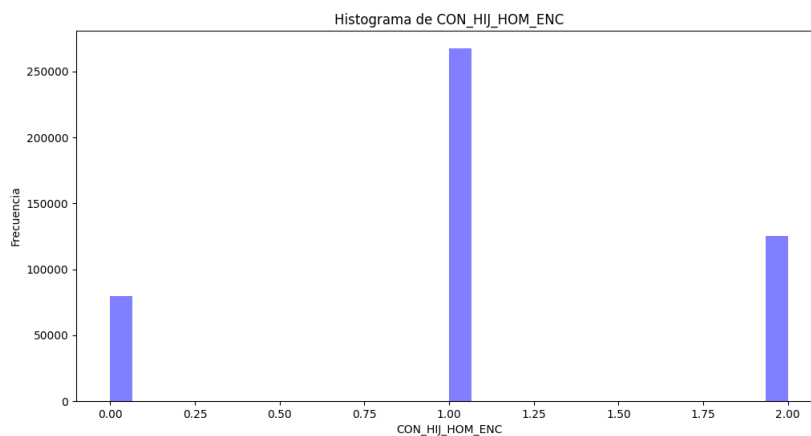


Figura 181: Frecuencia de la columna CONHIJHOMENC

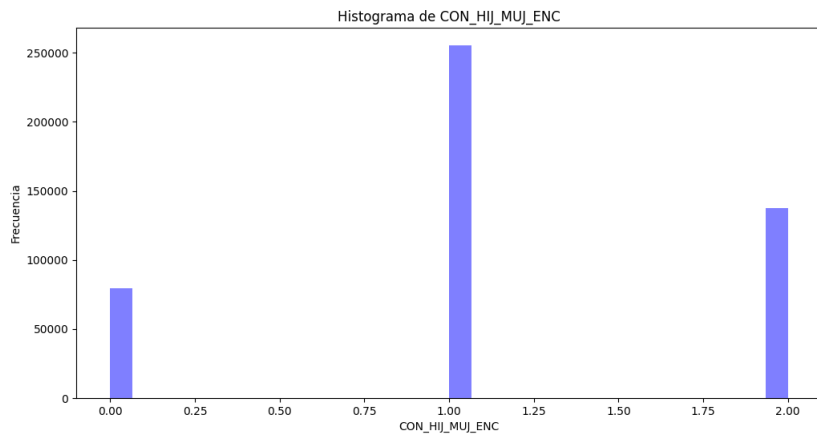


Figura 182: Frecuencia de la columna CONHIJMUJENC

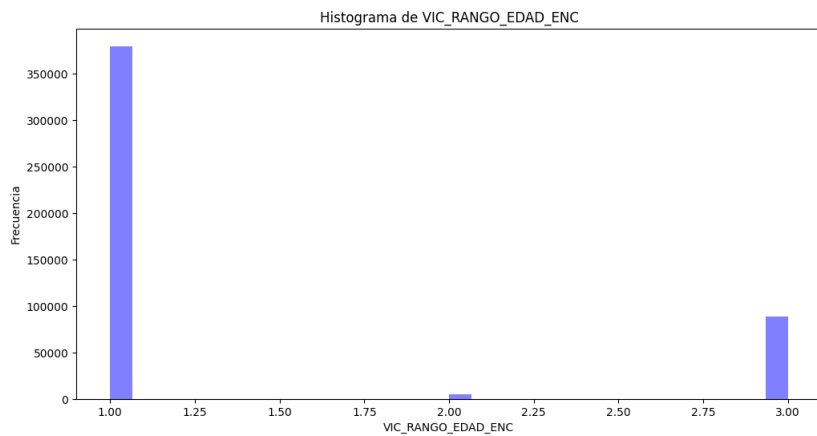


Figura 183: Frecuencia de la columna VICRANGOEDADENC

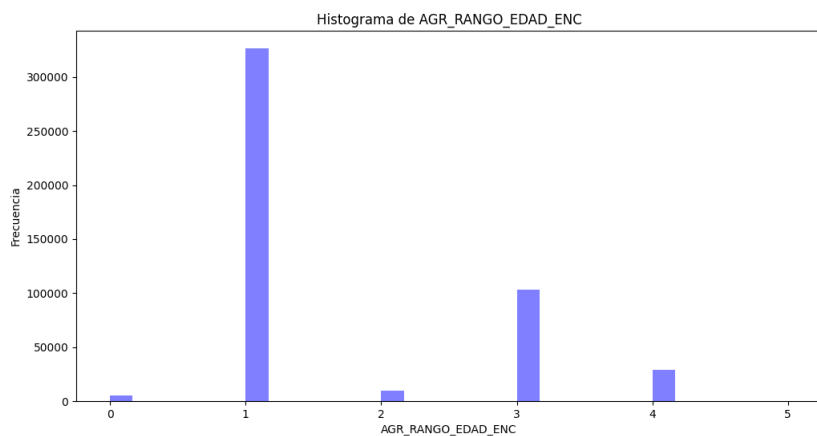


Figura 184: Frecuencia de la columna AGRRANGOEDADENC

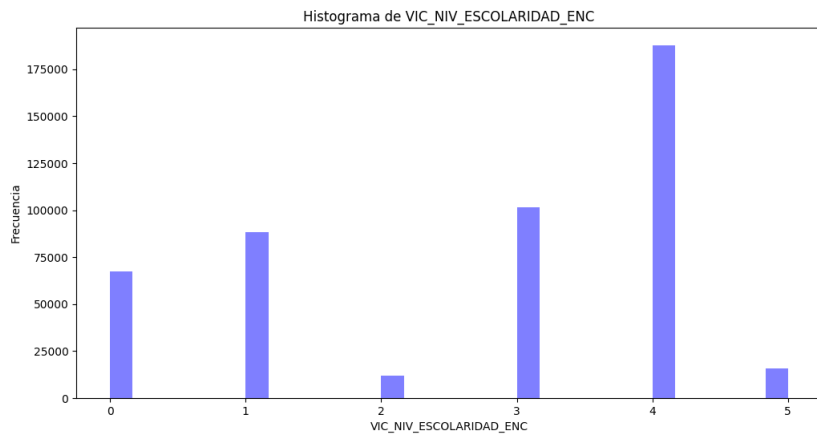


Figura 185: Frecuencia de la columna VICNIVESCULARIDADENC

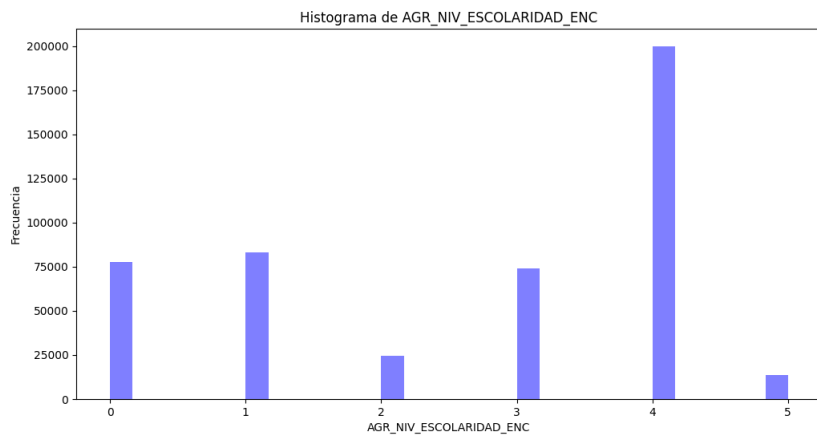


Figura 186: Frecuencia de la columna AGRNIVESCULARIDADENC

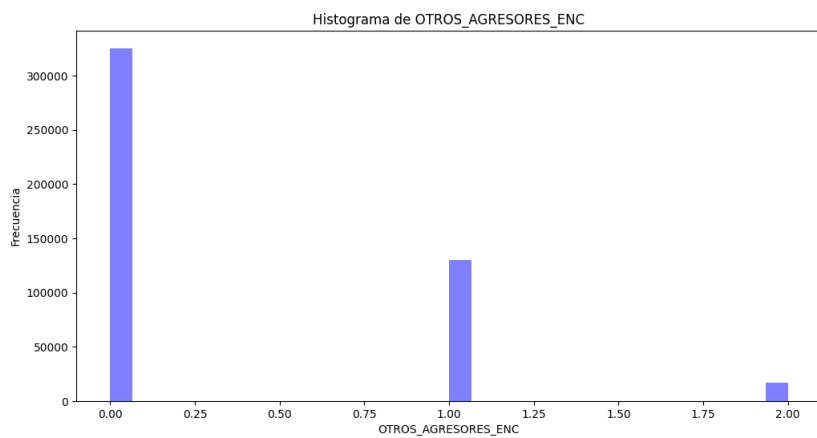


Figura 187: Frecuencia de la columna OTROSAGRESORENC

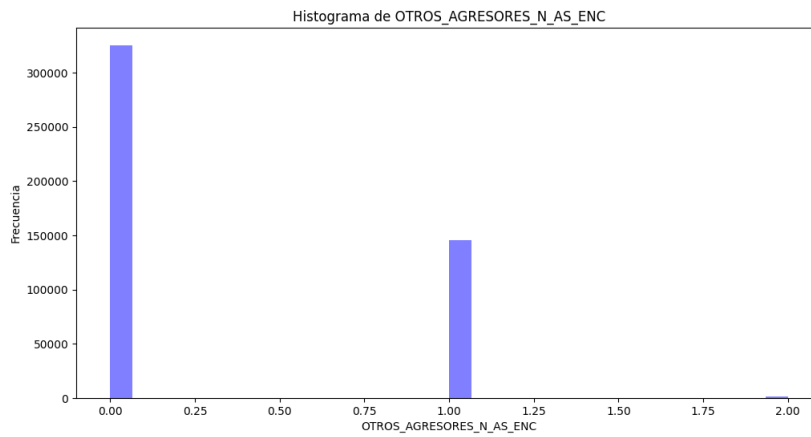


Figura 188: Frecuencia de la columna OTROSAGRESORESNASENC

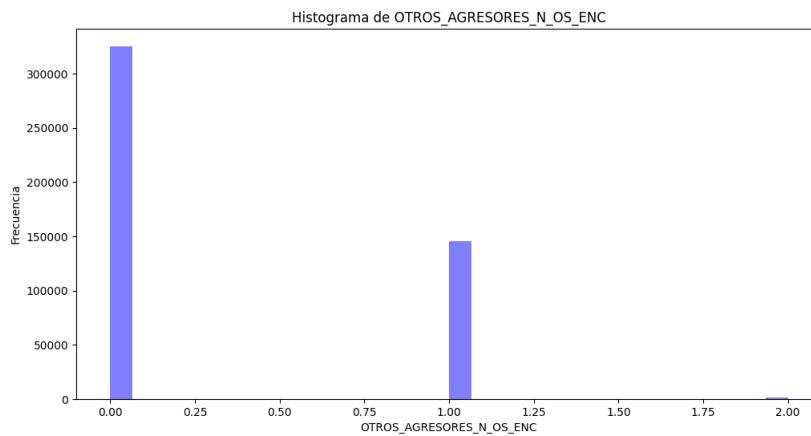


Figura 189: Frecuencia de la columna OTROSAGRESORESNNOSENC

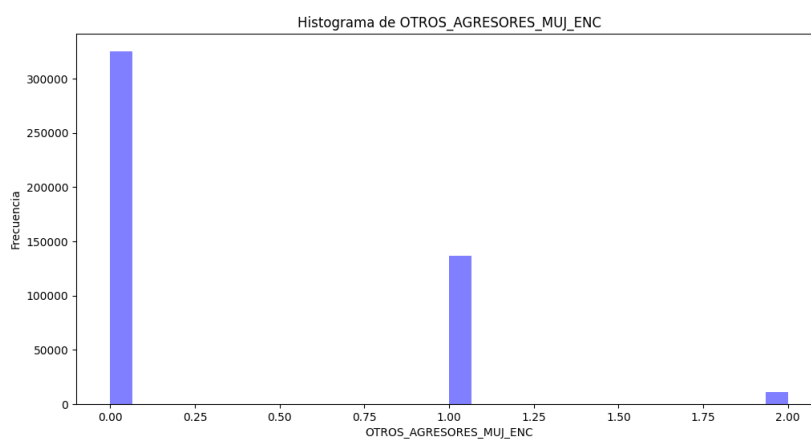


Figura 190: Frecuencia de la columna OTROSAGRESORESMUJENC

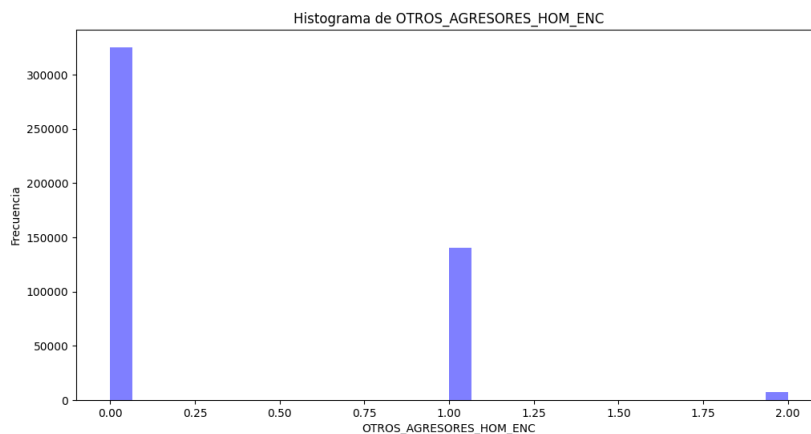


Figura 191: Frecuencia de la columna OTROSAGRESORESHOMENC

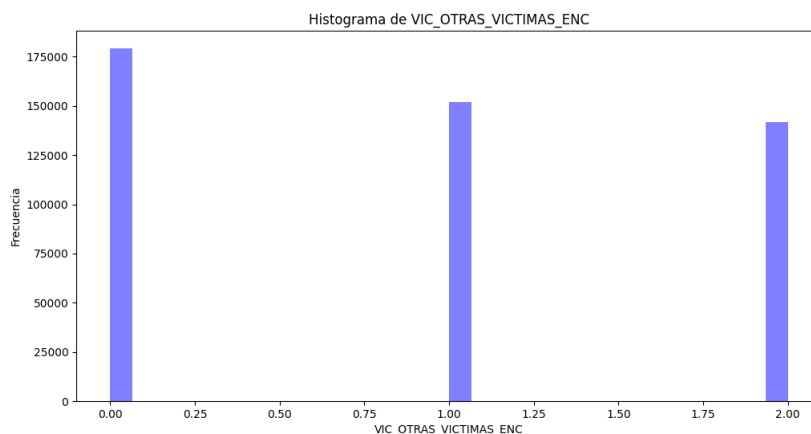


Figura 192: Frecuencia de la columna VICOTRASVICTIMASENC

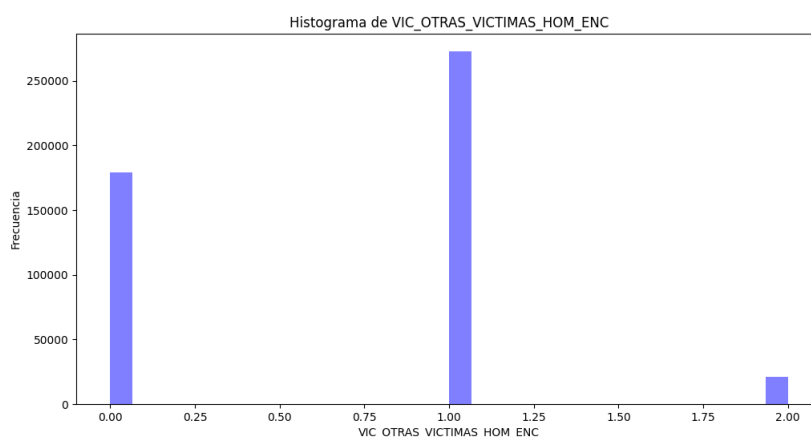


