
Primeros pasos de las PYMES en la Ciencia de Datos

Paula Camila González Ortega



UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería



**Primeros pasos de las PYMES en la Ciencia de
Datos**

Trabajo de graduación en modalidad de Trabajo Profesional presentado
por
Paula Camila González Ortega
para optar al grado académico de Licenciada en Ingeniería en Ciencias
de la Computación y Tecnologías de la Información

Guatemala,
2022

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería

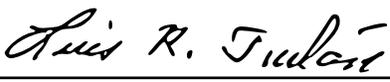


**Primeros pasos de las PYMES en la Ciencia de
Datos**

Trabajo de graduación en modalidad de Trabajo Profesional presentado
por
Paula Camila González Ortega
para optar al grado académico de Licenciada en Ingeniería en Ciencias
de la Computación y Tecnologías de la Información

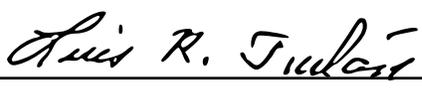
Guatemala,
2022

Vo.Bo.:

(f) 

Ing. Luis Roberto Furlán Collver

Tribunal Examinador:

(f) 

Ing. Luis Roberto Furlán Collver

(f) 

Ing. Eddy Omar Castro Jauregui

(f) 

Ing. Mario Francisco Barrientos Alejos

Fecha de aprobación: Guatemala, 12 de diciembre de 2022.

Luego de cursar múltiples cursos enfocados en la Ciencia de Datos y realizar prácticas labores como Analista de *“Business Intelligence”*, en este proyecto plasmé la importancia de la Ciencia de los Datos en las PYMES de Guatemala.

Ofrecer a las PYMES las herramientas gratuitas y accesibles desarrolladas en este proyecto es de gran satisfacción para mí, ya que actualmente las herramientas disponibles en el mercado requieren una membresía mensual elevada que frena a los empresarios de implementar esta ciencia en su estrategia de negocio. Espero que sea de gran ayuda para que las PYMES de mi país, Guatemala, comiencen a conocer y aprovechar los beneficios de la Ciencia de Datos que tanto me apasiona.

Este proyecto no hubiese sido posible sin todo el apoyo que he recibido de mis seres queridos. Agradezco a mis padres todo su apoyo desde que nací, cada palabra de aliento, su amor incondicional y sus consejos para ser mejor. Gracias especiales a mi papá por inspirarme a enfocar este proyecto en las PYMES. Agradezco a mi hermano por siempre hacerme feliz, recordarme lo importante de disfrutar cada momento de la vida y reconfortarme con sus genuinos abrazos. Agradezco a mi abuelita, mi “mimi”, por cuidarme desde mis primeros meses, buscar siempre mi bienestar y estar allí para escucharme cuando lo necesito. Agradezco a mi tío “chepito” por ayudarme en todo aquello que requiere de sus habilidades.

Al ingeniero Luis Furlán agradezco inmensamente todo su apoyo como asesor de este proyecto. Gracias por el tiempo que se tomó para darme sugerencias y orientarme en el tema, por compartirme tantos conocimientos y principalmente, por enseñarme la Ciencia de Datos al ser mi catedrático de “Data Science” en 2021, el curso que me trajo a esta rama de la Computación.

Doy gracias a amigas y amigos, con quienes compartí los cinco años de la carrera. Gracias por convertir esta etapa de mi vida en algo inolvidable, por su amistad incondicional y sus chistes científicos que nos dan sentido de pertenencia. Gracias a la Unidad de Vida Estudiantil por darme la oportunidad y el espacio de vivir mi pasión por la danza y el yoga. Gracias a cada uno de mis catedráticas y catedráticos, gracias al departamento de Computación y la Fundación de la Universidad del Valle de Guatemala, ya que sin ellos no estaría aquí a punto de culminar mis estudios como ingeniera.

Prefacio	III
Lista de figuras	VII
Lista de figuras	VII
Lista de cuadros	VIII
Lista de cuadros	VIII
Resumen	IX
1. Introducción	1
2. Antecedentes	2
3. Justificación	3
4. Objetivos	5
4.1. Objetivo general	5
4.2. Objetivos específicos	5
5. Marco teórico	6
5.1. La ciencia de paciencia: la Ciencia de Datos	6
5.1.1. ¿Cuál es la ciencia de esta ciencia?	6
5.1.2. Su evolución	6
5.1.3. Descifrando sus conceptos básicos	7
5.1.4. El valor de la Ciencia de Datos en los negocios	7
5.2. La Ciencia de Datos y otras ciencias	8
5.2.1. ¿Cómo implementar esta ciencia según la industria?	8
5.2.2. Economía, la ciencia hermana	8
5.2.3. Actividades principales	8
5.3. Pequeñas y medianas empresas	9
5.3.1. Distribución de las PYMES en Guatemala	9
5.3.2. ¿Por qué las PYMES son importantes en la economía de un país?	9
5.4. Caja de herramientas para la ciencia de datos	11
5.4.1. Requerimientos para la exitosa implementación	11
5.4.2. Herramientas disponibles en la red	11

