

**Universidad del Valle de Guatemala
Facultad de Business and Management School**



**PROPUESTA DE MODELO DETERMINÍSTICO PARA LA
SEGMENTACIÓN DE CLIENTES Y PERFILADO UTILIZANDO
ANÁLISIS DE DATOS Y MERCADEO DIGITAL**

Trabajo de graduación en modalidad de tesis presentado por

Raúl Christian Sánchez Ippisch

para optar por el grado académico de

Licenciado en International Marketing and Business Analytics

Guatemala,
2022

**Universidad del Valle de Guatemala
Facultad de Business and Management School**



**PROPUESTA DE MODELO DETERMINÍSTICO PARA LA
SEGMENTACIÓN DE CLIENTES Y PERFILADO UTILIZANDO
ANÁLISIS DE DATOS Y MERCADEO DIGITAL**

Trabajo de graduación en modalidad de tesis presentado por

Raúl Christian Sánchez Ippisch

para optar por el grado académico de

Licenciado en International Marketing and Business Analytics

Guatemala,
2022

HOJA DE APROBACIÓN

Vo.Bo. Asesor



Ing. José Vicente Herrera Juárez

Vo.Bo. Terna Examinadora



Ing. José Vicente Herrera Juárez
(ASESOR)



MSc. José Antonio Medrano



MSc. Carmen María López de Burmester

Fecha de Aprobación: Guatemala, 13 de Julio del 2022

PREFACIO

La elaboración de la presente tesis surge de un interés personal para crear relaciones entre las redes sociales y el análisis de datos. Durante el periodo de mi licenciatura, mostré un interés en el tema de análisis de datos y redes sociales para determinar nuevas maneras de crear buyer personas a través de un método nuevo el cual se puede implementar de manera fácil en cualquier empresa.

Agradezco al Ing. Vicente Herrera por las asesorías en la creación de la presente tesis y en la confianza concedida para ser parte de la presente investigación

Quisiera también agradecer a todas aquellas personas que me ayudaron a realizar mi carrera, familia y catedráticos. Sin su apoyo no hubiera sido posible crear esta investigación.

ÍNDICE

Hoja de aprobación	II
Prefacio	III
Listado de cuadros	V
Listado de figuras.....	V
Resumen.....	VIII
I. Introducción	1
II. Objetivos	2
III. Justificación.....	3
IV. Marco teórico.....	4
A. Modelo determinístico.....	4
B. Posibles modelos por utilizar para el análisis.....	5
C. Teoría de comportamiento del consumidor.....	7
D. Neuromarketing.....	7
E. Neuromarketing y redes sociales.....	10
F. Marketing digital	11
G. Buyer persona.....	11
H. Segmentación	12
I. Perfilado	17
J. Análisis de datos.....	17
K. Estadística.....	17
L. R Studio.....	17
V. Metodología	18
VI. Resultados.....	20
VII. Análisis de resultados.....	52
VIII. Conclusiones.....	59
IX. Recomendaciones.....	60
X. Bibliografía.....	61
XI. Anexo	63
A. Diagrama de flujo metodología.....	63
B. Análisis exploratorio	66
C. Buyer Persona Template.....	78
D. Código realizado en R studio	80
E. Entregable a cliente	93

LISTADO DE CUADROS

Cuadro 1. Ejemplo de cuadro de segmentación de mercado para un perfilado más exacto	16
Cuadro 2. Proyección financiera utilizando tasa de crecimiento anual de 5% para determinar la viabilidad de la implementación del modelo determinístico para su aplicación en empresas financieras.	49
Cuadro 3. Proyección financiera utilizando tasa de crecimiento anual de 10% para determinar la viabilidad de la implementación del modelo determinístico para su aplicación en empresas financieras.	50
Cuadro 4. Proyección financiera utilizando tasa de crecimiento anual de 15% para determinar la viabilidad de la implementación del modelo determinístico para su aplicación en empresas financieras.	51

LISTADO DE FIGURAS

Figura 1. Modelo de sistema determinístico y estocástico.....	4
Figura 2: Árbol de decisión.....	5
Figura 3. Plantilla utilizada para la determinación de Buyer Persona	12
Figura 4. Modelo de segmentación de árbol de decisión determinado con la variable “Propósito” para el perfilado de cliente.	20
Figura 5. Primer modelo de segmentación mediante clusterización de las variables numéricas para la comparación de variables en la base de datos.	21
Figura 6. Boxplot de la distribución entre los clústeres y la edad obtenido por el análisis de segmentación para el perfilado de clientes potenciales.....	22
Figura 7. Boxplot de la distribución entre los clústeres y el total de crédito adquirido obtenido por el análisis de segmentación para el perfilado de clientes potenciales.	23
Figura 8. Boxplot de la distribución entre los clústeres y la duración del crédito adquirido obtenido por el análisis de segmentación para el perfilado de clientes potenciales.	24
Figura 9. Scatterplot para la determinación de correlación entre las variables duración y edad.	25
Figura 10. Scatterplot para determinar la correlación entre las variables total de crédito adquirido y edad.	26
Figura 11. Scatterplot para determinar la correlación entre las variables total de crédito adquirido y edad.	27
Figura 12. Modelo de segmentación de árbol de decisión para el perfilado de cliente.	28

Figura 13. Segundo modelo de segmentación mediante clusterización de las variables de caracteres para la comparación de variables en la base de datos.	29
Figura 14. Boxplot de la distribución entre los clústeres y la duración del crédito adquirido obtenido por el análisis de segmentación para el perfilado de clientes potenciales.....	30
Figura 15. Boxplot de la distribución entre los clústeres y el total de crédito adquirido obtenido por el análisis de segmentación para el perfilado de clientes potenciales.....	31
Figura 16. Boxplot de la distribución entre los clústeres y la duración del crédito adquirido obtenido por el análisis de segmentación para el perfilado de clientes potenciales.....	32
Figura 17. Scatterplot para la determinación de correlación entre las variables duración y edad.	33
Figura 18. Scatterplot para determinar la correlación entre las variables total de crédito adquirido y edad.	34
Figura 19. Scatterplot para determinar la correlación entre las variables total de crédito adquirido y edad.	35
Figura 20. Modelo de segmentación mediante clusterización para la comparación de variables en la base de datos.....	36
Figura 21. Barplot mediante clusterización para analizar el género y la cantidad de préstamos adquiridos.	37
Figura 22. Barplot mediante clusterización para analizar el género y el estado laboral del cliente. ...	38
Figura 23. Modelo de segmentación mediante clusterización para la comparación de variables en la base de datos.....	39
Figura 24. Modelo de segmentación de árbol de decisión para el perfilado de cliente.	40
Figura 25. Buyer persona 1 basado en el árbol de decisión para la determinación del cliente potencial para la empresa.....	41
Figura 26. Buyer persona 2 basado en el árbol de decisión para la determinación del cliente potencial para la empresa.....	43
Figura 27. Buyer persona 2 basado en el árbol de decisión para la determinación del cliente potencial para la empresa.....	46
Figura 28. Histograma de distribución de edades de la base de datos analizada.	66
Figura 29. Histograma de la frecuencia del monto del crédito de la base de datos analizada.	66
Figura 31. Histograma de frecuencia de duración de crédito de la base de datos analizada.....	67
Figura 32. Relación entre el total de crédito y la duración de este según la base de datos analizada..	68
Figura 33. Relación entre la edad y el monto del crédito según la base de datos analizada.	68
Figura 34. Relación entre la edad y la duración del crédito de la base de datos analizada.....	69
Figura 35. Gráfica de barras de frecuencia del propósito de la adquisición de la deuda con el monto total de esta.	69
Figura 35. Gráfica de barras de frecuencia del propósito de la adquisición del crédito con el monto promedio de esta.....	70

Figura 36. Boxplot de la distribución entre edad y género.	71
Figura 37. Boxplot de la distribución entre el género y el monto del crédito.	71
Figura 38. Boxplot de la distribución entre género y duración del crédito.	72
Figura 39. Relación entre el monto de crédito y la duración de este según el género.	73
Figura 40. Relación entre la edad y el monto del crédito según el género.	74
Figura 41. Relación entre la edad y la duración del crédito según el género.	75
Figura 42. Gráfico de barras de la relación entre el propósito del crédito y el monto de este según el género.	76
Figura 43. Gráfico de barras de la frecuencia del propósito del crédito y el monto de este según el género.	76
Figura 44. Plantilla de Buyer Persona utilizada para los Buyer persona basados en los resultados obtenidos del análisis.	78
Figura 45. Formato de Excel entregable a cliente para recolección de leads.	93
Figura 46. Formato de Excel entregable a cliente dashboard	93

RESUMEN

El objetivo principal de la presente investigación fue proponer un modelo determinístico para la segmentación y perfilado de clientes utilizando análisis de datos y mercadeo digital empleando una base de datos pública. Para lo cual se utilizó la base de datos “German_credit_data.csv” con la finalidad de crear un proceso de perfilado para la optimización del proceso de adquisición de clientes mediante redes sociales. Dicha base de datos fue obtenida de la página Kaggle donde es posible la descarga de base de datos públicas. Se realizó una exploración de datos en R studio en la cual se obtuvieron la cantidad de filas y los diferentes tipos de variables dentro de la base de datos. Empleando estrategias de neuromarketing, perfilado y análisis de datos, se propuso el modelo determinístico presentado donde se determinó que las variables más significativas de la base de datos pública analizada son el género del cliente, la duración, el propósito y el monto total del crédito, obtenidos a partir del árbol de decisiones y el modelo de segmentación por clusterización. A partir del árbol de decisiones y el modelo de clusterización se determinó que los clientes potenciales prefieren adquirir un crédito para la compra de un auto dependiendo de la edad, la cuenta bancaria y si es dueño de un hogar o alquila vivienda. A partir de los datos obtenidos, se crearon buyer persona para un perfilado más preciso. Además, para determinar la viabilidad de la implementación del modelo determinístico para su aplicación en empresas financieras, se realizó una evaluación financiera con una proyección de ventas. Con este análisis se determinó que dicho modelo es viable para la implementación de un perfilado más exacto mediante modelos de árbol de decisión y de clusterización, teniendo un costo de \$7,500 mensuales, utilizando \$5,000 de pauta digital con un 3% de porcentaje de adquisición de clientes para la obtención de 1,000,000 de impresiones diarias. De esta manera se recomienda que para la aplicación del modelo determinístico propuesto en la presente investigación se utilice una base de datos adaptada para la segmentación de clientes de una empresa.

I. INTRODUCCIÓN

Las redes sociales se han vuelto un factor importante a nivel internacional en múltiples empresas por el hecho los recientes acontecimientos de pandemia por COVID-19. Se ha demostrado que no solamente es un portal en donde se pueden encontrar fotos, videos, posts o eventos de lo que está sucediendo mundialmente. Son plataformas con múltiples herramientas las cuales sirven para poder comercializar tus productos o servicios.

La pandemia logró que se aumentará un 27% el uso de las redes sociales a nivel mundial. Existen más de 4.600 millones de usuarios en las redes sociales. Usuarios que pueden ser capaces de conocer las diferentes empresas que existen a nivel internacional o nacional. Uno de los principales problemas a nivel internacional es el factor humano, en donde especialistas de marketing utilizan técnicas tradicionales para crear y segmentar perfiles los cuales pueden llegar a ser de utilidad, pero utilizando estas técnicas tradicionales, estudios de mercado, buyer personas, técnicas de prueba y error para encontrar si la publicidad es aceptada no es lo óptimo dentro estas plataformas ya que no se aprovecha en su totalidad la herramienta. (Juste, M.)

Las plataformas sociales proveen una cantidad de información, las cuales pueden ayudar a las empresas a generar diferentes técnicas de business intelligence con marketing digital; para poder encontrar la técnica que más se adapte a su cliente objetivo. A diferencia del mercadeo tradicional, el marketing digital provee bases de datos con variables de suma importancia para poder determinar (con diferentes métodos estadísticos y herramientas de análisis de datos) un modelo determinístico para encontrar el cliente objetivo de una empresa.

Instagram, Facebook, Google, Twitter, y Tik Tok han demostrado ser plataformas poderosas no solo para creadores de contenido, sino también para emprendedores y además empresarios. Plataformas en las cuales se pueden dar a conocer ya sea de manera orgánica o en publicidad pagada. Estas plataformas ofrecen a los empresarios recolectar su propia data para sus empresas. Ofreciendo una cantidad de herramientas para que puedan analizar la data estratégicamente con el fin de perfilar y segmentar clientes de una manera más eficiente y exacta para lograr minimizar costos y optimizar sus costos de publicidad.

El COVID-19 ha demostrado la importancia de las redes sociales, ha demostrado que todas las personas (sin importar el estrato social) tienen acceso a las redes sociales o acceso a internet. Ofreciéndole a las empresas una manera más efectiva y controlada de generar una segmentación más exacta para poder llegarle a todas aquellas personas las cuales puedan demostrar interés en los diferentes productos o servicios que las empresas ofrecen.

II. OBJETIVOS

A. General

Proponer un modelo determinístico para la segmentación y perfilado de clientes utilizando análisis de datos y mercadeo digital empleando una base de datos pública.

B. Específicos

1. Determinar y evaluar las variables más significativas de la base de datos de referencia utilizando procesos estadísticos para generar un modelo estadístico en R Studio.
2. Evaluar las variables de las bases de datos para determinar los tipos de clientes que la empresa puede adquirir.
3. Determinar la importancia de las variables edad, genero, trabajo, alquiler/dueño de vivienda, propósito de crédito y duración para el modelo determinístico utilizando el modelo de árbol de decisión.
4. Realizar una evaluación financiera con una proyección de ventas para determinar la viabilidad de la implementación del modelo determinístico para su aplicación en empresas financieras.

III. JUSTIFICACIÓN

La problemática generada por el COVID-19 ha sido un tema el cual ha afectado a miles de emprendimientos, negocios y empleados a nivel internacional. Empresarios guatemaltecos tuvieron que cerrar sus negocios debido al virus. Uno de los principales problemas es el pensamiento de los adultos mayores, quienes consideran que las redes sociales no tienen ningún impacto en sus negocios. Sin embargo, en la pandemia muchos de estos empresarios se dieron cuenta de la importancia de las redes sociales, no solamente en Guatemala sino también a nivel internacional. Los empresarios subestimaron las redes sociales y el análisis de datos, muchos no lo vieron como una oportunidad de crecimiento durante la pandemia.

Existen empresas que aprovecharon las herramientas de análisis de datos y pudieron crecer, recolectaron datos y se dieron cuenta de la gran capacidad que existe dentro de las redes sociales, no solamente para compartir el día a día de los usuarios, sino también como una oportunidad de crecimiento. Las redes sociales ofrecen diferentes herramientas para poder comercializar los diferentes productos o servicios de una empresa. Siendo una herramienta de recolección de datos para poder generar diferentes modelos determinísticos para segmentar y perfilar de una manera más eficaz a los posibles clientes de una empresa.

Es importante dar a conocer el poder que tienen las plataformas para recolectar datos importantes. Con estas bases de datos se pueden generar modelos determinísticos utilizando estadística descriptiva e inferencial para crear modelos matemáticos con programas (R Studio y Tableau) para un análisis extensivo de diferentes variables con las que se determinará diversos perfiles para empresas. Con el fin de ayudar a diferentes empresas a reducir costos, optimizando el presupuesto de las empresas para poder invertir de mejor manera dicho dinero.

De esta manera surge la necesidad de integrar business intelligence y marketing digital para hacer más eficiente la segmentación con modelos matemáticos aplicados en ambas áreas. Comprobando el efecto de estas plataformas para las empresas guatemaltecas con el fin de levantar las ventas y reducir costos de publicidad. Por lo que, en el presente trabajo se determinará un modelo determinístico dinámico en el cual se utilizan datos históricos para crear modelos de predicción para segmentaciones futuras.

IV. MARCO TEÓRICO

A. Modelo determinístico

Un modelo determinista es un modelo matemático el cual define que las mismas entradas o condiciones iniciales producirán invariablemente las mismas salidas o resultados. En donde no se contempla la existencia de azar o la existencia de incertidumbre en el proceso de modelado del modelo.

Este modelo se relaciona con la creación de entornos a través de la creación de simuladores por el estudio inicial de las diferentes situaciones hipotéticas para crear el sistema de gestión el cual ayudará a erradicar la propagación de errores en el modelo. Estos modelos solamente pueden ser aplicados para sistemas deterministas no caóticos. Estos modelos no pueden predecir la mayor parte de las características definidas. Entre más variables tenga el análisis de elementos los cuales se puedan utilizar para que el modelo se aproxime a un modelo probabilístico con un enfoque estocástico.

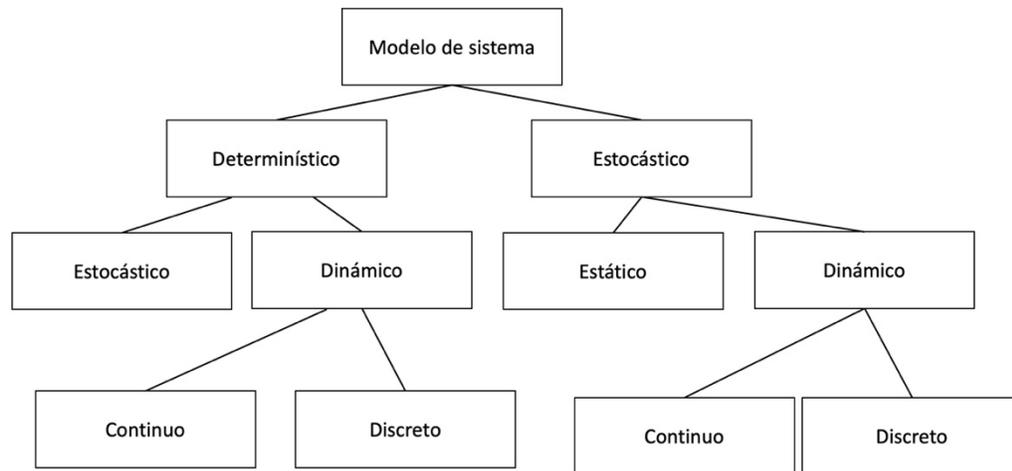


Figura 1. Modelo de sistema determinístico y estocástico

(Neetesuela, 2013)

B. Posibles modelos por utilizar para el análisis

1. Modelo árbol de decisión

El modelo o el algoritmo de clúster jerárquico agrupa todos los datos basados en la distancia que tiene cada uno y unificando todos los datos que son parecidos entre sí. Se pueden ver con una representación gráfica los diferentes elementos y cómo se relacionan entre sí.

Ejemplo:

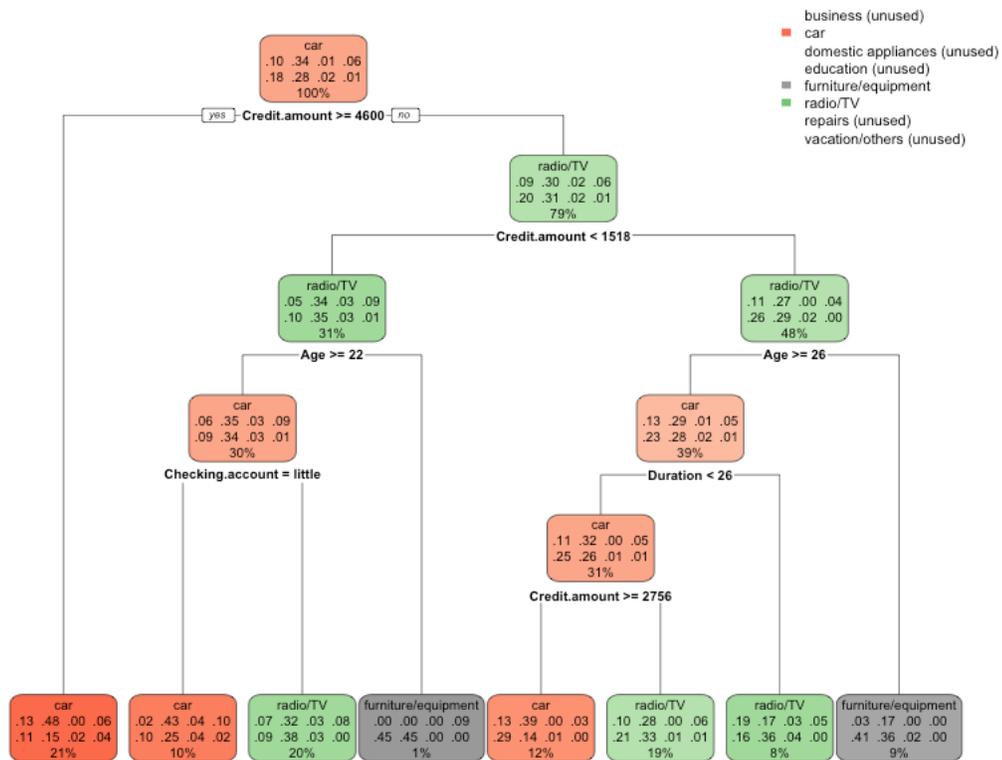


Figura 2: Árbol de decisión

2. Análisis estadístico

Con la idea de recopilar, interpretar y validar los datos. Con el análisis estadístico se pueden realizar varios análisis y operaciones estadísticas para cuantificar los datos y poder trasladarlos a un análisis más completo. Con la idea de poder correlacionar los datos de una manera más numérica.

La estadística es una de las herramientas fundamentales de la investigación científica, su uso es esencial especialmente cuando se trabajan problemas de investigación en los que se miden variables cuantitativas, y su importancia es mayor en los diseños experimentales, en los cuales se debe de controlar la validez de los resultados experimentales. La estadística permite conducir la experiencia de investigación desde el planteamiento del problema hasta la comprobación de hipótesis, y la elaboración de conclusiones. Por esta razón, para realizar una investigación se deben de tener claras las aplicaciones de las diversas técnicas y medidas estadísticas en función de las necesidades específicas del trabajo de investigación que se está realizando (Lizarzabury, 2013).

3. Medidas de tendencia central

Con el objetivo de resumir los datos obtenidos en un valor representativo adecuado se utilizan las siguientes medidas (Johnson & Kuby, 2012).

4. Análisis cualitativo

Utilizar este análisis para determinar por qué las personas adquieren el crédito con una financiera independiente. Con la idea de determinar a qué tipo de personas se les puede mandar la publicidad. También se busca encontrar la relación según las diferentes características de las personas contra el objetivo del estudio a analizar. Con el fin de identificar relaciones utilizando tablas de contingencia y también chi cuadrado. (Rodrigo, Joaquín Amat).

5. Análisis predictivo

Utilizando datos históricos de una financiera independiente para determinar patrones, con el fin de analizar las diferentes tendencias encontradas dentro del análisis. Este modelo se utiliza con el fin de encontrar variables que puedan ayudar a predecir el futuro de una empresa. Creando estrategias conforme el análisis realizado. Con la finalidad de mejorar el perfilado a realizar dentro del presente análisis. Creando buyer personas más precisos con datos reales.

C. Teoría de comportamiento del consumidor

La teoría del consumidor es definida por todos aquellos individuos que consumen algún bien o servicio con el objetivo de satisfacer sus necesidades. Ya sean desde gustos propios o preferencias del consumidor. Estas preferencias son las que guían la elección entre diversas combinaciones de bienes que existen dentro de la economía.

El consumidor genera una cesta de consumo en donde van incluidas listas de bienes y servicios que pueden ser del consumidor en algún determinado momento del tiempo, y en ciertas circunstancias en algún momento definido. Existen solamente dos tipos de cesta de consumo las cuales se definen por (x_1, x_2) y (y_1, y_2) donde x e y toman valores no negativos.

Con la finalidad de demostrar las decisiones del consumidor, se definieron dos axiomas los cuales funcionan para determinar las decisiones que puede tomar cada consumidor los cuales fueron desarrollados por economistas

1. Completitud: El consumidor se expresa sin indiferencia o preferencia entre cualquier cesta de bienes por muy similares o diferentes que sean.
2. Reflexividad: Cualquier cesta es igual de buena que la otra. Cualquier cesta sin importar si es preferida o indiferente entre sí misma asegura que cada cesta de bienes pertenece al conjunto formado por la cesta.
3. Transitividad: Si $(x_1, x_2) \succeq (y_1, y_2)$ y $(y_1, y_2) \succeq (z_1, z_2)$ se supone que $(x_1, x_2) \succeq (z_1, z_2)$. Lo cual se define que el consumidor piensa que la cesta X es igual de buena que la cesta Y y que la cesta Y es igual de buena como la Z, lo cual hace que la cesta X sea al menos igual de buena que la cesta Z.

D. Neuromarketing

1. Ciencia que estudia las emociones y el comportamiento de compra

Durante muchos años se ha estudiado el cerebro humano, donde se han generado varios desafíos, disciplinas que han abordado diferentes temas como, la psicología, psiquiatría, psicoanálisis y filosofía. A pesar de la tecnología y los grandes avances en el estudio del cerebro, aún tenemos varias preguntas las cuales no se han logrado responder a lo largo de los años. Preguntas como; ¿dónde reside nuestra conciencia en nuestro cerebro?, ¿qué significa la felicidad?, ¿cómo hace el ser humano para recordar tanta información?, pero la más importante para este estudio es, ¿por qué los consumidores prefieren cierta marca, servicio o

producto? El cerebro humano es un reto que evoluciona día a día con el mundo cambiante. (Carasila, & Milton, 2010)

Existen muchas preguntas dentro del neuromarketing como, ¿existe la manera de determinar las necesidades del consumidor, porque compra cierto producto y cómo podemos determinar una estrategia de mercadeo en base al análisis generado? Existe una manera, en donde nace el neuromarketing en donde se estudia el comportamiento del consumidor para entender con diferentes tecnologías se estudia el comportamiento del consumidor para llegar a entender porque realizan ciertas compras, o porque prefieren una marca, producto o servicio. (Carasila, & Milton, 2010)

Blue Brain es uno de los proyectos más grandes e importantes en el mundo. Con ayuda de IBM se creó un procesador el cual es capaz de simular 100 mil millones de neuronas. Con el fin de entender los comportamientos del consumidor con tecnología artificial. Esto se utilizó con el fin de entender de manera más amplia el comportamiento del consumidor y el proceso de compra de cada persona. Además, se utilizó para entender las emociones de compra de los consumidores. Desde donde mira la persona a la hora de querer comprar un producto hasta el interés puesto en ello. (Carasila, & Milton, 2010)

El Sistema nervioso central está constituido por el encéfalo y la médula espinal los cuales están protegidos por una membrana la cual se le llama duramadre, aracnoides y las meninges. El encéfalo está protegido por envolturas óseas, que son el cráneo y la columna vertebral. El cerebro es la parte que más grande del encéfalo. El cerebro se divide en dos hemisferios. El cerebro sistema nervioso central contiene una sustancia blanca y sustancia gris. El cual se divide en diferentes partes, frontal, parietal temporal y occipital. Esta parte del cerebro se encarga de transmitir información sensorial donde el componente básico es la neurona sensorial, la cual se conforma por un núcleo, cuerpo y una dendrita y el axón el cual está encargada de enviar la información en el cerebro. El mecanismo de este es una sinapsis. Existen más de 100 mil millones de neuronas las cuales están denominadas por compuestos de moléculas químicas que ayudan a transmitir fácilmente la información por el cerebro. Esta sinapsis genera imágenes las cuales se definen como experiencia en donde se concluye que la memoria y el conocimiento no se archivan directamente en el cerebro. Estas experiencias son generadas por las neuronas en su sinapsis. (Carasila, & Milton, 2010)

Se llega a la conclusión que la neurociencia estudia el sistema nervioso central, la cual estudia a detalle la estructura, patología, farmacología y las funciones químicas del sistema nervioso y de cómo diferentes elementos del sistema nervioso estudiando las maneras en las cuales el sistema nervioso da origen a la conducta de las personas. En resumen, la neurociencia estudia el cerebro a detalle con el sistema nervioso central, el cual da origen a la conducta de las personas. El estudio proviene no solamente para entender la conducta sino también para entender como interactúan las neuronas a nivel molecular. (Carasila, & Milton, 2010)

El comportamiento de compra del consumidor se comprende como los pasos que toman y que se desarrollan para llegar a la decisión final de compra. Existen varias propuestas de este proceso, pero se ha identificado un proceso en común a lo largo de varios estudios. El cual consiste en el despertar, identificar o reconocer las necesidades de recopilación y tratamiento de la información; formulación y la elección. (Carasila, & Milton, 2010)

Estos estudios han demostrado que dentro de la elección de compra del consumidor siempre está influido por algún factor interno o externo. El cual concluye en el modelo básico de decisión de compra del consumidor que incluye un comportamiento después de la compra o decisión tomada por el consumidor. La mayoría de las decisiones son tomadas por el factor económico del consumidor. El consumidor es descrito como una persona pasiva el cual se define por campañas promocionales de las empresas lo cual los hacen impulsivos e irracionales a la hora de comprar un producto. Desde un punto de vista emocional el consumidor está denominado por la impulsividad a la hora de realizar compras. La mayoría de las compras son realizadas por una base emocional los cuales compran dependiendo de su estado de ánimo o los sentimientos que tienen durante ese momento en específico. El cual no es definido como un comportamiento irracional. (Carasila, & Milton, 2010)