Figura 193: Frecuencia de la columna VICOTRASVICTIMASHOMENC

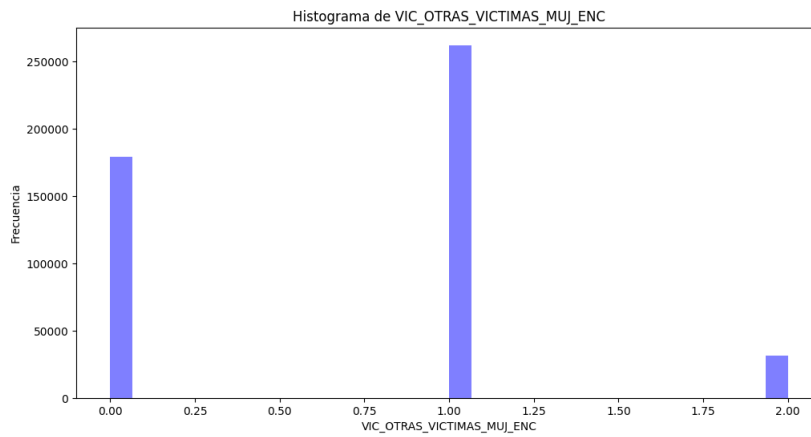


Figura 194: Frecuencia de la columna VICOTRASVICTIMASMUJENC

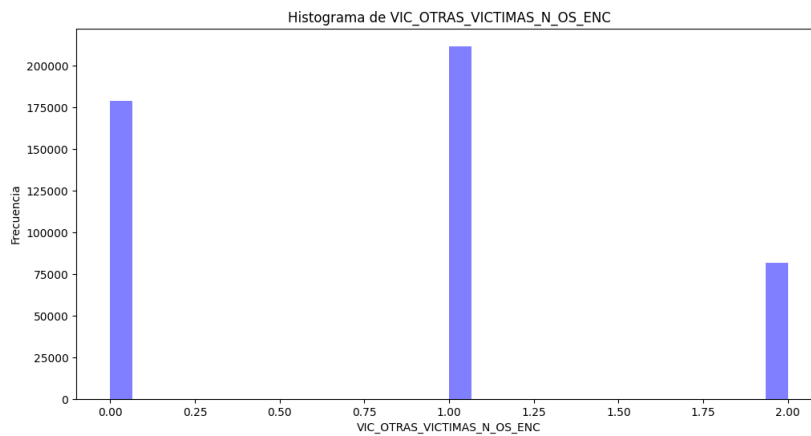


Figura 195: Frecuencia de la columna VICOTRASVICTIMASNOSENC

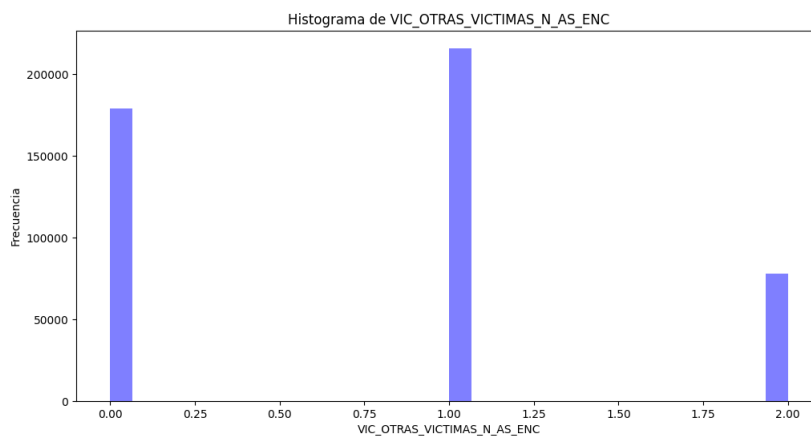


Figura 196: Frecuencia de la columna VICOTRASVICTIMASNASENC

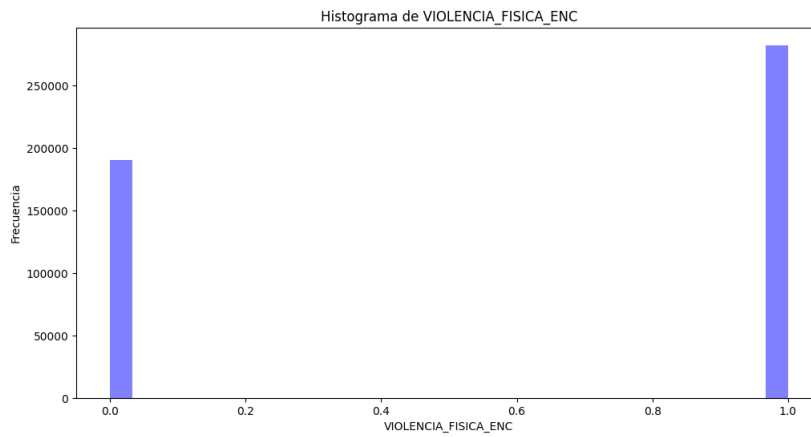


Figura 197: Frecuencia de la columna VIOLENCIAFISICAENC

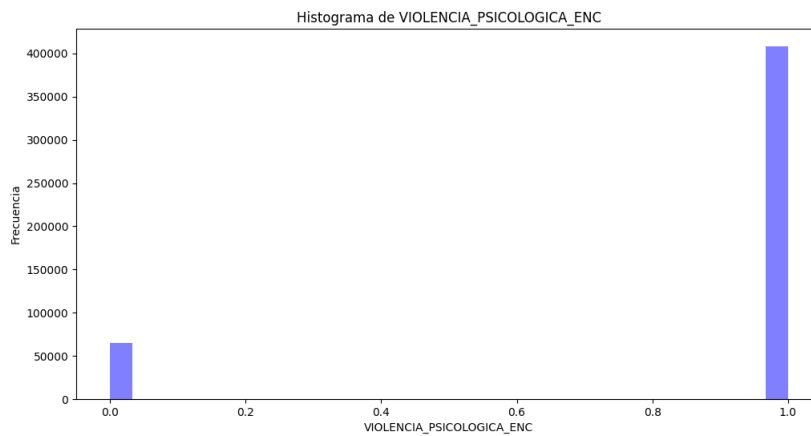


Figura 198: Frecuencia de la columna VIOLENCIAPSICOLOGICAENC

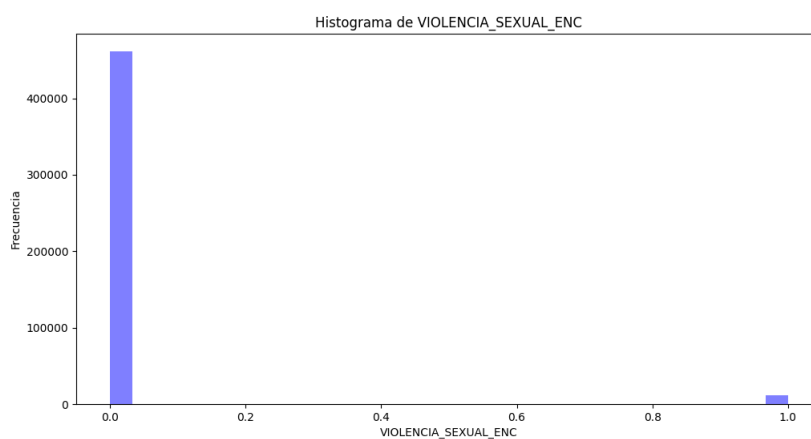


Figura 199: Frecuencia de la columna VIOLENCIASEXUALENC

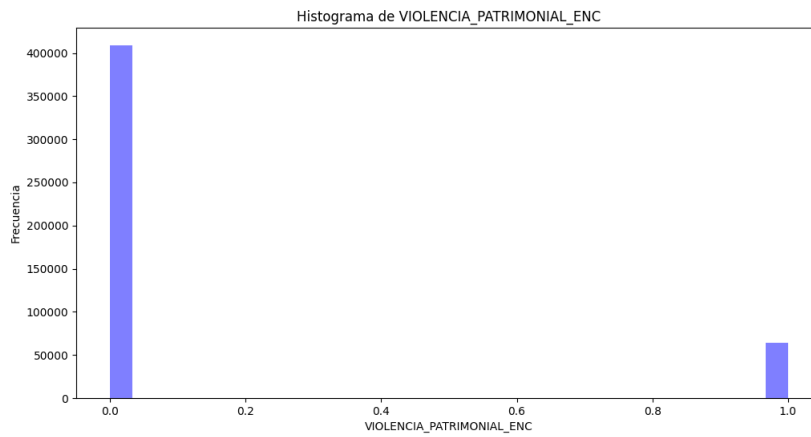


Figura 200: Frecuencia de la columna VIOLENCIAPATRIMONIALENC

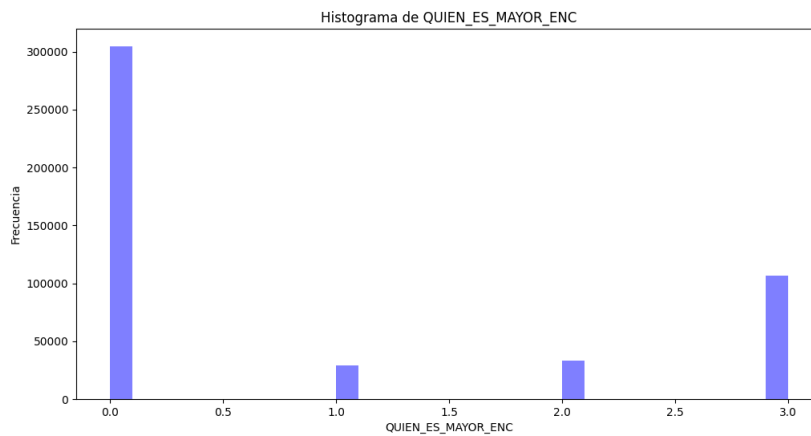


Figura 201: Frecuencia de la columna QUIENESMAYORENC

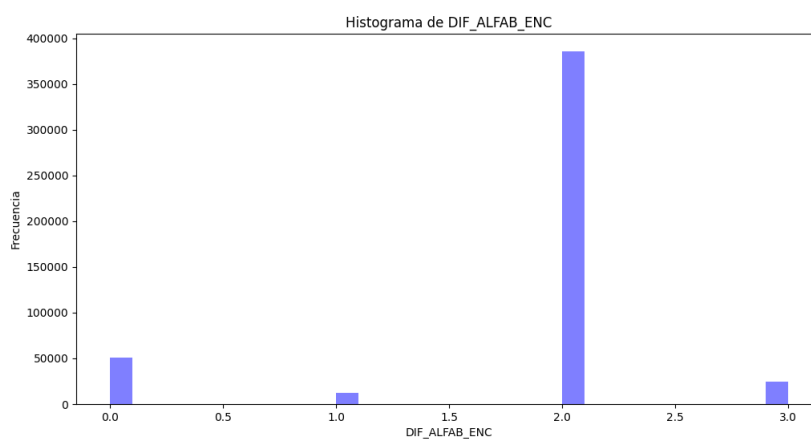


Figura 202: Frecuencia de la columna DIFALFABENC

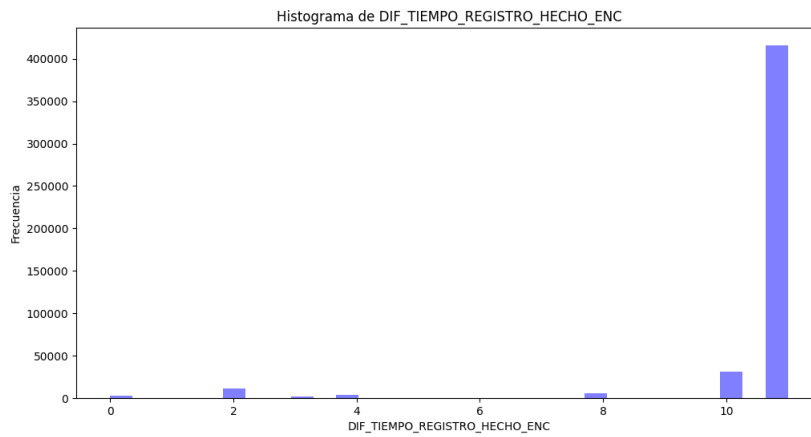


Figura 203: Frecuencia de la columna DIFTIEMPOREGISTROHECHOENC

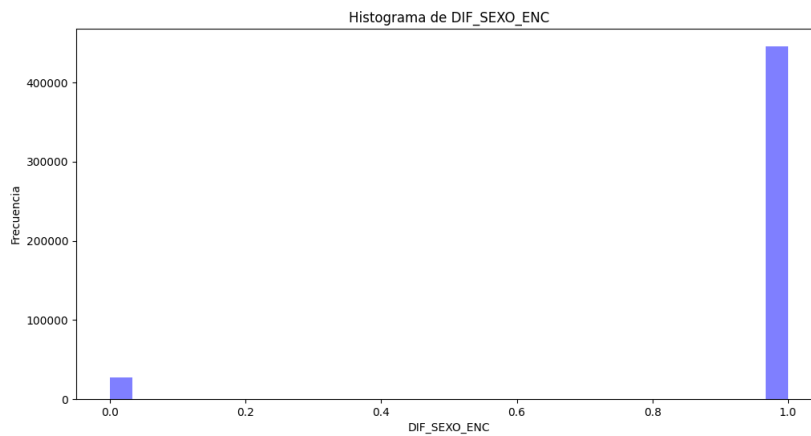


Figura 204: Frecuencia de la columna DIFSEXOENC

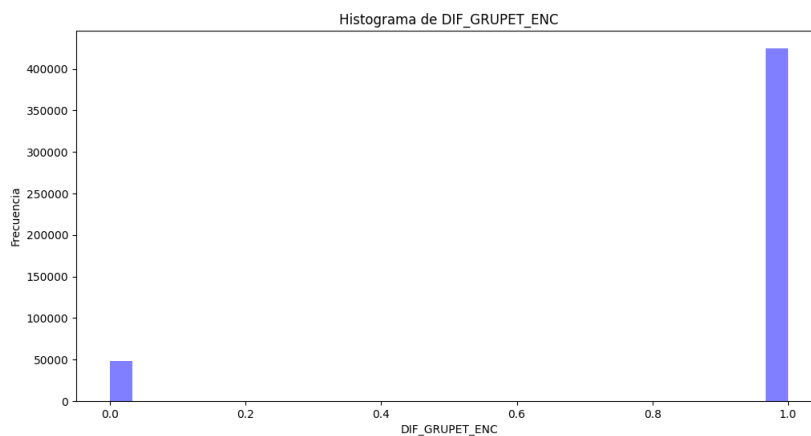


Figura 205: Frecuencia de la columna DIFGRUPETENC