5.5. Aprendiendo a caminar como científico de datos	16
5.5.1. Recopilación de los datos	16
5.5.2. Preparación de los datos	17
5.5.3. Exploración de los datos	17
5.5.4. Diseño del primer informe	19
6. Metodología	22
6.1. Planteamiento del problema	22
6.2. Tipo de investigación	23
6.3. Alcance	23
6.4. Técnica de investigación	23
7. Resultados	25
7.1. Encuesta	25
7.1.1. Convocatoria	25
7.1.2. Instrumento de investigación	25
7.1.3. Resultados	25
7.2. Grupo focal	31
7.2.1. Alcance e hipótesis	31
7.2.2. Información general	32
7.2.3. Participantes	32
7.2.4. Responsables	32
7.2.5. Convocatoria	32
7.2.6. Hallazgos**	32
7.3. Taller introductorio	33
7.3.1. Información general	33
7.3.2. Convocatoria	34
7.3.3. Hallazgos	34
8. Propuesta de solución	35
8.1. ¿Por qué un manual de Ciencia de los Datos?	35
8.2. Contenido del manual de Ciencia de Datos	35
8.3. Manual	36
8.4. Tablero	37
9. Conclusiones	42
10.Recomendaciones	43
11.Bibliografía	44
12.Anexos	48
12.1. Ilustraciones	48
12.1.1. Ilustración 2	49
12.1.2. Ilustración 3	50
12.1.3. Ilustración 4	51
12.1.4. Ilustración 5	51
12.1.5. Ilustración 6	52
12.1.6. Ilustración 7	53
12.1.7. Ilustración 8	54
12.2. Tablas	55
12.2.1. Tabla 2	55
12.2.2. Tabla 3	55
12.2.3. Tabla 4	56
12.2.4. Tabla 5	56

12.3. Recursos	60
12.3.1. Recurso 1	60
12.3.2. Recurso 2	60
12.3.3. Recurso 3	60
12.3.4. Recurso 4	60
12.3.5. Recurso 5	60
12.3.6. Recurso 6	60
Glosario	61

Lista de figuras

3.1. Gráfica elaborada por el autor sobre el aporte de las empresas al PIB de Guatemala	3
3.2. Gráfica elaborada por el autor sobre el aporte de las empresas a la Fuerza laboral de Guatemala	4
5.1. Gráfica elaborada por el autor sobre la distribución de las PYMES por sector económico	9
5.2. Gráfica elaborada por el autor sobre porcentaje de empresas importadoras.	10
5.3. Gráfica elaborada por el autor sobre porcentaje de empresas exportadoras.	10
5.4. Gráfica elaborada por el autor sobre las ventas anuales por el tamaño de las empresas.	11
5.5. Las 7 Librerías recomendadas según el proceso y lenguaje de programación utilizado.	12
5.6. Ambiente de trabajo de Power BI Desktop	13
5.7. Ambiente de trabajo de Tableau Public	14
5.8. Ambiente de trabajo de Excel	14
5.9. Ambiente de trabajo de Ingrofr.am	15
5.10. Ambiente de trabajo de Picktochart	15
8.1. Portada del manual de usuario elaborado para las PYMES	37
8.2. Reporte interactivo elaborado con los datos ejemplos	39
8.3. Reporte estático elaborado con los datos ejemplos	40
8.4. Modelo de regresión lineal para predecir la demanda del año ingresado	41
12.1. Disciplinas y habilidades involucradas en la Ciencia de Datos.	48
12.2. Disciplinas necesarias para la aplicación de la Ciencia de Datos.	49
12.3. Cinco etapas del ciclo de vida de la Ciencia de Datos: captar, mantener, procesar, analizar y comunicar.	50
12.4. Línea del tiempo de la Ciencia de Datos.	51
12.5. Diferencias entre la economía y la Ciencia de Datos.	51
12.6. Los procesos en la Ciencia de Datos.	52
12.7. Boletín digital utilizado en la convocatoria de PYMES para la encuesta.	53
12.8. Primera sección de la encuesta elaborada	54