Durante los años se han definido diferentes modelos que tratan de clasificar a los consumidores de una manera lógica para lograr entender las diferentes emociones que afectan el factor de compra a la hora de comprar un producto. La neurociencia aplicada en marketing aún sigue siendo un estudio en proceso para entender más a detalle a los consumidores. (Carasila, & Milton, 2010)

Las emociones son un factor clave para el marketing, esto se debe a que el proceso de compra se basa mucho en los sentimientos que siente el consumidor en el momento. Las emociones tienen un alto impacto en el marketing. Es considerado de suma importancia para entender las influencias internas y externas del consumidor. (Carasila, & Milton, 2010)

Las emociones pueden ser entendidas como una medida de objetivos que se cumplen por una interacción entre la persona y el medio ambiente. Las emociones son evolutivas que son cambiantes a lo largo de la vida de una persona, el cual genera cambios emocionales dependiendo de las experiencias emocionales que viva una persona a lo largo de su vida. Estas emociones tienen un impacto ambiental las cuales son estimuladas por la información recibida en los órganos sensoriales. La transmisión de información empieza los órganos sensoriales por la vía lemniscal alcanza la corteza sensorial primaria la cual llega a la amígdala la cual emite un juicio sobre el estímulo que siente la persona en el momento. La cual se denomina tálamo-córtico-amígdalar. Donde se determina que, dependiendo de la emoción del consumidor, será su comportamiento de

compra a la hora de adquirir algún producto o servicio con alguna determinada marca. (Carasila, & Milton, 2010)

E. Neuromarketing y redes sociales

La neurociencia ayuda a entender los diferentes campos científicos de la neurociencia enlazados con el marketing. Esto se da para conocer los diferentes efectos que tiene la publicidad en los consumidores. El neuromarketing se ha vuelto un tema de mucha importancia en el marketing ya que ayuda a comprender cómo reacciona el cerebro humano contra la publicidad generada por empresas. (Abuin, Dias, & García, 2020)

Especialistas de marketing utilizan mucho esta herramienta ya que ayuda a entender que es lo que buscan los consumidores. Con la información recolectada en los estudios de neuromarketing se ha logrado identificar nuevas maneras de cómo influir en el comportamiento de compra de los consumidores. Ayudando a las empresas a enfocarse en diferentes factores cerebrales que ayudan a los consumidores a tomar sus decisiones de compra a la hora de buscar un producto. (Abuin, Dias, & García, 2020)

El neuromarketing es también utilizado en las redes sociales, para entender qué tipo de contenido lanzar, el lenguaje a utilizar y las emociones que genera en línea. Con el fin de poder perfilar y analizar las diferentes emociones generadas para lograr analizar y mejorar la segmentación buscada por las empresas. (Abuin, Dias, & García, 2020)

Existen diferentes estudios los cuales fueron de utilidad para determinar la respuesta cerebral de los 35 usuarios contra los anuncios de televisión del Super Bowl. Lograron medir el nivel de satisfacción que generaban estos anuncios. Logrando identificar los diferentes gustos que existen entre personas a la hora de ver un anuncio. (Abuin, Dias, & García, 2020)

Cvikikj Michahelles analizó en el 2003 cómo el contenido y sus diferentes características mediante Facebook afectan a sus usuarios. El estudio se enfocó en el contenido, el día y la hora de la publicación, cuántos me gusta tenía, comentarios y la cantidad de veces que fue compartido el anuncio. Entre sus resultados encontraron que los temas de entretenimiento son los que más influyen en esta red social, ya que son las publicaciones que más interacción tienen los usuarios. (Abuin, Dias, & García, 2020)

Hay varias tendencias temáticas que se destacan en los análisis para determinar qué es lo que el consumidor busca en línea. Las posibles reacciones emocionales que demuestran los usuarios se han utilizado para identificar los posibles elementos en el anuncio que influyen en los usuarios. Con esto identificado se

logra crear un análisis para separar los posibles clientes que más se adapten al anuncio, logrando que las empresas logren segmentar de manera más específica a sus leads calificados. (Abuin, Dias, & García, 2020)

F. Marketing digital

El marketing digital es una herramienta necesaria para las empresas ya que ayuda a implementar las nuevas herramientas tecnológicas en una empresa. Con diferentes técnicas se han logrado identificar oportunidades de negocio en los mercados globales. Esto ayuda a que las empresas desarrollen diferentes estrategias de comunicación integradas con el marketing para poder lograr una segmentación del grupo objetivo de la empresa. (Borcio, Calle, & Zambrano, 2018)

Existen varias herramientas con las cuales se puede lograr el objetivo de marketing digital para las empresas. Empezando con Facebook, este ofrece herramientas de publicidad las cuales ayudan a pautar de una manera más específica en las redes sociales para llegarle al grupo objetivo de una empresa. Otras herramientas como Google ayudan a sincronizar a los usuarios y a las empresas de manera más rápida ya que la mayoría de las personas tienen un teléfono inteligente el cual permite que puedan observar los diferentes productos o servicios que ofrece una empresa. (Borcio, Calle, & Zambrano, 2018)

La analítica web es un área muy importante del mercadeo digital, ya que es un área en donde se logra recolectar datos de relevancia para las empresas. Logrando generar bases de datos en donde se almacenan para luego ser utilizados en análisis que ayudarán a determinar el grupo objetivo de cada empresa de una manera más específica. Utilizando Facebook ads manager, se puede extraer toda la data recolectada de diferentes anuncios para determinar si fueron atractivos para el consumidor. (Borcio, Calle, & Zambrano, 2018)

Utilizando Google Analytics se puede determinar el nivel de importancia que ha tenido una página web con sus diferentes campañas de tráfico (Facebook ads o Google ads) para un consumidor. Con esto se pueden generar cambios efectivos. Con estas herramientas se pueden medir las edades de los consumidores, el CPC (Costo por clic), el porcentaje de retención, las interacciones que ha logrado una página web o anuncio. Con esto se puede determinar una segmentación más exacta en el área de mercadeo, generando un costo beneficio del presupuesto para las campañas digitales. (Borcio, Calle, & Zambrano, 2018)

G. Buyer persona

El buyer persona es una manera de identificar al cliente ideal de una empresa para el uso de un producto o un servicio. Utilizando un personaje ficticio para determinar a nuestro cliente ideal. Se utilizan

datos como, su edad, sexo, profesión, ingresos, relaciones, entre otros factores de relevancia para identificar a los consumidores de un producto o servicio.

Ejemplo:

La imagen muestra una plantilla de Buyer Persona con el título '# CLIENTE IDEAL / BUYER PERSONA' en un encabezado verde. El formulario está dividido en varias secciones con iconos y el texto 'Escribe aquí':

- ANTECEDENTES** (icono de flechas): Escribe aquí.
- DEMOGRÁFICA** (icono de personas): Escribe aquí.
- OBJETIVOS** (icono de triángulo): Escribe aquí.
- NOMBRE AQUÍ** (icono de persona): Escribe aquí. Incluye una foto de un hombre sonriente.
- PASATIEMPOS E INTERESES** (icono de libro): Escribe aquí.
- DESAFÍOS** (icono de bombilla): Escribe aquí.
- OBJECIONES COMUNES** (icono de triángulo invertido): Escribe aquí.
- GRANDES TEMORES** (icono de bombilla): Escribe aquí.

En la parte inferior del formulario, se indica: 'Personaliza esta plantilla de lienzo de cliente ideal en el editor online IDXT.org' y el logo de 'startups'.

Figura 3. Plantilla utilizada para la determinación de Buyer Persona (Cliente ideal, 2018)

H. Segmentación

Una estrategia de segmentación dentro del marketing digital se logra a través de un análisis en donde se determina un grupo objetivo para una empresa. Esta segmentación se logra a través de un estudio en donde se determina cuáles son las características que relación al producto o servicio con el consumidor. Estas características varían desde el sexo, la edad, la orientación geográfica, el comportamiento de compra entre otras. (Bermúdez, 2018)

Dentro de las estrategias digitales se realizan diferentes análisis y estudios en donde se determina el público de una empresa. Con esto se encuentran las diferentes características y los diferentes intereses que nuestros clientes pueden llegar a tener en común con nuestra empresa. Normalmente una empresa ya tiene un grupo objetivo definido, pero en las redes sociales se pueden encontrar clientes potenciales para las empresas. (Bermúdez, 2018)

Existen muchos canales de comunicación digital es importante definir los diferentes canales para tener una mejor relación con nuestro público objetivo. Estos canales incluyen, redes sociales, email marketing, páginas web etc. Con estos canales se puede recolectar data la cual puede ser analizada después para definir de una manera más exacta nuestro grupo objetivo. (Bermúdez, 2018)

Utilizando publicidad online se puede lograr con un presupuesto conseguir visitas a nuestras plataformas. Utilizando técnicas de posicionamiento web SEO con el objetivo de poder alcanzar en los medios digitales a nuestro grupo objetivo. La publicidad digital nos ayuda a dar a conocer nuestros servicios y una audiencia más amplia, recolectando datos importantes para luego definir nuestro grupo objetivo en línea. Con esto logramos identificar diferentes características de nuestro público. (Bermúdez, 2018)

1. Segmentación del mercado

La segmentación del mercado es el procedimiento en el cual se divide un mercado en diferentes subconjuntos de consumidores que tienen necesidades o características comunes, y de seleccionar segmentos para llegar a ellos mediante una mezcla de marketing específica (Schiffman & Lazar, 2005).

La estrategia de segmentación permite que los productores eviten la competencia directa en el mercado debido a la diferenciación de sus ofertas, no solo en términos de precio, sino también en cuanto al estilo, el empaque, el atractivo promocional, el sistema de distribución y un mejor servicio. Con la aplicación de la segmentación se ha descubierto que los costos que implica la investigación, las corridas de producción más pequeñas y las campañas promocionales diferenciadas suelen compensarse con el incremento en las ventas. En la mayoría de los casos, los consumidores aceptan pagar más por productos que satisfagan más cercanamente sus necesidades específicas (Schiffman & Lazar, 2005).

La segmentación del mercado es solo el primer paso de una estrategia de marketing de tres fases. Después de segmentar el mercado en conglomerados homogéneos, se debe seleccionar uno o más segmentos para considerarlos como una meta. Para esto se debe de considerar una mezcla de marketing específica, es decir, una mezcla de producto, precio, canal o atractivo promocional específicos para cada segmento en particular. Por último, está el posicionamiento del producto, de manera que este se perciba por los consumidores de cada segmento como capaz de satisfacer sus necesidades mejor que la competencia (Schiffman & Lazar, 2005). En la actualidad, cada vez más negocios utilizan programas de base de datos para el marketing, con el fin de determinar quiénes son mejores clientes para, luego, dividir a todos sus clientes en segmentos (Schiffman & Lazar, 2005).

Los estudios de segmentación tienen la finalidad de descubrir las necesidades y los deseos de grupos de consumidores específicos, para desarrollar y promover bienes y servicios especializados que satisfagan las necesidades de cada grupo. Estos estudios también son usados como guías para el rediseño o reposicionamiento de un producto o para la adición de un nuevo segmento al mercado. Además, al llenar los huecos en las líneas de productos, los mercadólogos usan la investigación de segmentación para identificar cuáles son los canales más adecuados para la difusión de anuncios (Schiffman & Lazar, 2005).

2. Bases para la segmentación

El paso inicial en el desarrollo de una estrategia de segmentación consiste en seleccionar las bases más apropiadas sobre las cuales segmentar el mercado. Las bases más frecuentes para la segmentación son nueve categorías principales de características del consumidor. Entre ellas se encuentran los factores geográficos, demográficos, psicológicos, pictográficos, socioculturales, relacionados con el uso, de uso-situación, de beneficios deseados, así como formas híbridas de segmentación, como los perfiles demográficos-pictográficos, los factores geo-demográficos y los referentes a los valores y estilos de vida. Cada uno de los formatos híbridos emplea una combinación de bases diversas de segmentación para crear perfiles completos de segmentos particulares de los consumidores (Schiffman & Lazar, 2005).

- Segmentación geográfica

En esta segmentación el mercado se divide por localidades. La teoría indica que las personas que viven en una misma área comparten ciertas necesidades y deseos similares, y que esas necesidades son distintas a quienes viven en otras áreas. Es una estrategia útil ya que es relativamente fácil encontrar diferencias establecidas geográficamente para muchos productos. Además, es fácil alcanzar los segmentos geográficos utilizando los medios de comunicación locales (Schiffman & Lazar, 2005).

- Segmentación demográfica

Las características demográficas, como la edad, sexo, estado marital, ingresos, ocupación y educación, son las que emplean más a menudo como base para la segmentación. Esta estrategia ayuda a localizar un mercado meta. La información demográfica es la ruta más accesible y efectiva, en términos de costos, para identificar un mercado meta. Las variables demográficas son más fáciles de medir que otras variables de segmentación; están incluidas en los estudios pictográficos y socioculturales porque asignan mayor significado a los hallazgos. Estas variables revelan las tendencias continuas que apuntan hacia nuevas oportunidades de negocios, como los cambios en la distribución de edades, géneros e ingresos (Schiffman & Lazar, 2005).

- Segmentación psicológica

Estas características se refieren a las cualidades internas del consumidor individual. Las estrategias a menudo se basan en variables psicológicas específicas, como sus motivaciones, personalidad, percepciones, aprendizaje y actitudes (Schiffman & Lazar, 2005).

- Segmentación pictográfica

Este tipo de segmentación está estrechamente relacionada con la investigación psicológica, en especial con respecto a la medición de la personalidad y de las actitudes. Esta forma de investigación es un valioso instrumento de marketing ya que ayuda a identificar los segmentos de consumidores prometedores que quizá

respondan a mensajes de marketing específicos. El perfil pictográfico de un segmento de consumidores se visualiza como la combinación de las mediciones de actividades, intereses y opiniones de los consumidores (Schiffman & Lazar, 2005).

- Segmentación sociocultural

Las variables socioculturales ofrecen bases adicionales, por ejemplo, al subdividir en segmentos los mercados de consumidores con base en la etapa dentro del ciclo de vida familiar, la clase social, los valores culturales y fundamentales, las afiliaciones y la membresía transcultural (Schiffman & Lazar, 2005).

- Segmentación relacionada con el uso

Esta forma de segmentación divide a los consumidores en categorías mediante características de uso del producto, el servicio o la marca, como el nivel de uso, el nivel de conciencia y el grado de lealtad hacia la marca. La segmentación basada en el índice de uso distingue entre usuarios intensos, medianos, ligeros, y los no usuarios de dicha marca, producto o servicio (Schiffman & Lazar, 2005).

- Segmentación por la situación de uso

La ocasión o la situación de uso con frecuencia determina lo que los consumidores quieren o no comprar o consumir, por lo que se centran en la situación de uso como variable de segmentación (Schiffman & Lazar, 2005).

- Segmentación por beneficios

La segmentación por ventajas percibidas sirve para posicionar varias marcas dentro de una misma categoría de productos. El cambio en los estilos de vida representa un rol fundamental en la determinación de los beneficios del producto que son importantes para los consumidores, ofreciendo así oportunidades para crear nuevos productos o servicios (Schiffman & Lazar, 2005).

- Enfoques híbridos de segmentación

En lugar de depender de una sola base de segmentación, los mercadólogos utilizan diversas variables de segmentación para un mercado. Los enfoques pueden incluir los perfiles pictográfico-demográfico, la geodemografía, el sistema Vals y la segmentación inteligente de Yankelovich (Schiffman & Lazar, 2005).

Cuadro 1. Ejemplo de cuadro de segmentación de mercado para un perfilado más exacto

Segmentaciones de mercado	
Geográfica	Empresas que se encuentren en el país e Guatemala, más que todo en la Ciudad de Guatemala, en las zonas 10,11,12,14,15,16 y en Carretera a el Salvador. Personas de diferentes comunidades con un nivel socio económico medio-alto, las cuales puedan adquirir el servicio de las redes sociales.
Demográfica	Hombres y Mujeres, casados, divorciados o comprometidos y solteros. De edades de 20-55 años (con emprendimientos o empresas) de nacionalidades múltiples (Cualquier nacionalidad en Guatemala), que tengan una cultura de querer crecer profesional y empresarialmente. Que pertenezca a cualquier religión (Ya que la religión no es un factor determinante) con cualquier tipo de profesión (siempre y cuando tenga un emprendimiento o una empresa) con educación universitaria o mayor a universitario. Con ingresos económicos empresariales de más de Q50,000 mensuales para poder financiar el servicio de mercadeo digital y publicidad.
Pictográfica	Personas con personalidad de crecimiento, que tengan ganas de crecer. Con carácter fuerte. Con un estilo de vida en el cual se encarga de hacer o manejar una empresa para incrementar las ventas y su presencia en línea. Padres de familia que se encargan de compartir con la familia en el fin de semana. También personas que les gusta tener una vida sana en donde se preocupan por su bienestar y salud. Les gusta ejercitarse y además leer libros para crecer. Personas que invierten en diferentes proyectos en su tiempo libre. Pertenecen a la clase social Media alta. Con una actitud de empoderamiento para dar un ejemplo a seguir a su familia y a sus trabajadores. Le interesa mucho leer sobre el mercado y además de estudiar para crecer la empresa.
Comportamientos por cliente	Estos empresarios buscan en línea como mejorar su servicio y como comercializarlo en redes. Actualmente se enfoca en realizar búsquedas en motores de búsqueda para encontrar un servicio el cual pueda ayudarlos a mejorar su empresa. También son personas que trabajan con recomendaciones, ya sean de grupos de Facebook o de otro lugar. Frecuentemente utilizan estos medios para encontrar lo que necesitan con el objetivo de reducir costos. Adquieren servicios en el momento adecuado para la empresa. Son empresas fieles a los servicios que contratan ya que confían que estas personas van a lograr a mejorar la calidad de la empresa. Gastan entre Q2,000 a Q10,000 dependiendo del servicio y la necesidad de la empresa. Utilizan celulares inteligentes, tabletas y computadoras para realizar su búsqueda. Están conectados en horas laborales ya que en casa pasan tiempo con su familia.

I. Perfilado

Al momento de comercializar un producto o servicio es muy importante que las empresas se preocupen por determinar el segmento de la población en el que se encontrarán sus clientes y al cual estará dirigido su discurso, así sus energías y capital serán invertidos de manera eficiente; a esto se le llama perfilado.

J. Análisis de datos

El análisis de datos es una ciencia que se encarga de estudiar y examinar un conjunto de datos con el propósito de sacar conclusiones sobre la información de manera que se puedan tomar decisiones a partir de ellas, o simplemente ampliar los conocimientos sobre dicho tema (Nettleton, 2003).

El análisis y entendimiento de los datos de negocio es una parte fundamental de toda empresa en cualquier sector. Ya que de ello depende el rumbo de la economía y las tendencias de consumo en la sociedad, así como la medición de la rentabilidad, costos y competitividad de las empresas comerciales en el mercado (Nettleton, 2003).

El análisis de datos empieza con la recolección de datos utilizando diferentes instrumentos cualitativos o cuantitativos analizando toda la información recolectada en los estudios. Con la información recolectada se pueden emplear diferentes métodos de análisis para tomar decisiones que pueden llegar a beneficiar a una empresa. (Saraduy, 2007)

En la mercadotecnia se puede utilizar el análisis de datos para predecir el comportamiento que tendrá el consumidor con nuestra campaña de marketing. Además, nos sirve para recolectar información importante sobre nuestro consumidor. (Saraduy, 2007)

K. Estadística

La estadística es una herramienta que nos ayuda a determinar si nuestro análisis resuelve nuestro problema a estudiar. Se utiliza especialmente para poder trabajar problemas de investigación midiendo variables cuantitativas creando modelos experimentales para luego después comprobar si nuestra hipótesis es válida. Para poder realizar esta experimentación se requieren de diferentes técnicas estadísticas las cuales serán empleadas en el programa de R facilitando el estudio de las variables y de la base de datos. (Anderson, 2015)

L. R Studio

R studio es un programa en el cual se pueden realizar diversos análisis para determinar ciertos factores interesantes dentro de una base de datos.

V. METODOLOGÍA

Etapas 1: Determinación y evaluación de variables

- Utilizar base de datos histórica que provee el cliente.
- Utilizar base de datos de referencia para crear los modelos.
- Cargar la base de datos a R Studio.
- Utilizar códigos `str()`, `summary()` y `glimpse()` para revisar la estructura de la base de datos.

Etapas 2: Exploración y limpieza de datos

- Explorar los datos para determinar si existen variables las cuales necesitan ser eliminadas utilizando el código `glimpse()` para revisar que variables dentro del `data.frame` no son necesarios a utilizar.
- Utilizando R, realizar una limpieza de datos para determinar las variables de más importancia para el análisis. Utilizando `as.factor()` convertir las variables en factores.
- Identificar errores dentro de la base de datos para remover datos insignificantes para el análisis. Removiendo los valores NULL de la base datos.
- Crear diferentes tipos de gráficos (como: Boxplot, Histogramas, gráficos de barra, scatterplot) para determinar la correlación entre variables.

Etapas 3: Construcción de modelo de segmentación

- Crear modelos de clusterización mediante variables numéricas utilizando las librerías de `ggplot2`, `tidyR`, `Cluster rpart`, `rpart.plot` y `tidyverse`.
- Crear gráficos de las variables para determinar la relación entre sí
- Crear modelos de clusterización para la comparación de variables cualitativas utilizando las librerías de `ggplot2`, `tidyR`, `Cluster rpart`, `rpart.plot` y `tidyverse`.

Etapas 4: Construir modelo de árbol de decisión

- Crear modelo de decisión con base en un árbol de decisiones utilizando las librerías de `ggplot2`, `tidyR`, `Cluster rpart`, `rpart.plot` y `tidyverse`.

Etapas 5: Creación de buyer persona para perfilado de clientes

- Utilizar una plantilla para poder determinar perfiles de segmentación.
- Crear buyer persona mediante clusterización (Cuantitativa o cualitativa)
- Revisar en que los buyer persona estén acorde al árbol de decisión y a los clústeres.

Etapas 6: Creación de campañas

- Ingresar a Facebook ads o Google ads mediante un usuario.
- Crear campañas de segmentación por medio de Google ads y Facebook ads.
- Segmentar acorde a la información obtenida por medio de los buyer persona.

Etapas 7: Archivo de clientes potenciales

- Crear archivo de Excel de los clientes potenciales obtenidos por las campañas publicitarias.
- Exportarlo y enviarlo al cliente.

Etapas 8: Evaluación financiera

- Determinar la inversión total del negocio
- Calcular todos los costos fijos y gastos variables de la empresa
- Realizar una proyección financiera para determinar si es factible implementar el modelo determinístico utilizando evaluación de proyectos.
- Calcular VPN y TIRR del flujo para determinar si el negocio es rentable

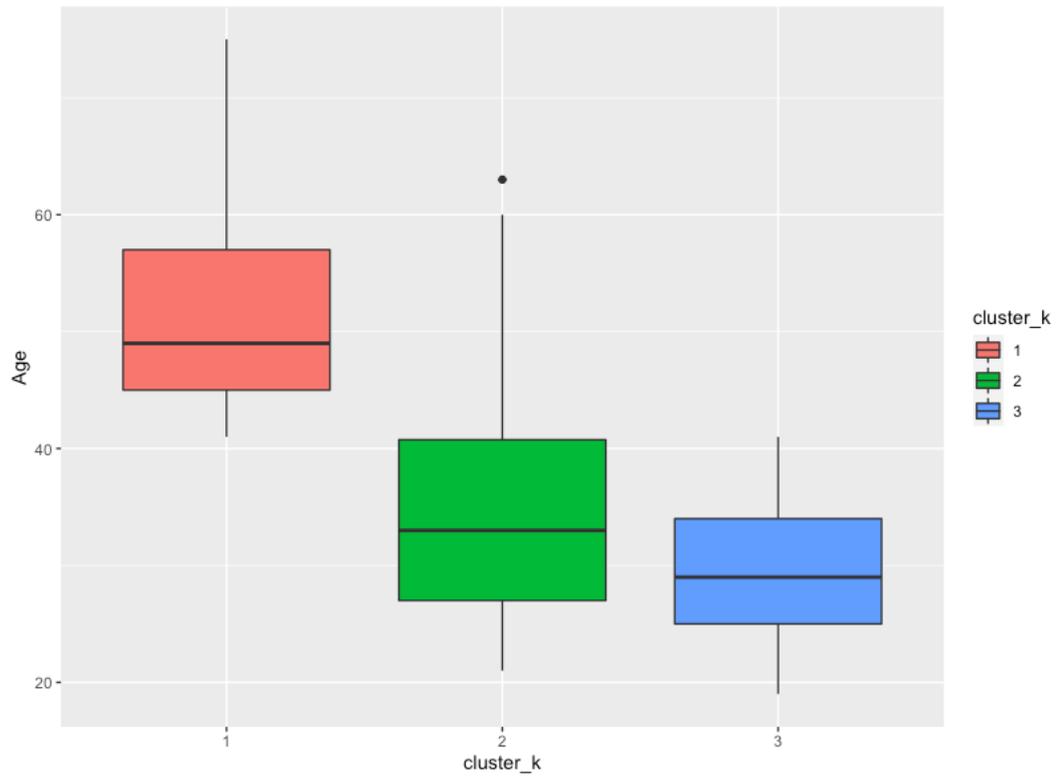
VI. RESULTADOS

Figura 4. Modelo de segmentación de árbol de decisión determinado con la variable “Propósito” para el perfilado de cliente.



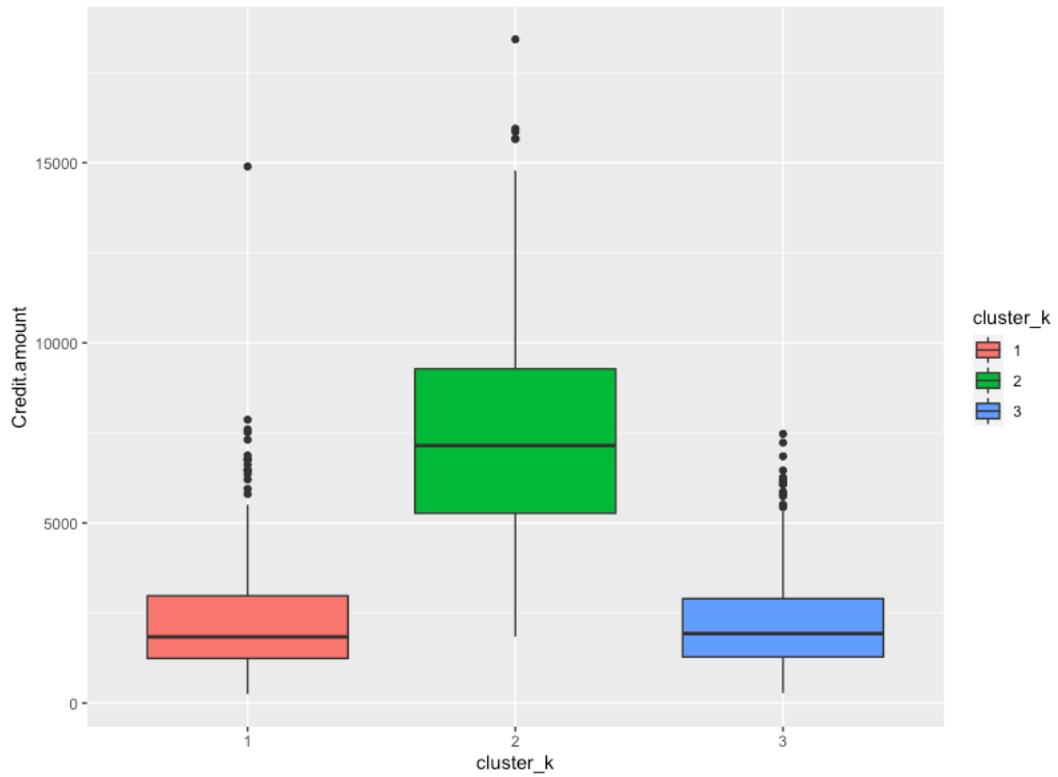
Nota: El árbol de decisión se obtuvo utilizando R studio con la base de datos “German_credit_data.csv”

Figura 6. Boxplot de la distribución entre los clústeres y la edad obtenido por el análisis de segmentación para el perfilado de clientes potenciales.