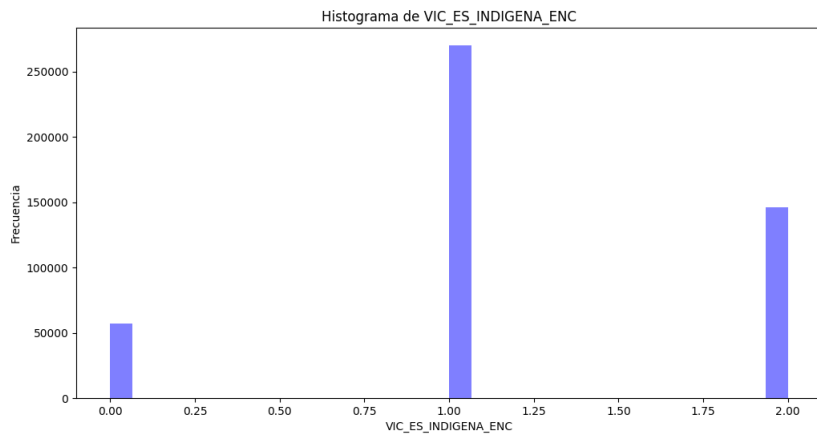


Figura 206: Frecuencia de la columna VICESINDIGENAENC

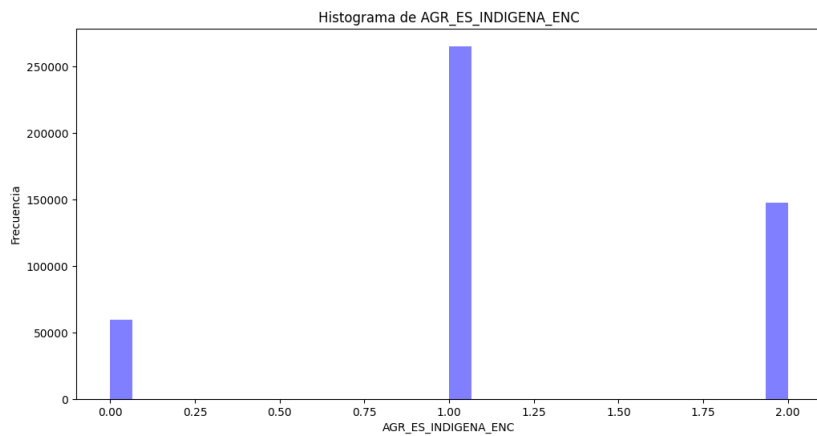


Figura 207: Frecuencia de la columna AGRESINDIGENAENC

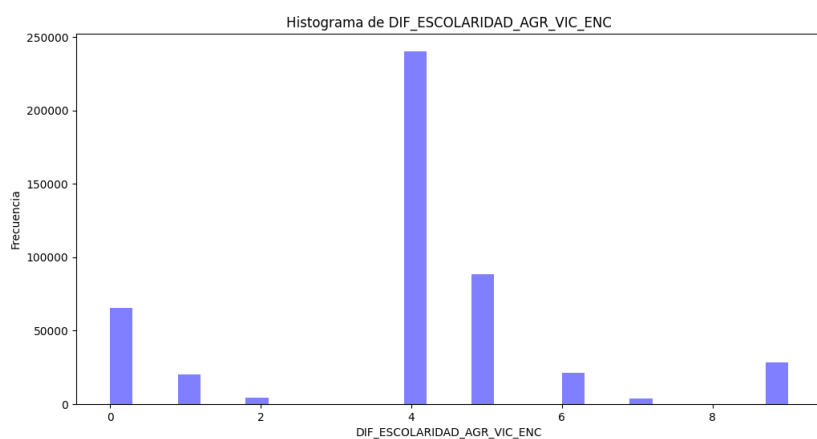


Figura 208: Frecuencia de la columna DIFESCOLARIDADAGRVICENC

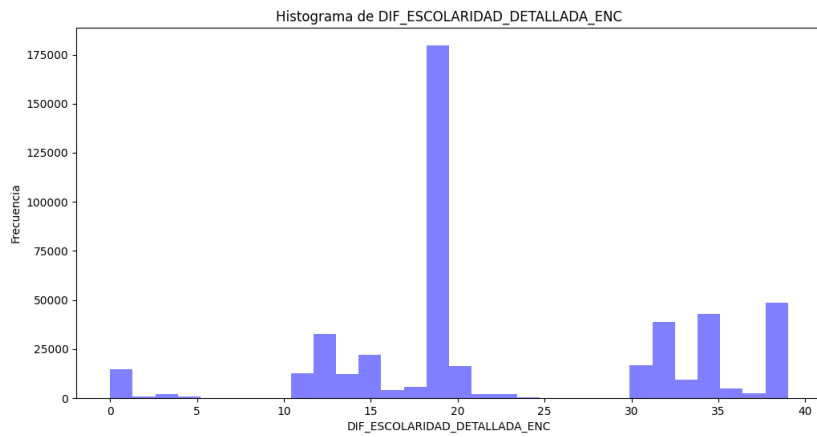


Figura 209: Frecuencia de la columna DIFESCOLARIDADDETALLADAENC

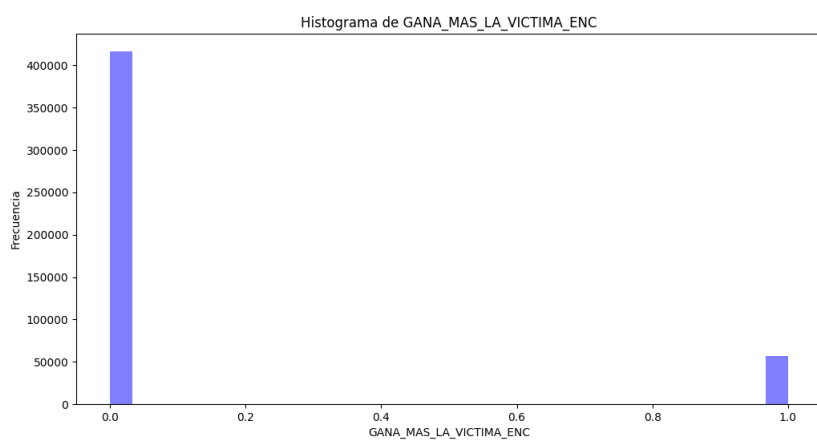


Figura 210: Frecuencia de la columna GANAMASLAVICTIMAENC

11.3. Análisis bivariados

Distribución de la edad de la víctima vs la edad del agresor

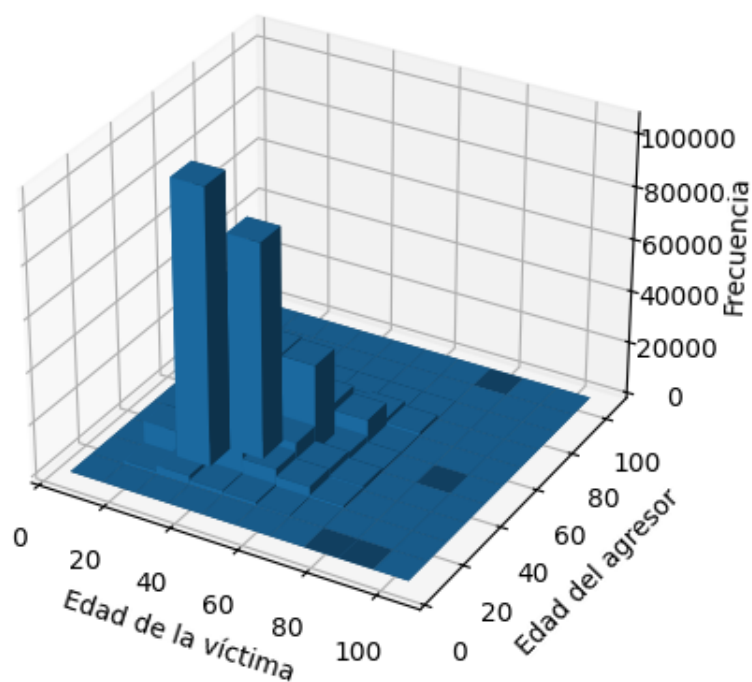


Figura 211: Distribución entre las edades de la víctima y el agresor

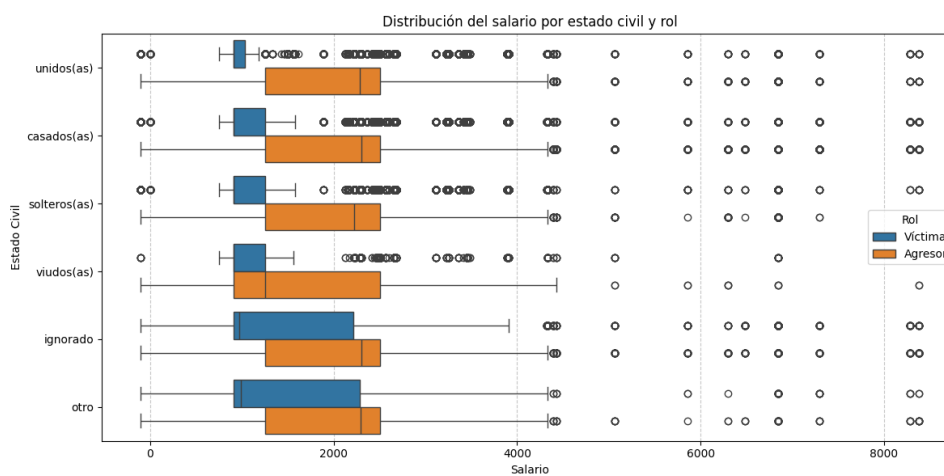


Figura 212: Diagrama de caja de caja y bigotes del estado civil vs salario de víctimas y agresores

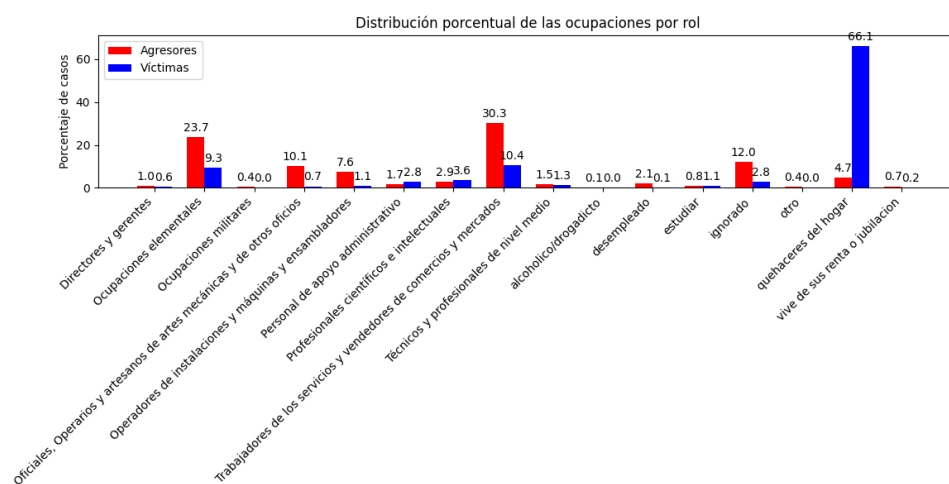


Figura 213: Distribución porcentual de las ocupaciones según el rol (víctima o agresor)

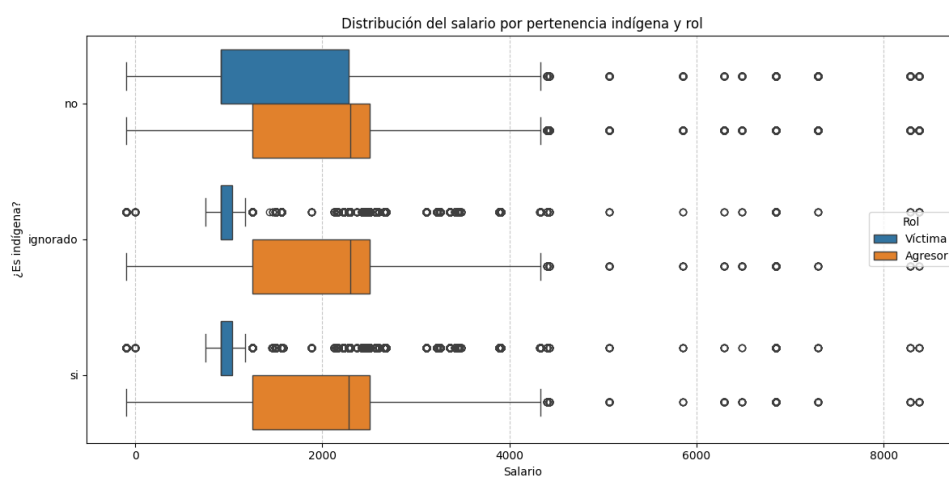


Figura 214: Salario según la pertenencia indígena, según el rol (víctima o agresor)

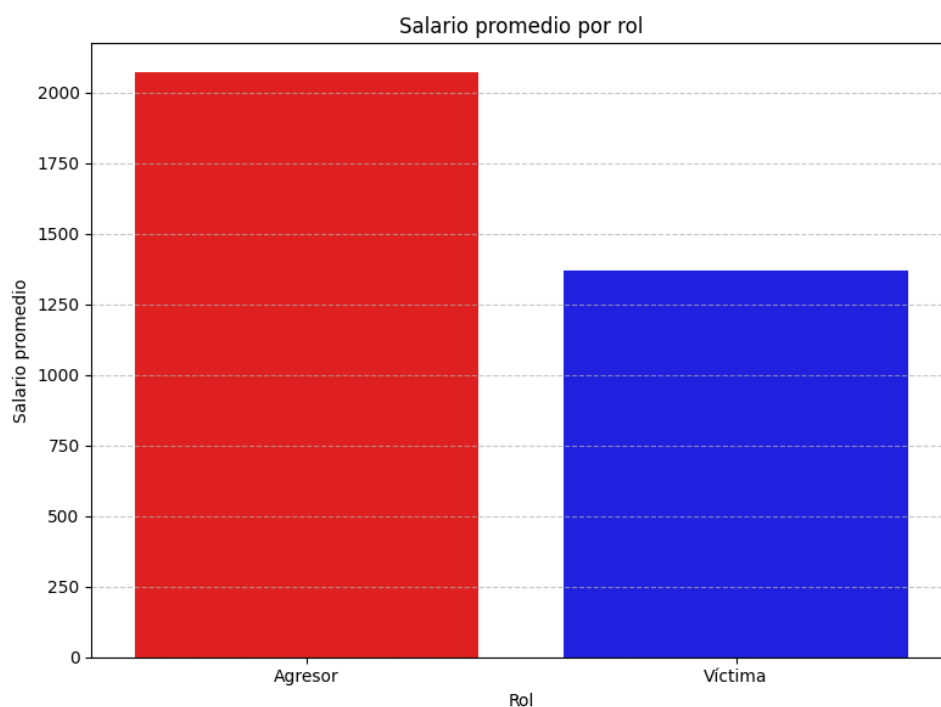


Figura 215: Salario promedio según el rol (víctima o agresor)