Lista de cuadros

5.1. Formas de usar la Ciencia de Datos en las empresas. 47	8
5.2. Los 5 IDEs más utilizados para la Ciencia de Datos. 24 28	12
5.3. Tabla comparativa, elaborada por el autor, de las herramientas más reconocidas para la Ciencia de Datos que no requieren experiencia en programación.	16
5.4. Conceptos de las estadísticas descriptivas básicas. 33 8	18
5.5. Pasos recomendados para realizar un informe de análisis de datos	20
12.1. Número de empresas registradas y activas clasificadas por tamaño y departamento. 35	55
12.2. Aporte al monto total de importaciones y exportaciones por tamaño de empresa. 38	55
12.3. Distribución porcentual por actividad económica de las MIPYES activas en 2017. 53	56
12.4. Respuestas de los estudiantes de administración y ciencia de datos, durante el focus group realizado.	59

El presente proyecto de graduación aborda la Ciencia de Datos desde una perspectiva empresarial para conocer situación de las PYMES guatemaltecas con respecto al tema, la importancia y beneficios de esta en un negocio. En el primer capítulo, se justifica la investigación llevada a cabo. El segundo capítulo encontrará el marco teórico, en el que los temas relacionados a la Ciencia de Datos y las PYMES se dividen en cinco grandes temas. Se describen los conceptos básicos de la Ciencia de Datos, su historia, los procesos involucrados, las aplicaciones en la vida real y algunas herramientas utilizadas para su implementación. También se desglosa la distribución de las PYMES en Guatemala y el papel que juegan en la economía del país.

El capítulo 3 describe la metodología de investigación para conocer la situación de las PYMES en la Ciencia de Datos. Es importante mencionar que el objetivo general era investigar la importancia de aplicar la Ciencia de Datos en los procesos de compra y venta de las PYMES. La investigación de campo constó de tres fases: una encuesta a un grupo muestra de las PYMES guatemaltecas, un grupo focal a estudiantes universitarios expertos en el tema y un taller introductorio para el grupo muestra de las PYMES. Con este último se obtuvieron comentarios y sugerencias sobre el tablero y manual desarrollados.

La explicación del proceso realizado para desarrollar el manual básico de la Ciencia de Datos y el tablero, con el cual se pretende que las PYMES den sus primeros pasos en dicha ciencia, se encuentra en el capítulo 4. Finalmente se comparten las conclusiones y recomendaciones obtenidas al culminar la investigación, destacando que los objetivos de proyecto si fueron cumplidos.

CAPÍTULO 1

Introducción

El término “Ciencia de Datos” cada vez gana más campo en las empresas de países desarrollados. Su implementación en las empresas produce mejores resultados a nivel administrativo y económico. Sin embargo, la Ciencia de Datos y los conceptos relacionados a ella son poco conocidos para la mayoría de las pequeñas y medianas empresas del país.

Actualmente, algunas empresas en Guatemala ya aprovechan la información que obtienen a través de la “Ciencia de Datos”. De manera contraria la mayoría de pequeñas y medianas empresas en el país al no implementar la Ciencia de Datos como parte de su logística, pierden oportunidades, estrategias e información valiosa del mercado.

La hipótesis del presente trabajo de investigación plantea que las PYMES en Guatemala aún no reconocen los beneficios del uso de la Ciencia de Datos. Sin embargo, muestran interés en conocer dicho método para potenciar sus ventas y comprender “científicamente” la situación real de sus empresas. Con un manual de la Ciencia de Datos es posible ayudar a las PYMES, a incorporar dicha ciencia en sus procesos de comercialización. Por lo tanto, como producto final se propone la creación de un “Manual básico de la Ciencia de Datos”, incluyendo algunos programas para computadora, que orientará a las PYMES a incorporar dicha ciencia, para aprovechar sus beneficios en el área comercial.

El artículo científico titulado “The future of data analysis”, escrito por John W. Tukey en la Universidad de Princeton y los laboratorios Bell, fue una de las primeras publicaciones que abordaron la ciencia de los datos. En la época en que Tukey publicó este artículo, las primeras computadoras ya se habían creado, pero los retos de cálculo y las nuevas oleadas de datos marcaron una reforma que originó varios términos, entre ellos el “BigData espacial”. [39] Dicho artículo científico enfatizó la necesidad de contar con herramientas y conocimientos para el procesamiento de grandes volúmenes de información. Turkey también indicó la importancia de declarar la adopción del análisis de datos como una ciencia experimental y dejar de verlo como un sistema lógico deductivo, cosa que hoy nos parecería obvia pero que en su momento era una revolución. [30]

Desde 1974 se han publicado muchos libros, artículos y reportes que aportan formas innovadoras para analizar los datos desde un enfoque más computacional que estadístico. Entre todas estas referencias sobre la Ciencia de Datos, es portante destacar la definición que crea el paradigma de este tema: “Ciencia de datos es el vínculo entre las metodologías tradicionales de estadística, tecnología de cómputo moderna y el conocimiento de expertos específico para convertir datos en información y conocimiento”. [17]

Si los datos se interpretan de manera correcta, a través de los análisis es posible ofrecer grandes beneficios para las organizaciones, instituciones y sociedades. Esto debido a que la ciencia de datos revela tendencias y genera información que las empresas pueden utilizar para tomar mejores decisiones y crear productos y servicios más innovadores. [40]

El último informe del Sistema Nacional de Información MIPYME Guatemala [35], elaborado por la Oficina Coordinadora de Estadísticas de la MIPYME, señala que las PYMES en Guatemala aportan el 35 % del Producto Interno Bruto – PIB – (Figura 3.1) y el 85 % de la fuerza laboral del país (Figura 3.2).