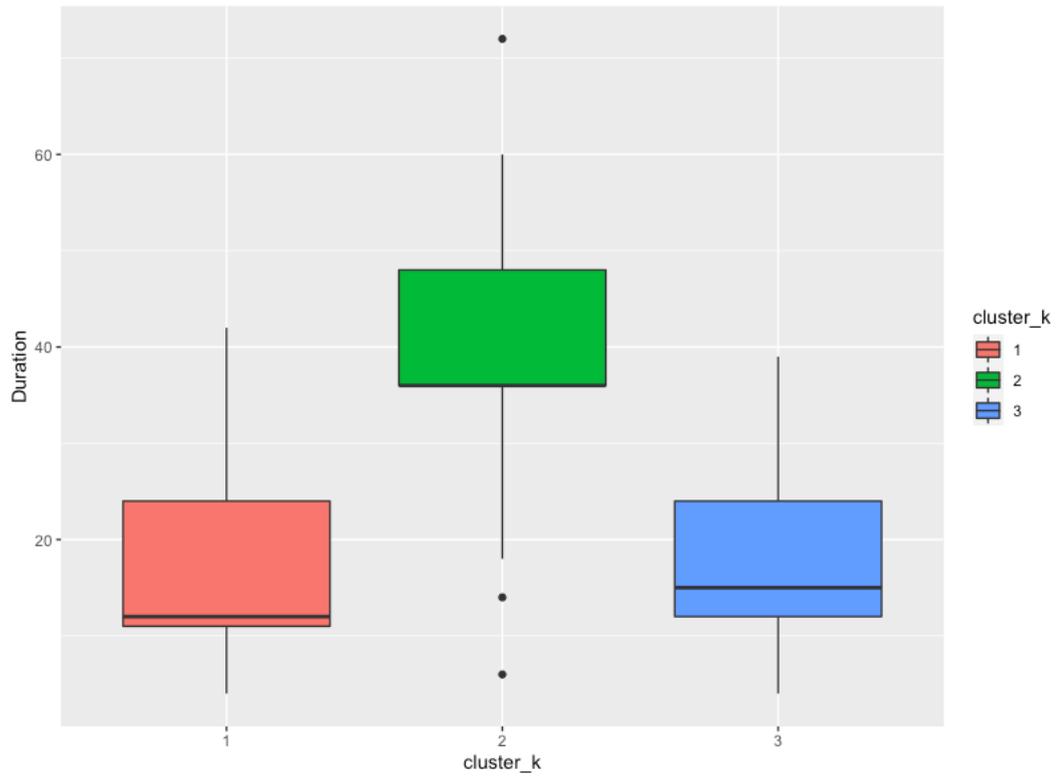
Nota: El Boxplot presentado se obtuvo utilizando R Studio con la base de datos “German_credit_data.csv” a partir del primer modelo de segmentación mediante clusterización.

Figura 7. Boxplot de la distribución entre los clústeres y el total de crédito adquirido obtenido por el análisis de segmentación para el perfilado de clientes potenciales.



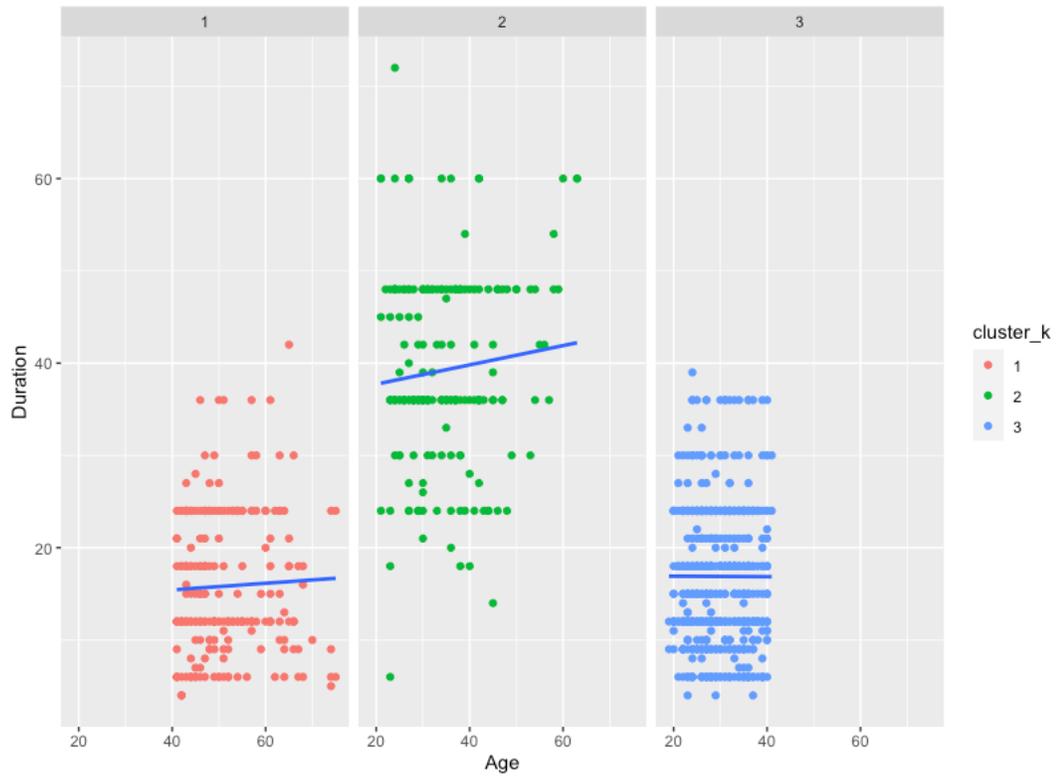
Nota: El Boxplot presentado se obtuvo utilizando R Studio con la base de datos “German_credit_data.csv” del primer modelo de segmentación mediante clusterización.

Figura 8. Boxplot de la distribución entre los clústeres y la duración del crédito adquirido obtenido por el análisis de segmentación para el perfilado de clientes potenciales.



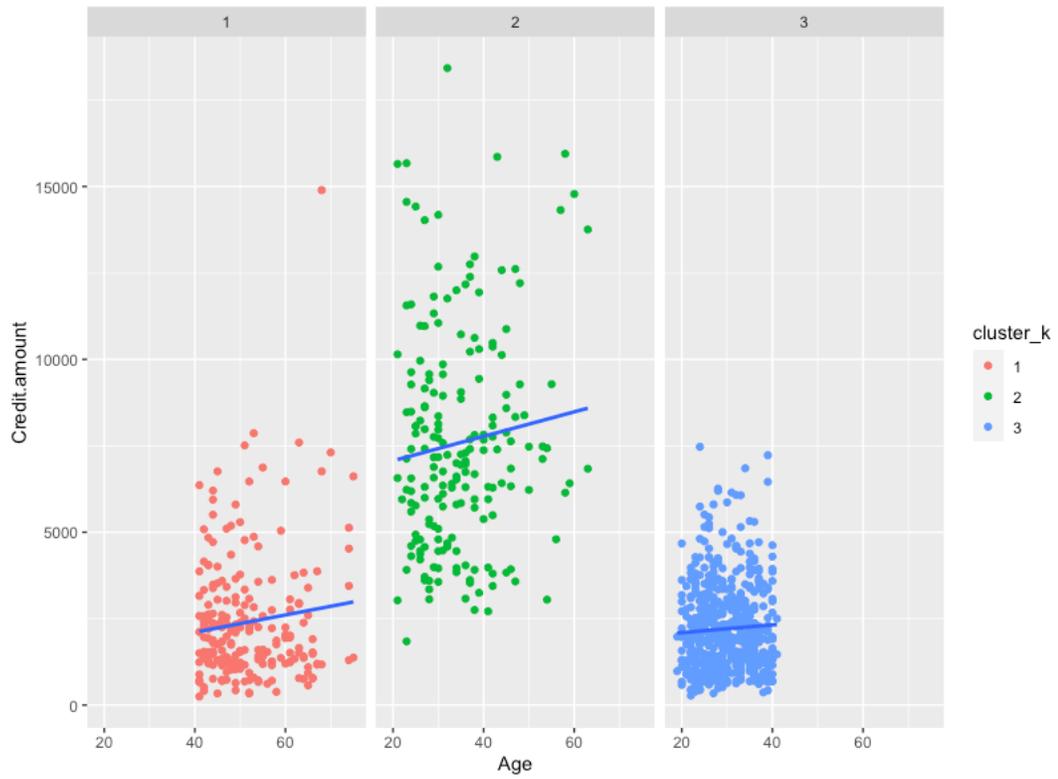
Nota: El Boxplot presentado se obtuvo utilizando R Studio con la base de datos “German_credit_data.csv” del primer modelo de segmentación mediante clusterización.

Figura 9. Scatterplot para la determinación de correlación entre las variables duración y edad.



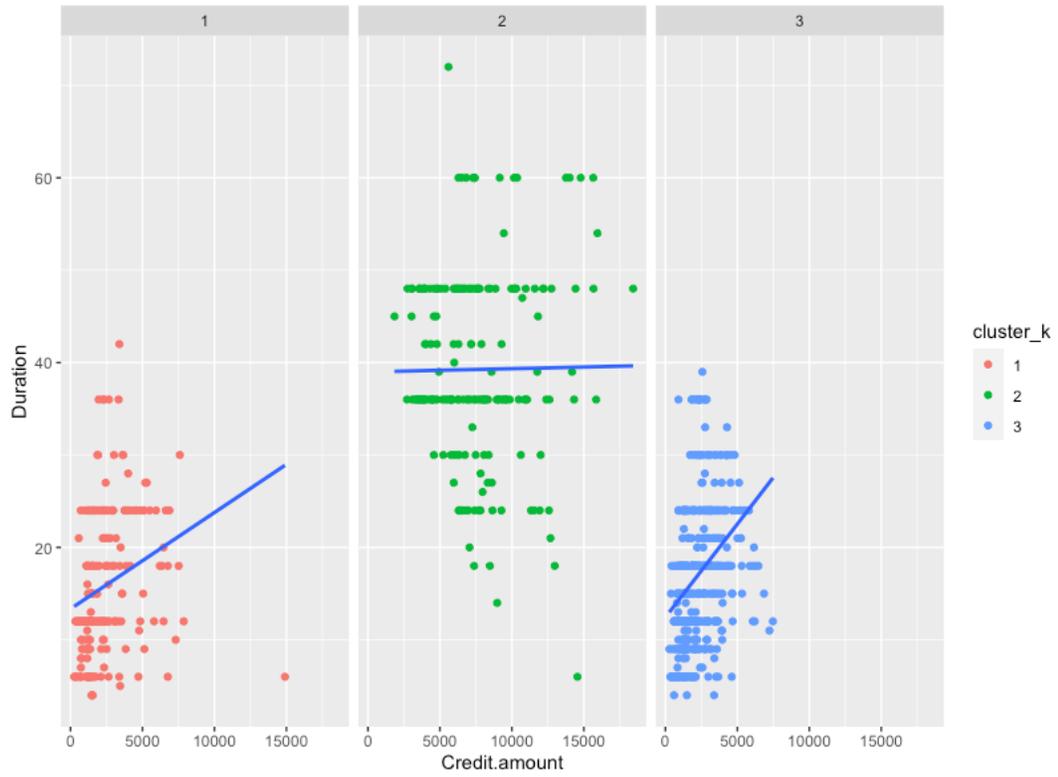
Nota: El Scatterplot presentado se obtuvo utilizando R Studio con la base de datos "German_credit_data.csv" del primer modelo de segmentación mediante clusterización.

Figura 10. Scatterplot para determinar la correlación entre las variables total de crédito adquirido y edad.



Nota: El Scatterplot presentado se obtuvo utilizando R Studio con la base de datos "German_credit_data.csv" del primer modelo de segmentación mediante clusterización.

Figura 11. Scatterplot para determinar la correlación entre las variables total de crédito adquirido y edad.



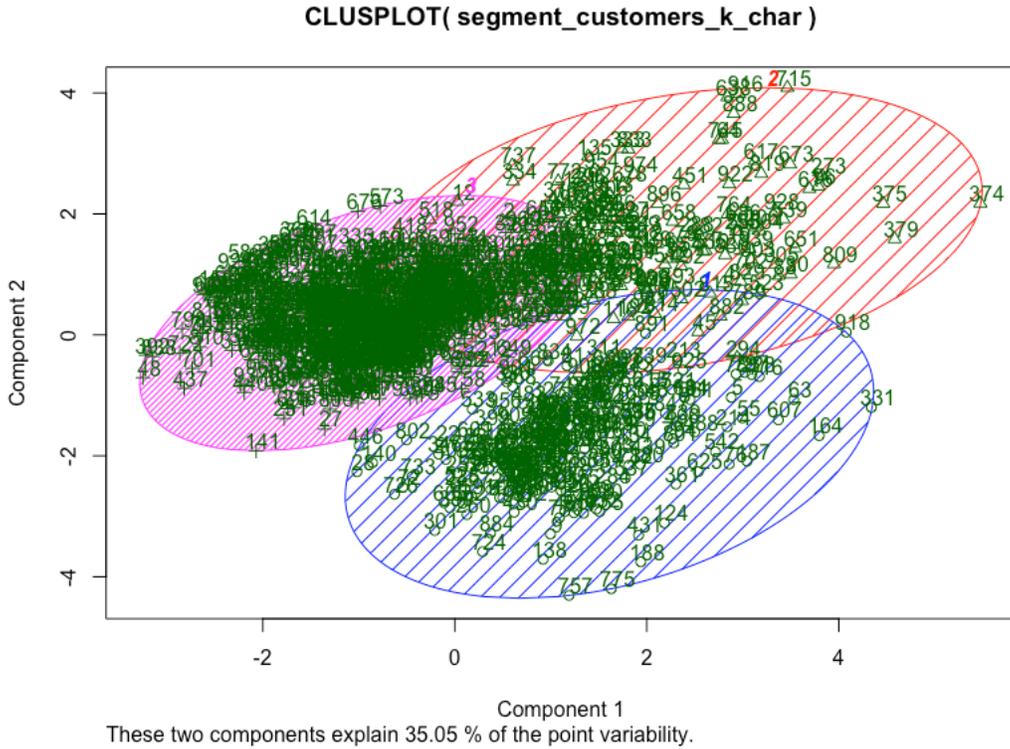
Nota: El Scatterplot presentado se obtuvo utilizando R Studio con la base de datos “German_credit_data.csv” del primer modelo de segmentación mediante clusterización.

Figura 12. Modelo de segmentación de árbol de decisión para el perfilado de cliente.



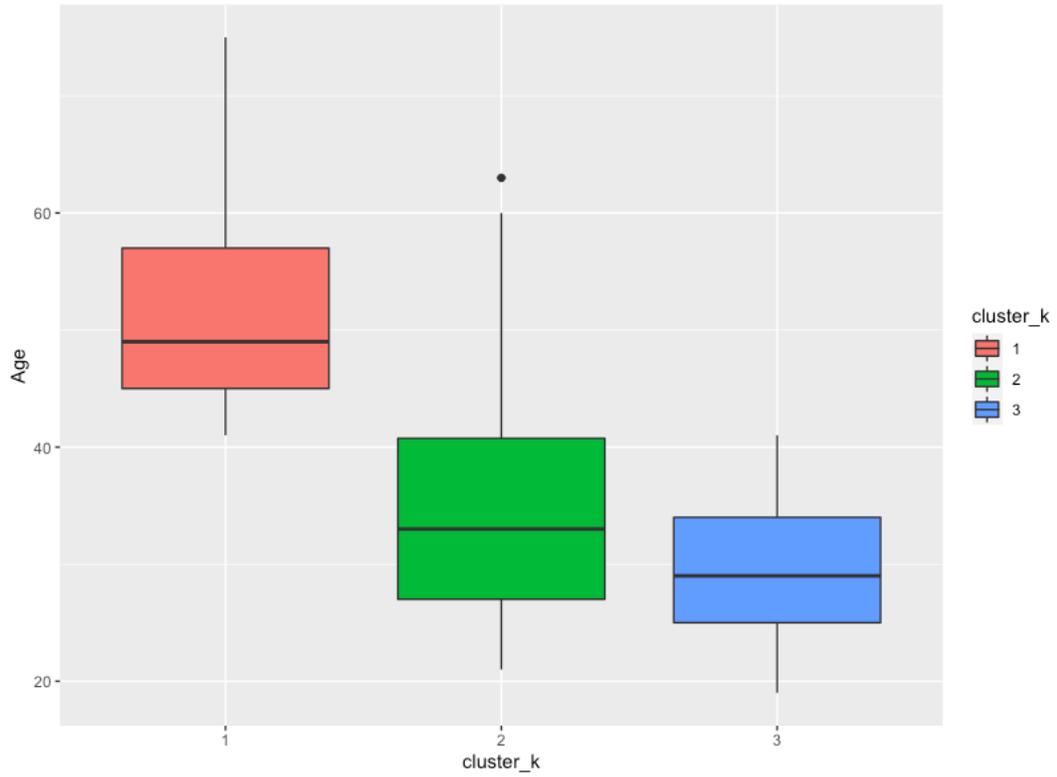
Nota: El árbol de decisión fue realizado utilizando el clúster de la Figura 5.

Figura 13. Segundo modelo de segmentación mediante clusterización de las variables de caracteres para la comparación de variables en la base de datos.



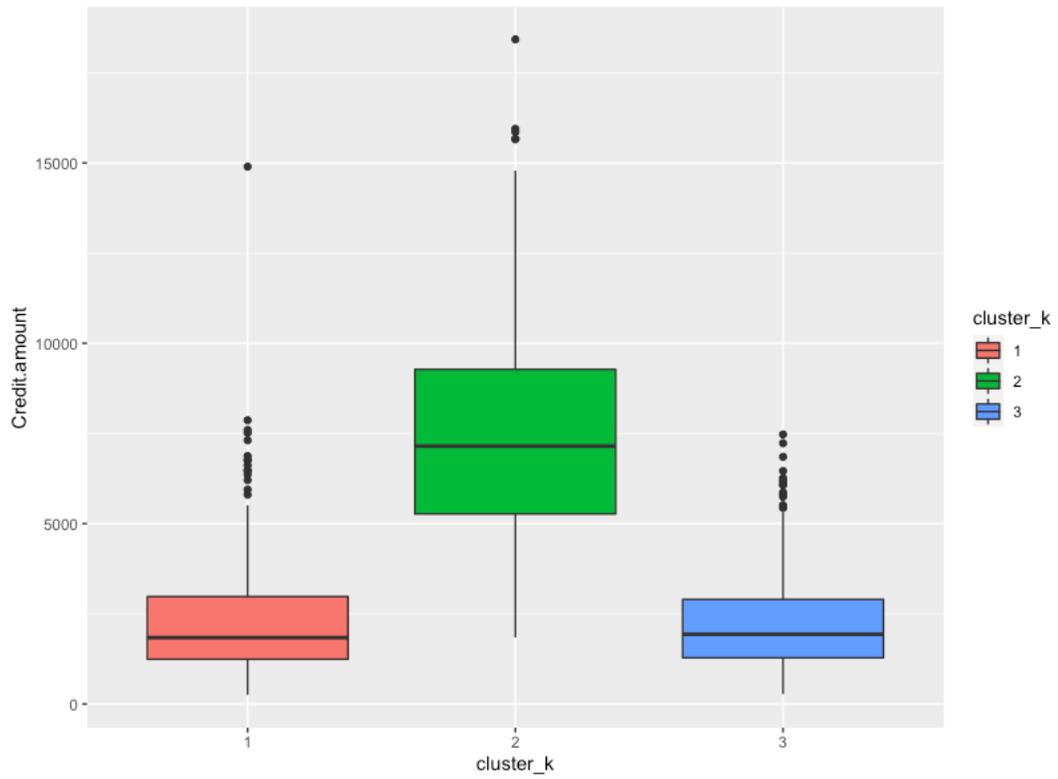
Nota: Las variables representadas en el modelo de segmentación de clusterización son: edad, género, trabajo, compra/alquiler, cuenta ahorro, cuenta monetaria, total de crédito, duración y propósito, de la base de datos "German_credit_data.csv".

Figura 14. Boxplot de la distribución entre los clústeres y la duración del crédito adquirido obtenido por el análisis de segmentación para el perfilado de clientes potenciales.



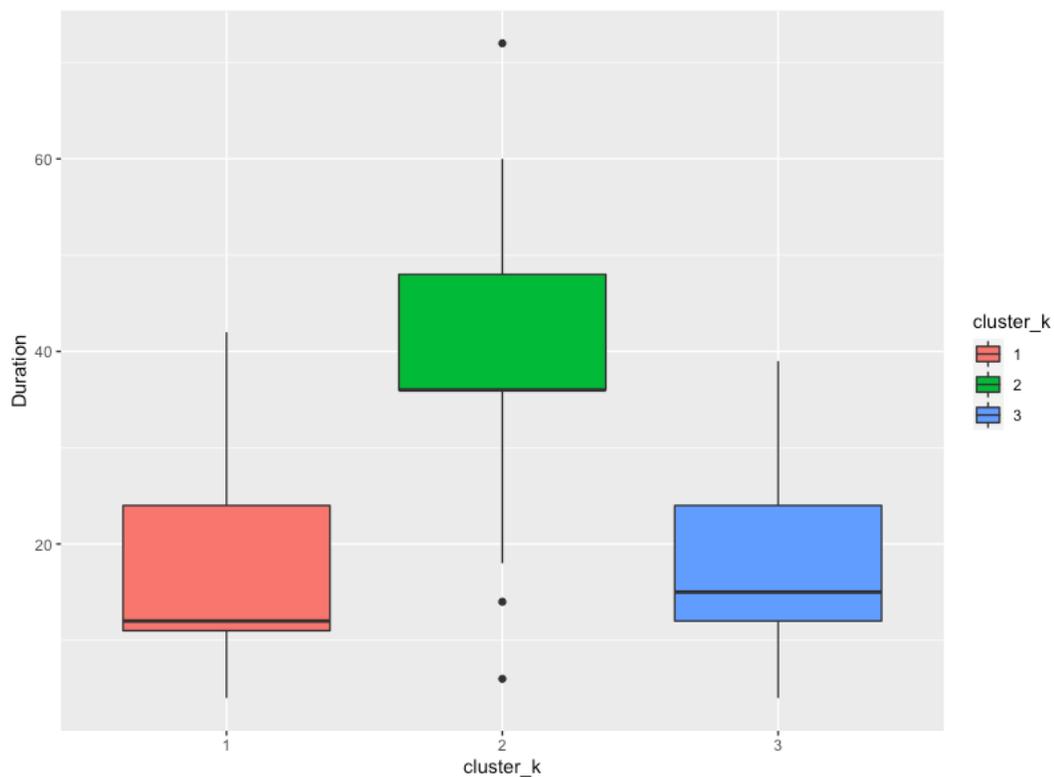
Nota: El Boxplot presentado se obtuvo utilizando R Studio con la base de datos “German_credit_data.csv” del primer modelo de segmentación mediante clusterización.

Figura 15. Boxplot de la distribución entre los clústeres y el total de crédito adquirido obtenido por el análisis de segmentación para el perfilado de clientes potenciales.



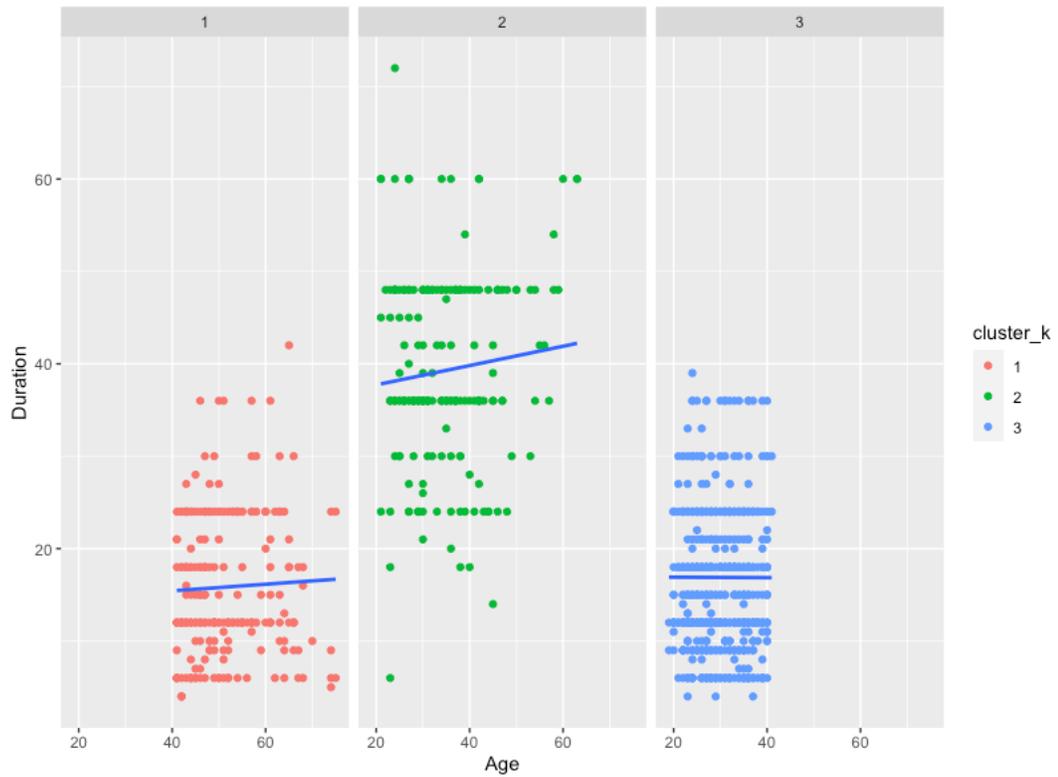
Nota: El Boxplot presentado se obtuvo utilizando R Studio con la base de datos “German_credit_data.csv” del primer modelo de segmentación mediante clusterización.

Figura 16. Boxplot de la distribución entre los clústeres y la duración del crédito adquirido obtenido por el análisis de segmentación para el perfilado de clientes potenciales.



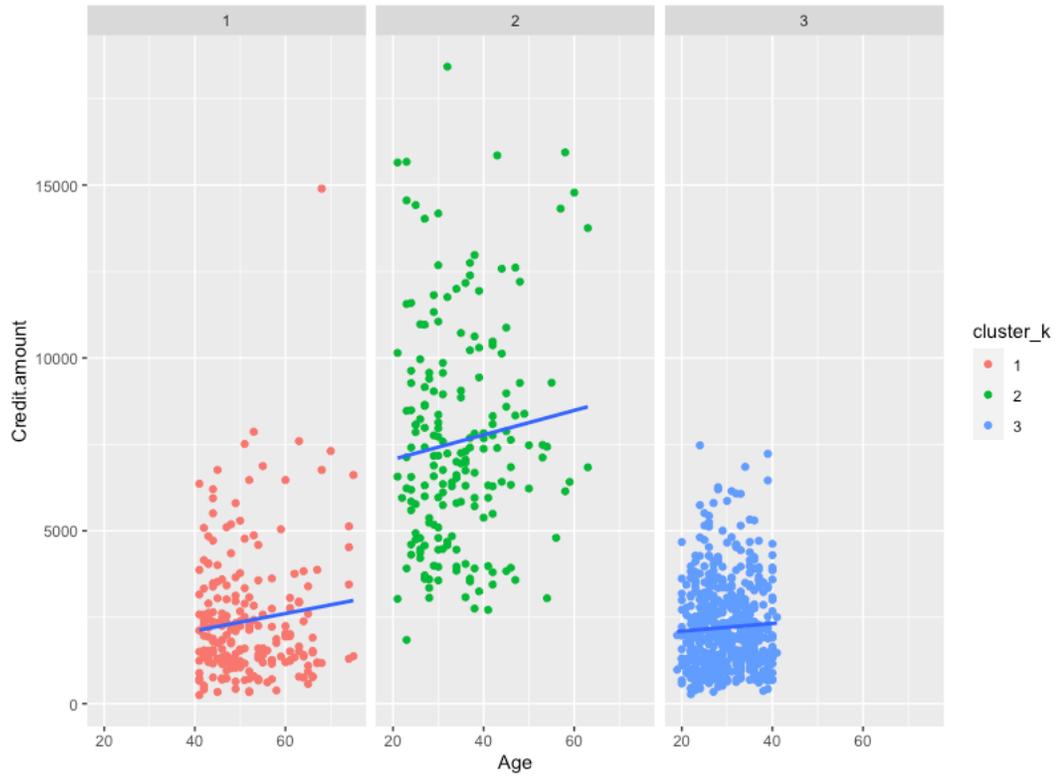
Nota: El Boxplot presentado se obtuvo utilizando R Studio con la base de datos “German_credit_data.csv” del primer modelo de segmentación mediante clusterización.

Figura 17. Scatterplot para la determinación de correlación entre las variables duración y edad.



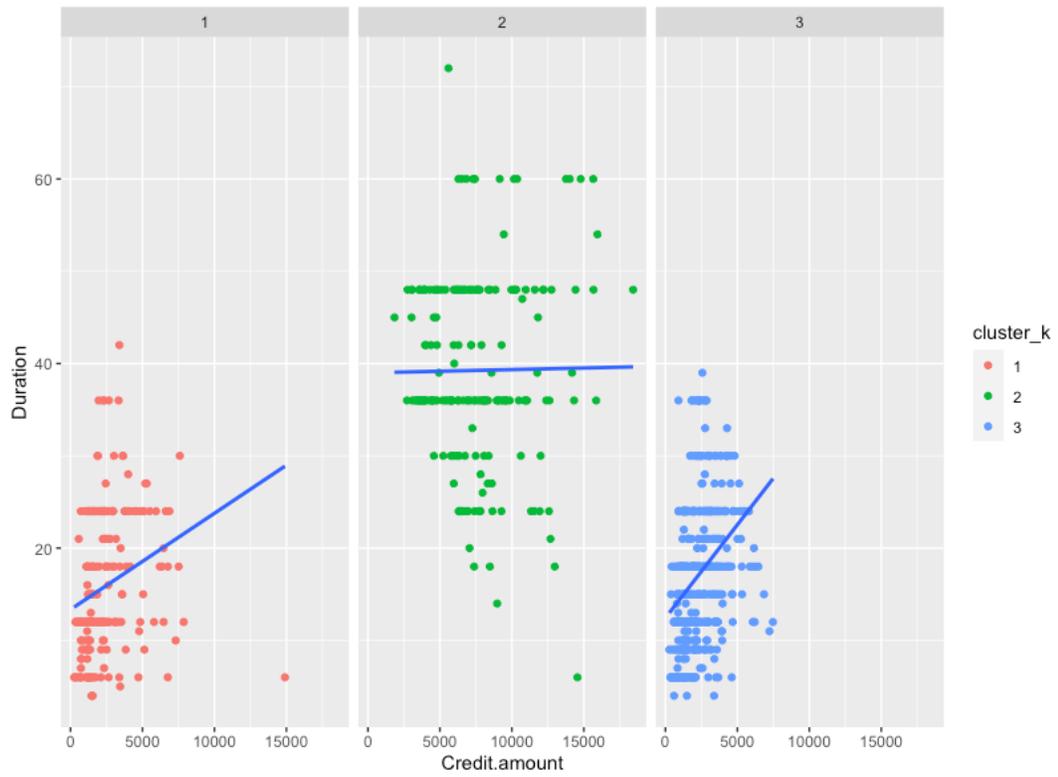
Nota: El Scatterplot presentado se obtuvo utilizando R Studio con la base de datos "German_credit_data.csv" del segundo modelo de segmentación mediante clusterización.