11.4. Tablas de contingencia

Cuadro 39: Tabla de contingencia entre AGR_ES_INDIGENA y AGR_ALFAB mostrando frecuencias absolutas y relativas

AGR_ES_INDIGENA AGR_ALFAB	ignorado	no	si
alfabeta	59313	298652	139351
analfabeta	9861	27032	42394
ignorado	4912	5395	3762

Cuadro 40: Tabla de contingencia entre AGR_NIV_ESCOLARIDAD y AGR_ALFAB mostrando frecuencias absolutas y relativas

AGR_NIV_ESCOLARIDAD AGR_ALFAB	basicos	diversificado	ignorado	ninguno	primaria	universidad
alfabeta	96768	101470	17374	16420	249118	16166
analfabeta	0	0	0	79287	0	0
ignorado	0	0	14069	0	0	0

Cuadro 41: Tabla de contingencia entre AGR_SEXO y AGR_ALFAB mostrando frecuencias absolutas y relativas

AGR_SEXO AGR_ALFAB	hombres	mujeres
alfabeta	435165	62151
analfabeta	64756	14531
ignorado	12008	2061

Cuadro 42: Tabla de contingencia entre VIOLENCIA_FISICA y AGR_ALFAB mostrando frecuencias absolutas y relativas

VIOLENCIA_FISICA AGR_ALFAB	no presente	presente
alfabeta	213881	283435
analfabeta	29987	49300
ignorado	5959	8110

Cuadro 43: Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y AGR_ALFAB mostrando frecuencias absolutas y relativas

VIOLENCIA_PATRIMONIAL AGR_ALFAB	no presente	presente
alfabeta	430370	66946
analfabeta	69412	9875
ignorado	13012	1057

Cuadro 44: Tabla de contingencia entre **VIOLENCIA_PSICOLOGICA** y **AGR_ALFAB** mostrando frecuencias absolutas y relativas

VIOLENCIA_PSICOLOGICA AGR_ALFAB	no presente	presente
alfabeta	64381	432935
analfabeta	13865	65422
ignorado	3003	11066

Cuadro 45: Tabla de contingencia entre **VIOLENCIA_SEXUAL** y **AGR_ALFAB** mostrando frecuencias absolutas y relativas

VIOLENCIA_SEXUAL AGR_ALFAB	no presente	presente
alfabeta	484800	12516
analfabeta	77449	1838
ignorado	13733	336

Cuadro 46: Tabla de contingencia entre **AGR_TRABAJA** y **AGR_EST_CIV** mostrando frecuencias absolutas y relativas

AGR_TRABAJA AGR_EST_CIV	ignorado	no	si
casados(as)	6622	37599	203013
ignorado	7714	13592	60613
otro	441	2545	9930
solteros(as)	2418	24584	70510
unidos(as)	3450	23263	122630
viudos(as)	37	804	907

Cuadro 47: Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y AGR_EST_CIV mostrando frecuencias absolutas y relativas

DIF_ESCOLARIDAD_AGR_VIC AGR_EST_CIV	-1	-2	-3	-4	0	1	2	3	4	ignorado
casados(as)	35452	11025	2355	144	126898	44626	11062	2118	117	13437
ignorado	12475	4032	846	64	38129	12166	3297	511	17	10382
otro	2152	723	185	12	6006	2180	586	120	10	942
solteros(as)	14099	4286	1007	68	47489	19295	5006	1033	52	5177
unidos(as)	21254	5910	1140	59	76098	29092	6533	1182	43	8032
viudos(as)	424	191	94	8	699	169	49	13	0	101

Cuadro 48: Tabla de contingencia entre GANA_MAS_LA_VICTIMA y AGR_EST_CIV mostrando frecuencias absolutas y relativas

GANA_MAS_LA_VICTIMA AGR_EST_CIV	no	si
casados(as)	208173	39061
ignorado	68100	13819
otro	10227	2689
solteros(as)	80673	16839
unidos(as)	128428	20915
viudos(as)	1389	359

Cuadro 49: Tabla de contingencia entre VIC_TRABAJA y AGR_EST_CIV mostrando frecuencias absolutas y relativas

VIC_TRABAJA AGR_EST_CIV	ignorado	no	si
casados(as)	1266	152974	92994
ignorado	929	48355	32635
otro	90	7305	5521
solteros(as)	908	60801	35803
unidos(as)	765	103489	45089
viudos(as)	14	1038	696

Cuadro 50: Tabla de contingencia entre AGR_NIV_ESCOLARIDAD y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas

AGR_NIV_ESCOLARIDAD AGR_ES_INDIGENA	basicos	diversificado	ignorado	ninguno	primaria	universidad
ignorado	10266	10208	9051	13215	29847	1499
no	63259	71635	15041	34157	134341	12646
si	23243	19627	7351	48335	84930	2021

Cuadro 51: Tabla de contingencia entre AGR_SEXO y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas

AGR_SEXO AGR_ES_INDIGENA	hombres	mujeres
ignorado	65612	8474
no	283763	47316
si	162554	22953

Cuadro 52: Tabla de contingencia entre VIOLENCIA_FISICA y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas

VIOLENCIA_FISICA AGR_ES_INDIGENA	no presente	presente
ignorado	31352	42734
no	152270	178809
si	66205	119302

Cuadro 53: Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas

VIOLENCIA_PATRIMONIAL AGR_ES_INDIGENA	no presente	presente
ignorado	67869	6217
no	284329	46750
si	160596	24911

Cuadro 54: Tabla de contingencia entre VIOLENCIA_PSICOLOGICA y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas

VIOLENCIA_PSICOLOGICA AGR_ES_INDIGENA	no presente	presente
ignorado	12402	61684
no	38971	292108
si	29876	155631

Cuadro 55: Tabla de contingencia entre VIOLENCIA_SEXUAL y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas

VIOLENCIA_SEXUAL AGR_ES_INDIGENA	no presente	presente
ignorado	72485	1601
no	322245	8834
si	181252	4255

Cuadro 56: Tabla de contingencia entre AGR_SEXO y AGR_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas

AGR_SEXO AGR_NIV_ESCOLARIDAD	hombres	mujeres
basicos	84957	11811
diversificado	88090	13380
ignorado	27118	4325
ninguno	79068	16639
primaria	218700	30418
universidad	13996	2170

Cuadro 57: Tabla de contingencia entre VIOLENCIA_FISICA y AGR_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas

VIOLENCIA_FISICA AGR_NIV_ESCOLARIDAD	no presente	presente
basicos	41554	55214
diversificado	48883	52587
ignorado	13698	17745
ninguno	36240	59467
primaria	100281	148837
universidad	9171	6995

Cuadro 58: Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y AGR_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas

VIOLENCIA_PATRIMONIAL AGR_NIV_ESCOLARIDAD	no presente	presente
basicos	83551	13217
diversificado	86710	14760
ignorado	28695	2748
ninguno	84079	11628
primaria	216090	33028
universidad	13669	2497

Cuadro 59: Tabla de contingencia entre VIOLENCIA_PSICOLOGICA y AGR_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas

VIOLENCIA_PSICOLOGICA AGR_NIV_ESCOLARIDAD	no presente	presente
basicos	12151	84617
diversificado	10457	91013
ignorado	5891	25552
ninguno	16953	78754
primaria	34637	214481
universidad	1160	15006

Cuadro 60: Tabla de contingencia entre VIOLENCIA _SEXUAL y AGR _NIV _ESCOLARIDAD mostrando frecuencias absolutas y relativas

VIOLENCIA _SEXUAL AGR _NIV _ESCOLARIDAD	no presente	presente
basicos	94259	2509
diversificado	98952	2518
ignorado	30630	813
ninguno	93449	2258
primaria	242992	6126
universidad	15700	466

Cuadro 61: Tabla de contingencia entre VIOLENCIA _FISICA y AGR _SEXO mostrando frecuencias absolutas y relativas

VIOLENCIA _FISICA AGR _SEXO	no presente	presente
hombres	202913	309016
mujeres	46914	31829

Cuadro 62: Tabla de contingencia entre VIOLENCIA _PATRIMONIAL y AGR _SEXO mostrando frecuencias absolutas y relativas

VIOLENCIA _PATRIMONIAL AGR _SEXO	no presente	presente
hombres	442717	69212
mujeres	70077	8666

Cuadro 63: Tabla de contingencia entre VIOLENCIA _PSICOLOGICA y AGR _SEXO mostrando frecuencias absolutas y relativas

VIOLENCIA _PSICOLOGICA AGR _SEXO	no presente	presente
hombres	73548	438381
mujeres	7701	71042

Cuadro 64: Tabla de contingencia entre VIOLENCIA _SEXUAL y AGR _SEXO mostrando frecuencias absolutas y relativas

VIOLENCIA _SEXUAL AGR _SEXO	no presente	presente
hombres	497647	14282
mujeres	78335	408

Cuadro 65: Tabla de contingencia entre DIF _ESCOLARIDAD _AGR _VIC y AGR _TRABAJA mostrando frecuencias absolutas y relativas

DIF _ESCOLARIDAD _AGR _VIC AGR _TRABAJA	-1	-2	-3	-4	0	1	2	3	4	ignorado
ignorado	1760	561	203	9	5613	1451	353	64	12	10656
no	19921	7062	1973	146	47884	14709	4134	801	31	5726
si	64175	18544	3451	200	241822	91368	22046	4112	196	21689

Cuadro 66: Tabla de contingencia entre VIC _TRABAJA y AGR _TRABAJA mostrando frecuencias absolutas y relativas

VIC _TRABAJA AGR _TRABAJA	ignorado	no	si
ignorado	1356	12798	6528
no	957	51229	50201
si	1659	309935	156009

Cuadro 67: Tabla de contingencia entre AGR _TRABAJA y GANA _MAS _LA _VICTIMA mostrando frecuencias absolutas y relativas

AGR _TRABAJA GANA _MAS _LA _VICTIMA	ignorado	no	si
no	20682	54413	421895
si	0	47974	45708

Cuadro 68: Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y GANA_MAS_LA_VICTIMA mostrando frecuencias absolutas y relativas

DIF_ESCOLARIDAD_AGR_VIC GANA_MAS_LA_VICTIMA	-1	-2	-3	-4	0	1	2	3	4	ignorado
no	63856	16482	3174	183	254234	96632	23490	4501	227	34211
si	22000	9685	2453	172	41085	10896	3043	476	12	3860

Cuadro 69: Tabla de contingencia entre VIC_TRABAJA y GANA_MAS_LA_VICTIMA mostrando frecuencias absolutas y relativas

VIC_TRABAJA GANA_MAS_LA_VICTIMA	ignorado	no	si
no	3972	362391	130627
si	0	11571	82111

Cuadro 70: Tabla de contingencia entre AGR_ES_INDIGENA y AGR_ALFAB mostrando frecuencias absolutas y relativas

AGR_ES_INDIGENA AGR_ALFAB	ignorado	no	si
alfabeta	10.041614	50.561394	23.591943
analfabeta	1.669454	4.576482	7.177249
ignorado	0.831595	0.913366	0.636902

Cuadro 71: Tabla de contingencia entre AGR_NIV_ESCOLARIDAD y AGR_ALFAB mostrando frecuencias absolutas y relativas

AGR_NIV_ESCOLARIDAD AGR_ALFAB	basicos	diversificado	ignorado	ninguno	primaria	universidad
alfabeta	16.382696	17.178739	2.941396	2.779885	42.175353	2.736883
analfabeta	0.000000	0.000000	0.000000	13.423186	0.000000	0.000000
ignorado	0.000000	0.000000	2.381863	0.000000	0.000000	0.000000

Cuadro 72: Tabla de contingencia entre AGR_SEXO y AGR_ALFAB mostrando frecuencias absolutas y relativas

AGR_SEXO AGR_ALFAB	hombres	mujeres
alfabeta	73.672868	10.522083
analfabeta	10.963106	2.460079
ignorado	2.032939	0.348925

Cuadro 73: Tabla de contingencia entre VIOLENCIA_FISICA y AGR_ALFAB mostrando frecuencias absolutas y relativas

VIOLENCIA_FISICA AGR_ALFAB	no presente	presente
alfabeta	36.209775	47.985176
analfabeta	5.076760	8.346426
ignorado	1.008851	1.373012

Cuadro 74: Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y AGR_ALFAB mostrando frecuencias absolutas y relativas

VIOLENCIA_PATRIMONIAL AGR_ALFAB	no presente	presente
alfabeta	72.861080	11.333871
analfabeta	11.751361	1.671825
ignorado	2.202915	0.178949

Cuadro 75: Tabla de contingencia entre VIOLENCIA_PSICOLOGICA y AGR_ALFAB mostrando frecuencias absolutas y relativas

VIOLENCIA_PSICOLOGICA AGR_ALFAB	no presente	presente
alfabeta	10.899619	73.295331
analfabeta	2.347326	11.075859
ignorado	0.508404	1.873459

Cuadro 76: Tabla de contingencia entre VIOLENCIA_SEXUAL y AGR_ALFAB mostrando frecuencias absolutas y relativas

VIOLENCIA_SEXUAL AGR_ALFAB	no presente	presente
alfabeta	82.076008	2.118942
analfabeta	13.112015	0.311171
ignorado	2.324979	0.056884

Cuadro 77: Tabla de contingencia entre AGR_TRABAJA y AGR_EST_CIV mostrando frecuencias absolutas y relativas

AGR_TRABAJA AGR_EST_CIV	ignorado	no	si
casados(as)	1.121096	6.365462	34.369836
ignorado	1.305970	2.301108	10.261702
otro	0.074661	0.430865	1.681136
solteros(as)	0.409364	4.162039	11.937251
unidos(as)	0.584081	3.938396	20.761099
viudos(as)	0.006264	0.136116	0.153554

Cuadro 78: Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y AGR_EST_CIV mostrando frecuencias absolutas y relativas