Aporte de las empresas al PIB de Guatemala

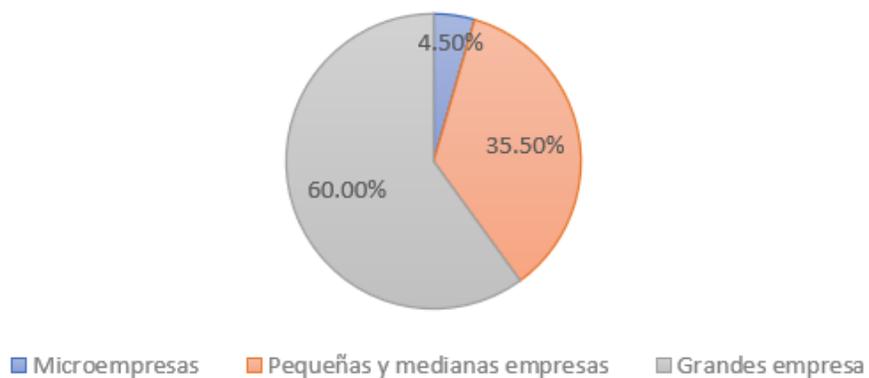


Figura 3.1: Gráfica elaborada por el autor sobre el aporte de las empresas al PIB de Guatemala

Aporte de las empresas a la fuerza laboral de Guatemala



Figura 3.2: Gráfica elaborada por el autor sobre el aporte de las empresas a la Fuerza laboral de Guatemala

Otro factor tomado en cuenta para el enfoque de esta investigación es que el número de empresas PYMES activas en Guatemala ascendió de 372,779 en 2015, a 481,570 en 2017. [53]

Sin embargo, los efectos negativos de la pandemia COVID-19 sobre el sector económico han afectado más a las pequeñas y medianas empresas en comparación con las grandes corporaciones. Los resultados de la encuesta sobre el "seguimiento" de COVID-19 para cuatro países de la región [55] (El Salvador, Guatemala, Honduras y Nicaragua) ayudan a comprender mejor cómo están cambiando las cosas para las empresas de dichos países.

Contar con un manual de Ciencia de Datos les permitirá a las PYMES utilizar herramientas para analizar los datos de sus ventas de forma eficiente, para encontrar oportunidades, puntos de mejora y potenciar sus ingresos. [31] [41]

4.1. Objetivo general

Investigar la importancia de aplicar la Ciencia de Datos en los procesos de compra y venta de las PYMES.

4.2. Objetivos específicos

- Crear un manual básico, práctico y accesible para las PYMES sobre los principios fundamentales de la Ciencia de Datos.
- Elaborar un conjunto de datos de ejemplo, que las PYMES puedan replicar con los datos de sus compras y ventas, con el fin de utilizarlo en su primer análisis de datos.
- Desarrollar para las PYMES un tablero gratuito de la Ciencia de Datos enfocado en la limpieza y exploración de datos.
- Recomendar los requerimientos básicos de equipo, como una computadora con alta capacidad de memoria, conexión a internet y datos por analizar.
- Ofrecer un glosario de términos, métodos y modelos utilizados durante la preparación, análisis y predicción de datos.
- Describir los métodos aplicados para el manejo y predicción de datos.
- Producir herramientas de instrucción del método, tanto escritos como audiovisuales, que instruyan a las PYMES para que puedan realizar un análisis de información a través de la Ciencia de Datos.

5.1. La ciencia de paciencia: la Ciencia de Datos

5.1.1. ¿Cuál es la ciencia de esta ciencia?