Figura 18. Scatterplot para determinar la correlación entre las variables total de crédito adquirido y edad.



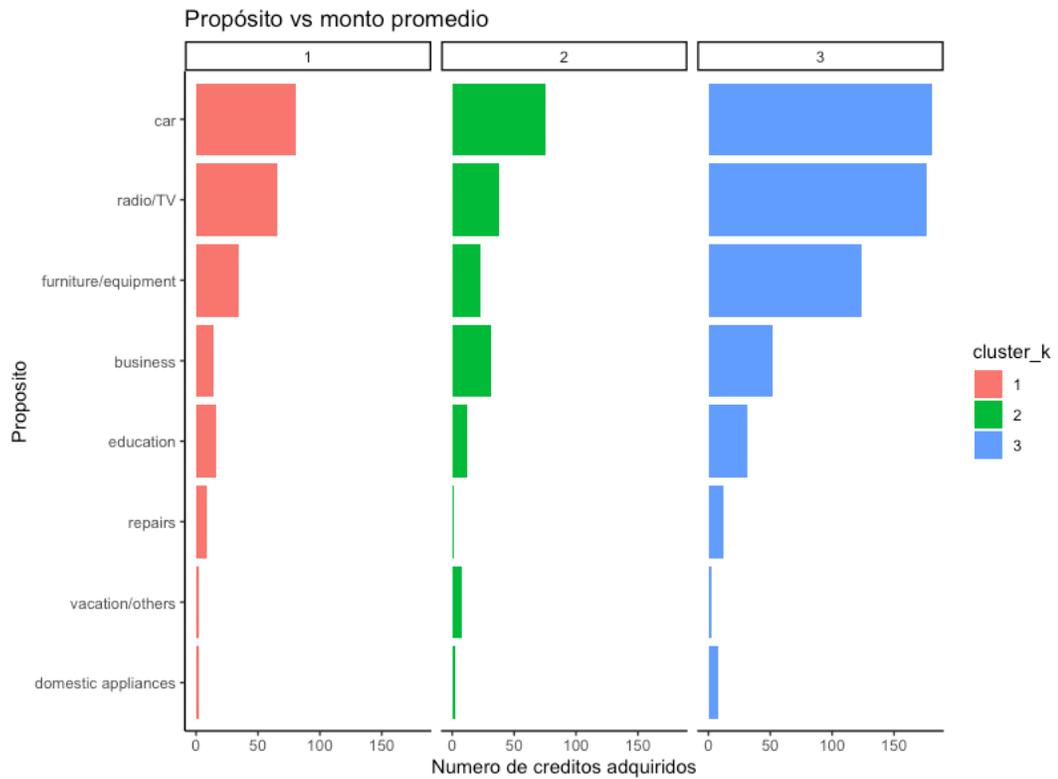
Nota: El Scatterplot presentado se obtuvo utilizando R Studio con la base de datos "German_credit_data.csv" del segundo modelo de segmentación mediante clusterización.

Figura 19. Scatterplot para determinar la correlación entre las variables total de crédito adquirido y edad.



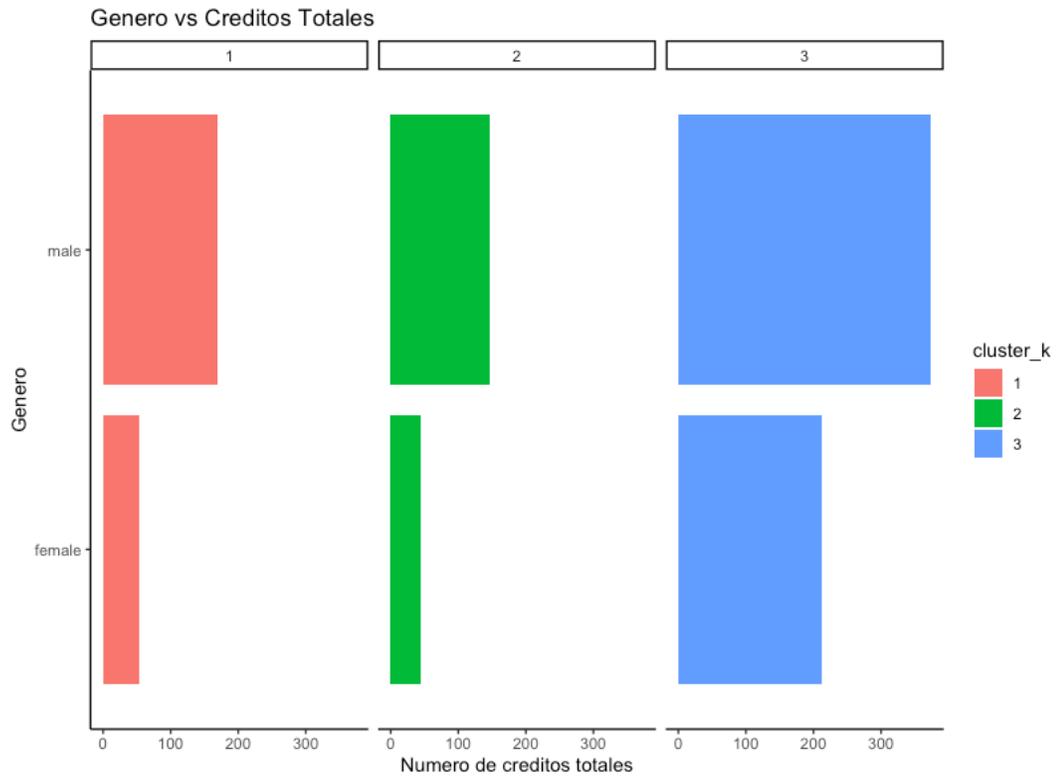
Nota: El Scatterplot presentado se obtuvo utilizando R Studio con la base de datos “German_credit_data.csv” del segundo modelo de segmentación mediante clusterización.

Figura 20. Modelo de segmentación mediante clusterización para la comparación de variables en la base de datos.



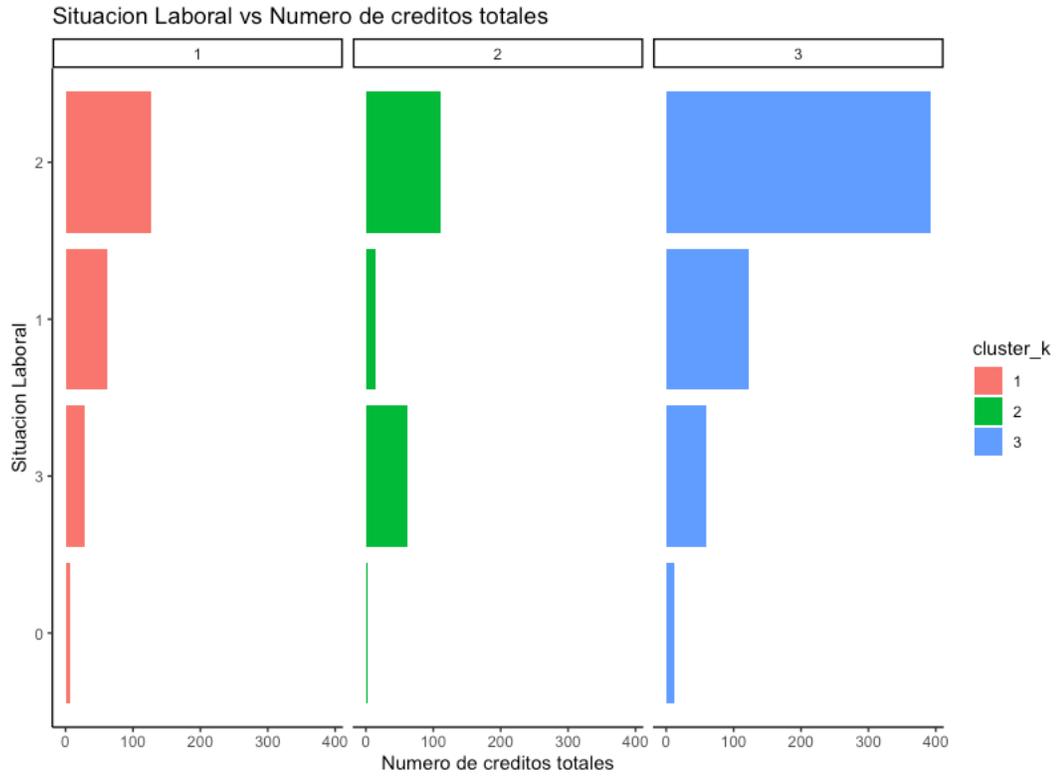
Nota: El Scatterplot presentado se obtuvo utilizando R Studio con la base de datos "German_credit_data.csv" del segundo modelo de segmentación mediante clusterización.

Figura 21. Barplot mediante clusterización para analizar el género y la cantidad de préstamos adquiridos.



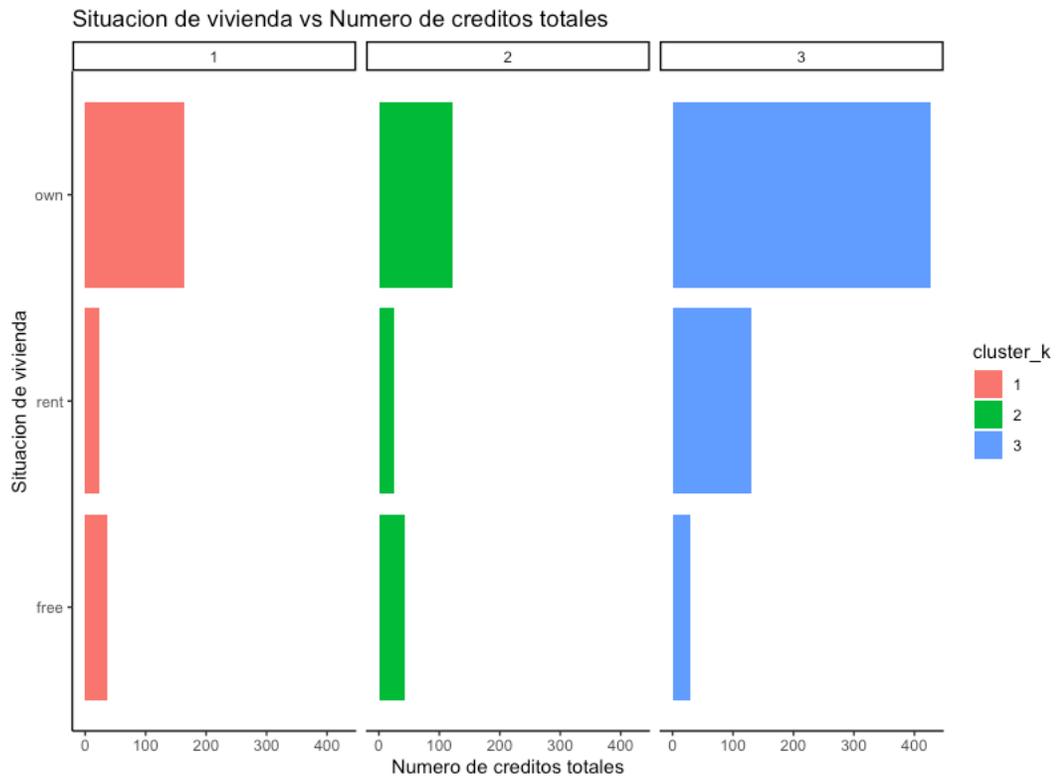
Nota: El barplot presentado se obtuvo utilizando R Studio con la base de datos "German_credit_data.csv" del segundo modelo de segmentación mediante clusterización.

Figura 22. Barplot mediante clusterización para analizar el género y el estado laboral del cliente.



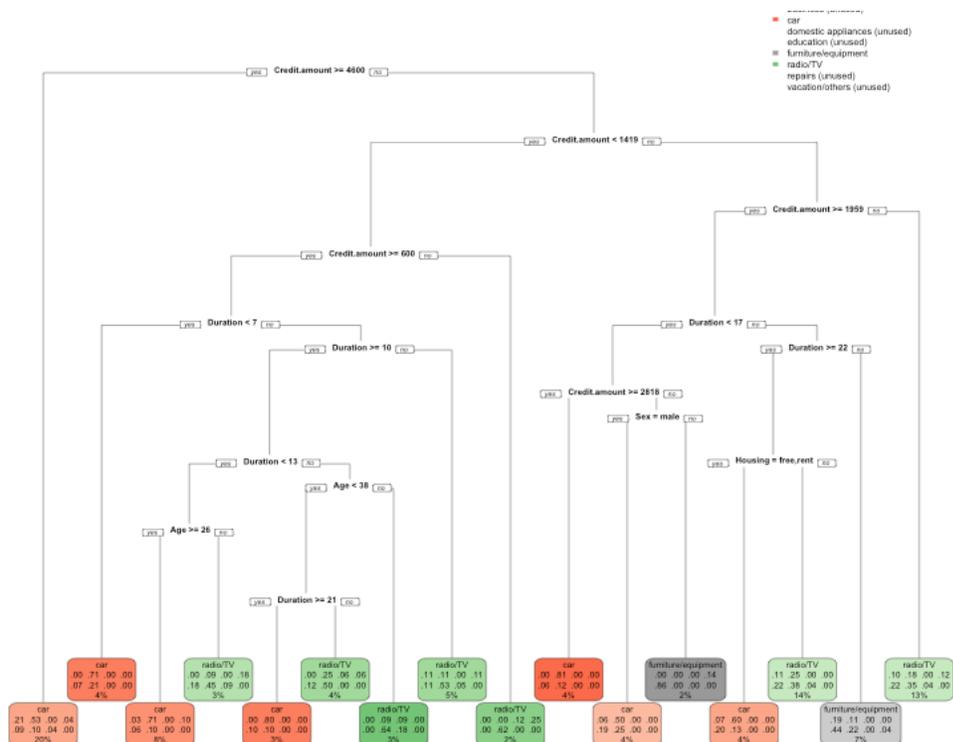
Nota: Las variables representadas en el modelo de segmentación de clusterización son: edad, género, trabajo, compra/alquiler, cuenta ahorro, cuenta monetaria, total de crédito, duración y propósito, de la base de datos "German_credit_data.csv".

Figura 23. Modelo de segmentación mediante clusterización para la comparación de variables en la base de datos.



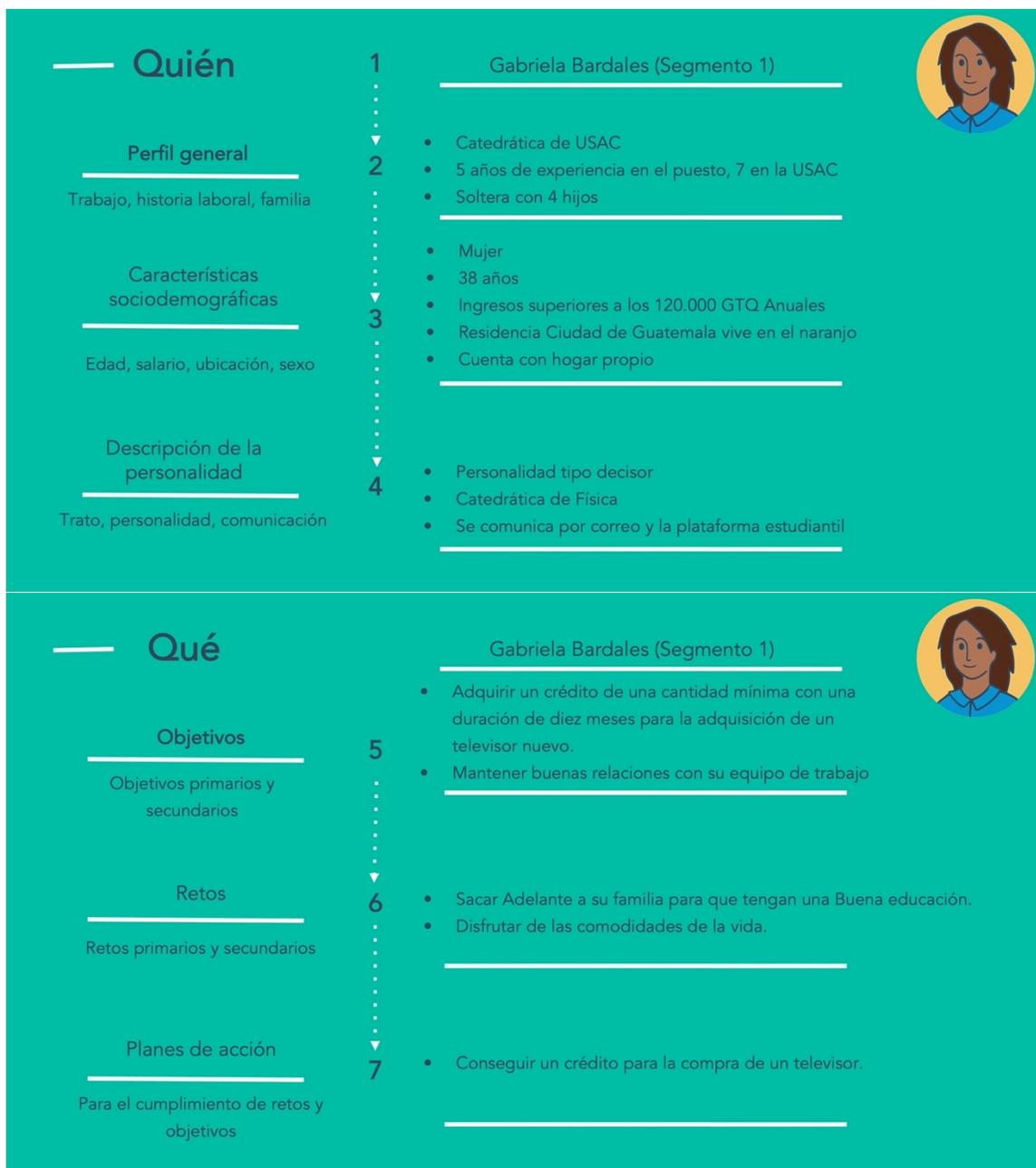
Nota: El barplot presentado se obtuvo utilizando R Studio con la base de datos "German_credit_data.csv" del segundo modelo de segmentación mediante clusterización

Figura 24. Modelo de segmentación de árbol de decisión para el perfilado de cliente.



Nota: Las variables representadas en el modelo de segmentación de clusterización son: edad, género, trabajo, compra/alquiler, cuenta ahorro, cuenta monetaria, total de crédito, duración y propósito, de la base de datos "German_credit_data.csv".

Figura 25. Buyer persona 1 basado en el árbol de decisión para la determinación del cliente potencial para la empresa.



— Por qué

Comentarios

Testimonios sobre retos y objetivos

Áreas de oportunidad

Argumentos para no mantener un relación comercial con nosotros

8

- «Gracias al buen trabajo que e realizado como catedrática durante los años me gustaría poder adquirir un crédito para poder adquirir un nuevo televisor para mi familia ..»

9

- Nuevos televisores al mejor precio!
- Tenemos las tasas de interés mas bajas en todo el mercado.

Gabriela Bardales (Segmento 1)



— Cómo

Mensaje de marketing

Respuesta a la problemática del cliente

Mensaje de ventas

Respuesta de ventas para llegar al cliente

10

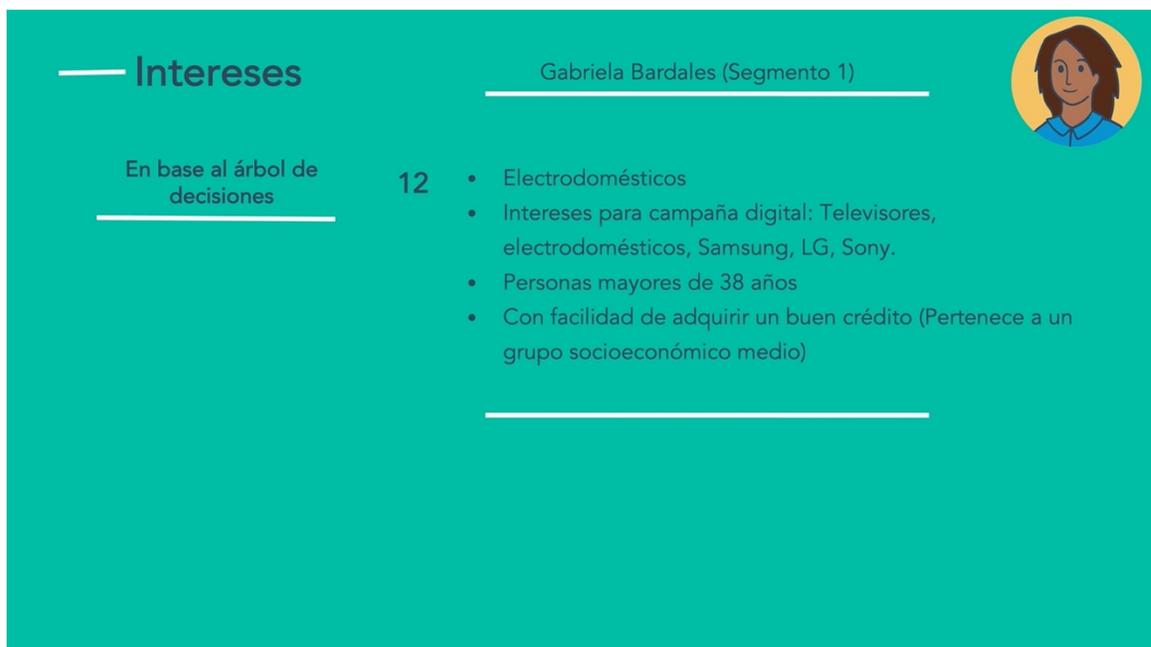
- Compra tu televisor mediante un crédito con baja tasa de interés y con un plazo corto!

11

- Contamos con una tasa de interés baja a la hora de adquirir un crédito con nosotros para que puedas adquirir tus electrodomésticos favoritos.

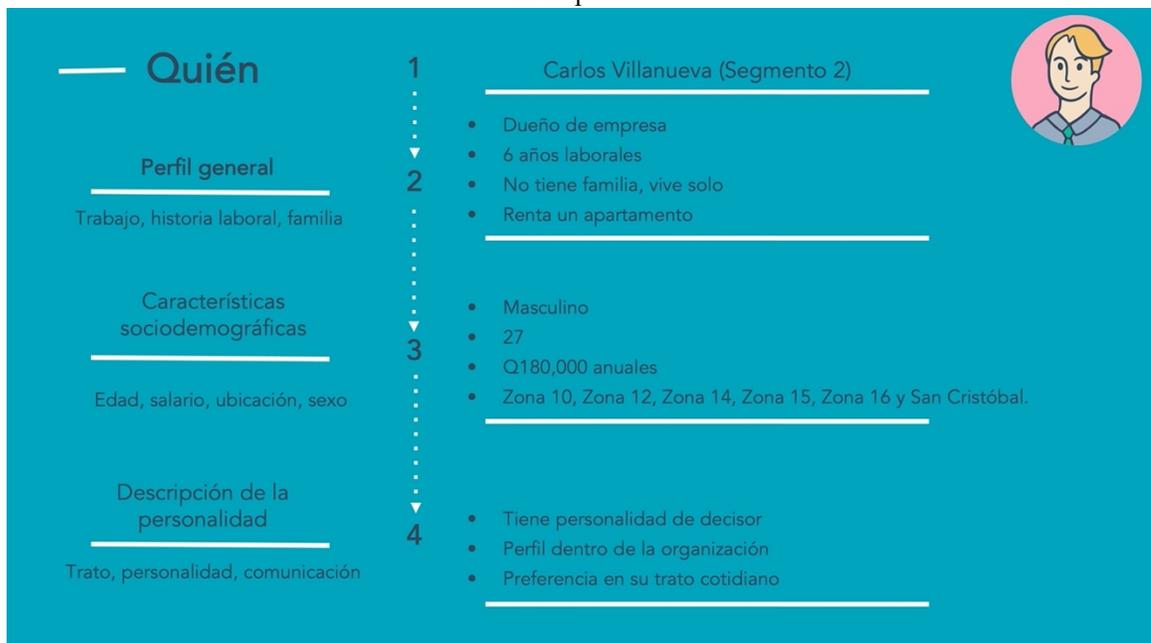
Gabriela Bardales (Segmento 1)





Nota: El buyer persona fue obtenido a partir de la información de los diferentes árboles de decisión presentados en las Figuras 4, 12 y 24. También se utilizaron los segmentos obtenidos mediante clusterización en las Figuras 5 y 13.

Figura 26. Buyer persona 2 basado en el árbol de decisión para la determinación del cliente potencial para la empresa.



Qué

Objetivos

Objetivos primarios y secundarios

5

- Comprar un auto para utilizar y poderse movilizar dentro de la ciudad para el mejor manejo de su empresa
- Adquirir un crédito para la compra del auto

Retos

Retos primarios y secundarios

6

- Continuar su negocio de transportación dentro del país
- Comprarse el carro de sus sueños

Planes de acción

Para el cumplimiento de retos y objetivos

7

- Adquirir el carro de su sueños mediante un crédito.



Carlos Villanueva (Segmento 2)

Por qué

Comentarios

Testimonios sobre retos y objetivos

8

- Mi objetivo principal es comprarme el carro de mis sueños
- Me gustaría poder adquirir el carro de mis sueños mediante un crédito para mejorar mi récord crediticio.

Áreas de oportunidad

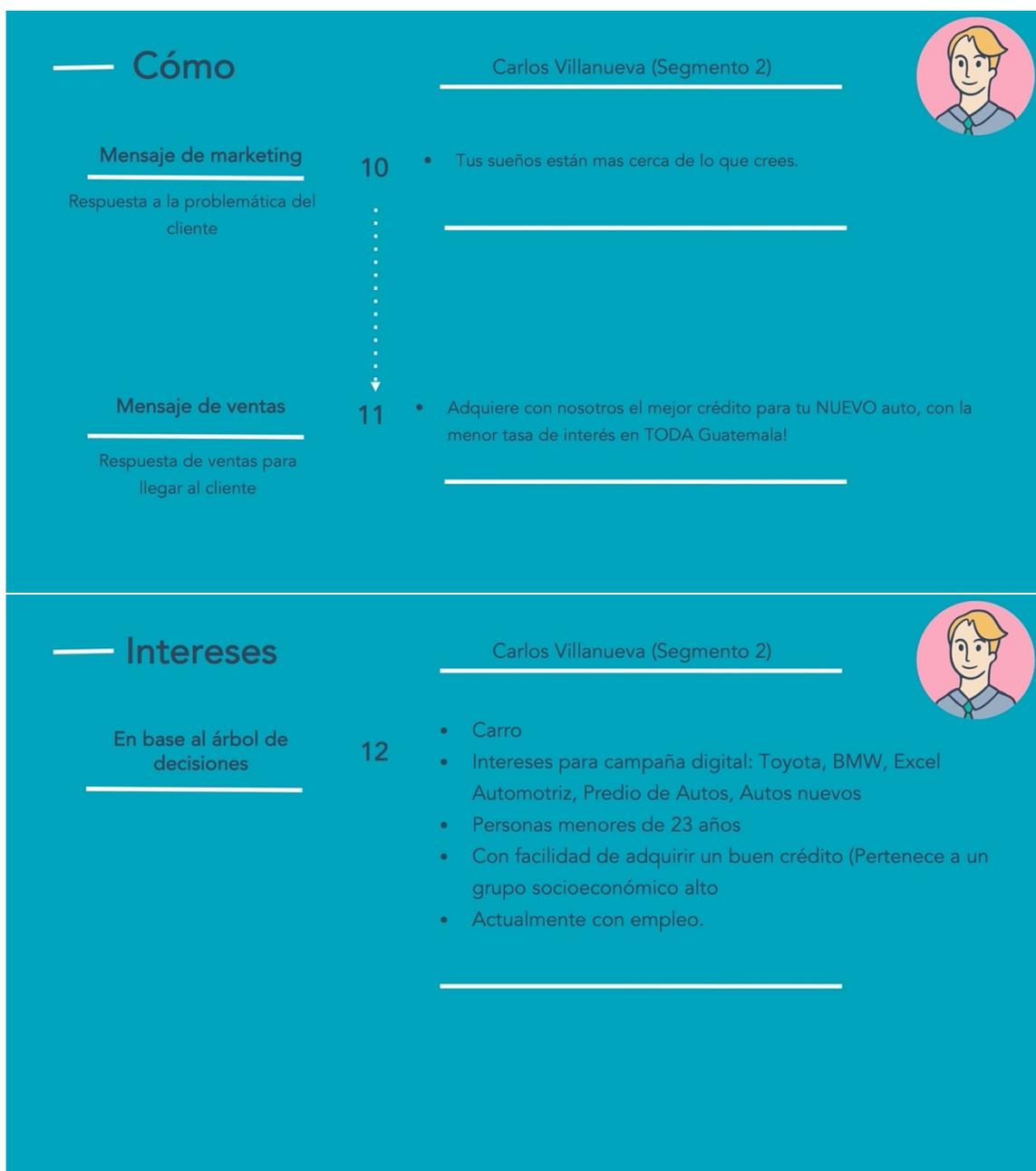
Argumentos para no mantener un relación comercial con nosotros

9

- Las tasas de interés en los bancos son muy altas y no atraen para adquirir el préstamo.