DIF_ESCOLARIDAD_AGR_VIC AGR_EST_CIV	-1	-2	-3	-4	0	1	2	3	4	ignorado
casados(as)	6.001977	1.866518	0.398698	0.024379	21.483666	7.555124	1.872782	0.358575	0.019808	2.274867
ignorado	2.112001	0.682612	0.143227	0.010835	6.455190	2.059688	0.558178	0.086512	0.002878	1.757659
otro	0.364331	0.122403	0.031320	0.002032	1.016808	0.369071	0.099209	0.020316	0.001693	0.159479
solteros(as)	2.386942	0.725614	0.170484	0.011512	8.039826	3.266618	0.847509	0.174886	0.008804	0.876459
unidos(as)	3.598275	1.000555	0.193001	0.009989	12.883292	4.925238	1.106028	0.200111	0.007280	1.359807
viudos(as)	0.071783	0.032336	0.015914	0.001354	0.118340	0.028611	0.008296	0.002201	0.000000	0.017099

Cuadro 79: Tabla de contingencia entre GANA_MAS_LA_VICTIMA y AGR_EST_CIV mostrando frecuencias absolutas y relativas

GANA_MAS_LA_VICTIMA AGR_EST_CIV	no	si
casados(as)	35.243418	6.612976
ignorado	11.529241	2.339539
otro	1.731418	0.455244
solteros(as)	13.657834	2.850821
unidos(as)	21.742693	3.540882
viudos(as)	0.235156	0.060778

Cuadro 80: Tabla de contingencia entre VIC_TRABAJA y AGR_EST_CIV mostrando frecuencias absolutas y relativas

VIC_TRABAJA AGR_EST_CIV	ignorado	no	si
casados(as)	0.214332	25.898299	15.743763
ignorado	0.157278	8.186438	5.525063
otro	0.015237	1.236727	0.934698
solteros(as)	0.153723	10.293530	6.061401
unidos(as)	0.129514	17.520553	7.633509
viudos(as)	0.002370	0.175732	0.117832

Cuadro 81: Tabla de contingencia entre AGR_NIV_ESCOLARIDAD y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas

AGR_NIV_ESCOLARIDAD AGR_ES_INDIGENA	basicos	diversificado	ignorado	ninguno	primaria	universidad
ignorado	1.738020	1.728201	1.532323	2.237282	5.053058	0.253779
no	10.709666	12.127712	2.546422	5.782736	22.743756	2.140951
si	3.935010	3.322826	1.244515	8.183053	14.378538	0.342153

Cuadro 82: Tabla de contingencia entre AGR_SEXO y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas

AGR_SEXO AGR_ES_INDIGENA	hombres	mujeres
ignorado	11.108026	1.434637
no	48.040706	8.010537
si	27.520180	3.885913

Cuadro 83: Tabla de contingencia entre VIOLENCIA_FISICA y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas

VIOLENCIA_FISICA AGR_ES_INDIGENA	no presente	presente
ignorado	5.307853	7.234811
no	25.779113	30.272131
si	11.208420	20.197673

Cuadro 84: Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas

VIOLENCIA_PATRIMONIAL AGR_ES_INDIGENA	no presente	presente
ignorado	11.490133	1.052530
no	48.136529	7.914714
si	27.188694	4.217400

Cuadro 85: Tabla de contingencia entre VIOLENCIA_PSICOLOGICA y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas

VIOLENCIA_PSICOLOGICA AGR_ES_INDIGENA	no presente	presente
ignorado	2.099642	10.443021
no	6.597740	49.453504
si	5.057968	26.348126

Cuadro 86: Tabla de contingencia entre VIOLENCIA_SEXUAL y AGR_ES_INDIGENA mostrando frecuencias absolutas y relativas

VIOLENCIA_SEXUAL AGR_ES_INDIGENA	no presente	presente
ignorado	12.271616	0.271047
no	54.555659	1.495585
si	30.685727	0.720366

Cuadro 87: Tabla de contingencia entre AGR_SEXO y AGR_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas

AGR_SEXO AGR_NIV_ESCOLARIDAD	hombres	mujeres
basicos	14.383109	1.999587
diversificado	14.913522	2.265217
ignorado	4.591042	0.732217
ninguno	13.386109	2.816961
primaria	37.025625	5.149728
universidad	2.369505	0.367378

Cuadro 88: Tabla de contingencia entre VIOLENCIA_FISICA y AGR_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas

VIOLENCIA_FISICA AGR_NIV_ESCOLARIDAD	no presente	presente
basicos	7.035038	9.347658
diversificado	8.275828	8.902911
ignorado	2.319054	3.004205
ninguno	6.135385	10.067686
primaria	16.977443	25.197910
universidad	1.552638	1.184244

Cuadro 89: Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y AGR_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas

VIOLENCIA_PATRIMONIAL AGR_NIV_ESCOLARIDAD	no presente	presente
basicos	14.145075	2.237621
diversificado	14.679890	2.498849
ignorado	4.858026	0.465233
ninguno	14.234465	1.968605
primaria	36.583755	5.591597
universidad	2.314144	0.422739

Cuadro 90: Tabla de contingencia entre VIOLENCIA_PSICOLOGICA y AGR_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas

VIOLENCIA_PSICOLOGICA AGR_NIV_ESCOLARIDAD	no presente	presente
basicos	2.057148	14.325548
diversificado	1.770356	15.408382
ignorado	0.997339	4.325920
ninguno	2.870121	13.332950
primaria	5.863999	36.311354
universidad	0.196386	2.540496

Cuadro 91: Tabla de contingencia entre VIOLENCIA _SEXUAL y AGR _NIV _ESCOLARIDAD mostrando frecuencias absolutas y relativas

VIOLENCIA _SEXUAL AGR _NIV _ESCOLARIDAD	no presente	presente
basicos	15.957926	0.424770
diversificado	16.752445	0.426294
ignorado	5.185619	0.137640
ninguno	15.820794	0.382276
primaria	41.138229	1.037124
universidad	2.657990	0.078893

Cuadro 92: Tabla de contingencia entre VIOLENCIA _FISICA y AGR _SEXO mostrando frecuencias absolutas y relativas

VIOLENCIA _FISICA AGR _SEXO	no presente	presente
hombres	34.352907	52.316006
mujeres	7.942479	5.388608

Cuadro 93: Tabla de contingencia entre VIOLENCIA _PATRIMONIAL y AGR _SEXO mostrando frecuencias absolutas y relativas

VIOLENCIA _PATRIMONIAL AGR _SEXO	no presente	presente
hombres	74.951411	11.717501
mujeres	11.863945	1.467143

Cuadro 94: Tabla de contingencia entre VIOLENCIA _PSICOLOGICA y AGR _SEXO mostrando frecuencias absolutas y relativas

VIOLENCIA _PSICOLOGICA AGR _SEXO	no presente	presente
hombres	12.451581	74.217332
mujeres	1.303769	12.027318

Cuadro 95: Tabla de contingencia entre VIOLENCIA _SEXUAL y AGR _SEXO mostrando frecuencias absolutas y relativas

VIOLENCIA _SEXUAL AGR _SEXO	no presente	presente
hombres	84.250989	2.417924
mujeres	13.262013	0.069074

Cuadro 96: Tabla de contingencia entre DIF _ESCOLARIDAD _AGR _VIC y AGR _TRABAJA mostrando frecuencias absolutas y relativas

DIF _ESCOLARIDAD _AGR _VIC AGR _TRABAJA	-1	-2	-3	-4	0	1	2	3	4	ignorado
ignorado	0.297966	0.094977	0.034368	0.001524	0.950274	0.245652	0.059762	0.010835	0.002032	1.804047
no	3.372599	1.195587	0.334026	0.024718	8.106699	2.490215	0.699881	0.135608	0.005248	0.969404
si	10.864744	3.139475	0.584250	0.033860	40.940150	15.468483	3.732359	0.696156	0.033183	3.671919

Cuadro 97: Tabla de contingencia entre VIC _TRABAJA y AGR _TRABAJA mostrando frecuencias absolutas y relativas

VIC _TRABAJA AGR _TRABAJA	ignorado	no	si
ignorado	0.229569	2.166685	1.105182
no	0.162019	8.673003	8.498964
si	0.280867	52.471592	26.412120

Cuadro 98: Tabla de contingencia entre AGR _TRABAJA y GANA _MAS _LA _VICTIMA mostrando frecuencias absolutas y relativas

AGR _TRABAJA GANA _MAS _LA _VICTIMA	ignorado	no	si
no	3.501436	9.212050	71.426274
si	0.000000	8.121936	7.738305

Cuadro 99: Tabla de contingencia entre DIF _ESCOLARIDAD _AGR _VIC y GANA _MAS _LA _VICTIMA mostrando frecuencias absolutas y relativas

DIF _ESCOLARIDAD _AGR _VIC GANA _MAS _LA _VICTIMA	-1	-2	-3	-4	0	1	2	3	4	ignorado
no	10.810738	2.790381	0.537354	0.030982	43.041485	16.359672	3.976826	0.762013	0.038431	5.791878
si	3.724571	1.639658	0.415290	0.029119	6.955637	1.844679	0.515176	0.080586	0.002032	0.653493

Cuadro 100: Tabla de contingencia entre VIC_TRABAJA y GANA_MAS_LA_VICTIMA mostrando frecuencias absolutas y relativas

VIC_TRABAJA GANA_MAS_LA_VICTIMA	ignorado	no	si
no	0.672454	61.352324	22.114981
si	0.000000	1.958955	13.901285

Cuadro 101: Tabla de contingencia entre CANTIDAD_HIJOS y VIC_ALFAB mostrando frecuencias absolutas y relativas

CANTIDAD_HIJOS VIC_ALFAB	demasiados hijos	hijo unico	hijos medios	ignorado	muchos hijos	sin hijos
alfabeta	1.052022	19.045934	28.085469	15.061997	7.817029	9.329543
analfabeta	1.365902	2.624130	5.329523	3.198222	4.118529	2.145014
ignorado	0.007957	0.076015	0.088205	0.560379	0.039277	0.054853

Cuadro 102: Tabla de contingencia entre VIC_ES_INDIGENA y VIC_ALFAB mostrando frecuencias absolutas y relativas

VIC_ES_INDIGENA VIC_ALFAB	ignorado	no	si
alfabeta	9.417579	50.577478	20.396938
analfabeta	2.313128	5.998591	10.469601
ignorado	0.219411	0.432897	0.174378

Cuadro 103: Tabla de contingencia entre VIC_NIV_ESCOLARIDAD y VIC_ALFAB mostrando frecuencias absolutas y relativas

VIC_NIV_ESCOLARIDAD VIC_ALFAB	basicos	diversificado	ignorado	ninguno	primaria	universidad
alfabeta	14.672949	18.142387	2.042419	2.401163	39.770296	3.362780
analfabeta	0.000000	0.000000	0.000000	18.781320	0.000000	0.000000
ignorado	0.000000	0.000000	0.826686	0.000000	0.000000	0.000000

Cuadro 104: Tabla de contingencia entre VIC_TRABAJA y VIC_ALFAB mostrando frecuencias absolutas y relativas

VIC_TRABAJA VIC_ALFAB	ignorado	no	si
alfabeta	0.360098	48.431955	31.599940
analfabeta	0.063995	14.506697	4.210628
ignorado	0.248361	0.372626	0.205698

Cuadro 105: Tabla de contingencia entre AGR_EST_CIV y VIC_EST_CIV mostrando frecuencias absolutas y relativas

AGR_EST_CIV VIC_EST_CIV	casados(as)	ignorado	otro	solteros(as)	unidos(as)	viudos(as)
casados(as)	36.360112	0.965341	0.141872	3.210242	1.411951	0.092606
ignorado	0.875444	11.348600	0.065349	0.398529	0.607613	0.023363
otro	0.079740	0.051298	1.704330	0.093791	0.106489	0.005079
solteros(as)	3.030785	0.802814	0.161680	11.277325	1.786440	0.105981
unidos(as)	1.258397	0.645028	0.094130	1.253149	21.204493	0.045880
viudos(as)	0.251916	0.055699	0.019300	0.275618	0.166590	0.023025

Cuadro 106: Tabla de contingencia entre AGR_TRABAJA y VIC_EST_CIV mostrando frecuencias absolutas y relativas

AGR_TRABAJA VIC_EST_CIV	ignorado	no	si
casados(as)	1.329333	6.860660	33.992131
ignorado	0.806200	2.341062	10.171635
otro	0.060101	0.357728	1.622897
solteros(as)	0.652816	3.775192	12.737018
unidos(as)	0.622342	3.732867	20.145868
viudos(as)	0.030643	0.266476	0.495029

Cuadro 107: Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y VIC_EST_CIV mostrando frecuencias absolutas y relativas

DIF_ESCOLARIDAD_AGR_VIC VIC_EST_CIV	-1	-2	-3	-4	0	1	2	3	4	ignorado
casados(as)	5.838604	1.744623	0.368055	0.023194	21.646193	7.808056	1.967928	0.360776	0.017607	2.407089
ignorado	2.086268	0.678211	0.137471	0.009311	6.367493	2.044620	0.564442	0.085327	0.002709	1.343047
otro	0.330302	0.110721	0.022855	0.001524	0.951289	0.374997	0.098532	0.018623	0.002032	0.129852
solteros(as)	2.740777	0.923016	0.235833	0.017776	8.194903	3.021474	0.709700	0.147967	0.008126	1.165452
unidos(as)	3.477903	0.953152	0.184366	0.007449	12.539616	4.717000	1.069798	0.196048	0.007788	1.347956
viudos(as)	0.061455	0.020316	0.004063	0.000846	0.297627	0.238203	0.081602	0.033860	0.002201	0.051975

Cuadro 108: Tabla de contingencia entre GANA_MAS_LA_VICTIMA y VIC_EST_CIV mostrando frecuencias absolutas y relativas

GANA_MAS_LA_VICTIMA VIC_EST_CIV	no	si
casados(as)	35.177391	7.004734
ignorado	11.054697	2.264201
otro	1.602074	0.438653
solteros(as)	14.587622	2.577403
unidos(as)	21.076672	3.424405
viudos(as)	0.641303	0.150845

Cuadro 109: Tabla de contingencia entre VIC_REL_AGR y VIC_EST_CIV mostrando frecuencias absolutas y relativas

VIC_REL_AGR VIC_EST_CIV	conviviente	esposos(as)	ex-conyuges	hermanos(as)	hijastros(as)	hijos(as)	nietos(as)	otro pariente	padres/madres	suegros(as)
casados(as)	1.260260	31.345654	1.470698	1.567367	0.045033	0.310494	0.023702	3.485691	2.034632	0.638595
ignorado	5.501530	0.175393	6.214278	0.162019	0.038769	0.364161	0.026241	0.631146	0.152369	0.052990
otro	0.034537	0.009311	1.685030	0.049266	0.004063	0.023194	0.000339	0.133915	0.080755	0.020316
solteros(as)	4.580376	0.384985	3.445228	2.278591	0.212639	1.319683	0.071783	3.256630	1.301060	0.314049
unidos(as)	18.692777	0.609814	2.270296	0.601857	0.037584	0.215179	0.010497	1.472391	0.446779	0.143904
viudos(as)	0.022686	0.004063	0.014560	0.055022	0.003725	0.009311	0.001354	0.241081	0.361622	0.078724