Una definición de Ciencia de Datos es que es una ciencia multidisciplinaria que requiere y desarrolla diferentes habilidades durante su aplicación (Ver [Ilustración 1](#) y [Ilustración 2](#) en anexos). Algunas disciplinas involucradas en la Ciencia de Datos son las ciencias sociales, la estadística, la matemática, la mercadotecnia y la computación. [\[49\]](#)

Hal Varian (2009), financiero central de Google, define a la Ciencia de Datos como “la habilidad de tomar datos, ser capaz de entenderlos, procesarlos, extraer valor de ellos y comunicarlo” (Ver [Ilustración 3](#) en anexos). Las empresas que la implementan reportan altas ganancias de sus inversiones, crecimiento de ventas, eficiencia operacional, rápido ciclo de comercialización, mayor satisfacción de sus clientes y mejores experiencias de venta. [\[25\]](#)

5.1.2. Su evolución

Esta ciencia nació en 1962 cuando el matemático John W. Tukey predijo el efecto positivo en el avance del análisis de datos a causa de la fusión entre la estadística y las computadoras. Publicó su visión sobre el tema en el artículo “The Future of Data Analysis” [\[50\]](#), en el que establece el análisis de datos como una ciencia empírica que involucra la observación, resumen y visualización de los datos para comprender mejor una situación. En 1977 se crea la Asociación Internacional de Computación Estadística (IASC) con la misión de relacionar los métodos tradicionales de la estadística con la tecnología moderna y el conocimiento de expertos para convertir los datos en información y conocimiento (Ver [Ilustración 4](#) en anexos).

En 1994, la revista Bloomberg Businessweek [\[17\]](#) publicó un artículo sobre algunas empresas que comenzaron a recopilar grandes cantidades de información personal para diseñar mejores campañas de publicidad. DJ Patil y Jeff Hammerbacher, de LinkedIn y Facebook, popularizaron el término “científico de datos” en 2008 a través de internet y conferencias.

La Ciencia de Datos, durante más de 30 años, se ha expandido hasta involucrar gobiernos,

empresas y organizaciones de todo el mundo. Cada día más profesionales se involucran en el campo, desde ingenieros hasta biólogos, genetistas e incluso astrónomos. Un ejemplo de ello es la creación de nuevos tratamientos para enfermedades, la clasificación de especies nuevas y la predicción precisa del calentamiento global. [36]

5.1.3. Descifrando sus conceptos básicos

El término “**Big Data**”, traducido como “**Macrodatos**”, se refiere a la ciencia encargada de analizar enormes volúmenes de datos estructurados o no estructurados. Los estructurados son aquellos datos organizados y listos para manipularlos (datos cuantitativos). Los datos no estructurados son aquellos que requieren ser procesados para analizarlos (datos cualitativos) [21].

Otros términos relacionados son el “**Business Intelligence**” y el “**Business Analytics**”, conocidos como la “Inteligencia de negocios” y “Análisis empresarial” respectivamente. El último término consiste en el análisis de datos estructurados enfocados al futuro para buscar tendencias e indicadores macroeconómicos. La inteligencia de negocios se concentra en datos estructurados del pasado para corregir errores en la toma de decisiones [21].

Gracias a la inteligencia de negocios, la Coca-Cola Bottling Company maximizó su eficiencia operativa al automatizar el proceso para generar informes, ahorrando así más de 260 horas al año.

La minería de datos, del inglés “**Data Mining**”, es una técnica cuantitativa utilizada por la inteligencia de negocios que permite procesar, identificar y extraer grandes cantidades de datos con el fin de identificar patrones y tendencias [48].

Otro término utilizado es “**Machine Learning**” (aprendizaje automático), que consiste en la implementación de modelos para hacer predicciones, clasificaciones y buscar nuevas relaciones entre los datos. Un ejemplo es cómo Netflix predice las películas y series que pueden interesarle a un usuario según su historial.

La ingeniería de datos, conocido como “**Data Engineering**” se basa en extraer, transformar y cargar los datos de forma eficiente y segura, para posteriormente sacarles provecho [22].

El término “**Data Analytics**”, traducido como “analítica de datos”, abarca desde la adquisición hasta la modelización de los datos, incluyendo la transformación y el resumen de estos [22].

5.1.4. El valor de la Ciencia de Datos en los negocios

De acuerdo con Stedman (2020) [45], la Ciencia de Datos convierte los datos en información valiosa para las empresas porque permite identificar tendencias y conocimientos útiles en la toma de decisiones comerciales, realización de planificación estratégica y desarrollo de nuevos proyectos. Toda empresa que implementa la Ciencia de Datos logra incrementar su productividad laboral en un 40% y entrar a un proceso de cultura ágil porque el personal puede usar el tiempo de manera más eficiente [1]. La información obtenida a través de los análisis y reportes es esencial para mejorar los planes de mercadeo y ventas para las empresas, además de generar ventajas sobre la competencia del negocio [11].

5.2. La Ciencia de Datos y otras ciencias

5.2.1. ¿Cómo implementar esta ciencia según la industria?

A continuación, se muestra una tabla con ejemplos del uso de la Ciencia de Datos en diferentes sectores de la industria.