Carlos Villanueva (Segmento 2)



Nota: El buyer persona fue obtenido a partir de la información de los diferentes árboles de decisión presentados en las Figuras 4, 12 y 24. También se utilizaron los segmentos obtenidos mediante clusterización en las Figuras 5 y 13.

Figura 27. Buyer persona 2 basado en el árbol de decisión para la determinación del cliente potencial para la empresa.



— Por qué

Comentarios

Testimonios sobre retos y objetivos

Áreas de oportunidad

Argumentos para no mantener un relación comercial con nosotros

Joselin Alvizurez (Segmento 3)



8

- Me gustaria poder adquirir un credito para poder comprar los muebles mas bonitos para mi hogar.
- Me gustaria poder obtener una promocion dentro de mi trabajo para ganar mas dinero.

9

- Las tasas de interés en los bancos son muy altas y no atraen para adquirir el préstamo.

— Cómo

Mensaje de marketing

Respuesta a la problemática del cliente

Mensaje de ventas

Respuesta de ventas para llegar al cliente

Joselin Alvizurez (Segmento 3)



10

- ¡Aprovecha tu credito! Compra y equipa tu hogar con los muebles mas modernos del mercado.

11

- Adquiere con nosotros el mejor crédito para tus nuevos muebles, con la menor tasa de interés en TODA Guatemala!

Intereses

Joselin Alvirez (Segmento 3)



En base al árbol de decisiones

10

- Posición laboral: Secretaria
- Intereses para campaña digital: Muebles nuevos, mueblesm Ikea, O, Kalea, Walmart, hogar, familia.
- Personas mayores e iguales a 35 años
- Con facilidad de adquirir un buen crédito (Pertenece a un grupo socioeconómico bajo.
- Actualmente con empleo.

Nota: El buyer persona fue obtenido a partir de la información de los diferentes árboles de decisión presentados en las Figuras 4, 12 y 24. También se utilizaron los segmentos obtenidos mediante clusterización en las Figuras 5 y 13.

Cuadro 2. Proyección financiera utilizando tasa de crecimiento anual de 5% para determinar la viabilidad de la implementación del modelo determinístico para su aplicación en empresas financieras.

FLUJOS	Tasa de crecimiento anual											
	5%											
	Año en numero Año Calendario	Y0 2022	Y1 2023	Y2 2024	Y3 2025	Y4 2026	Y5 2027	Y6 2028	Y7 2029	Y8 2030	Y9 2031	Y10 2032
Ingresos												
Paquete de 10,000 interesados		\$ 450,000.00	\$ 472,500.00	\$ 496,125.00	\$ 520,931.25	\$ 546,977.81	\$ 574,326.70	\$ 603,043.04	\$ 633,195.19	\$ 664,854.95	\$ 698,097.70	
*Pauta incluida/ 10% de comision		\$ (300,000.00)	\$ (315,000.00)	\$ (330,750.00)	\$ (347,287.50)	\$ (364,651.88)	\$ (382,884.47)	\$ (402,028.69)	\$ (422,130.13)	\$ (443,236.63)	\$ (465,398.46)	
Paquete de 20,000 Prospectos		\$ 870,000.00	\$ 913,500.00	\$ 959,175.00	\$ 1,007,133.75	\$ 1,057,490.44	\$ 1,110,364.96	\$ 1,165,883.21	\$ 1,224,177.37	\$ 1,285,386.24	\$ 1,349,655.55	
*Pauta incluida/ 10% de comision		\$ (600,000.00)	\$ (630,000.00)	\$ (661,500.00)	\$ (694,575.00)	\$ (729,303.75)	\$ (765,768.94)	\$ (804,057.38)	\$ (844,260.25)	\$ (886,473.27)	\$ (930,796.93)	
Paquete de 30,000 Prospectos		\$ 1,308,000.00	\$ 1,373,400.00	\$ 1,442,070.00	\$ 1,514,173.50	\$ 1,589,882.18	\$ 1,669,376.28	\$ 1,752,845.10	\$ 1,840,487.35	\$ 1,932,511.72	\$ 2,029,137.31	
*Pauta incluida/ 10% de comision		\$ (900,000.00)	\$ (945,000.00)	\$ (992,250.00)	\$ (1,041,862.50)	\$ (1,093,955.63)	\$ (1,148,653.41)	\$ (1,206,086.08)	\$ (1,266,390.38)	\$ (1,329,709.90)	\$ (1,396,195.39)	
\$2,500 por llevar a 10,000 interesados con +/- 10% de incertidumbre mensual		\$ 828,000.00	\$ 869,400.00	\$ 912,870.00	\$ 958,513.50	\$ 1,006,439.18	\$ 1,056,761.13	\$ 1,109,599.19	\$ 1,165,079.15	\$ 1,223,333.11	\$ 1,284,499.76	
*Pauta no incluida en servicio												
TOTAL INGRESOS		\$ 828,000.00	\$ 869,400.00	\$ 912,870.00	\$ 958,513.50	\$ 1,006,439.18	\$ 1,056,761.13	\$ 1,109,599.19	\$ 1,165,079.15	\$ 1,223,333.11	\$ 1,284,499.76	
*Pauta incluida/ 10% de comision		\$ 82,800.00	\$ 86,940.00	\$ 91,287.00	\$ 95,851.35	\$ 100,643.92	\$ 105,676.11	\$ 110,959.92	\$ 116,507.91	\$ 122,333.31	\$ 128,449.98	
TOTAL CON COMISION		\$ 910,800.00	\$ 956,340.00	\$ 1,004,157.00	\$ 1,054,364.85	\$ 1,107,083.09	\$ 1,162,437.25	\$ 1,220,559.11	\$ 1,281,587.06	\$ 1,345,666.42	\$ 1,412,949.74	
COSTOS Y GASTOS												
Costos Fijos												
Sueldos		\$ 271,621.45	\$ 285,202.52	\$ 299,462.65	\$ 314,435.78	\$ 330,157.57	\$ 346,665.45	\$ 363,998.72	\$ 382,198.66	\$ 401,308.59	\$ 421,374.02	
Internet		\$ 1,200.00	\$ 1,260.00	\$ 1,323.00	\$ 1,389.15	\$ 1,458.61	\$ 1,531.54	\$ 1,608.11	\$ 1,688.52	\$ 1,772.95	\$ 1,861.59	
Hosting		\$ 1,200.00	\$ 1,260.00	\$ 1,323.00	\$ 1,389.15	\$ 1,458.61	\$ 1,531.54	\$ 1,608.11	\$ 1,688.52	\$ 1,772.95	\$ 1,861.59	
Pagina Web		\$ 1,200.00	\$ 1,260.00	\$ 1,323.00	\$ 1,389.15	\$ 1,458.61	\$ 1,531.54	\$ 1,608.11	\$ 1,688.52	\$ 1,772.95	\$ 1,861.59	
Contabilidad		\$ 12,000.00	\$ 12,600.00	\$ 13,230.00	\$ 13,891.50	\$ 14,586.08	\$ 15,315.38	\$ 16,081.15	\$ 16,885.21	\$ 17,729.47	\$ 18,615.94	
Oficina		\$ 24,000.00	\$ 25,200.00	\$ 26,460.00	\$ 27,783.00	\$ 29,172.15	\$ 30,630.76	\$ 32,162.30	\$ 33,770.41	\$ 35,458.93	\$ 37,231.88	
Mantenimiento		\$ 1,200.00	\$ 1,260.00	\$ 1,323.00	\$ 1,389.15	\$ 1,458.61	\$ 1,531.54	\$ 1,608.11	\$ 1,688.52	\$ 1,772.95	\$ 1,861.59	
Servicios Basicos		\$ 3,600.00	\$ 3,780.00	\$ 3,969.00	\$ 4,167.45	\$ 4,375.82	\$ 4,594.61	\$ 4,824.34	\$ 5,065.56	\$ 5,318.84	\$ 5,584.78	
Costos Variables												
Comision de venta		\$ 41,400.00	\$ 43,470.00	\$ 45,643.50	\$ 47,925.68	\$ 50,321.96	\$ 52,838.06	\$ 55,479.96	\$ 58,253.96	\$ 61,166.66	\$ 64,224.99	
Gasto Publicidad		\$ 12,000.00	\$ 12,600.00	\$ 13,230.00	\$ 13,891.50	\$ 14,586.08	\$ 15,315.38	\$ 16,081.15	\$ 16,885.21	\$ 17,729.47	\$ 18,615.94	
Capasitaciones		\$ 4,000.00	\$ 4,200.00	\$ 4,410.00	\$ 4,630.50	\$ 4,862.03	\$ 5,105.13	\$ 5,360.38	\$ 5,628.40	\$ 5,909.82	\$ 6,205.31	
TOTAL COSTOS Y GASTOS		\$ 373,421.45	\$ 392,092.52	\$ 411,697.15	\$ 432,282.01	\$ 453,896.11	\$ 476,590.91	\$ 500,420.46	\$ 525,441.48	\$ 551,713.56	\$ 579,299.23	
EBITDA		\$ 537,378.55	\$ 564,247.48	\$ 592,459.85	\$ 622,082.84	\$ 653,186.98	\$ 685,846.33	\$ 720,138.65	\$ 756,145.58	\$ 793,952.86	\$ 833,650.51	
Depresacion de Computadoras		\$ (3,000.00)	\$ (3,150.00)	\$ (3,307.50)	\$ (3,472.88)	\$ (3,646.52)	\$ (3,828.84)	\$ (4,020.29)	\$ (4,221.30)	\$ (4,432.37)	\$ (4,653.98)	
EBT		\$ 534,378.55	\$ 561,097.48	\$ 589,152.35	\$ 618,609.97	\$ 649,540.47	\$ 682,017.49	\$ 716,118.36	\$ 751,924.28	\$ 789,520.50	\$ 828,996.52	
Impuestos		\$ (90,844.35)	\$ (95,386.57)	\$ (100,155.90)	\$ (105,163.69)	\$ (110,421.88)	\$ (115,942.97)	\$ (121,740.12)	\$ (127,827.13)	\$ (134,218.48)	\$ (140,929.41)	
Utilidad		\$ 443,534.20	\$ 465,710.91	\$ 488,996.45	\$ 513,446.27	\$ 539,118.59	\$ 566,074.52	\$ 594,378.24	\$ 624,097.15	\$ 655,302.01	\$ 688,067.11	
INVERSION		\$ (882,000.00)										
Cashflow												
Utilidad Neta		\$ 443,534.20	\$ 465,710.91	\$ 488,996.45	\$ 513,446.27	\$ 539,118.59	\$ 566,074.52	\$ 594,378.24	\$ 624,097.15	\$ 655,302.01	\$ 688,067.11	
Depreciacion		\$ (3,000.00)	\$ (3,150.00)	\$ (3,307.50)	\$ (3,472.88)	\$ (3,646.52)	\$ (3,828.84)	\$ (4,020.29)	\$ (4,221.30)	\$ (4,432.37)	\$ (4,653.98)	
TOTAL CASHFLOW		\$ 446,534.20	\$ 468,860.91	\$ 492,303.95	\$ 516,919.15	\$ 542,765.11	\$ 569,903.36	\$ 598,398.53	\$ 628,318.46	\$ 659,734.38	\$ 692,721.10	
EBITDA		\$ 537,378.55	\$ 564,247.48	\$ 592,459.85	\$ 622,082.84	\$ 653,186.98	\$ 685,846.33	\$ 720,138.65	\$ 756,145.58	\$ 793,952.86	\$ 833,650.51	
CashFlow		\$ 446,534.20	\$ 468,860.91	\$ 492,303.95	\$ 516,919.15	\$ 542,765.11	\$ 569,903.36	\$ 598,398.53	\$ 628,318.46	\$ 659,734.38	\$ 692,721.10	
FreeCashFlow		\$ (882,000.00)	\$ 537,378.55	\$ 564,247.48	\$ 592,459.85	\$ 622,082.84	\$ 653,186.98	\$ 685,846.33	\$ 720,138.65	\$ 756,145.58	\$ 793,952.86	\$ 833,650.51
TIR												65%
VPN		\$	\$ 2,324,114.08									
WACC												15.03%

Nota: La proyección financiera fue realizadas utilizando como precio base \$2500. Utilizando \$5000 de pauta mensual para alcanzar 30,000 interesados a un costo de \$0.17 por interesado.

Cuadro 3. Proyección financiera utilizando tasa de crecimiento anual de 10% para determinar la viabilidad de la implementación del modelo determinístico para su aplicación en empresas financieras.

FLUJOS	Tasa de crecimiento anual											
	Año en numero Año Calendario	10%										
		Y0 2022	Y1 2023	Y2 2024	Y3 2025	Y4 2026	Y5 2027	Y6 2028	Y7 2029	Y8 2030	Y9 2031	Y10 2032
Ingresos												
Paquete de 10,000 Interesados	\$ 450,000.00	\$ 495,000.00	\$ 544,500.00	\$ 598,950.00	\$ 658,845.00	\$ 724,729.50	\$ 797,202.45	\$ 876,922.70	\$ 964,614.96	\$ 1,061,076.46		
*Pauta incluida/ 10% de comision	\$ (300,000.00)	\$ (330,000.00)	\$ (363,000.00)	\$ (399,300.00)	\$ (439,230.00)	\$ (483,153.00)	\$ (531,468.30)	\$ (584,615.13)	\$ (643,076.64)	\$ (707,384.31)		
Paquete de 20,000 Prospectos	\$ 870,000.00	\$ 957,000.00	\$ 1,052,700.00	\$ 1,157,970.00	\$ 1,273,767.00	\$ 1,401,143.70	\$ 1,541,258.07	\$ 1,695,383.88	\$ 1,864,922.26	\$ 2,051,414.49		
*Pauta incluida/ 10% de comision	\$ (600,000.00)	\$ (660,000.00)	\$ (726,000.00)	\$ (798,600.00)	\$ (878,460.00)	\$ (966,306.00)	\$ (1,062,936.60)	\$ (1,169,230.26)	\$ (1,286,153.29)	\$ (1,414,768.61)		
Paquete de 30,000 Prospectos	\$ 1,308,000.00	\$ 1,438,800.00	\$ 1,582,680.00	\$ 1,740,948.00	\$ 1,915,042.80	\$ 2,106,547.08	\$ 2,317,201.79	\$ 2,548,921.97	\$ 2,803,814.16	\$ 3,084,195.58		
*Pauta incluida/ 10% de comision	\$ (900,000.00)	\$ (990,000.00)	\$ (1,089,000.00)	\$ (1,197,900.00)	\$ (1,317,690.00)	\$ (1,449,459.00)	\$ (1,594,404.90)	\$ (1,753,845.39)	\$ (1,929,229.93)	\$ (2,122,152.92)		
\$2,500 por llevar a 10,000 interesados con +/- 10% de incertidumbre mensual	\$ 828,000.00	\$ 910,800.00	\$ 1,001,880.00	\$ 1,102,068.00	\$ 1,212,274.80	\$ 1,333,502.28	\$ 1,466,852.51	\$ 1,613,537.76	\$ 1,774,891.53	\$ 1,952,380.69		
*Pauta no incluida en servicio												
TOTAL INGRESOS	\$ 828,000.00	\$ 910,800.00	\$ 1,001,880.00	\$ 1,102,068.00	\$ 1,212,274.80	\$ 1,333,502.28	\$ 1,466,852.51	\$ 1,613,537.76	\$ 1,774,891.53	\$ 1,952,380.69		
*Pauta incluida/ 10% de comision	\$ 82,800.00	\$ 91,080.00	\$ 100,188.00	\$ 110,206.80	\$ 121,227.48	\$ 133,350.23	\$ 146,685.25	\$ 161,353.78	\$ 177,489.15	\$ 195,238.07		
TOTAL CON COMISION	\$ 910,800.00	\$ 1,001,880.00	\$ 1,102,068.00	\$ 1,212,274.80	\$ 1,333,502.28	\$ 1,466,852.51	\$ 1,613,537.76	\$ 1,774,891.53	\$ 1,952,380.69	\$ 2,147,618.76		
COSTOS Y GASTOS												
Costos Fijos												
Sueldos	\$ 271,621.45	\$ 298,783.60	\$ 328,661.96	\$ 361,528.15	\$ 397,680.97	\$ 437,449.06	\$ 481,193.97	\$ 529,313.37	\$ 582,244.70	\$ 640,469.17		
Internet	\$ 1,200.00	\$ 1,320.00	\$ 1,452.00	\$ 1,597.20	\$ 1,756.92	\$ 1,932.61	\$ 2,125.87	\$ 2,338.46	\$ 2,572.31	\$ 2,829.54		
Hosting	\$ 1,200.00	\$ 1,320.00	\$ 1,452.00	\$ 1,597.20	\$ 1,756.92	\$ 1,932.61	\$ 2,125.87	\$ 2,338.46	\$ 2,572.31	\$ 2,829.54		
Pagina Web	\$ 1,200.00	\$ 1,320.00	\$ 1,452.00	\$ 1,597.20	\$ 1,756.92	\$ 1,932.61	\$ 2,125.87	\$ 2,338.46	\$ 2,572.31	\$ 2,829.54		
Contabilidad	\$ 12,000.00	\$ 13,200.00	\$ 14,520.00	\$ 15,972.00	\$ 17,569.20	\$ 19,326.12	\$ 21,258.73	\$ 23,384.61	\$ 25,723.07	\$ 28,295.37		
Oficina	\$ 24,000.00	\$ 26,400.00	\$ 29,040.00	\$ 31,944.00	\$ 35,138.40	\$ 38,652.24	\$ 42,517.46	\$ 46,769.21	\$ 51,446.13	\$ 56,590.74		
Mantenimiento	\$ 1,200.00	\$ 1,320.00	\$ 1,452.00	\$ 1,597.20	\$ 1,756.92	\$ 1,932.61	\$ 2,125.87	\$ 2,338.46	\$ 2,572.31	\$ 2,829.54		
Servicios Basicos	\$ 3,600.00	\$ 3,960.00	\$ 4,356.00	\$ 4,791.60	\$ 5,270.76	\$ 5,797.84	\$ 6,377.62	\$ 7,015.38	\$ 7,716.92	\$ 8,488.61		
Costos Variables												
Comision de venta	\$ 41,400.00	\$ 45,540.00	\$ 50,094.00	\$ 55,103.40	\$ 60,613.74	\$ 66,675.11	\$ 73,342.63	\$ 80,676.89	\$ 88,744.58	\$ 97,619.03		
Gasto Publicidad	\$ 12,000.00	\$ 13,200.00	\$ 14,520.00	\$ 15,972.00	\$ 17,569.20	\$ 19,326.12	\$ 21,258.73	\$ 23,384.61	\$ 25,723.07	\$ 28,295.37		
Capasitaciones	\$ 4,000.00	\$ 4,400.00	\$ 4,840.00	\$ 5,324.00	\$ 5,856.40	\$ 6,442.04	\$ 7,086.24	\$ 7,794.87	\$ 8,574.36	\$ 9,431.79		
TOTAL COSTOS Y GASTOS	\$ 373,421.45	\$ 410,763.60	\$ 451,839.96	\$ 497,023.95	\$ 546,726.35	\$ 601,398.98	\$ 661,538.88	\$ 727,692.77	\$ 800,462.04	\$ 880,508.25		
EBITDA	\$ 537,378.55	\$ 591,116.40	\$ 650,228.04	\$ 715,250.85	\$ 786,775.93	\$ 865,453.53	\$ 951,998.88	\$ 1,047,198.77	\$ 1,151,918.64	\$ 1,267,110.51		
Deprecacion de Computadoras	\$ (3,000.00)	\$ (3,300.00)	\$ (3,630.00)	\$ (3,993.00)	\$ (4,392.30)	\$ (4,831.53)	\$ (5,314.68)	\$ (5,846.15)	\$ (6,430.77)	\$ (7,073.84)		
EBT	\$ 534,378.55	\$ 587,816.40	\$ 646,598.04	\$ 711,257.85	\$ 782,383.63	\$ 860,622.00	\$ 946,684.20	\$ 1,041,352.62	\$ 1,145,487.88	\$ 1,260,036.67		
Impuestos	\$ (90,844.35)	\$ (99,928.79)	\$ (109,921.67)	\$ (120,913.83)	\$ (133,005.22)	\$ (146,305.74)	\$ (160,936.31)	\$ (177,029.94)	\$ (194,732.94)	\$ (214,206.23)		
Utilidad	\$ 443,534.20	\$ 487,887.61	\$ 536,676.38	\$ 590,344.01	\$ 649,378.42	\$ 714,316.26	\$ 785,747.88	\$ 864,322.67	\$ 950,754.94	\$ 1,045,830.43		
INVERSION	\$ (882,000.00)											
Cashflow												
Utilidad Neta	\$ 443,534.20	\$ 487,887.61	\$ 536,676.38	\$ 590,344.01	\$ 649,378.42	\$ 714,316.26	\$ 785,747.88	\$ 864,322.67	\$ 950,754.94	\$ 1,045,830.43		
Depreciacion	\$ (3,000.00)	\$ (3,300.00)	\$ (3,630.00)	\$ (3,993.00)	\$ (4,392.30)	\$ (4,831.53)	\$ (5,314.68)	\$ (5,846.15)	\$ (6,430.77)	\$ (7,073.84)		
TOTAL CASHFLOW	\$ 446,534.20	\$ 491,187.61	\$ 540,306.38	\$ 594,337.01	\$ 653,770.72	\$ 719,147.79	\$ 791,062.57	\$ 870,168.82	\$ 957,185.70	\$ 1,052,904.28		
EBITDA	\$ 537,378.55	\$ 591,116.40	\$ 650,228.04	\$ 715,250.85	\$ 786,775.93	\$ 865,453.53	\$ 951,998.88	\$ 1,047,198.77	\$ 1,151,918.64	\$ 1,267,110.51		
CashFlow	\$ 446,534.20	\$ 491,187.61	\$ 540,306.38	\$ 594,337.01	\$ 653,770.72	\$ 719,147.79	\$ 791,062.57	\$ 870,168.82	\$ 957,185.70	\$ 1,052,904.28		
FreeCashFlow	\$ (882,000.00)	\$ 537,378.55	\$ 591,116.40	\$ 650,228.04	\$ 715,250.85	\$ 786,775.93	\$ 865,453.53	\$ 951,998.88	\$ 1,047,198.77	\$ 1,151,918.64	\$ 1,267,110.51	
TIR		70%										
VPN	\$	2,969,782.62										
WACC		15.03%										

Nota: Nota: La proyección financiera fue realizadas utilizando como precio base \$2500. Utilizando \$5000 de pauta mensual para alcanzar \$30,000 interesados a un costo de \$0.17 por interesado.

Cuadro 4. Proyección financiera utilizando tasa de crecimiento anual de 15% para determinar la viabilidad de la implementación del modelo determinístico para su aplicación en empresas financieras.

FLUJOS	Tasa de crecimiento anual 15%																			
	Año en numero	Y0	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10								
	Año Calendario	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032								
Ingresos																				
Paquete de 10,000 interesados	\$	450,000.00	\$	517,500.00	\$	595,125.00	\$	684,393.75	\$	787,052.81	\$	905,110.73	\$	1,040,877.34	\$	1,197,008.95	\$	1,376,560.29	\$	1,583,044.33
*Pauta incluida/ 10% de comision	\$	(300,000.00)	\$	(345,000.00)	\$	(396,750.00)	\$	(456,262.50)	\$	(524,701.88)	\$	(603,407.16)	\$	(693,918.23)	\$	(798,005.96)	\$	(917,706.86)	\$	(1,055,362.89)
Paquete de 20,000 Prospectos	\$	870,000.00	\$	1,000,500.00	\$	1,150,575.00	\$	1,323,161.25	\$	1,521,635.44	\$	1,749,880.75	\$	2,012,362.87	\$	2,314,217.30	\$	2,661,349.89	\$	3,060,552.37
*Pauta incluida/ 10% de comision	\$	(600,000.00)	\$	(690,000.00)	\$	(793,500.00)	\$	(912,525.00)	\$	(1,049,403.75)	\$	(1,206,814.31)	\$	(1,387,836.46)	\$	(1,596,011.93)	\$	(1,835,413.72)	\$	(2,110,725.78)
Paquete de 30,000 Prospectos	\$	1,308,000.00	\$	1,504,200.00	\$	1,729,830.00	\$	1,989,304.50	\$	2,287,700.18	\$	2,630,855.20	\$	3,025,483.48	\$	3,479,306.00	\$	4,001,201.90	\$	4,601,382.19
*Pauta incluida/ 10% de comision	\$	(900,000.00)	\$	(1,035,000.00)	\$	(1,190,250.00)	\$	(1,368,787.50)	\$	(1,574,105.63)	\$	(1,810,221.47)	\$	(2,081,754.69)	\$	(2,394,017.89)	\$	(2,753,120.58)	\$	(3,166,088.66)
\$2,500 por llevar a 10,000 interesados con +/- 10% de incertidumbre mensual	\$	828,000.00	\$	952,200.00	\$	1,095,030.00	\$	1,259,284.50	\$	1,448,177.18	\$	1,665,403.75	\$	1,915,214.31	\$	2,202,496.46	\$	2,532,870.93	\$	2,912,801.57
*Pauta no incluida en servicio																				
TOTAL INGRESOS	\$	828,000.00	\$	952,200.00	\$	1,095,030.00	\$	1,259,284.50	\$	1,448,177.18	\$	1,665,403.75	\$	1,915,214.31	\$	2,202,496.46	\$	2,532,870.93	\$	2,912,801.57
*Pauta incluida/ 10% de comision	\$	82,800.00	\$	95,220.00	\$	109,503.00	\$	125,928.45	\$	144,817.72	\$	166,540.38	\$	191,521.43	\$	220,249.65	\$	253,287.09	\$	291,280.16
TOTAL CON COMISION	\$	910,800.00	\$	1,047,420.00	\$	1,204,533.00	\$	1,385,212.95	\$	1,592,994.89	\$	1,831,944.13	\$	2,106,735.75	\$	2,422,746.11	\$	2,786,158.02	\$	3,204,081.73
COSTOS Y GASTOS																				
Costos Fijos																				
Sueldos	\$	271,621.45	\$	312,364.67	\$	359,219.37	\$	413,102.27	\$	475,067.62	\$	546,327.76	\$	628,276.92	\$	722,518.46	\$	830,896.23	\$	955,530.66
Internet	\$	1,200.00	\$	1,380.00	\$	1,587.00	\$	1,825.05	\$	2,098.81	\$	2,413.63	\$	2,775.67	\$	3,192.02	\$	3,670.83	\$	4,221.45
Hosting	\$	1,200.00	\$	1,380.00	\$	1,587.00	\$	1,825.05	\$	2,098.81	\$	2,413.63	\$	2,775.67	\$	3,192.02	\$	3,670.83	\$	4,221.45
Pagina Web	\$	1,200.00	\$	1,380.00	\$	1,587.00	\$	1,825.05	\$	2,098.81	\$	2,413.63	\$	2,775.67	\$	3,192.02	\$	3,670.83	\$	4,221.45
Contabilidad	\$	12,000.00	\$	13,800.00	\$	15,870.00	\$	18,250.50	\$	20,988.08	\$	24,136.29	\$	27,756.73	\$	31,920.24	\$	36,708.27	\$	42,214.52
Oficina	\$	24,000.00	\$	27,600.00	\$	31,740.00	\$	36,501.00	\$	41,976.15	\$	48,272.57	\$	55,513.46	\$	63,840.48	\$	73,416.55	\$	84,429.03
Mantenimiento	\$	1,200.00	\$	1,380.00	\$	1,587.00	\$	1,825.05	\$	2,098.81	\$	2,413.63	\$	2,775.67	\$	3,192.02	\$	3,670.83	\$	4,221.45
Servicios Basicos	\$	3,600.00	\$	4,140.00	\$	4,761.00	\$	5,475.15	\$	6,296.42	\$	7,240.89	\$	8,327.02	\$	9,576.07	\$	11,012.48	\$	12,664.35
Costos Variables																				
Comision de venta	\$	41,400.00	\$	47,610.00	\$	54,751.50	\$	62,964.23	\$	72,408.86	\$	83,270.19	\$	95,760.72	\$	110,124.82	\$	126,643.55	\$	145,640.08
Gasto Publicidad	\$	12,000.00	\$	13,800.00	\$	15,870.00	\$	18,250.50	\$	20,988.08	\$	24,136.29	\$	27,756.73	\$	31,920.24	\$	36,708.27	\$	42,214.52
Capasitaciones	\$	4,000.00	\$	4,600.00	\$	5,290.00	\$	6,083.50	\$	6,996.03	\$	8,045.43	\$	9,252.24	\$	10,640.08	\$	12,236.09	\$	14,071.51
TOTAL COSTOS Y GASTOS	\$	373,421.45	\$	429,434.67	\$	493,849.87	\$	567,927.35	\$	653,116.45	\$	751,083.92	\$	863,746.51	\$	993,308.48	\$	1,142,304.76	\$	1,313,650.47
EBITDA	\$	537,378.55	\$	617,985.33	\$	710,683.13	\$	817,285.60	\$	939,878.44	\$	1,080,860.21	\$	1,242,989.24	\$	1,429,437.62	\$	1,643,853.27	\$	1,890,431.26
Deprecacion de Computadoras	\$	(3,000.00)	\$	(3,450.00)	\$	(3,967.50)	\$	(4,562.63)	\$	(5,247.02)	\$	(6,034.07)	\$	(6,939.18)	\$	(7,980.06)	\$	(9,177.07)	\$	(10,553.63)
EBT	\$	534,378.55	\$	614,535.33	\$	706,715.63	\$	812,722.98	\$	934,631.42	\$	1,074,826.13	\$	1,236,050.06	\$	1,421,457.56	\$	1,634,676.20	\$	1,879,877.63
Impuestos	\$	(90,844.35)	\$	(104,471.01)	\$	(120,141.66)	\$	(138,162.91)	\$	(158,887.34)	\$	(182,720.44)	\$	(210,128.51)	\$	(241,647.79)	\$	(277,894.95)	\$	(319,579.20)
Utilidad	\$	443,534.20	\$	510,064.32	\$	586,573.97	\$	674,560.07	\$	775,744.08	\$	892,105.69	\$	1,025,921.55	\$	1,179,809.78	\$	1,356,781.24	\$	1,560,298.43
INVERSION	\$	(882,000.00)																		
Cashflow																				
Utilidad Neta	\$	443,534.20	\$	510,064.32	\$	586,573.97	\$	674,560.07	\$	775,744.08	\$	892,105.69	\$	1,025,921.55	\$	1,179,809.78	\$	1,356,781.24	\$	1,560,298.43
Depreciacion	\$	(3,000.00)	\$	(3,450.00)	\$	(3,967.50)	\$	(4,562.63)	\$	(5,247.02)	\$	(6,034.07)	\$	(6,939.18)	\$	(7,980.06)	\$	(9,177.07)	\$	(10,553.63)
TOTAL CASHFLOW	\$	446,534.20	\$	513,514.32	\$	590,541.47	\$	679,122.69	\$	780,991.10	\$	898,139.76	\$	1,032,860.73	\$	1,187,789.84	\$	1,365,958.31	\$	1,570,852.06
EBITDA	\$	537,378.55	\$	617,985.33	\$	710,683.13	\$	817,285.60	\$	939,878.44	\$	1,080,860.21	\$	1,242,989.24	\$	1,429,437.62	\$	1,643,853.27	\$	1,890,431.26
CashFlow	\$	446,534.20	\$	513,514.32	\$	590,541.47	\$	679,122.69	\$	780,991.10	\$	898,139.76	\$	1,032,860.73	\$	1,187,789.84	\$	1,365,958.31	\$	1,570,852.06
FreeCashFlow	\$	(882,000.00)	\$	537,378.55	\$	617,985.33	\$	710,683.13	\$	817,285.60	\$	939,878.44	\$	1,080,860.21	\$	1,242,989.24	\$	1,429,437.62	\$	1,643,853.27
TIR																				
VPN	\$																			
WACC																				

Nota: La proyección financiera fue realizadas utilizando como precio base \$2500. Utilizando \$5000 de pauta mensual para alcanzar 30,000 interesados a un costo de \$.17 por interesado.