Cuadro 110: Tabla de contingencia entre VIC_TRABAJA y VIC_EST_CIV mostrando frecuencias absolutas y relativas

VIC_TRABAJA VIC_EST_CIV	ignorado	no	si
casados(as)	0.183520	25.796042	16.202563
ignorado	0.283406	7.810765	5.224727
otro	0.009142	1.145983	0.885601
solteros(as)	0.101241	10.880489	6.183296
unidos(as)	0.091083	17.130319	7.279675
viudos(as)	0.004063	0.547681	0.240404

Cuadro 111: Tabla de contingencia entre CANTIDAD_HIJOS y VIC_ES_INDIGENA mostrando frecuencias absolutas y relativas

CANTIDAD_HIJOS VIC_ES_INDIGENA	demasiados hijos	hijo unico	hijos medios	ignorado	muchos hijos	sin hijos
ignorado	0.264106	2.439425	3.629087	2.929545	1.304446	1.383509
no	0.927080	12.683181	19.980802	11.361636	5.770377	6.285891
si	1.234695	6.623473	9.893308	4.529417	4.900012	3.860010

Cuadro 112: Tabla de contingencia entre VIC_NIV_ESCOLARIDAD y VIC_ES_INDIGENA mostrando frecuencias absolutas y relativas

VIC_NIV_ESCOLARIDAD VIC_ES_INDIGENA	basicos	diversificado	ignorado	ninguno	primaria	universidad
ignorado	1.531984	1.685030	0.828379	2.809681	4.823828	0.271217
no	10.031286	13.465172	1.480179	7.003210	22.283264	2.745856
si	3.109678	2.992185	0.560548	11.369593	12.663204	0.345708

Cuadro 113: Tabla de contingencia entre VIC_TRABAJA y VIC_ES_INDIGENA mostrando frecuencias absolutas y relativas

VIC_TRABAJA VIC_ES_INDIGENA	ignorado	no	si
ignorado	0.150845	8.333728	3.465544
no	0.409534	32.854105	23.745327
si	0.112076	22.123446	8.805395

Cuadro 114: Tabla de contingencia entre CANTIDAD_HIJOS y VIC_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas

CANTIDAD_HIJOS VIC_NIV_ESCOLARIDAD	demasiados hijos	hijo unico	hijos medios	ignorado	muchos hijos	sin hijos
basicos	0.053837	3.937549	5.212368	2.573679	0.981594	1.913922
diversificado	0.039954	5.279241	6.722005	3.219553	0.855297	2.026336
ignorado	0.021670	0.437976	0.535492	1.467481	0.144920	0.261566
ninguno	1.448181	3.038912	6.006887	3.807697	4.461359	2.419448
primaria	0.856313	8.116180	13.738759	7.132046	5.394026	4.532973
universidad	0.005925	0.936222	1.287686	0.620141	0.137640	0.375166

Cuadro 115: Tabla de contingencia entre VIC_TRABAJA y VIC_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas

VIC_TRABAJA VIC_NIV_ESCOLARIDAD	ignorado	no	si
basicos	0.046049	9.279600	5.347299
diversificado	0.078385	7.269178	10.794824
ignorado	0.312695	1.666407	0.890003
ninguno	0.086004	16.242178	4.854302
primaria	0.133915	28.035356	11.601024
universidad	0.015406	0.818559	2.528815

Cuadro 116: Tabla de contingencia entre AGR_EST_CIV y VIC_REL_AGR mostrando frecuencias absolutas y relativas

AGR_EST_CIV VIC_REL_AGR	casados(as)	ignorado	otro	solteros(as)	unidos(as)	viudos(as)
conviviente	1.556702	5.693007	0.065349	4.212659	18.545826	0.018623
esposos(as)	31.603326	0.163204	0.007111	0.260551	0.493336	0.001693
ex-conyuges	1.709070	6.178895	1.680120	3.069724	2.447043	0.015237
hermanos(as)	1.401116	0.311002	0.074153	2.157881	0.744745	0.025226
hijastros(as)	0.095654	0.078047	0.007449	0.046557	0.110213	0.003894
hijos(as)	0.950612	0.166590	0.055869	0.578155	0.409026	0.081771
nietos(as)	0.055699	0.010158	0.004402	0.034029	0.016253	0.013375
otro pariente	3.122207	0.730016	0.220766	3.482813	1.556194	0.108859
padres/madres	0.882385	0.272063	0.059762	2.557426	0.587805	0.017776
suegros(as)	0.479623	0.265799	0.011682	0.108859	0.373134	0.009481

Cuadro 117: Tabla de contingencia entre AGR_TRABAJA y VIC_REL_AGR mostrando frecuencias absolutas y relativas

AGR_TRABAJA VIC_REL_AGR	ignorado	no	si
conviviente	0.974653	3.624008	25.493506
esposos(as)	0.891866	3.821410	27.815945
ex-conyuges	0.635378	2.255397	12.209314
hermanos(as)	0.216364	1.381139	3.116620
hijastros(as)	0.016083	0.085834	0.239896
hijos(as)	0.080755	0.816020	1.345247
nietos(as)	0.004910	0.074491	0.054514
otro pariente	0.453043	3.405782	5.362028
padres/madres	0.172685	1.385202	2.819331
suegros(as)	0.055699	0.484702	0.708176

Cuadro 118: Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y VIC_REL_AGR mostrando frecuencias absolutas y relativas

DIF_ESCOLARIDAD_AGR_VIC VIC_REL_AGR	-1	-2	-3	-4	0	1	2	3	4	ignorado
conviviente	4.090764	1.080295	0.198418	0.008126	15.719892	5.688775	1.242822	0.208068	0.006433	1.848572
esposos(as)	4.421235	1.241637	0.221104	0.012867	17.421175	5.929348	1.366240	0.218734	0.008465	1.688416
ex-conyuges	2.530169	0.828379	0.159479	0.011512	7.290002	2.510869	0.661450	0.098870	0.005587	1.003772
hermanos(as)	0.757612	0.251070	0.057223	0.004402	2.460418	0.661789	0.156263	0.027765	0.001354	0.336227
hijastros(as)	0.083634	0.031828	0.007280	0.002201	0.124773	0.036907	0.011682	0.002370	0.000000	0.041140
hijos(as)	0.586281	0.268338	0.077370	0.004910	0.761844	0.186398	0.053837	0.009142	0.000508	0.293395
nietos(as)	0.041478	0.015914	0.007111	0.000508	0.038600	0.004571	0.002032	0.000169	0.000000	0.023533
otro pariente	1.659127	0.633346	0.204851	0.014729	3.904197	1.430405	0.424263	0.101241	0.008126	0.840568
padres/madres	0.263090	0.051805	0.010158	0.000508	1.774250	1.406026	0.457106	0.136455	0.006772	0.271047
suegros(as)	0.101918	0.027426	0.009650	0.000339	0.501971	0.349263	0.116308	0.039785	0.003217	0.098701

Cuadro 119: Tabla de contingencia entre GANA_MAS_LA_VICTIMA y VIC_REL_AGR mostrando frecuencias absolutas y relativas

GANA_MAS_LA_VICTIMA VIC_REL_AGR	no	si
conviviente	26.247731	3.844435
esposos(as)	27.522889	5.006332
ex-conyuges	12.130252	2.969838
hermanos(as)	3.828690	0.885432
hijastros(as)	0.298643	0.043171
hijos(as)	1.930174	0.311848
nietos(as)	0.120033	0.013882
otro pariente	7.565620	1.655233
padres/madres	3.468930	0.908288
suegros(as)	1.026797	0.221781

Cuadro 120: Tabla de contingencia entre VIC_TRABAJA y VIC_REL_AGR mostrando frecuencias absolutas y relativas

VIC_TRABAJA VIC_REL_AGR	ignorado	no	si
conviviente	0.145428	21.401218	8.545521
esposos(as)	0.131884	20.453314	11.944023
ex-conyuges	0.075169	7.814659	7.210262
hermanos(as)	0.031320	2.711319	1.971483
hijastros(as)	0.010327	0.236680	0.094807
hijos(as)	0.148644	1.422278	0.671100
nietos(as)	0.009311	0.095146	0.029458
otro pariente	0.087866	5.839112	3.293875
padres/madres	0.026241	2.492415	1.858561
suegros(as)	0.006264	0.845139	0.397175

Cuadro 121: Tabla de contingencia entre CANTIDAD_HIJOS y VIC_TRABAJO mostrando frecuencias absolutas y relativas

CANTIDAD_HIJOS VIC_TRABAJO	demasiados hijos	hijo unico	hijos medios	ignorado	muchos hijos	sin hijos
ignorado	0.005587	0.073137	0.094977	0.422739	0.030982	0.045033
no	1.671317	14.062288	20.361216	11.853110	7.795528	7.567821
si	0.748977	7.610654	13.047004	6.544749	4.148326	3.916556

Cuadro 122: Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y VIC_TRABAJO mostrando frecuencias absolutas y relativas

DIF_ESCOLARIDAD_AGR_VIC VIC_TRABAJO	-1	-2	-3	-4	0	1	2	3	4	ignorado
ignorado	0.044187	0.014390	0.005418	0.000339	0.165066	0.047573	0.013375	0.002370	0.000677	0.379060
no	7.992591	1.947951	0.372965	0.017776	32.606421	12.739388	3.048562	0.590852	0.027426	3.967346
si	6.498530	2.467698	0.574261	0.041986	17.225635	5.417389	1.430066	0.249377	0.012359	2.098965

Cuadro 123: Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y VIOLENCIA_FISICA mostrando frecuencias absolutas y relativas

VIOLENCIA_PATRIMONIAL VIOLENCIA_FISICA	no presente	presente
no presente	36.312539	5.982847
presente	50.502817	7.201797

Cuadro 124: Tabla de contingencia entre VIOLENCIA_PSICOLOGICA y VIOLENCIA_FISICA mostrando frecuencias absolutas y relativas

VIOLENCIA_PSICOLOGICA VIOLENCIA_FISICA	no presente	presente
no presente	1.432436	40.862949
presente	12.322914	45.381701

Cuadro 125: Tabla de contingencia entre VIOLENCIA_SEXUAL y VIOLENCIA_FISICA mostrando frecuencias absolutas y relativas

VIOLENCIA_SEXUAL VIOLENCIA_FISICA	no presente	presente
no presente	41.511194	0.784192
presente	56.001808	1.702806

Cuadro 126: Tabla de contingencia entre VIOLENCIA _PATRIMONIAL y VIOLENCIA _PSICOLOGICA mostrando frecuencias absolutas y relativas

VIOLENCIA _PATRIMONIAL VIOLENCIA _PSICOLOGICA	no presente	presente
no presente	12.297857	1.457492
presente	74.517499	11.727151

Cuadro 127: Tabla de contingencia entre VIOLENCIA _SEXUAL y VIOLENCIA _PSICOLOGICA mostrando frecuencias absolutas y relativas

VIOLENCIA _SEXUAL VIOLENCIA _PSICOLOGICA	no presente	presente
no presente	13.245253	0.510097
presente	84.267749	1.976901

Cuadro 128: Tabla de contingencia entre VIOLENCIA _PATRIMONIAL y VIOLENCIA _SEXUAL mostrando frecuencias absolutas y relativas

VIOLENCIA _PATRIMONIAL VIOLENCIA _SEXUAL	no presente	presente
no presente	84.992178	12.520824
presente	1.823178	0.663820

Cuadro 129: Tabla de contingencia entre CANTIDAD _HIJOS y VIC _ALFAB mostrando frecuencias absolutas y relativas

CANTIDAD _HIJOS VIC _ALFAB	demasiados hijos	hijo unico	hijos medios	ignorado	muchos hijos	sin hijos
alfabeta	6214	112499	165893	88967	46173	55107
analfabeta	8068	15500	31480	18891	24327	12670
ignorado	47	449	521	3310	232	324

Cuadro 130: Tabla de contingencia entre VIC_ES_INDIGENA y VIC_ALFAB mostrando frecuencias absolutas y relativas

VIC_ES_INDIGENA VIC_ALFAB	ignorado	no	si
alfabeta	55627	298747	120479
analfabeta	13663	35432	61841
ignorado	1296	2557	1030

Cuadro 131: Tabla de contingencia entre VIC_NIV_ESCOLARIDAD y VIC_ALFAB mostrando frecuencias absolutas y relativas

VIC_NIV_ESCOLARIDAD VIC_ALFAB	basicos	diversificado	ignorado	ninguno	primaria	universidad
alfabeta	86669	107162	12064	14183	234912	19863
analfabeta	0	0	0	110936	0	0
ignorado	0	0	4883	0	0	0

Cuadro 132: Tabla de contingencia entre VIC_TRABAJA y VIC_ALFAB mostrando frecuencias absolutas y relativas

VIC_TRABAJA VIC_ALFAB	ignorado	no	si
alfabeta	2127	286074	186652
analfabeta	378	85687	24871
ignorado	1467	2201	1215

Cuadro 133: Tabla de contingencia entre AGR_EST_CIV y VIC_EST_CIV mostrando frecuencias absolutas y relativas

AGR_EST_CIV VIC_EST_CIV	casados(as)	ignorado	otro	solteros(as)	unidos(as)	viudos(as)
casados(as)	214769	5702	838	18962	8340	547
ignorado	5171	67033	386	2354	3589	138
otro	471	303	10067	554	629	30
solteros(as)	17902	4742	955	66612	10552	626
unidos(as)	7433	3810	556	7402	125249	271
viudos(as)	1488	329	114	1628	984	136

Cuadro 134: Tabla de contingencia entre AGR_TRABAJA y VIC_EST_CIV mostrando frecuencias absolutas y relativas

AGR_TRABAJA VIC_EST_CIV	ignorado	no	si
casados(as)	7852	40524	200782
ignorado	4762	13828	60081
otro	355	2113	9586
solteros(as)	3856	22299	75234
unidos(as)	3676	22049	118996
viudos(as)	181	1574	2924

Cuadro 135: Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y VIC_EST_CIV mostrando frecuencias absolutas y relativas

DIF_ESCOLARIDAD_AGR_VIC VIC_EST_CIV	-1	-2	-3	-4	0	1	2	3	4	ignorado
casados(as)	34487	10305	2174	137	127858	46120	11624	2131	104	14218
ignorado	12323	4006	812	55	37611	12077	3334	504	16	7933
otro	1951	654	135	9	5619	2215	582	110	12	767
solteros(as)	16189	5452	1393	105	48405	17847	4192	874	48	6884
unidos(as)	20543	5630	1089	44	74068	27862	6319	1158	46	7962
viudos(as)	363	120	24	5	1758	1407	482	200	13	307

Cuadro 136: Tabla de contingencia entre GANA_MAS_LA_VICTIMA y VIC_EST_CIV mostrando frecuencias absolutas y relativas