Industria	Implementación
Marketing	Identificar patrones en el comportamiento del consumidor para incrementar la tasa de clientes leales y alcance.
Publicidad	Campañas promocionales personalizadas cuyos enfoques son más eficientes.
Finanzas	Identificación de comportamientos sospechosos para la detección de fraudes y gestión de riesgos. Predecir el comportamiento de la bolsa.
Manufactura	Crear un algoritmo para anticipar fallos en equipos, mejorar los inventarios, planificar la producción y organizar los centros de trabajo
Logística	Optimizar las cadenas de suministro, ofrecer servicios con alto rendimiento y generar análisis predictivos de fallos, tiempos o tráfico.
Salud	Detectar tendencias en diagnósticos e incrementar los índices de salud preventiva
Cloud Computing	Predecir el escalamiento del servicio y mejorar las migraciones o cargas de trabajo al presupuesto la oferta y consumo.

Tabla 5.1: Formas de usar la Ciencia de Datos en las empresas. [47]

5.2.2. Economía, la ciencia hermana

Tanto la Ciencia de Datos como la economía son ciencias que se apoyan en la estadística para encontrar soluciones a problemas cuantitativos a través de modelos, en conjunto con habilidades analíticas. La principal diferencia entre dichas ciencias es el enfoque de su investigación: la economía fija su atención en la causalidad y la Ciencia de Datos en la predicción. [9] (Ver [Ilustración 5](#) en anexos).

La Ciencia de Datos complementa a la economía y viceversa. La economía ofrece los conocimientos en finanzas y comportamientos de los consumidores mientras que la Ciencia de Datos brinda las herramientas que permiten el manejo de grandes volúmenes de datos sin la complejidad de muchos modelos matemáticos que los economistas aplican. [9]

5.2.3. Actividades principales

El 80 % del tiempo en la Ciencia de Datos es utilizado para la preparación. Dicho proceso consiste en la creación, limpieza y estructuración de los conjuntos de datos a analizar. Esta tarea que consume mucho tiempo permite que los datos sean consistentes y adecuados para los siguientes procesos. [7] (Ver [Ilustración 6](#) en anexos)

5.3. Pequeñas y medianas empresas

5.3.1. Distribución de las PYMES en Guatemala

De acuerdo con el informe realizado por el Ministerio de economía (2015) [35], Guatemala es el departamento que cuenta con más empresas a nivel nacional ya que en este se registra el 44.3 % de la distribución total de las PYMES. Luego se encuentra Quetzaltenango con el 5.6 % y Escuintla con el 3.9 %. (Ver [Tabla 2](#) en anexos)

En la Figura 5.1 se presentan los datos, publicados en el Informe de Situación y Evolución del Sector MIPYME de Guatemala 2015-2017 [53], sobre las actividades económicas realizadas por las MIPYMES. En ella se observa que gran parte de las PYMES se dedican a actividades inmobiliarias (32 %) y comercio, reparación de vehículos, efectos personales y enseres domésticos (29 %). Únicamente el 2 % de las PYMES activas en 2015 pertenecen al sector de agricultura, ganadería, silvicultura, pesca y minería. (Ver [Tabla 4](#) en anexos)

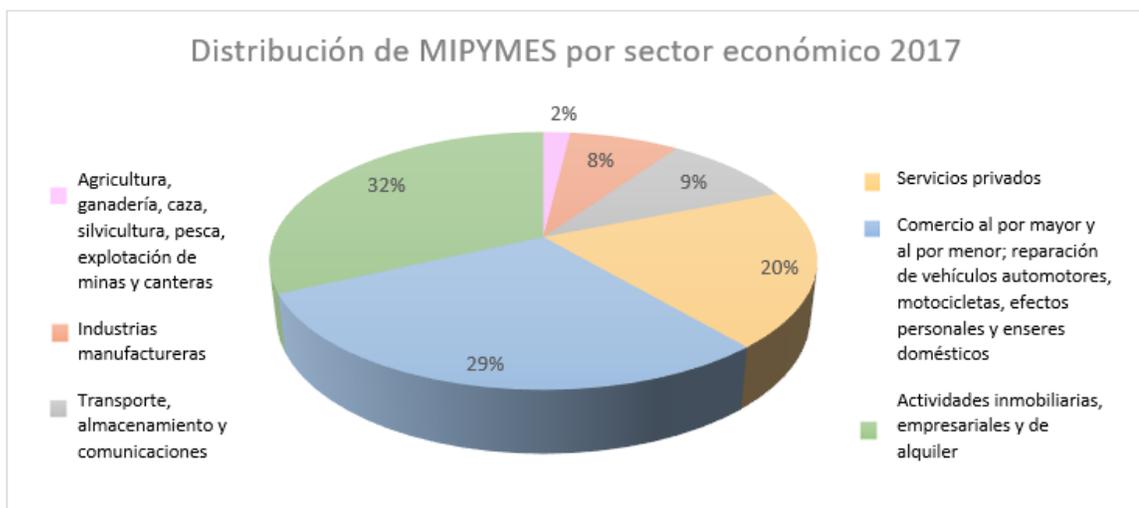


Figura 5.1: Gráfica elaborada por el autor sobre la distribución de las PYMES por sector económico.