VII. ANÁLISIS DE RESULTADOS

El objetivo principal de la presente investigación fue proponer un modelo determinístico para la segmentación y perfilado de clientes utilizando análisis de datos y mercadeo digital empleando una base de datos pública. Para lo cual se utilizó la base de datos “German_credit_data.csv” con la finalidad de crear un proceso de perfilado para la optimización del proceso de adquisición de clientes mediante redes sociales. Dicha base de datos fue obtenida de la página Kaggle donde es posible la descarga de base de datos públicas. Luego de encontrar la base de datos adecuada para el análisis, se cargaron los datos en el programa de R studio, como se puede observar en el Anexo D. Se realizó una exploración de datos en la cual se obtuvieron la cantidad de filas y los diferentes tipos de variables dentro de la base de datos utilizando los códigos `str()`, `summary()` y `glimpse()`.

En la exploración de datos se determinaron las relaciones entre variables utilizando los gráficos, histograma, boxplot, scatterplot y gráfico de barras. Con los cuales se estableció la relación entre las variables de más relevancia las cuales fueron edad, género, monto adquirido, duración de crédito (en meses) y su situación laboral, como se puede observar en el Anexo B.

A partir de la información obtenida de la exploración de datos fue posible obtener información relevante para la creación de los distintos modelos para la determinación de segmentos. Determinando las variables relacionadas entre sí. Se obtuvo el modelo de clusterización y de árbol de decisiones utilizando las librerías `tidyR`, `Cluster rpart`, `rpart.plot` y `tidyverse`.

En el primer árbol de decisiones (Figura 4) se determinó que no hay una preferencia del propósito de adquisición del crédito. Mostrándose valores similares en porcentajes para la adquisición de un crédito para un carro, tv o radio y muebles para el hogar.

En la Figura 4, en la sección de Resultados, en la sección de Resultados se observa un árbol de decisiones el cual es utilizado para las posibles decisiones que un cliente potencial podría tomar al momento de decidir adquirir un préstamo mediante probabilidades. Según el árbol de decisiones, dependiendo de la cantidad de crédito a obtener el cliente potencial tomará la decisión de adquirir un préstamo para la adquisición de un carro. Si no coincide con la cantidad de crédito a adquirir el cliente potencial decidirá optar por una televisión. Si la persona es mayor o igual a 22 años decidirá irse por un carro. Pero esto depende si la persona posee una cuenta monetaria. Si la persona no tiene una cuenta monetaria optará por un radio o televisor. Si la persona es mayor o igual a 26 años optará por un carro. Las cifras en los cuadros de decisión representan las probabilidades de que la persona decida adquirir el crédito para un diferente propósito.

En la Figura 5, en la sección de Resultados, se observa un modelo de segmentación mediante clusterización para la comparación de variables en la base de datos. El cual se utiliza para indicar si las variables están adecuadamente asignadas. En el modelo se determinó que los clústeres están claramente identificados. El clúster 1 tiene cualidades parecidas con el clúster 2 al igual que el clúster 3, como se esperaba. Esto indica que los clústeres tienen similitud entre sí.

En la Figura 6, en la sección de Resultados, se muestra el Boxplot de la comparación de edad de las personas que solicitan el crédito para determinar los diferentes segmentos mediante diferentes edades. Con esto se determina que se pueden realizar segmentos por medio de edades. De igual forma, en la Figura 9 se determina que mediante la duración del crédito se puede determinar otro segmento, utilizando duración de crédito y además la edad con la finalidad de crear segmentos de mercado basado en clusters.

En la Figura 11, en la sección de Resultados, se observa que existe una relación entre el clúster 1 y el clúster 2 de la duración del crédito y la edad del cliente. Mientras que en el clúster 3 no existe una relación entre edad y duración del crédito. Asimismo, con la Figura 10 se determinó que existe una relación entre la cantidad del crédito y la edad de las personas mientras que en el clúster 3 no existe una relación entre sí.

La Figura 12, en la sección de Resultados, es un árbol de decisiones basado en los datos obtenidos mediante el clúster. En este modelo se observa que, con los diferentes segmentos separados, el árbol de decisión cambia de cómo se presenta el modelo en la Figura 4. Este modelo permite determinar si el cliente tiene una casa propia y no un alquiler decide adquirir un crédito para un auto. Si el cliente cuenta con una renta, se categoriza mediante la edad. Si su edad es mayor o igual a 23 optará por un auto. Si es mayor, optará por decoraciones para su hogar. Si menor o igual a 23, el árbol decide verificar si tiene cuenta monetaria, si cuenta con ella entonces optará por un auto, y si no cuenta con una cuenta monetaria entonces optará por un radio o televisor. Todas las decisiones se basan entre los diferentes propósitos con sus diferentes probabilidades de decisión.

En la Figura 13, en la sección de Resultados, se observa un modelo de segmentación mediante clusterización para la comparación de variables en la base de datos. El cual se utiliza para indicar si las variables en el modelo comparten propiedades, determinando que los clústeres comparten propiedades entre sí. Con el clúster 1, 2 y 3, se puede relacionar que, los clientes que tienen cierta edad y si cuentan con trabajo, usualmente están interesados en un carro.

En la Figura 14, se realizó un boxplot para mostrar un resumen entre 3 clústeres de segmentación para la comparación de la edad. En la primera caja de bigotes se observa una asimetría hacia la derecha, indicando que los clientes dentro de esta son de mayor edad. En este no se presentan datos atípicos. En el segundo segmento se observa una asimetría a la derecha, con una media con una buena distribución entre los datos.

Se observa un dato atípico, que a pesar de ser los que pertenecen a una edad media de igual manera existen personas de mayor edad dentro del segmento 2. En el segmento 3 se observa que la mayoría es de menor edad, siendo una gráfica con asimetría a la derecha y con menor desviación estándar.

En la Figura 15 se observa que existen más datos atípicos entre los segmentos. En el segmento 1, existe una asimetría hacia la derecha, indicando que estos clientes optan por adquirir un crédito por menor valor. Mientras que otros datos atípicos indican que hay clientes dispuestos a optar por un crédito de mayor valor. En el segundo clúster se observa que la gráfica es asimétrica, en donde se presenta una mayor desviación estándar a comparación de los demás segmentos, indicando que las personas del segundo segmento están dispuestas a agarrar un crédito de mayor valor. De igual manera se pueden observar datos atípicos en esta, indicando que estos clientes están dispuestos a adquirir un crédito de mayor valor. En el segmento 3 se observa una asimetría a la derecha, un segmento similar al segmento 1, en donde se observa que las personas se comportan de igual manera que en el segmento 1.

En la Figura 16, se presenta la relación entre la duración del préstamo y entre los segmentos de clúster. Se observa que en el segmento 1, la mayoría prefiere una duración de crédito menor, debido a que hay una asimetría hacia la derecha. Estos están dispuestos a optar por un crédito por menor tiempo. En el segmento 2 podemos observar que la mayoría de los clientes están dispuestos a optar por un crédito por mayor tiempo, aunque existen datos atípicos entre menor tiempo y mayor tiempo de adquisición de crédito. En el clúster 3 observamos un comportamiento similar al segmento 1, en donde se observa que la mayoría de las personas está dispuesta a una media de menor tiempo.

En la Figura 17, se realizó un scatterplot para la determinación de la correlación entre las variables duración y edad para los diferentes segmentos. En el segmento 1 se determinó que las personas de mayor edad prefieren adquirir un crédito por mayor duración creando una relación entre sí. En el segmento dos observamos que las personas de menor y mayor edad prefieren adquirir un crédito por mayor duración sin importar la edad. Mientras que para el segmento 3 no existe una correlación de las variables duración y edad entre los clusters.

De igual manera, en la Figura 18, se realizó un scatterplot para determinar la correlación entre las variables total de crédito y la edad entre los segmentos. En el segmento 1 se observa que existe una relación entre la edad y el total de crédito adquirido. Determinando que entre mayor la edad menor es la adquisición total del crédito. En el segmento 2 se observa que entre menor la edad mayor la cantidad total de adquisición de crédito, en donde se observa una relación entre si. En el segmento 3, se observa que las edades entre 20y 40 prefieren adquirir un crédito de menor cantidad.

La Figura 19 presenta relación con la Figura 18, demostrando que las personas de mayor edad prefieren un crédito de menor tiempo y de mayor valor, mientras que en el segmento dos se observa que no existe relación entre la duración y la cantidad total de crédito adquirido, ya que estos clientes prefieren un crédito de mayor valor. En la Figura 19 se realizó un scatterplot para determinar si los segmentos están relacionados entre las variables duración y total de crédito adquirido. Se observa que las personas del segmento 1 tienen una menor duración de crédito por una cantidad menor de tiempo. Mientras que, en el segmento 2 no existe la relación entre las variables. En el segmento 3 observamos que, si existe una relación entre las variables, la duración del crédito es menor al igual que la cantidad de crédito adquirido.

En la Figura 19 se presenta que el segmento 1 prefiere una menor duración con un menor crédito siendo estos los de mayor edad, de acuerdo con lo determinado en la Figura 18. Con la Figura 20 se determinó que este segmento prefiere adquirir un crédito para un carro de menor valor, una radio o muebles para su hogar. En la Figura 18 se observa que las personas de menor edad prefieren adquirir un crédito mayor con cualquier duración del crédito, siempre escogiendo un carro, radio o tv, o muebles para el hogar siempre teniendo en cuenta que el valor del crédito es mayor para este segmento. El segmento 3 es el que más adquiere préstamos, con una edad menor al igual que la duración y el total del crédito adquirido, este segmento prefiere adquirir un carro, una radio o tv, o muebles para su hogar.

Por otro lado, se determinó la relación entre el total de préstamos y el género de los clientes, presentado en la Figura 21. De esta manera se determinó que en el segmento 1 la mayoría de los clientes son hombres, los cuales adquieren un carro; mientras que las mujeres prefieren adquirir un crédito para vacaciones, como se puede observar en la Figura 42 en la parte de anexos. En el segmento 2 podemos observar que, de igual manera, la mayoría de los hombres prefiere adquirir un crédito. En la Figura 42 se observa que la mayoría de los hombres prefiere adquirir un crédito para un carro o para su negocio. Mientras que en el segmento 3 de la Figura 21 se muestra que es donde se encuentra la mayoría de los clientes hombres o mujeres indicando que la mayoría de los hombres y mujeres que adquieren un préstamo residen dentro de este segmento, el cual como vimos en la Figura 18 son de menor edad. En la Figura 20 vemos que adquieren un crédito para un auto de igual manera que un crédito para un radio o una tv.

Para determinar la situación laboral de los clientes se realizó el gráfico presentado en la Figura 22. En este caso podemos observar un comportamiento parecido para la variable de situación laboral. En donde la mayoría de las personas cuentan con un trabajo promedio adquiriendo una cantidad de préstamos similar por segmento. En el segmento 1 se observa que la mayoría de las personas cuentan con un trabajo que requiere menor educación. Mientras que en el segmento dos se puede observar que hay personas que tienen un trabajo que requiere de educación elevada. Por otro lado, en el segmento 3 podemos observar que existen clientes con un trabajo de una calidad mucho menor de los cuales adquieren una mayor cantidad de préstamos.

En la Figura 23, en la parte de resultados, se realizó un gráfico para determinar la situación de vivienda por segmento. En el segmento 1 podemos observar que la mayoría cuenta con una propiedad, mientras que la segunda mayoría no cuenta con hogar ni con renta. En el segmento 2 podemos observar que la mayoría cuenta con un hogar, pero al igual que en el primer segmento no cuentan con un hogar. En segmento 3 podemos observar que la mayoría de los clientes cuenta con un hogar propio, mientras que en este caso existen clientes que cuentan con alquiler de hogar.

La Figura 24, en la sección de Resultados, es un árbol de decisiones basado en los datos obtenidos mediante el clúster 2 de caracteres. En este modelo se observa que, con los diferentes segmentos separados, el árbol de decisión cambia de cómo se presenta el modelo en las Figuras 12 y 4. Este modelo permite determinar las decisiones basadas en diferentes variables utilizadas en la base de datos. Determinando que si la persona cuenta con un crédito mayor o igual a 4600 esta persona optará directamente por adquirir un auto. Si no cuenta con esta cantidad de crédito la persona optará directamente por evaluar si cuenta con una cantidad de crédito mayor o igual a 1419, si cuenta con esta cantidad de crédito se volverá a evaluar si la persona cuenta con una capacidad de crédito mayor o igual a 600; mientras que si no cuenta con esta cantidad optará por la compra de un televisor. Si cuenta con esta cantidad se evaluará si la persona decide adquirir un crédito por una duración igual o mayor a 7 meses. Si cuenta con este requisito optará directamente por adquirir un crédito para el automóvil. Si no cuenta con este requisito verificará si cuente con un requisito mayor o igual 10 meses, si no cuenta con este requisito la persona adquirirá un crédito para la compra de una radio o tv. Si cuenta con este requisito se verificará si la persona quisiera adquirir el crédito por un plazo mayor o igual a 13 meses. Si cumplen con el requisito se verificará la edad, si esta es mayor o igual a 26 años entonces comprará el auto, si no lo cumple optará por una radio o tv. Con respecto a la rama del árbol de duración, menor o igual a 13 meses se determinará la edad de la persona. Si esta persona es menor a 38 años se verificará si la duración del crédito es mayor o igual a 21 meses, si es mayor, la persona adquiriera un crédito para un auto, si no cumple la persona adquirirá un préstamo para un televisor. De igual manera, se pueden observar todas las diferentes decisiones que podría tomar un cliente mediante diferentes factores para los diferentes propósitos existentes dentro de la base de datos.

Mediante los datos obtenidos en el análisis exploratorio y en los modelos se realizó buyer personas para determinar un cliente potencial para la empresa. En la Figura 25 se observa el perfilado del cliente Gabriela Bardales, la cual es una catedrática de 38 años que decide adquirir un crédito para adquirir un televisor. Debido a que mediante los árboles de decisiones y los clústeres se determinó que el segmento dos prefiere adquirir un crédito para comprar un televisor, por lo que se decidió realizar un buyer persona basado en dichos datos, con el objetivo de atraer clientes potenciales para la adquisición de un televisor mediante un crédito. Se determinaron diversos intereses los cuales podrían ser utilizados en las campañas digitales mediante Google ads y Facebook ads.

En la Figura 26 se realizó un buyer persona llamado Carlos Villanueva, basado en los árboles de decisión y los segmentos obtenidos por clúster de los cuales determinaron que era más probable que una persona adquiriera un auto utilizando un crédito. En el buyer persona se determinó la edad, los posibles lugares de residencia y diversos intereses los cuales pueden ser utilizados para la creación de campañas digitales para que la empresa financiera pueda adquirir clientes con el modelo realizado mediante un árbol de decisiones y los clústeres. Tomando como referencia el segmento 2.

En la Figura 27, de igual manera se realizó un buyer persona llamado Joselin Alvizures, basado en los árboles de decisión y en los segmentos obtenidos por medio de clusterización de los cuales determinaron que era más probable que una persona adquiriera muebles para su hogar utilizando un crédito. En el buyer persona se determinó la edad, los posibles lugares de residencia y diversos intereses los cuales pueden ser utilizados para la creación de campañas digitales para que la empresa financiera pueda adquirir clientes con el modelo realizado mediante un árbol de decisiones y los clústeres. Tomando como referencia el segmento 3.

A partir de los buyer persona obtenidos del modelo determinístico se pueden crear campañas publicitarias por medio de Google Ads y Facebook ads en el apartado de segmentación. Creando una campaña de clientes potenciales en Facebook y en Google ads de smart ads con las características del buyer persona determinado, en el apartado de segmentación en ambas plataformas. A partir de las campañas es posible la obtención de un archivo descargable con la información de los clientes potenciales (como se puede observar en el Anexo E en las Figuras 46 y 47). Con estas campañas se obtendrán datos de clientes en relación con los modelos realizados.

Para determinar la viabilidad de la implementación del modelo determinístico para su aplicación en empresas financieras, se realizó una evaluación financiera con una proyección de ventas. Con este análisis se determinó que dicho modelo es viable para la implementación de un perfilado más exacto mediante modelos de árbol de decisión y de clusterización. Se crearon 3 paquetes viables para la implementación del modelo en empresas medianas y grandes. El cual tiene un costo mínimo de \$7,500 mensuales con el objetivo de encontrar variables necesarias dentro de una empresa para la implementación del modelo. Utilizando \$5,000 de pauta digital con un 3% de porcentaje de adquisición de clientes con el objetivo de obtener 1,000,000 de impresiones diarias para alcanzar al cliente potencial y que este adquiriera un crédito. Con un costo de \$0.17 por cliente adquirido con la meta de alcanzar al menos 30,000 clientes potenciales durante el mes que adquieran el producto o servicio de la empresa.

De esta manera se determinó que el modelo determinístico propuesto en la presente investigación es eficaz y viable para la segmentación y perfilado de clientes al utilizar análisis de datos y mercadeo digital. Pudiendo crear buyer personas de clientes potenciales para las empresas financieras a partir del modelo

presentado. Por lo que se recomienda que para la aplicación de este modelo se tenga una base de datos de al menos 1000 datos y 10 variables.

El modelo determinístico determinado en la presente investigación es aplicable en los diferentes mercados solamente si las empresas recolectan datos sobre su cliente. Con el objetivo principal de comercializar sus productos o servicios en línea, aplicándolo dentro del ámbito de mercadeo digital.

VIII. CONCLUSIONES

2. Se determinó que las variables más significativas de la base de datos pública analizada son el género del cliente, la duración, el propósito y el monto total del crédito, obtenidos a partir del árbol de decisiones y el modelo de segmentación por clusterización.
3. Según las variables determinadas con el árbol de decisión y el modelo de clusterización, los clientes potenciales prefieren adquirir un crédito para la compra de un auto dependiendo de la edad, la cuenta bancaria y si es dueño de un hogar o alquila vivienda.
4. Se determinó que el modelo determinístico es eficaz con la creación de buyer persona utilizando el árbol de decisión y el modelo de clusterización para un perfilado de cliente más preciso.
5. Se determinó que el modelo determinístico presentado en la investigación es viable para la implementación de un perfilado más exacto mediante modelos de árbol de decisión y clusterización, teniendo un costo de \$7,500 mensuales, utilizando \$5,000 de pauta digital con un 3% de porcentaje de adquisición de clientes para la obtención de 1,000,000 de impresiones diarias.
6. Mediante clusterización se pueden determinar segmentos de mercado para la clasificación de nivel socio económico de un cliente.

IX. RECOMENDACIONES

1. Realizar estudio de perfilado de clientes con una identidad financiera con datos recolectados en la ciudad de Guatemala para tener un dato más exacto del país al cual se está tratando de estudiar.
2. Realizar pruebas en Google Ads y Facebook ads para obtener resultados de buyer personas realizados en el análisis.
3. Realizar un análisis con una base de datos con mayor cantidad de filas para hacer más eficiente el perfilado.
4. Realizar artes para determinar si el mensaje de mercadeo realizado en los buyer persona tiene relevancia.
5. Crear más de dos perfiles utilizando más arboles de decisión para determinar más clientes potenciales en las redes sociales.
6. Aplicar el modelo en la carrera de International Marketing and business analytics para poner en práctica lo aprendido durante la carrera.

X. BIBLIOGRAFÍA

- Abuin, N., Dias, J., & García, D. (2020, July 21). *Neuromarketing as an Emotional Connection Tool Between Organizations and Audiences in Social Networks. A Theoretical Review*. Retrieved from Frontiers in Psychology : <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01787/full>
- Anderson, S. W. (2015). *Estadística para los negocios y economía 12a edición*. Mexico: CENAGE LEARNING .
- Bermudez, P. (2018, Febero 28). *La segmentación en marketing digital*. Retrieved from Pedro Bermudez : <https://pedrobermudeztaavera.com/segmentacion-en-marketing-digital/>
- Borcio, K., Calle, J., & Zambrano, M. (2018, Julio 1). *EL MARKETING DIGITAL*. Retrieved from Scielo SLD: <http://scielo.sld.cu/pdf/rus/v10n4/2218-3620-rus-10-04-103.pdf>
- Carasila Coca; Milton Andrés. (2010). *NEUROMARKETING: Las emociones y el comportamiento de compra perspectiva*. Bolivia: Universidad Católica Boliviana San Pablo Retrieved from PERSPECTIVAS: <https://www.redalyc.org/pdf/4259/425942454003.pdf>
- Hernández, S. (2010). *Población y Muestra*. Hidalgo: Universidad Autónoma del Estado de Hidalgo
- Johnson, R., & Kubys, P. (2012). *Estadística Elemental*. México D.F.: Cengage Learning.
- Juste, M. “La Pandemia Dispara El Uso De Las Redes Sociales, Un 27% m.” *Expansi, Expansión*, 9 Feb. 2021, <https://www.expansion.com/economia-digital/innovacion/2021/02/10/6022c89de5fdea59448b459b.html>.
- Lizarzabury, L. (2013). Uso de la estadística en trabajos de investigación en la Universidad San Pedro. *Memorias del II congreso de Investigación, Ciencia y Tecnología*, 935-939.
- McLachlan, Stacey. “How to Create a Buyer Persona (Free Buyer/Audience Persona Template).” *Social Media Marketing & Management Dashboard*, 9 Nov. 2021, <https://blog.hootsuite.com/buyer-persona/>.
- Neetesuola. (2013, June 25). ¿Que es un modelo? Neetesuola. Retrieved June 20, 2022, from <https://neetesuola.org/que-es-un-modelo/>
- Plantillas de buyer persona (cliente ideal) Editables Online. EDIT.org in english. (2018, June 1). Retrieved June 20, 2022, from <https://edit.org/es/blog/plantillas-buyer-persona-cliente-ideal>
- Ríos, S. (1995). *Modelización*. Alianza Universidad.
- Rodrigo, Joaquín Amat. “Test Estadísticos Para Variables Cualitativas: Test Exacto De Fisher, Chi-Cuadrado De Pearson, McNemar Y Q-Cochran.” *Test Estadísticos Para Variables Cualitativas: Test Exacto De Fisher, Chi-Cuadrado De Pearson, McNemar y Q-Cochran*, https://www.cienciadedatos.net/documentos/22.2_test_exacto_de_fisher_chi-cuadrado_de_pearson_mcnemar_qcochran.

Saraduy, Y. (2007, Septiembre 01). *El análisis de información y las investigaciones cuantitativa y cualitativa*. Retrieved from SciELO: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-34662007000300020

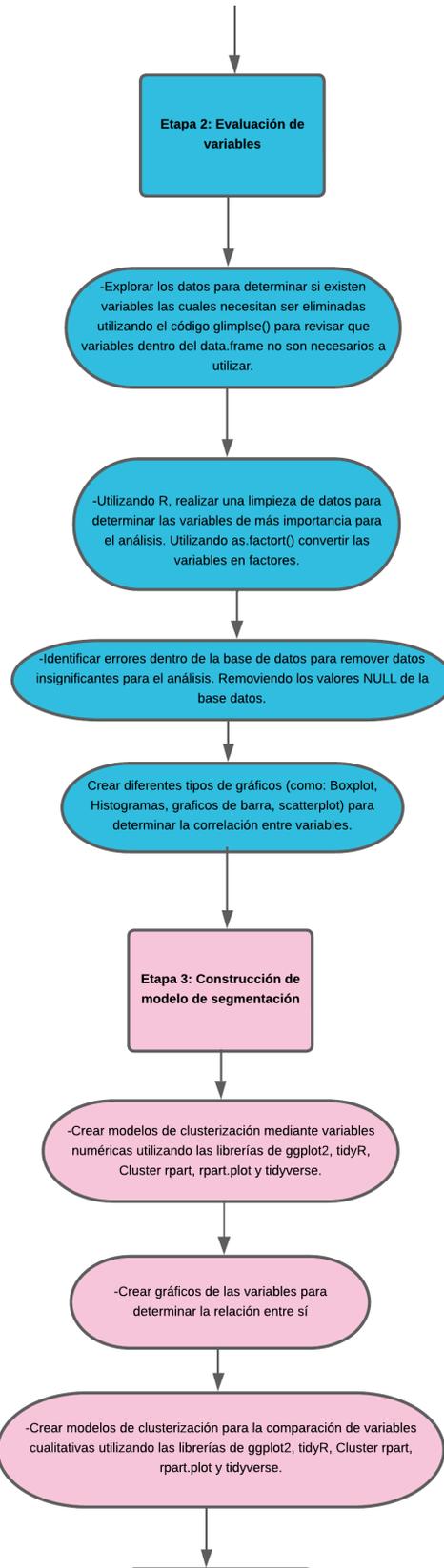
Shiffman, L., & Lazar, L. (2005). *Comportamiento del consumidor*. Mexico: Pearson Educación .