GANA_MAS_LA_VICTIMA VIC_EST_CIV	no	si
casados(as)	207783	41375
ignorado	65297	13374
otro	9463	2591
solteros(as)	86165	15224
unidos(as)	124494	20227
viudos(as)	3788	891

Cuadro 137: Tabla de contingencia entre VIC_REL_AGR y VIC_EST_CIV mostrando frecuencias absolutas y relativas

VIC_REL_AGR VIC_EST_CIV	conviviente	esposos(as)	ex-conyuges	hermanos(as)	hijastros(as)	hijos(as)	nietos(as)	otro pariente	padres/madres	suegros(as)
casados(as)	7444	185150	8687	9258	266	1834	140	20589	12018	3772
ignorado	32496	1036	36706	957	229	2151	155	3728	900	313
otro	204	55	9953	291	24	137	2	791	477	120
solteros(as)	27055	2274	20350	13459	1256	7795	424	19236	7685	1855
unidos(as)	110413	3602	13410	3555	222	1271	62	8697	2639	850
viudos(as)	134	24	86	325	22	55	8	1424	2136	465

Cuadro 138: Tabla de contingencia entre VIC_TRABAJA y VIC_EST_CIV mostrando frecuencias absolutas y relativas

VIC_TRABAJA VIC_EST_CIV	ignorado	no	si
casados(as)	1084	152370	95704
ignorado	1674	46136	30861
otro	54	6769	5231
solteros(as)	598	64268	36523
unidos(as)	538	101184	42999
viudos(as)	24	3235	1420

Cuadro 139: Tabla de contingencia entre CANTIDAD_HIJOS y VIC_ES_INDIGENA mostrando frecuencias absolutas y relativas

CANTIDAD_HIJOS VIC_ES_INDIGENA	demasiados hijos	hijo unico	hijos medios	ignorado	muchos hijos	sin hijos
ignorado	1560	14409	21436	17304	7705	8172
no	5476	74916	118021	67110	34084	37129
si	7293	39123	58437	26754	28943	22800

Cuadro 140: Tabla de contingencia entre VIC_NIV_ESCOLARIDAD y VIC_ES_INDIGENA mostrando frecuencias absolutas y relativas

VIC_NIV_ESCOLARIDAD VIC_ES_INDIGENA	basicos	diversificado	ignorado	ninguno	primaria	universidad
ignorado	9049	9953	4893	16596	28493	1602
no	59252	79535	8743	41366	131621	16219
si	18368	17674	3311	67157	74798	2042

Cuadro 141: Tabla de contingencia entre VIC_TRABAJA y VIC_ES_INDIGENA mostrando frecuencias absolutas y relativas

VIC_TRABAJA VIC_ES_INDIGENA	ignorado	no	si
ignorado	891	49225	20470
no	2419	194060	140257
si	662	130677	52011

Cuadro 142: Tabla de contingencia entre CANTIDAD_HIJOS y VIC_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas

CANTIDAD_HIJOS VIC_NIV_ESCOLARIDAD	demasiados hijos	hijo unico	hijos medios	ignorado	muchos hijos	sin hijos
basicos	318	23258	30788	15202	5798	11305
diversificado	236	31183	39705	19017	5052	11969
ignorado	128	2587	3163	8668	856	1545
ninguno	8554	17950	35481	22491	26352	14291
primaria	5058	47940	81151	42127	31861	26775
universidad	35	5530	7606	3663	813	2216

Cuadro 143: Tabla de contingencia entre VIC_TRABAJA y VIC_NIV_ESCOLARIDAD mostrando frecuencias absolutas y relativas

VIC_TRABAJA VIC_NIV_ESCOLARIDAD	ignorado	no	si
basicos	272	54812	31585
diversificado	463	42937	63762
ignorado	1847	9843	5257
ninguno	508	95938	28673
primaria	791	165597	68524
universidad	91	4835	14937

Cuadro 144: Tabla de contingencia entre AGR_EST_CIV y VIC_REL_-AGR mostrando frecuencias absolutas y relativas

AGR_EST_CIV VIC_REL_AGR	casados(as)	ignorado	otro	solteros(as)	unidos(as)	viudos(as)
conviviente	9195	33627	386	24883	109545	110
esposos(as)	186672	964	42	1539	2914	10
ex-conyuges	10095	36497	9924	18132	14454	90
hermanos(as)	8276	1837	438	12746	4399	149
hijastros(as)	565	461	44	275	651	23
hijos(as)	5615	984	330	3415	2416	483
nietos(as)	329	60	26	201	96	79
otro pariente	18442	4312	1304	20572	9192	643
padres/madres	5212	1607	353	15106	3472	105
suegros(as)	2833	1570	69	643	2204	56

Cuadro 145: Tabla de contingencia entre AGR_TRABAJA y VIC_REL_-AGR mostrando frecuencias absolutas y relativas

AGR_TRABAJA VIC_REL_AGR	ignorado	no	si
conviviente	5757	21406	150583
esposos(as)	5268	22572	164301
ex-conyuges	3753	13322	72117
hermanos(as)	1278	8158	18409
hijastros(as)	95	507	1417
hijos(as)	477	4820	7946
nietos(as)	29	440	322
otro pariente	2676	20117	31672
padres/madres	1020	8182	16653
suegros(as)	329	2863	4183

Cuadro 146: Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y VIC_REL_AGR mostrando frecuencias absolutas y relativas

DIF_ESCOLARIDAD_AGR_VIC VIC_REL_AGR	-1	-2	-3	-4	0	1	2	3	4	ignorado
conviviente	24163	6381	1172	48	92853	33602	7341	1229	38	10919
esposos(as)	26115	7334	1306	76	102902	35023	8070	1292	50	9973
ex-conyuges	14945	4893	942	68	43060	14831	3907	584	33	5929
hermanos(as)	4475	1483	338	26	14533	3909	923	164	8	1986
hijastros(as)	494	188	43	13	737	218	69	14	0	243
hijos(as)	3463	1585	457	29	4500	1101	318	54	3	1733
nietos(as)	245	94	42	3	228	27	12	1	0	139
otro pariente	9800	3741	1210	87	23061	8449	2506	598	48	4965
padres/madres	1554	306	60	3	10480	8305	2700	806	40	1601
suegros(as)	602	162	57	2	2965	2063	687	235	19	583

Cuadro 147: Tabla de contingencia entre GANA_MAS_LA_VICTIMA y VIC_REL_AGR mostrando frecuencias absolutas y relativas

GANA_MAS_LA_VICTIMA VIC_REL_AGR	no	si
conviviente	155038	22708
esposos(as)	162570	29571
ex-conyuges	71650	17542
hermanos(as)	22615	5230
hijastros(as)	1764	255
hijos(as)	11401	1842
nietos(as)	709	82
otro pariente	44688	9777
padres/madres	20490	5365
suegros(as)	6065	1310

Cuadro 148: Tabla de contingencia entre VIC_TRABAJA y VIC_REL_AGR mostrando frecuencias absolutas y relativas

VIC_TRABAJA VIC_REL_AGR	ignorado	no	si
conviviente	859	126411	50476
esposos(as)	779	120812	70550
ex-conyuges	444	46159	42589
hermanos(as)	185	16015	11645
hijastros(as)	61	1398	560
hijos(as)	878	8401	3964
nietos(as)	55	562	174
otro pariente	519	34490	19456
padres/madres	155	14722	10978
suegros(as)	37	4992	2346

Cuadro 149: Tabla de contingencia entre CANTIDAD_HIJOS y VIC_TRABAJA mostrando frecuencias absolutas y relativas

CANTIDAD_HIJOS VIC_TRABAJA	demasiados hijos	hijo unico	hijos medios	ignorado	muchos hijos	sin hijos
ignorado	33	432	561	2497	183	266
no	9872	83062	120268	70013	46046	44701
si	4424	44954	77065	38658	24503	23134

Cuadro 150: Tabla de contingencia entre DIF_ESCOLARIDAD_AGR_VIC y VIC_TRABAJA mostrando frecuencias absolutas y relativas

DIF_ESCOLARIDAD_AGR_VIC VIC_TRABAJA	-1	-2	-3	-4	0	1	2	3	4	ignorado
ignorado	261	85	32	2	975	281	79	14	4	2239
no	47210	11506	2203	105	192597	75248	18007	3490	162	23434
si	38385	14576	3392	248	101747	31999	8447	1473	73	12398

Cuadro 151: Tabla de contingencia entre VIOLENCIA _PATRIMONIAL y VIOLENCIA _FISICA mostrando frecuencias absolutas y relativas

VIOLENCIA _PATRIMONIAL VIOLENCIA _FISICA	no presente	presente
no presente	214488	35339
presente	298306	42539

Cuadro 152: Tabla de contingencia entre VIOLENCIA _PSICOLOGICA y VIOLENCIA _FISICA mostrando frecuencias absolutas y relativas

VIOLENCIA _PSICOLOGICA VIOLENCIA _FISICA	no presente	presente
no presente	8461	241366
presente	72788	268057

Cuadro 153: Tabla de contingencia entre VIOLENCIA _SEXUAL y VIOLENCIA _FISICA mostrando frecuencias absolutas y relativas

VIOLENCIA _SEXUAL VIOLENCIA _FISICA	no presente	presente
no presente	245195	4632
presente	330787	10058

Cuadro 154: Tabla de contingencia entre VIOLENCIA _PATRIMONIAL y VIOLENCIA _PSICOLOGICA mostrando frecuencias absolutas y relativas

VIOLENCIA _PATRIMONIAL VIOLENCIA _PSICOLOGICA	no presente	presente
no presente	72640	8609
presente	440154	69269

Cuadro 155: Tabla de contingencia entre VIOLENCIA_SEXUAL y VIOLENCIA_PSICOLOGICA mostrando frecuencias absolutas y relativas

VIOLENCIA_SEXUAL VIOLENCIA_PSICOLOGICA	no presente	presente
no presente	78236	3013
presente	497746	11677

Cuadro 156: Tabla de contingencia entre VIOLENCIA_PATRIMONIAL y VIOLENCIA_SEXUAL mostrando frecuencias absolutas y relativas

VIOLENCIA_PATRIMONIAL VIOLENCIA_SEXUAL	no presente	presente
no presente	502025	73957
presente	10769	3921