El sector económico con menor participación de las MIPYMES es el de agricultura, ganadería, caza, silvicultura, pesca, explotación de minas y canteras. De forma contraria las actividades inmobiliarias, empresariales y de alquiler son aquellas en las que la mayoría, exactamente el 32 %, de las MIPYMES participaron en 2017. [53]

5.3.2. ¿Por qué las PYMES son importantes en la economía de un país?

Para 2015 se registró que 35,948 empresas importaban productos. De estas empresas importadoras el 71.23 % fueron pequeñas empresas, 20.4 % microempresas y 5.29 % medianas empresas (Figura 5.2).

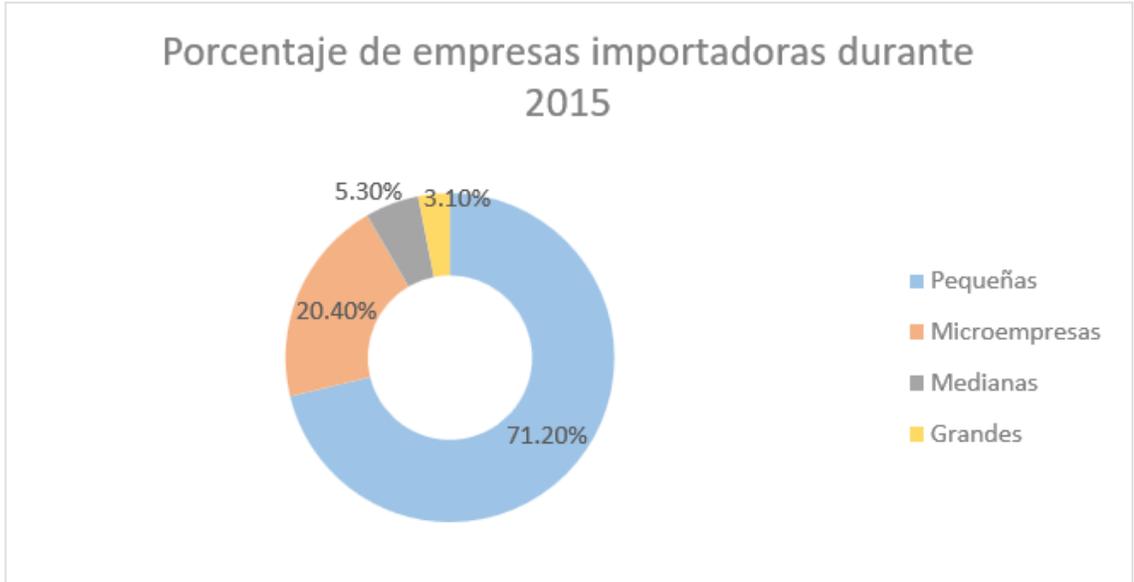


Figura 5.2: Gráfica elaborada por el autor sobre porcentaje de empresas importadoras. [38]

Según la OCSE-MIPYME [38], en 2015 el número de empresas exportadoras en Guatemala ascendió a 4,444. Las microempresas fueron las empresas más exportadoras de ese año (34.40%), seguido por las pequeñas empresas con una actividad exportadora del 34.30% y las medianas empresas con el 17.40% (Figura 5.3).

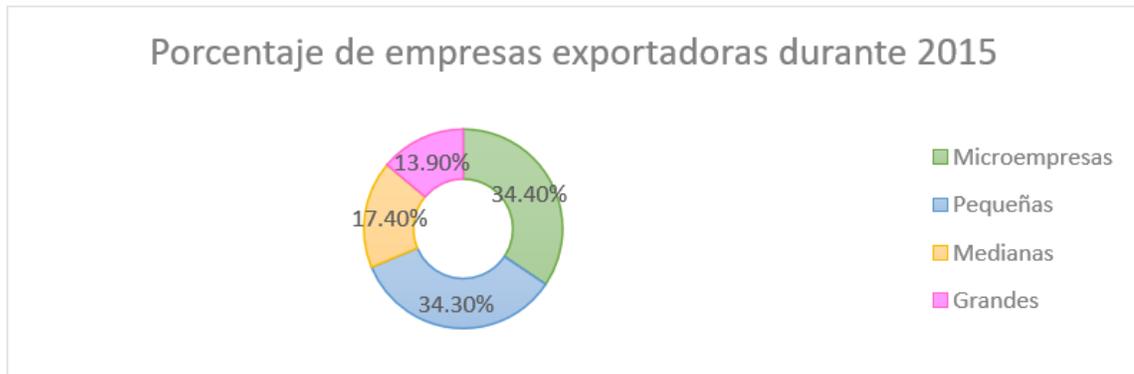


Figura 5.3: Gráfica elaborada por el autor sobre porcentaje de empresas exportadoras. [38]