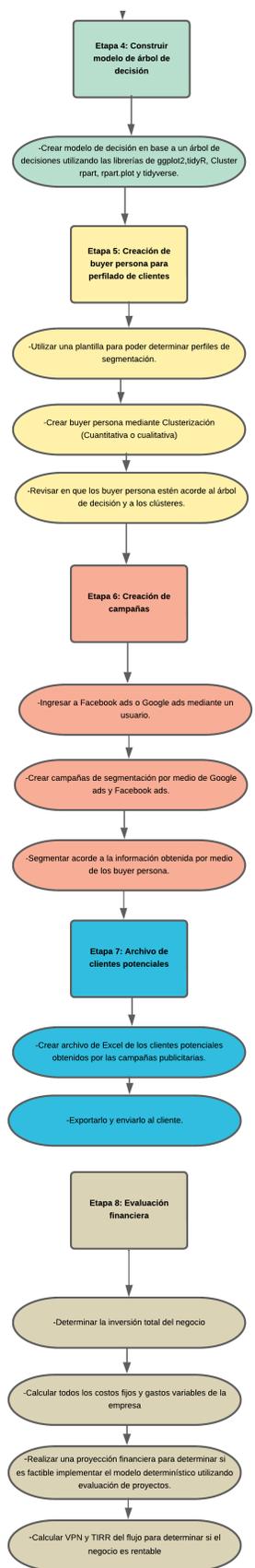
UCI, M. (2016, March 12). Credit Risk - Germ German Credit Risk. Kaggle. Retrieved June 20, 2022, from <https://www.kaggle.com/datasets/uciml/german-credit>

XI. ANEXO

A. Diagrama de flujo metodología

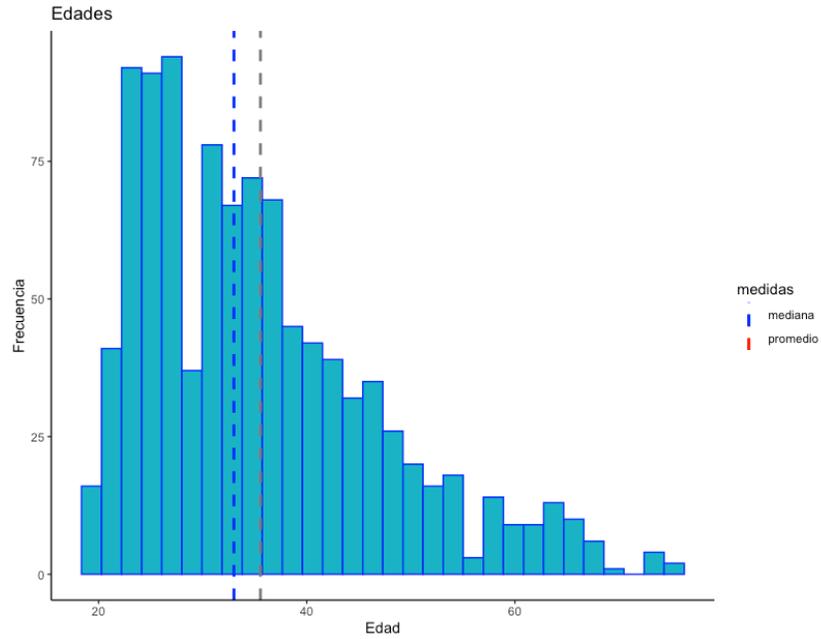






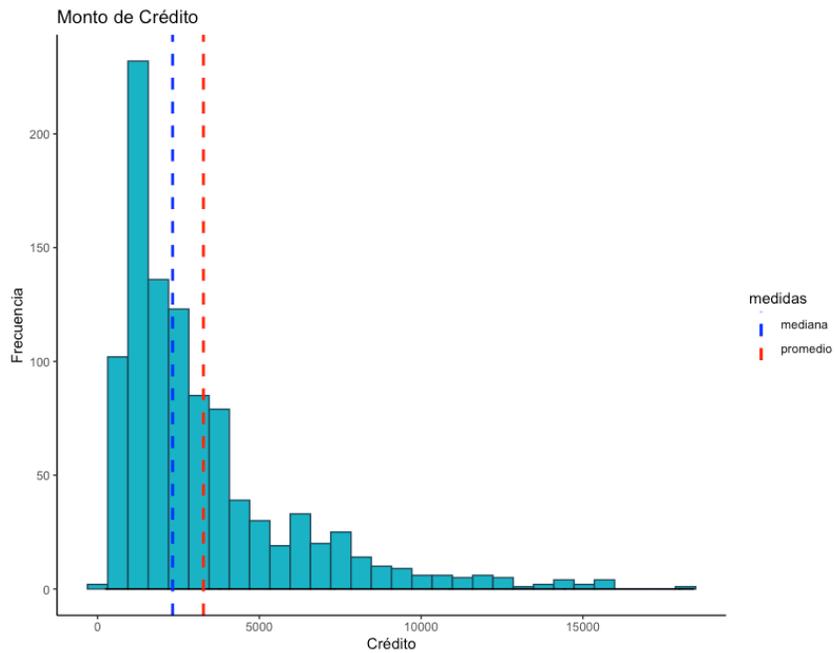
B. Análisis exploratorio

Figura 28. Histograma de distribución de edades de la base de datos analizada.



Nota: El histograma se realizó con la base de datos “German_credit_data.csv”.

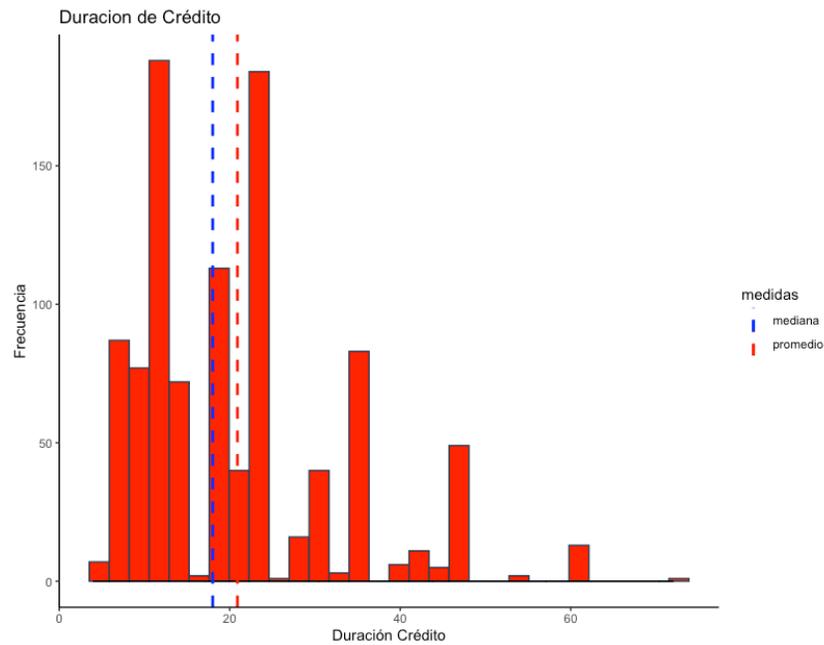
Figura 29. Histograma de la frecuencia del monto del crédito de la base de datos analizada.



Nota: El histograma se realizó con la base de datos “German_credit_data.csv”.

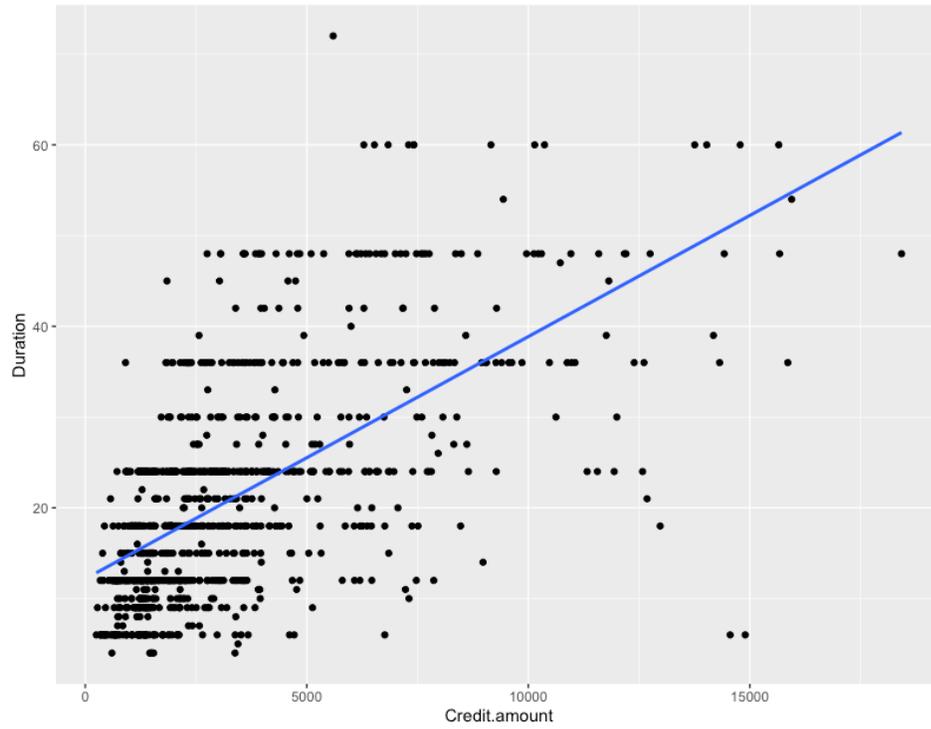
La Figura 28 muestra el histograma de las variables de edad en donde se observó que se presenta un sesgo a la derecha, lo cual demuestra que hay valores más separados de la media y que el rango de edad es de 25 a 35 años. A partir de las Figuras 28, 29 y 30 se determinó entre menor es la edad de la persona, menor es el monto total del crédito y menor es la duración del crédito. Como se puede observar en la Figura 32 se comprueba que si existe una relación entre el total de crédito adquirido por el cliente y la duración de este. Sin embargo, en la Figura 33 se puede observar que no existe una relación entre la edad del cliente y el monto del crédito lo cual indica que no existe una relevancia entre dichas variables. De igual forma, se observa que en la Figura 32 que la edad no es una variable relevante en relación con la duración total del crédito.

Figura 31. Histograma de frecuencia de duración de crédito de la base de datos analizada.



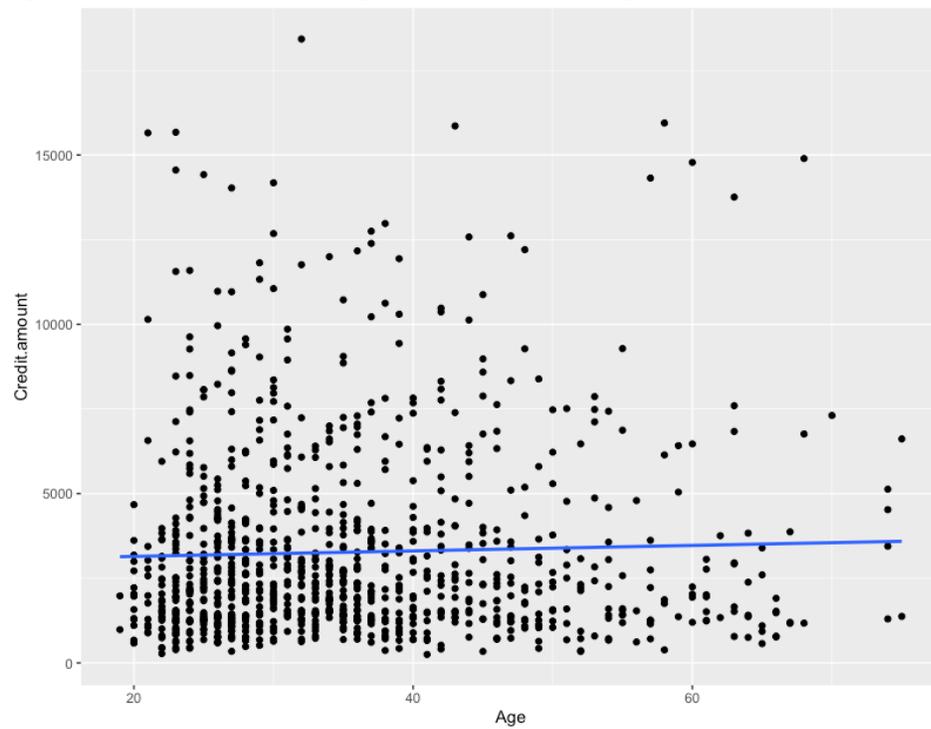
Nota: El histograma se realizó con la base de datos "German_credit_data.csv".

Figura 32. Relación entre el total de crédito y la duración de este según la base de datos analizada.



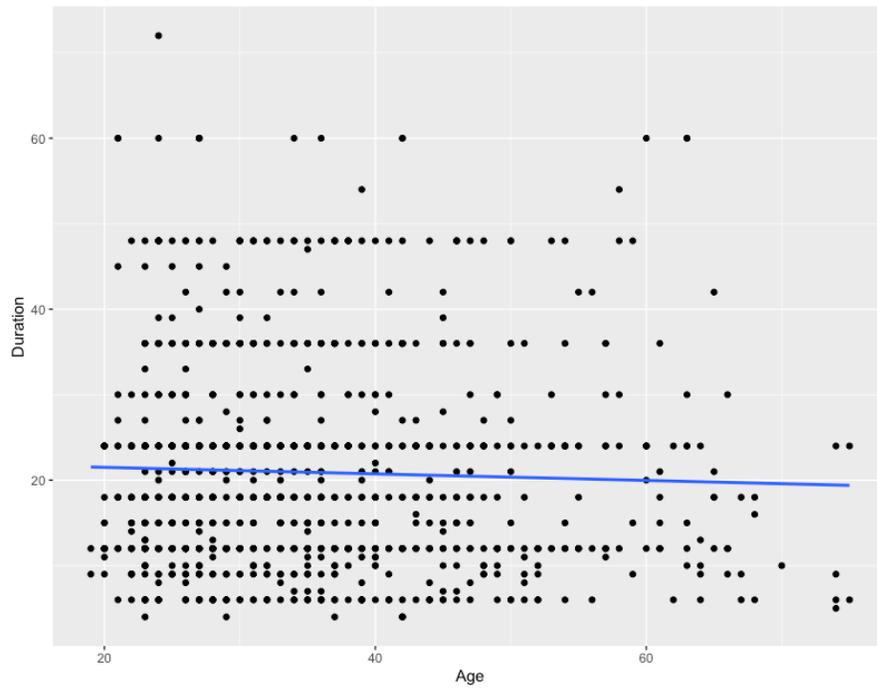
Nota: La Figura se obtuvo con la base de datos "German_credit_data.csv".

Figura 33. Relación entre la edad y el monto del crédito según la base de datos analizada.



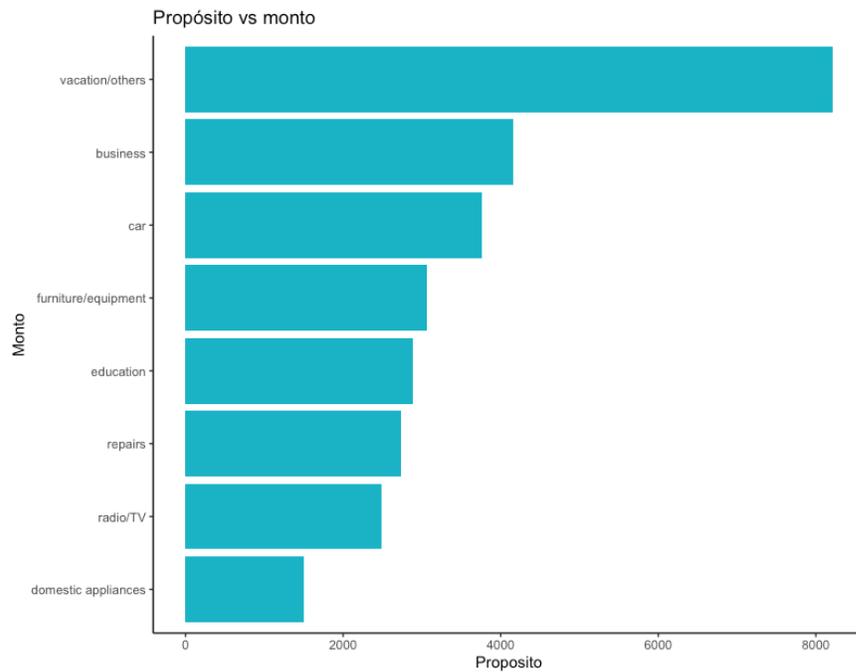
Nota: La Figura se obtuvo con la base de datos "German_credit_data.csv".

Figura 34. Relación entre la edad y la duración del crédito de la base de datos analizada.



Nota: La Figura se obtuvo con la base de datos “German_credit_data.csv”.

Figura 35. Gráfica de barras de frecuencia del propósito de la adquisición de la deuda con el monto total de esta.



En la Figura 34 se observa el propósito de la adquisición del crédito y el monto total de este determinando que las personas adquieren un crédito mayor por actividades de ocio. Por otro lado, en la Figura 35 se presenta la cantidad de personas que adquieren un crédito según el propósito de este, donde se puede observar que una mayor cantidad de personas adquieren el crédito para productos de uso diario y de vivienda.

Figura 35. Gráfica de barras de frecuencia del propósito de la adquisición del crédito con el monto promedio de esta.

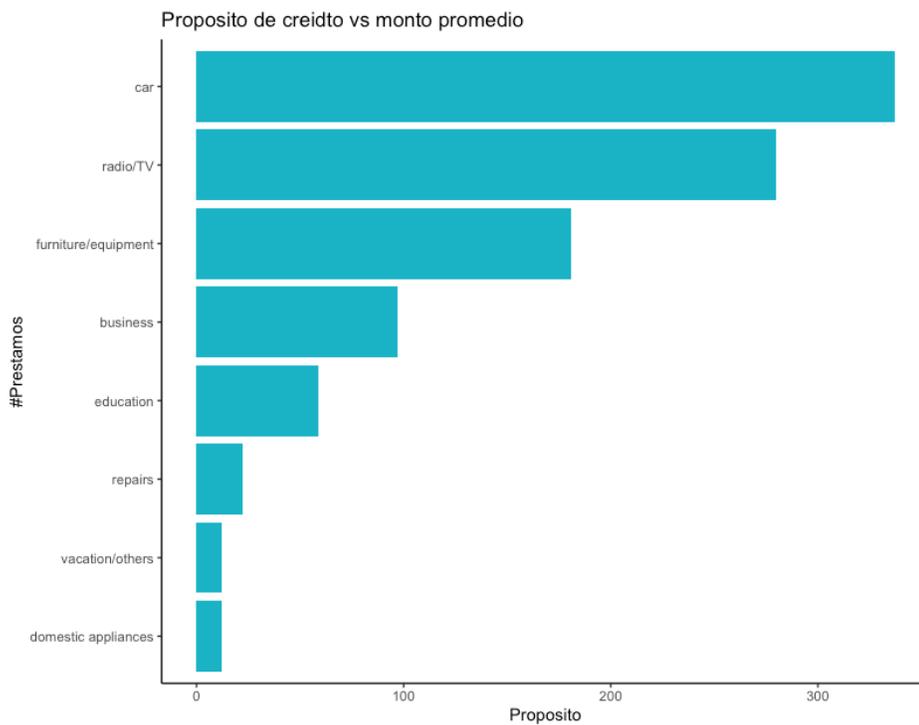


Figura 36. Boxplot de la distribución entre edad y género.

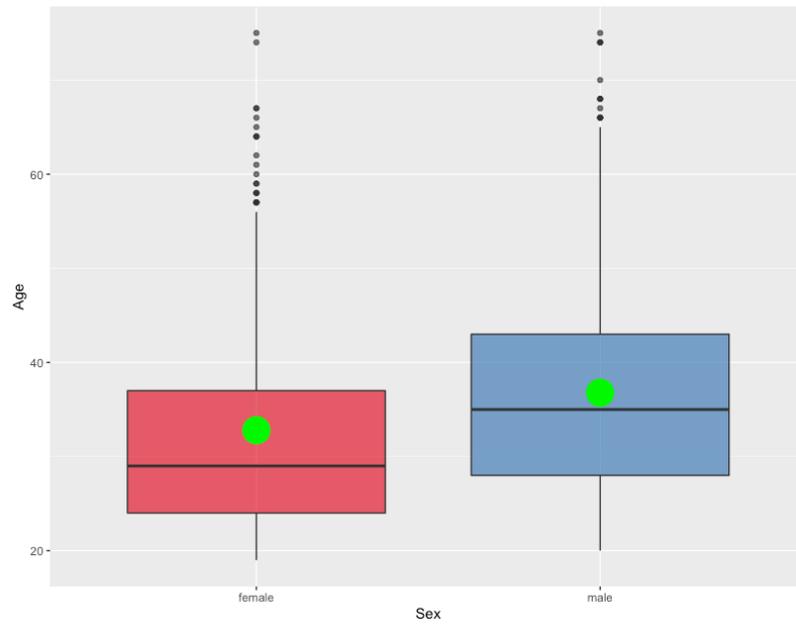


Figura 37. Boxplot de la distribución entre el género y el monto del crédito.

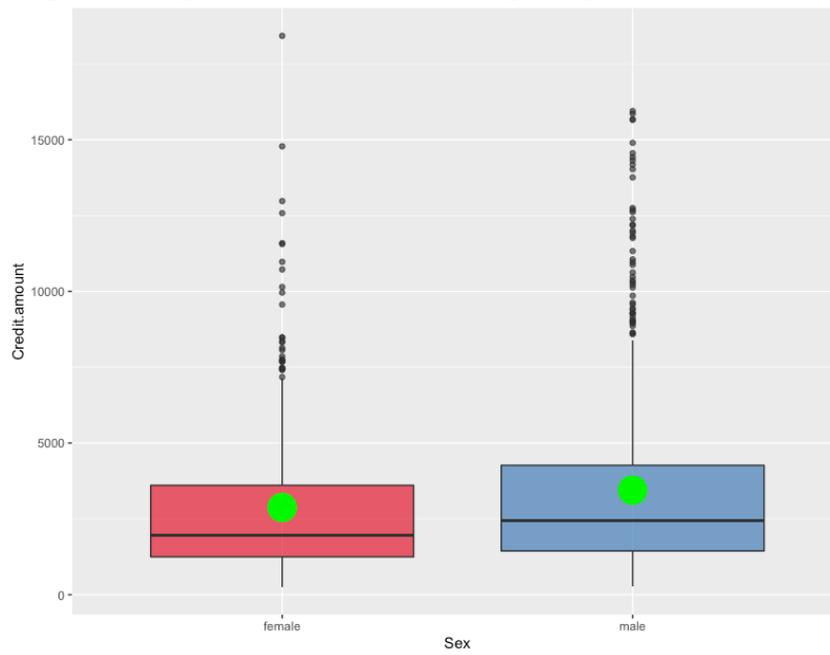
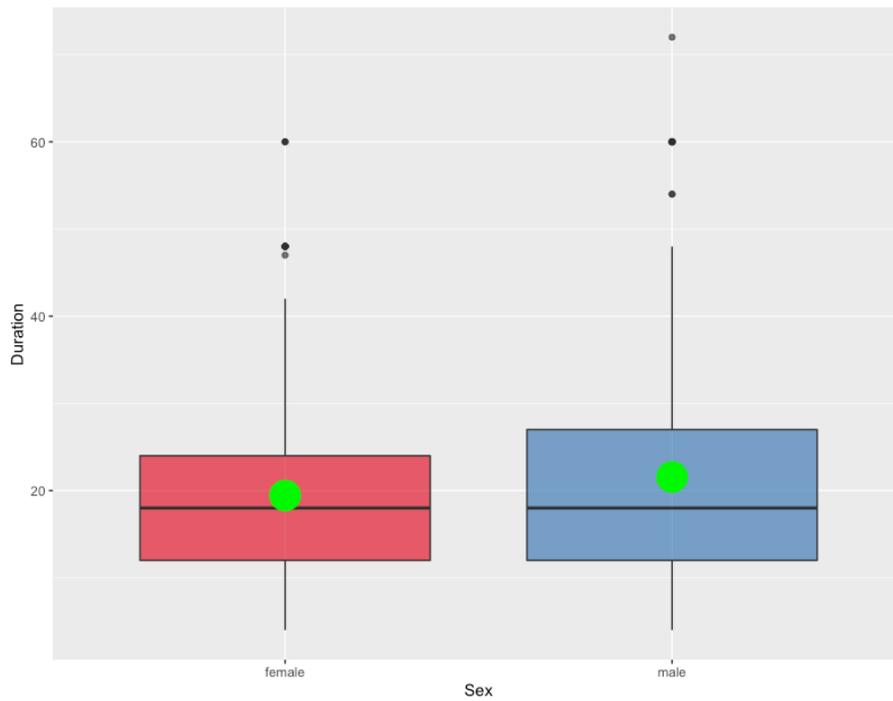


Figura 38. Boxplot de la distribución entre género y duración del crédito.



A partir de la Figura 36 se determinó que las mujeres de menor edad adquieren más créditos a comparación de las mujeres de mayor edad; mientras que en promedio los hombres de mayor edad son los que adquieren más créditos. Con la Figura 37 fue posible determinar que los hombres adquieren créditos de mayor monto, comparado con las mujeres. Además, en la Figura 38 se observa que no hay diferencia de promedio de duración de crédito según el género.

Figura 39. Relación entre el monto de crédito y la duración de este según el género.

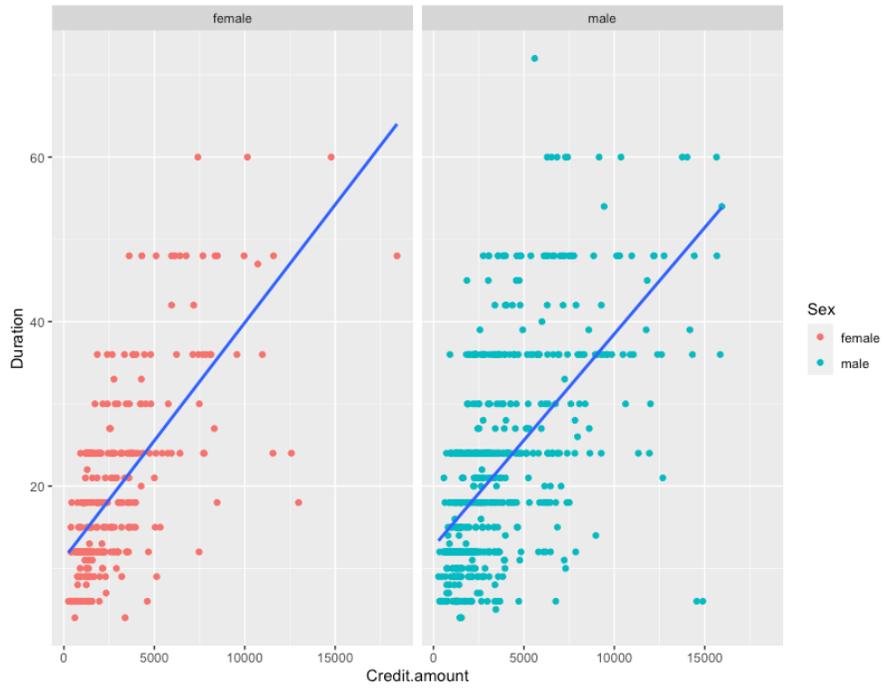


Figura 40. Relación entre la edad y el monto del crédito según el género.

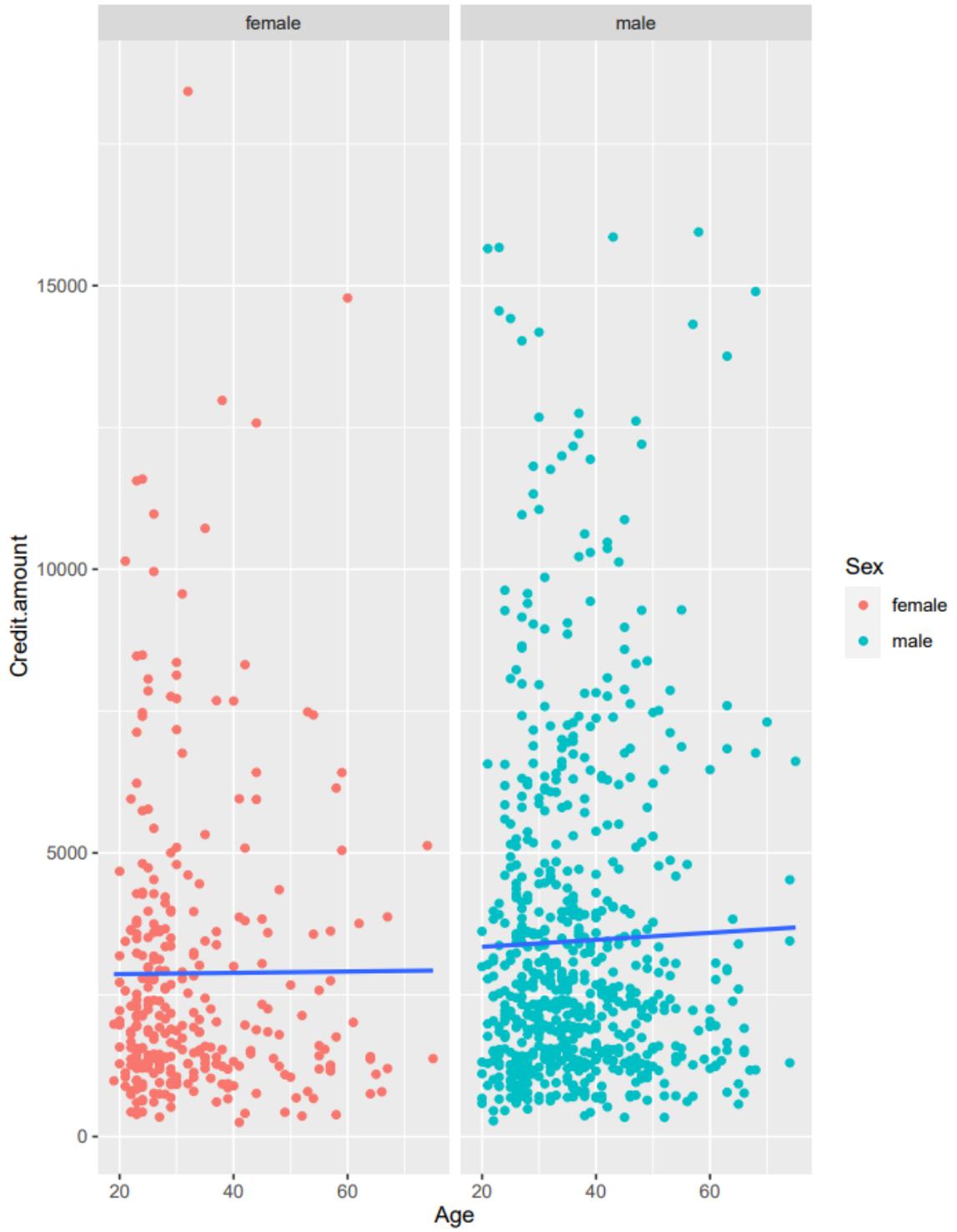


Figura 41. Relación entre la edad y la duración del crédito según el género.