11.5. Pruebas chi-cuadrada

Cuadro 157: Resultados de la prueba chi-cuadrado para mujeres

Antecedente	Consecuente	valor-p	chi2	dof	min esperado	filas	años
AGR_ALFAB	AGR_EST_CIV	0.000000	100.698723	4.000000	6.038494	274145	16
AGR_ALFAB	AGR_ES_INDIGENA	0.000000	721.938633	1.000000	665.988358	277587	16
AGR_ALFAB	AGR_GRUPET	0.000000	768.626606	5.000000	2.633154	295231	15
AGR_ALFAB	AGR_NIV_ESCOLARIDAD	0.000000	14309.924806	4.000000	74.301063	306953	16
AGR_ALFAB	AGR_OFICIO	0.000000	3043.706005	14.000000	2.246465	234650	13
AGR_ALFAB	AGR_SEXO	0.000000	177.737746	1.000000	129.414718	316288	16
AGR_ALFAB	AGR_TRABAJA	0.000000	62.651777	1.000000	311.060261	309632	16
AGR_ALFAB	CANTIDAD_HIJOS	0.000000	256.798565	4.000000	44.503610	265467	16
AGR_ALFAB	DIF_ALFAB	0.000000	7820.767935	2.000000	118.529923	315394	16
AGR_ALFAB	GANA_MAS_LA_VICTIMA	0.000000	61.559654	1.000000	285.692569	316288	16
AGR_ALFAB	OTROS_AGRESORES	0.007651	8.227449	1.000000	139.660948	100445	12
AGR_ALFAB	QUIEN_ES_MAYOR	0.000000	30.646057	2.000000	162.791630	301199	16
AGR_ALFAB	VIC_ALFAB	0.000000	3243.889731	1.000000	387.394359	315394	16
AGR_ALFAB	VIC_EST_CIV	0.000000	88.282900	4.000000	9.405628	274867	16
AGR_ALFAB	VIC_ES_INDIGENA	0.000000	767.557449	1.000000	650.888526	279251	16
AGR_ALFAB	VIC_GRUPET	0.000000	782.380118	5.000000	2.800017	295847	15
AGR_ALFAB	VIC_NACIONAL	0.014947	5.934083	1.000000	17.662513	313827	16
AGR_ALFAB	VIC_NIV_ESCOLARIDAD	0.000000	3102.702672	4.000000	84.389275	310235	16
AGR_ALFAB	VIC_OTRAS_VICTIMAS	0.003091	8.947869	1.000000	590.099616	198982	16
AGR_ALFAB	VIC_RANGO_EDAD	0.000000	98.229250	1.000000	455.936471	316288	16
AGR_ALFAB	VIC_REL_AGR	0.000000	359.613999	9.000000	1.450217	237262	12
AGR_ALFAB	VIC_TRABAJA	0.000000	293.627286	1.000000	739.061480	315326	16
AGR_ALFAB	VIOLENCIA_FISICA	0.000061	16.429340	1.000000	959.494069	316288	16
AGR_ALFAB	VIOLENCIA_PSICOLOGICA	0.000000	40.190149	1.000000	303.809140	316288	16
AGR_ESCOLARIDAD	AGR_ES_INDIGENA	0.000000	1233.726736	23.000000	2.266082	220742	13
AGR_ESCOLARIDAD	VIC_ES_INDIGENA	0.000000	1278.320014	23.000000	2.251339	221828	13
AGR_ESCOLARIDAD	VIC_TRABAJA	0.000000	1173.938774	23.000000	2.452868	200553	11
AGR_ESCOLARIDAD	VIOLENCIA_FISICA	0.000000	158.767936	23.000000	2.684650	225760	12
AGR_EST_CIV	AGR_ES_INDIGENA	0.000000	95.541065	4.000000	16.458575	244290	16
AGR_EST_CIV	AGR_NIV_ESCOLARIDAD	0.000000	308.985145	16.000000	1.399973	226666	13
AGR_EST_CIV	AGR_SEXO	0.000000	1095.653189	4.000000	2.504531	278646	16
AGR_EST_CIV	AGR_TRABAJA	0.000000	785.569418	4.000000	6.149923	271813	16
AGR_EST_CIV	DIF_ALFAB	0.000000	108.095330	8.000000	2.443175	273448	16
AGR_EST_CIV	GANA_MAS_LA_VICTIMA	0.000000	74.395170	4.000000	5.747648	278646	16
AGR_EST_CIV	OTROS_AGRESORES	0.000000	91.496479	4.000000	2.392851	84897	12
AGR_EST_CIV	QUIEN_ES_MAYOR	0.000000	937.230914	8.000000	3.413188	263701	16
AGR_EST_CIV	VIC_ALFAB	0.000000	123.913131	4.000000	8.449081	277519	16
AGR_EST_CIV	VIC_ES_INDIGENA	0.000000	93.806485	4.000000	15.295639	245937	16
AGR_EST_CIV	VIC_NIV_ESCOLARIDAD	0.000000	323.465056	16.000000	1.539766	245306	14
AGR_EST_CIV	VIC_OTRAS_VICTIMAS	0.000001	39.418624	4.000000	12.651967	179015	16
AGR_EST_CIV	VIC_RANGO_EDAD	0.000000	334.595098	4.000000	9.588437	278646	16
AGR_EST_CIV	VIC_TRABAJA	0.000000	82.887803	4.000000	14.726615	277700	16
AGR_EST_CIV	VIOLENCIA_FISICA	0.000000	188.983730	4.000000	18.237174	278646	16
AGR_EST_CIV	VIOLENCIA_PATRIMONIAL	0.000190	23.884785	4.000000	5.831789	278646	16
AGR_EST_CIV	VIOLENCIA_PSICOLOGICA	0.000002	31.777883	4.000000	6.276952	278646	16
AGR_ES_INDIGENA	GANA_MAS_LA_VICTIMA	0.000000	128.018725	1.000000	785.490669	282259	16
AGR_GRUPET	AGR_ES_INDIGENA	0.000000	17702.500000	4.000000	7.511082	282259	16
AGR_GRUPET	AGR_SEXO	0.016864	14.219016	5.000000	1.520553	239315	12
AGR_GRUPET	AGR_TRABAJA	0.000000	58.768694	5.000000	2.843717	307967	16
AGR_GRUPET	DIF_ALFAB	0.000000	610.993922	10.000000	1.157055	237501	12
AGR_GRUPET	GANA_MAS_LA_VICTIMA	0.000000	174.139372	5.000000	2.914220	317321	16
AGR_GRUPET	QUIEN_ES_MAYOR	0.005529	24.994627	10.000000	1.523129	298901	16
AGR_GRUPET	VIC_ALFAB	0.000000	1361.164304	5.000000	4.055996	315909	16
AGR_GRUPET	VIC_ES_INDIGENA	0.000000	13465.525240	5.000000	7.123523	280728	16
AGR_GRUPET	VIC_OTRAS_VICTIMAS	0.000000	85.757549	5.000000	5.611844	186982	15
AGR_GRUPET	VIC_TRABAJA	0.000000	607.334869	5.000000	6.785825	316279	16
AGR_GRUPET	VIOLENCIA_FISICA	0.000000	198.814908	5.000000	9.216375	317321	16
AGR_GRUPET	VIOLENCIA_PATRIMONIAL	0.000000	105.920475	5.000000	2.844158	317321	16
AGR_GRUPET	VIOLENCIA_PSICOLOGICA	0.000000	97.178188	5.000000	3.167170	317321	16
AGR_NACIONAL	AGR_ES_INDIGENA	0.000000	34.581494	1.000000	20.239976	273716	16
AGR_NACIONAL	AGR_NIV_ESCOLARIDAD	0.000330	20.908777	4.000000	2.281372	281347	15
AGR_NACIONAL	VIC_ALFAB	0.023863	6.062259	1.000000	16.699453	310089	16
AGR_NACIONAL	VIC_ES_INDIGENA	0.018451	5.992730	1.000000	24.198371	274594	16
AGR_NACIONAL	VIC_NIV_ESCOLARIDAD	0.000025	27.410306	4.000000	2.716262	304095	16
AGR_NACIONAL	VIC_TRABAJA	0.020526	5.371984	1.000000	27.320185	310514	16
AGR_NIV_ESCOLARIDAD	AGR_ES_INDIGENA	0.000000	1094.639299	4.000000	182.908294	270512	16
AGR_NIV_ESCOLARIDAD	DIF_ALFAB	0.000000	6793.386760	8.000000	28.896881	306246	16
AGR_NIV_ESCOLARIDAD	GANA_MAS_LA_VICTIMA	0.000000	324.338998	4.000000	74.616615	306953	16
AGR_NIV_ESCOLARIDAD	OTROS_AGRESORES	0.003598	15.870324	4.000000	25.838369	98459	12
AGR_NIV_ESCOLARIDAD	QUIEN_ES_MAYOR	0.000000	56.989775	8.000000	39.859339	294103	16
AGR_NIV_ESCOLARIDAD	VIC_ES_INDIGENA	0.000000	1145.345756	4.000000	183.332701	272114	16
AGR_NIV_ESCOLARIDAD	VIC_OTRAS_VICTIMAS	0.000905	19.264407	4.000000	132.002068	194809	16
AGR_NIV_ESCOLARIDAD	VIOLENCIA_FISICA	0.000000	103.868487	4.000000	222.794949	306953	16
AGR_NIV_ESCOLARIDAD	VIOLENCIA_PATRIMONIAL	0.000113	23.901806	4.000000	74.307022	306953	16
AGR_NIV_ESCOLARIDAD	VIOLENCIA_PSICOLOGICA	0.000000	120.554249	4.000000	75.781006	306953	16

Cuadro 160: Resultados del entrenamiento y evaluación de modelos para la relación entre agresor y víctima

			Entrenamiento						Evaluación					
Cantidad de Clases	Sem.Bla	Inicialización	Log-Likelihood	Entropía	NEC	AIC	BIC	Probabilidades Posteriores	Log-Likelihood	Entropía	NEC	AIC	BIC	Probabilidades Posteriores
2	42	huang	-1045453.229794	0.8109720	0.000000	2008926.459588	2091031.581934	0.939354	-448464.262305	0.819513	0.000000	89049.524610	897045.094002	0.939347
3	42	huang	-1045352.374381	0.8109720	0.000000	2008935.474992	2091051.232501	0.939342	-448330.267817	0.819512	0.000000	890691.615533	896836.469721	0.939335
4	42	huang	-404555.443106	0.865880	0.000000	1884056.636212	1881966.070904	0.971217	-102223.636085	0.866254	0.000000	896457.232170	886886.410054	0.971787
5	42	huang	-404249.194497	0.880886	0.000000	1884548.288994	1880810.994850	0.967978	-403369.806506	0.881353	0.000000	886389.613103	886631.366493	0.967473
6	42	huang	-2020695.315580	0.126443	0.000000	1841450.631150	1841763.758198	0.949845	-204616.744659	0.125882	0.000000	789292.469319	789583.157486	0.950022
7	42	huang	-202030.801721	0.8808167	0.000000	1841431.211541	1844489.863653	0.967067	-205127.636900	0.881041	0.000000	790325.207380	790602.230253	0.967478
8	42	huang	-404179.465344	0.133115	0.000000	1839039.930600	1839435.910070	0.939343	-204522.166226	0.132725	0.000000	788184.972453	788771.250022	0.939694
9	42	huang	-19466.419435	0.145669	0.000000	1839018.288600	1839491.079207	0.939364	-204018.618878	0.144621	0.000000	788127.237556	788561.880021	0.939477
2	73	huang	-1045453.342933	0.8109723	0.000000	2008926.685965	2091031.728211	0.939354	-448464.257966	0.819515	0.000000	89049.715932	897045.285224	0.939347
3	73	huang	-1042929.270912	0.8109713	0.000000	2005886.559825	2086044.123344	0.939333	-447211.979443	0.819489	0.000000	894453.958885	894598.812074	0.939334
4	73	huang	-404317.571395	0.100139	0.000000	1892705.142790	1892005.257482	0.952035	-405330.559430	0.109845	0.000000	813301.318800	811491.457844	0.953632
5	73	huang	-403965.513231	0.882441	0.000000	1879877.036462	1880230.632327	0.967160	-403330.516459	0.882911	0.000000	886111.403817	886352.596708	0.966672
6	73	huang	-202678.822494	0.881591	0.000000	1841411.746450	1845702.272827	0.967322	-205421.235992	0.881982	0.000000	790992.471984	791192.180161	0.967306
7	73	huang	-202035.542335	0.877390	0.000000	1841414.108460	1844508.732881	0.968199	-205121.917997	0.877712	0.000000	790313.835994	790651.828867	0.967002
8	73	huang	-202149.681005	0.877827	0.000000	184296.1363211	1843381.532305	0.968628	-204835.183480	0.878380	0.000000	789700.366990	790136.644549	0.968237
9	73	huang	-404179.530615	0.137062	0.000000	1837623.239229	1838216.908787	0.939345	-203724.414468	0.136465	0.000000	787559.482336	787842.620011	0.939230
2	42	random	-1045453.229794	0.8109720	0.000000	2008926.459588	2091031.581934	0.939354	-448464.262305	0.819513	0.000000	89049.524610	897045.094002	0.939347
3	42	random	-4061839.815398	0.853956	0.000000	1923709.830795	1923867.194314	0.980269	-412260.238675	0.853753	0.000000	824550.477350	824605.231438	0.980265
5	42	random	-404149.782375	0.101627	0.000000	1892439.564550	1892612.170436	0.950067	-405515.467299	0.101459	0.000000	811081.734507	811323.158078	0.950054
6	42	random	-202414.834326	0.877917	0.000000	1844889.849101	1843201.936000	0.967870	-205273.810615	0.878184	0.000000	790606.811229	790995.719406	0.967636
7	42	random	-403965.514298	0.880439	0.000000	1879241.285455	1879607.936467	0.967362	-402889.587556	0.880939	0.000000	886531.917512	886489.100851	0.967631
8	42	random	-409974.238725	0.129706	0.000000	1840028.477451	1840448.640835	0.953251	-404307.848057	0.129136	0.000000	788695.770809	789082.040578	0.953513
9	42	random	-202028.597730	0.873340	0.000000	1841334.7195450	1841819.860817	0.970393	-204488.626476	0.873422	0.000000	789067.252952	789501.135217	0.970844
2	73	random	-1045453.342933	0.8109723	0.000000	2008926.685965	2091031.728211	0.939354	-448464.257966	0.819515	0.000000	89049.715932	897045.285224	0.939347
3	73	random	-4061839.816093	0.853960	0.000000	1923709.862196	1923967.1305705	0.980265	-412260.232185	0.853743	0.000000	824551.644770	824605.898400	0.980263
4	73	random	-4061539.354771	0.853893	0.000000	1923118.700341	1923328.784324	0.980264	-412122.816797	0.853688	0.000000	824245.637505	824478.576370	0.980266
5	73	random	-402770.1164775	0.880434	0.000000	1845452.239551	1845714.935436	0.967752	-405417.478807	0.880840	0.000000	790884.957014	791126.818035	0.967449
6	73	random	-403941.860331	0.892144	0.000000	1870943.721161	188025.848809	0.967256	-403024.686985	0.892577	0.000000	806109.357971	806399.461147	0.966771
7	73	random	-403571.402671	0.881671	0.000000	1879212.810341	1879509.453553	0.967243	-402834.498513	0.882174	0.000000	805738.597026	806076.989899	0.966833
8	73	random	-40351.1483905	0.111224	0.000000	183910.9467911	183922.3157305	0.939136	-404179.629513	0.110956	0.000000	789229.924104	789601.251075	0.939182
9	73	random	-109593.937982	0.143012	0.000000	183996.787364	184044.0566321	0.942153	-404127.255872	0.142705	0.000000	788584.451754	789193.140009	0.942235
2	42	cas	-1045452.963065	0.8109689	0.000000	2008925.926129	2091030.968476	0.939350	-448464.266784	0.819480	0.000000	89049.673567	897045.422600	0.939353
3	42	cas	-1042929.271492	0.8109714	0.000000	2005886.642993	2086044.116503	0.939333	-447211.970271	0.819480	0.000000	894453.940043	894598.704631	0.939334
4	42	cas	-4061539.342462	0.853960	0.000000	1923118.674124	1923328.588918	0.980269	-412122.798400	0.853684	0.000000	824245.596790	824478.555584	0.980271
5	42	cas	-403985.406568	0.881752	0.000000	1878020.993130	1880083.590001	0.967425	-403020.640495	0.882266	0.000000	806102.689910	806345.134773	0.966932
6	42	cas	-202107.7636222	0.857299	0.000000	1878015.272645	1878330.399683	0.973916	-202253.967722	0.857037	0.000000	804566.703544	804855.610721	0.973900
7	42	cas	-202150.7456549	0.890453	0.000000	184326.4113899	1843631.761310	0.980114	-204024.008799	0.891136	0.000000	789918.617598	790256.610171	0.980494
8	42	cas	-2021410.943683	0.881246	0.000000	184290.1897325	1843322.956730	0.969522	-204832.594138	0.881326	0.000000	789745.185276	790131.465945	0.969295
9	42	cas	-202002.2357222	0.127046	0.000000	184009.4471645	184058.7102902	0.940937	-204177.679706	0.126450	0.000000	789725.852332	789160.989187	0.940446
2	73	cas	-1045452.963065	0.8109689	0.000000	2008925.926129	2091030.968476	0.939350	-448464.266784	0.819480	0.000000	89049.673567	897045.422600	0.939353
3	73	cas	-1042929.250676	0.8109710	0.000000	2005886.581353	2086044.064872	0.939334	-447211.983786	0.819487	0.000000	894453.967573	894598.826661	0.939335
4	73	cas	-4061526.889991	0.854492	0.000000	192300.2170992	1923302.246475	0.980104	-412136.282470	0.854211	0.000000	824312.564939	824505.703724	0.980107
5	73	cas	-4040016.863029	0.873292	0.000000	188008.2170408	188033.4735923	0.980730	-403370.216242	0.873614	0.000000	806190.412484	806431.830903	0.980305
6	73	cas	-404172.4375244	0.889259	0.000000	188550.6700489	188582.8175206	0.964545	-404087.142346	0.889662	0.000000	808534.294493	810123.929648	0.964136
7	73	cas	-402175.4352419	0.867115	0.000000	184305.7917028	184394.7353440	0.972028	-404948.247916	0.867049	0.000000	789966.415832	790304.408704	0.972233
8	73	cas	-202095.936656	0.128600	0.000000	1841996.793711	1842416.943006	0.948031	-204625.199056	0.129122	0.000000	789330.398112	789716.676881	0.947529
9	73	cas	-401970.666930	0.135206	0.000000	1839303.733859	1839976.424417	0.947015	-204137.107796	0.134727	0.000000	789364.221191	789708.783856	0.947540

*Resultados del análisis de clases latentes con diferentes configuraciones. Se presentan los promedios de las métricas clave como el likelihood logarítmico, entropía promedio, NEC, AIC, BIC y probabilidades máximas posteriores tanto para los conjuntos de entrenamiento que fueron obtenidos utilizando cross validation como de evaluación. Estos resultados permiten comparar la calidad del ajuste del modelo bajo distintas condiciones y por ende elegir el mejor modelo. Basado en todas estas métricas obtenidas se eligió el modelo con inicialización huang, 4 clases latentes y semilla de 42, ya que en general presentaba unas métricas bastante balanceadas.

11.7. Código fuente para todo el trabajo

El código fuente donde se realizó todo el trabajo se encuentra disponible en el siguiente repositorio de GitHub <https://github.com/lgarciap/ViolenciaIntrafamiliar>