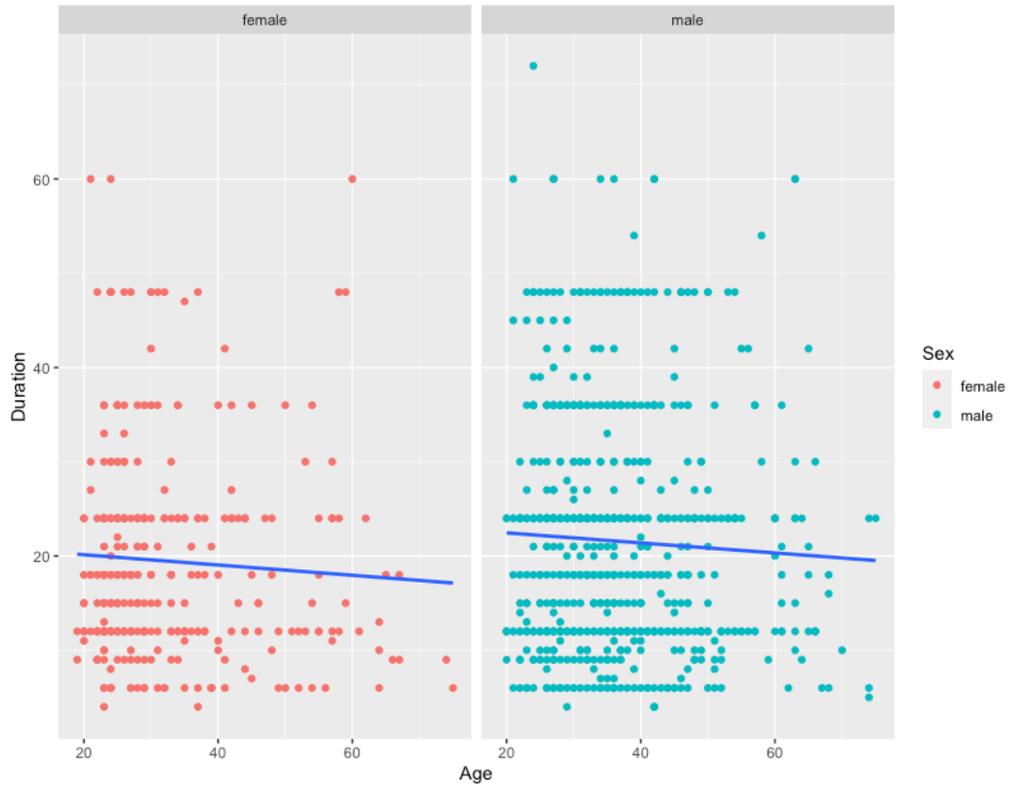


Figura 42. Gráfico de barras de la relación entre el propósito del crédito y el monto de este según el género.

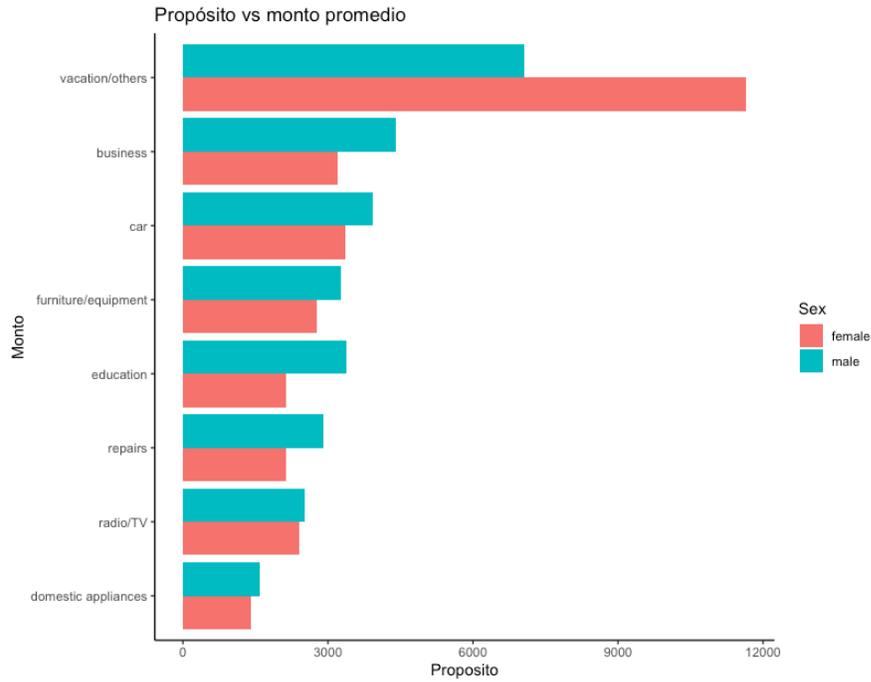
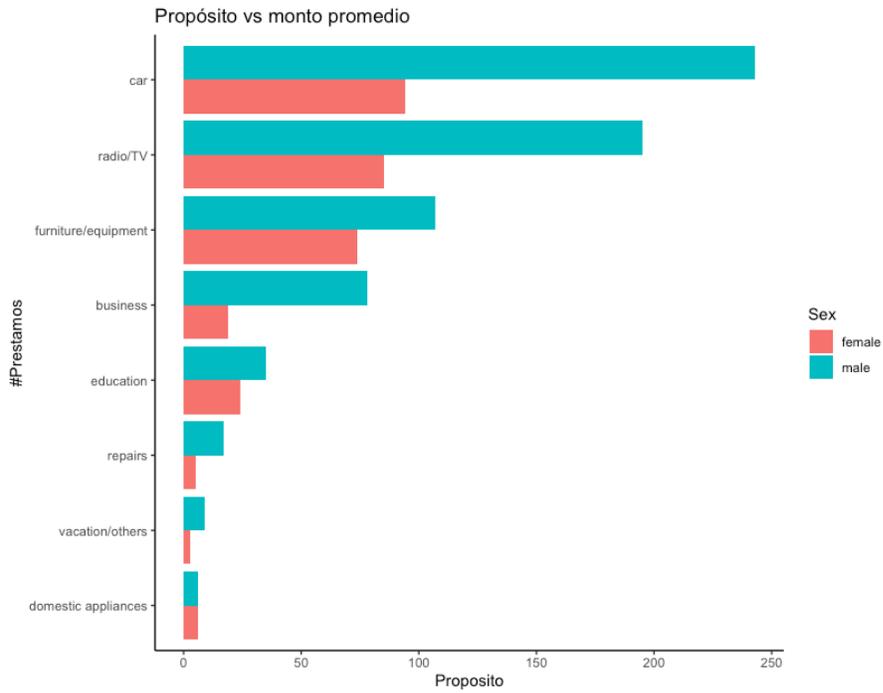


Figura 43. Gráfico de barras de la frecuencia del propósito del crédito y el monto de este según el género.



En la Figura 38 se observa la relación entre el monto del crédito y la duración de este según el género determinando que sí existe una relación entre dichas variables, siendo mayor la duración mientras mayor sea el monto del crédito indistintamente del género. Mientras, que en la Figura 40 se observó que no existe una relación entre la edad del cliente y el monto del crédito según el género. De igual manera, se presentó el mismo comportamiento en la Figura 41 demostrando que no hay relación entre la edad del cliente y la duración del crédito según el género. Por otro lado, en la Figura 42 se observa que las mujeres prefieren obtener un crédito para actividades de ocio, mientras que los hombres prefieren obtener un crédito para adquirir un carro, para sus hogares o para sus empresas. Además, en la Figura 43 se observa que en promedio los hombres adquieren más créditos que las mujeres.

— Por qué

Comentarios

Testimonios sobre retos y objetivos

Áreas de oportunidad

Argumentos para no mantener un relación comercial con nosotros

8

- Testimonio sobre retos
- Testimonio sobre objetivos

9

- Razones para no adquirir nuestro producto o servicio
- Razones para no adquirir nuestro producto o servicio

Nombre del buyer persona



— Cómo

Mensaje de marketing

Respuesta a la problemática del cliente

Mensaje de ventas

Respuesta de ventas para llegar al cliente

10

- Mensaje de marketing

11

- Mensaje de ventas

Nombre del buyer persona



D. Código realizado en R studio

```
#Trabajo de Graduacion
#Raul Sanchez Ippisch
#Titulo: PROPUESTA DE MODELO DETERMINÍSTICO PARA LA SEGMENTACIÓN DE CLIENTES
Y PERFILADO UTILIZANDO ANÁLISIS DE DATOS Y MERCADEO DIGITAL

library(dplyr)
library(ggplot2)
library(tidyr)
library(purrr)
library(cluster)
library(weights)
library(rpart)
library(rpart.plot)
library(tidyverse)
library(ggdendro)

setwd("~/Desktop/Base de datos Nueva tesis")

tesis <- read.csv("german_credit_data.csv")

#Exploracion de Datos

str(tesis)
summary(tesis)
glimpse(tesis)

#distribucion de edades

Edades_Dist = tesis %>%
  mutate(promedio = mean(Age),
         mediana = median(Age)) %>%
  ggplot(aes(x=Age))+
  geom_histogram(color="blue", fill="#19B6C6")+
  geom_vline(aes(xintercept=promedio, color = "Promedio"),
            linetype="dashed", size = 1)+
  geom_vline(aes(xintercept=mediana, color="mediana"),
            linetype="dashed", size = 1)+
  labs(title="Edades",x="Edad", y = "Frecuencia")+
  theme_classic()+
  scale_color_manual(name = "medidas", values = c(mediana = "blue", promedio
= "red"))

Edades_Dist

#Monto de credito

Dist_credit = tesis %>%
  mutate(promedio = mean(Credit.amount),
         mediana = median(Credit.amount)) %>%
```

```

ggplot(aes(x=Credit.amount))+
geom_histogram(color="#254A5B", fill="#19B6C6")+
geom_density(alpha=0.6)+
geom_vline(aes(xintercept=promedio, color = "promedio"),
           linetype="dashed", size = 1)+
geom_vline(aes(xintercept=mediana, color="mediana"),
           linetype="dashed", size = 1)+
labs(title="Monto de Crédito",x="Crédito", y = "Frecuencia")+
theme_classic()+
scale_color_manual(name = "medidas", values = c(mediana = "blue", promedio
= "red"))

```

Dist_credit

#Duracion de credito

```

Dist_duracion = tesis %>%
  mutate(promedio = mean(Duration),
         mediana = median(Duration)) %>%
  ggplot(aes(x=Duration))+
  geom_histogram(color="#254A5B", fill="red")+
  geom_density(alpha=0.6)+
  geom_vline(aes(xintercept=promedio, color = "promedio"),
            linetype="dashed", size = 1)+
  geom_vline(aes(xintercept=mediana, color="mediana"),
            linetype="dashed", size = 1)+
  labs(title="Duracion de Crédito",x="Duración Crédito", y = "Frecuencia")+
  theme_classic()+
  scale_color_manual(name = "medidas", values = c(mediana = "blue", promedio
= "red"))

```

Dist_duracion

#Monto y credito Relacion

```

Relacion_duracion_monto = tesis %>%
  ggplot(aes(x=Credit.amount, y=Duration)) +
  geom_point()+
  geom_smooth(method=lm, se=FALSE)

```

Relacion_duracion_monto

#Relacion credito y dutacion

```

Relacion_edad_monto = tesis %>%
  ggplot(aes(x=Age, y=Credit.amount)) +
  geom_point()+
  geom_smooth(method=lm, se=FALSE)

```

Relacion_edad_monto

#edad y duracion

```

Relacion_edad_duracion = tesis %>%
  ggplot(aes(x=Age, y=Duration)) +
  geom_point()+
  geom_smooth(method=lm, se=FALSE)

Relacion_edad_duracion

#Proposito y monto

Proposito_monto = tesis %>%
  dplyr::group_by(Purpose) %>%
  dplyr::summarise(Credit.amount = mean(Credit.amount)) %>%
  arrange(-Credit.amount) %>%
  ggplot()+
  geom_bar(aes(x = reorder(Purpose, Credit.amount), y = Credit.amount), stat =
"identity", fill="#19B6C6")+
  coord_flip()+
  labs(title="Propósito vs monto",x="Monto", y = "Proposito")+
  theme_classic()

Proposito_monto

#Prestamos segun proposito

Proposito_cantidad = tesis %>%
  dplyr::group_by(Purpose) %>%
  dplyr::summarise(Cantidad = n()) %>%
  ggplot()+
  geom_bar(aes(x = reorder(Purpose, Cantidad), y = Cantidad), stat =
"identity", fill="#19B6C6")+
  coord_flip()+
  labs(title="Proposito de Crédito vs monto promedio",x="#Prestamos", y =
"Proposito")+
  theme_classic()

Proposito_cantidad

#Exploracion de datos segun el sexo y la edad

#Distribucion segun edad

Dist_edades_SEX = tesis %>%
  ggplot(aes(x=Sex, y = Age, fill = Sex))+
  geom_boxplot(alpha=0.7) +
  stat_summary(fun=mean, geom="point", shape=20, size=14, color="green",
fill="red")+
  theme(legend.position="none") +
  scale_fill_brewer(palette="Set1")

Dist_edades_SEX

#Distribucion de edad y sexo s

```

```

Dist_credito_SEX = tesis %>%
  ggplot(aes(x=Sex, y = Credit.amount, fill = Sex))+
  geom_boxplot(alpha=0.7) +
  stat_summary(fun=mean, geom="point", shape=20, size=14, color="green",
fill="red")+
  theme(legend.position="none") +
  scale_fill_brewer(palette="Set1")

Dist_credito_SEX

#Duracion de credito

Dist_duracion_SEX = tesis %>%
  ggplot(aes(x=Sex, y = Duration, fill = Sex))+
  geom_boxplot(alpha=0.7) +
  stat_summary(fun=mean, geom="point", shape=20, size=14, color="green",
fill="red")+
  theme(legend.position="none") +
  scale_fill_brewer(palette="Set1")

Dist_duracion_SEX

#Credito y relacion

Relacion_duracion_monto_SEX = tesis %>%
  ggplot(aes(x=Credit.amount, y=Duration)) +
  geom_point(aes(colour = Sex))+
  geom_smooth(method=lm, se=F)+
  facet_wrap(~ Sex)

geom_line(aes(linetype = variable))

Relacion_duracion_monto_SEX

#edad y monto

Relacion_edad_monto_Sex = tesis %>%
  ggplot(aes(x=Age, y=Credit.amount)) +
  geom_point(aes(colour = Sex))+
  geom_smooth(method=lm, se=F)+
  facet_wrap(~ Sex)

Relacion_edad_monto_Sex

#edad y duracion

Relacion_edad_duracion_Sex = tesis %>%
  ggplot(aes(x=Age, y=Duration)) +
  geom_point(aes(colour = Sex))+
  geom_smooth(method=lm, se=F)+
  facet_wrap(~ Sex)

Relacion_edad_duracion_Sex

```

```

#proposito y monto relacionado

Proposito_monto_Sex = tesis %>%
  dplyr::group_by(Purpose, Sex) %>%
  dplyr::summarise(Credit.amount = mean(Credit.amount)) %>%
  ggplot()+
  geom_bar(aes(x = reorder(Purpose, Credit.amount), y = Credit.amount,
fill=Sex), stat = "identity", position=position_dodge()+
  coord_flip()+
  labs(title="Propósito vs monto promedio",x="Monto", y = "Proposito")+
  theme_classic()

Proposito_monto_Sex

#Prestamos segun proposito

Proposito_cantidad_Sex = tesis %>%
  dplyr::group_by(Purpose, Sex) %>%
  dplyr::summarise(Cantidad = n()) %>%
  ggplot()+
  geom_bar(aes(x = reorder(Purpose, Cantidad), y = Cantidad, fill=Sex), stat
= "identity", position=position_dodge()+
  coord_flip()+
  labs(title="Propósito vs monto promedio",x="#Prestamos", y = "Proposito")+
  theme_classic()

Proposito_cantidad_Sex

#Arbol de decisiones

modeloDT<-rpart(Purpose~., method="class", data=select(tesis, -X))
prp(modeloDT) #grafica basica
rpart.plot(modeloDT, yesno = 2, type = 0)

predDT<-predict(modeloDT, newdata=tesis, type="class")
PredDT2 = as.character(predDT)
PredDT2 = ifelse(PredDT2 == 1,"SI","NO")
resultsDT<-table(tesis$Purpose, PredDT2)
resultsDT
prop.table(resultsDT)

zp = prune(modeloDT, cp = 0.01)
rpart.plot(zp)

#Calculo el acurracy

```

```

accuracyDT<-sum(diag(resultsDT))/sum(resultsDT)
accuracyDT

#Clusteriazcion de Segmentos

Columnas = colnames(tesis)
Tipos = sapply(tesis, class)

nombres_DF = as.data.frame(cbind(Columnas, Tipos))

variables_para_dummificar = nombres_DF%>%
  filter(Tipos == "character")

for(i in 1:nrow(variables_para_dummificar)) {
  columna_ref = as.character(variables_para_dummificar[i, 1])
  temporal = select(tesis, columna_ref)
  colnames(temporal) = "REF"
  temporal$REF = as.factor(temporal$REF)
  temporal_dummy = dummify(temporal$REF)
  nombres_dummy = colnames(temporal_dummy)
  nombres_dummy = paste(columna_ref, nombres_dummy, sep = "-")
  colnames(temporal_dummy) = nombres_dummy
  archivo = if(i == 1){ archivo = temporal_dummy} else {archivo =
cbind(archivo, temporal_dummy)}
}

archivo = as.data.frame(tesis)

datos_clusters1 = select(tesis, Age,Credit.amount, Duration)
datos_clusters = scale(datos_clusters1)

datos_clusters_char1 = cbind(datos_clusters1, tesis)
datos_clusters_char1 = as.numeric(as.factor(datos_clusters_char1))
datos_clusters_char = scale(datos_clusters_char1)
datos_clusters_char <- data.frame(datos_clusters_char)

tot_withinss <- map_dbl(2:10, function(k){
  model <- kmeans(x = datos_clusters, centers = k)
  model$tot.withinss
})

elbow_df <- data.frame(
  k = 2:10 ,

```

```

    tot_withinss = tot_withinss
  )

codo = ggplot(elbow_df, aes(x = k, y = tot_withinss)) +
  geom_line() +
  scale_x_continuous(breaks = 2:10)

sil_width <- map_dbl(2:10, function(k){
  model <- pam(x = datos_clusters, k = k)
  model$silinfo$avg.width
})

sil_df <- data.frame(
  k = 2:10,
  sil_width = sil_width
)

silueta = ggplot(sil_df, aes(x = k, y = sil_width)) +
  geom_line() +
  scale_x_continuous(breaks = 2:10)

codo
silueta

set.seed(42)

model_customers <- kmeans(datos_clusters, centers = 3)

cluster_k <- as.data.frame(model_customers$cluster)
colnames(cluster_k) = "cluster_k"

segment_customers_k = bind_cols(datos_clusters1, cluster_k) %>%
  mutate(cluster_k = as.factor(cluster_k))

Conteo_clusters1 = segment_customers_k %>%
  dplyr::group_by(cluster_k) %>%
  dplyr::summarise(conteo = n())

Conteo_clusters1

clusplot(segment_customers_k, segment_customers_k$cluster_k, shape=TRUE,
color=TRUE, labels=2, shade = T)

Resumen_K = segment_customers_k %>%
  group_by(cluster_k) %>%

```

```

summarise_all(funs(mean(.)))

Resumen_K

ggplot(segment_customers_k, aes(x=cluster_k, y=Age, fill=cluster_k)) +
  geom_boxplot()

ggplot(segment_customers_k, aes(x=cluster_k, y=Credit.amount,
fill=cluster_k)) +
  geom_boxplot()

ggplot(segment_customers_k, aes(x=cluster_k, y=Duration, fill=cluster_k)) +
  geom_boxplot()

#Relacion de clusters utilizando Edad y duracion del prestamo

Scatterplot_clusters_edad_duracion1 = segment_customers_k %>%
  ggplot(aes(x=Age, y=Duration)) +
  geom_point(aes(colour = cluster_k))+
  geom_smooth(method=lm, se=F)+
  facet_wrap(~ cluster_k)

Scatterplot_clusters_edad_duracion1

#Comparación scatterplot edad y credito

Scatterplot_clusters_eedad_credito1 = segment_customers_k %>%
  ggplot(aes(x=Age, y=Credit.amount)) +
  geom_point(aes(colour = cluster_k))+
  geom_smooth(method=lm, se=F)+
  facet_wrap(~ cluster_k)

Scatterplot_clusters_eedad_credito1

#Comparación scatterplot edad y credito

Scatterplot_clusters_duracion_credito1 = segment_customers_k %>%
  ggplot(aes(x=Credit.amount, y=Duration)) +
  geom_point(aes(colour = cluster_k))+
  geom_smooth(method=lm, se=F)+
  facet_wrap(~ cluster_k)

Scatterplot_clusters_duracion_credito1

#Cluster1 con Arbol

tesisconcluster = cbind (tesis, cluster_k)

```

```

glimpse(tesisconcluster)

modeloDT5<-rpart(Purpose~., method="class", data=select(tesisconcluster, -
Credit.amount))
prp(modeloDT5)
rpart.plot(modeloDT5, yesno = 2, type = 0)

predDT6<-predict(modeloDT5, newdata=tesisconcluster, type="class")
PredDT5 = as.character(predDT)
PredDT5 = ifelse(PredDT5 == 1,"SI","NO")
resultsDT6<-table(tesis$Purpose, PredDT5)
resultsDT6
prop.table(resultsDT6)

zp5 = prune(modeloDT5, cp = 0.015)
rpart.plot(zp5)

#Calculo el accuracy
accuracyDT2<-sum(diag(resultsDT6))/sum(resultsDT6)
accuracyDT2

#-----CLUSTER CON CARACTERES

tot_withinss <- map_dbl(2:10, function(k){
  model <- kmeans(x = datos_clusters_char, centers = k)
  model$tot.withinss
})

elbow_df <- data.frame(
  k = 2:10 ,
  tot_withinss = tot_withinss
)

ggplot(elbow_df, aes(x = k, y = tot_withinss)) +
  geom_line() +
  scale_x_continuous(breaks = 2:10)

sil_width <- map_dbl(2:10, function(k){
  model <- pam(x = datos_clusters_char, k = k)
  model$silinfo$avg.width
})

sil_df <- data.frame(
  k = 2:10,
  sil_width = sil_width
)

```

```

ggplot(sil_df, aes(x = k, y = sil_width)) +
  geom_line() +
  scale_x_continuous(breaks = 2:10)

set.seed(42)

model_customers_char <- kmeans(datos_clusters_char, centers = 3)

cluster_k_char <- as.data.frame(model_customers_char$cluster)
colnames(cluster_k_char) = "cluster_k"

segment_customers_k_char = bind_cols(tesis, cluster_k) %>%
  mutate(cluster_k = as.factor(cluster_k)) %>%
  dplyr::rename(cluster_k = cluster_k)

Conteo_clusters1_char = segment_customers_k_char %>%
  dplyr::group_by(cluster_k) %>%
  dplyr::summarise(conteo = n())

Conteo_clusters1_char

clusplot(segment_customers_k_char, segment_customers_k_char$cluster_k,
shape=TRUE, color=TRUE, labels=2, shade = T)

Resumen_K_char = segment_customers_k_char %>%
  group_by(cluster_k) %>%
  summarise_all(funs(mean(.)))

Resumen_K_char

#Segun edad
ggplot(segment_customers_k_char, aes(x=cluster_k, y=Age, fill=cluster_k)) +
  geom_boxplot()

#Segun el credito
ggplot(segment_customers_k_char, aes(x=cluster_k, y=Credit.amount,
fill=cluster_k)) +
  geom_boxplot()

#Duracion de prestamo
ggplot(segment_customers_k_char, aes(x=cluster_k, y=Duration,
fill=cluster_k)) +
  geom_boxplot()

#scatterplot de duracion de credito y edad

```

```

Scatterplot_clusters_edad_duracion1_char = segment_customers_k_char %>%
  ggplot(aes(x=Age, y=Duration)) +
  geom_point(aes(colour = cluster_k))+
  geom_smooth(method=lm, se=F)+
  facet_wrap(~ cluster_k)

Scatterplot_clusters_edad_duracion1_char

#Scatter plot de edad y total de credito

Scatterplot_clusters_eedad_credito1_char = segment_customers_k_char %>%
  ggplot(aes(x=Age, y=Credit.amount)) +
  geom_point(aes(colour = cluster_k))+
  geom_smooth(method=lm, se=F)+
  facet_wrap(~ cluster_k)

Scatterplot_clusters_eedad_credito1_char

#relacion entre duracion y total de credito

Scatterplot_clusters_duracion_credito1_char = segment_customers_k_char %>%
  ggplot(aes(x=Credit.amount, y=Duration)) +
  geom_point(aes(colour = cluster_k))+
  geom_smooth(method=lm, se=F)+
  facet_wrap(~ cluster_k)

Scatterplot_clusters_duracion_credito1_char

#cantidad de prestamos segun Proposito

Proposito_cantidad_Purpose_char = segment_customers_k_char %>%
  dplyr::group_by(Purpose, cluster_k) %>%
  dplyr::summarise(Cantidad = n()) %>%
  ggplot()+
  geom_bar(aes(x = reorder(Purpose, Cantidad), y = Cantidad, fill=cluster_k),
  stat = "identity", position=position_dodge()+
  coord_flip()+
  labs(title="Propósito vs monto promedio",x="#Prestamos", y = "Proposito")+
  theme_classic()+
  facet_wrap(~ cluster_k)

Proposito_cantidad_Purpose_char

#cantidad de prestamos segun Genero

Proposito_cantidad_Sex_char = segment_customers_k_char %>%
  dplyr::group_by(Sex, cluster_k) %>%
  dplyr::summarise(Cantidad = n()) %>%
  ggplot()+
  geom_bar(aes(x = reorder(Sex, Cantidad), y = Cantidad, fill=cluster_k),
  stat = "identity", position=position_dodge()+

```

```

coord_flip()+
labs(title="Propósito vs monto promedio",x="#Prestamos", y = "Proposito")+
theme_classic()+
facet_wrap(~ cluster_k)

Proposito_cantidad_Sex_char

#cantida de prestamos segun treabajo

Proposito_cantidad_Job_char = segment_customers_k_char %>%
  dplyr::group_by(Job, cluster_k) %>%
  dplyr::summarise(Cantidad = n()) %>%
  ggplot()+
  geom_bar(aes(x = reorder(Job, Cantidad), y = Cantidad, fill=cluster_k),
stat = "identity", position=position_dodge())+
  coord_flip()+
  labs(title="Job vs Cluster",x="Job", y = "#Prestamos")+
  theme_classic()+
  facet_wrap(~ cluster_k)

Proposito_cantidad_Job_char

#Cantidad de prestamos vs housing

Proposito_cantidad_housing_char = segment_customers_k_char %>%
  dplyr::group_by(Housing, cluster_k) %>%
  dplyr::summarise(Cantidad = n()) %>%
  ggplot()+
  geom_bar(aes(x = reorder(Housing, Cantidad), y = Cantidad, fill=cluster_k),
stat = "identity", position=position_dodge())+
  coord_flip()+
  labs(title="Housing vs Cluster",x="Housing", y = "#Prestamos")+
  theme_classic()+
  facet_wrap(~ cluster_k)

Proposito_cantidad_housing_char

#-----CLUSTER CON CARACTERES

#Cluster2 con Arbol

dataset2 = subset(tesis, cluster_k_char == 2)

tesisconcluster = cbind (tesis, dataset2)

glimpse(tesisconcluster)

modeloDT7<-rpart(Purpose~., method="class", data=select(dataset2, -X))
prp(modeloDT7)
rpart.plot(modeloDT7, yesno = 2, type = 0)

```

```
predDT6<-predict(modeloDT5, newdata=tesisconcluster, type="class")
PredDT5 = as.character(predDT)
PredDT5 = ifelse(PredDT5 == 1,"SI","NO")
resultsDT6<-table(tesis$Purpose, PredDT5)
resultsDT6
prop.table(resultsDT6)
```

```
zp7 = prune(modeloDT5, cp = 0.01)
rpart.plot(zp7)
```

```
#Calculo el accuracy
accuracyDT2<-sum(diag(resultsDT6))/sum(resultsDT6)
accuracyDT2
```