Las grandes empresas fueron las empresas que más aportaron al monto total de las exportaciones a nivel nacional (Q62,891.1 millones). Su aporte fue el 49.5%, seguido por el 23.7% de las pequeñas empresas y el 15.9% de las medianas empresas. El monto de las importaciones totales fue Q111,332.7 millones. Para este valor, las grandes empresas presentaron el aporte más significativo (79.3%), seguido por las medianas empresas con el 10.5%, las pequeñas empresas con el 6.3% y las microempresas con el 2.7% (Ver [Tabla 3](#) en anexos).

El viceministerio de desarrollo de la MIPYME (2018) [53], reportó que durante el 2015 el monto de las ventas anuales realizadas por las empresas de Guatemala ascendió a Q500,213,307,681.11. En estas ventas, el mayor peso económico es el de las grandes empresas equivalente a Q325,035,775,733.00 seguido por el de las medianas empresas con el Q76,560,261,447.00. En la siguiente gráfica se detalla el porcentaje de ventas anuales para cada tamaño de empresa (Figura 5.4).

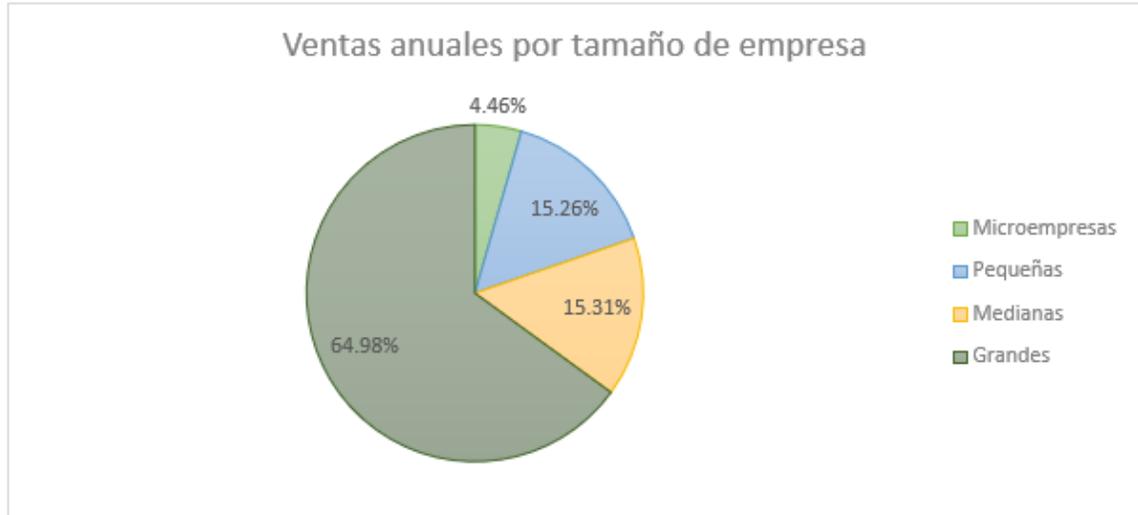


Figura 5.4: Gráfica elaborada por el autor sobre las ventas anuales por el tamaño de las empresas. [38]

5.4. Caja de herramientas para la ciencia de datos

5.4.1. Requerimientos para la exitosa implementación

Para implementar la Ciencia de Datos, además de tecnología básica como una computadora con acceso a internet, se requieren ciertas habilidades y conocimientos.

Las personas expertas en esta ciencia deben contar con conocimientos en programación, estadística y bases de datos. Python, R y SQL son los lenguajes de programación que todo científico de datos debe manejar. [51]

5.4.2. Herramientas disponibles en la red

Entornos de desarrollo integrado – IDE-

Un entorno de desarrollo integrado, -IDE- por sus siglas en inglés, está formado por un editor de código, herramientas de compilación y un depurador. Permiten escribir, probar y depurar código de forma sencilla. Se detalla las 5 herramientas más utilizadas cuando se empieza a experimentar con la Ciencia de Datos. [18]

Jupyter Notebook	PyCharm	Google Colab	Visual Studio Code	Rstudio
Es una aplicación web intuitiva que permite codificar en Julia, Python y R.	Ofrece finalización de código inteligente, inspecciones de código y resaltado de errores.	Permite programar y ejecutar código de Python desde la nube, con un navegador. No requiere configuración y ofrece la opción de trabajar colaborativamente	Este IDE permite escribir y ejecutar código en todos los lenguajes de programación. Ofrece la opción de editar código para cuadernos de Jupyter.	Interfaz gratuita para interactuar con programas en R y llevar a cabo análisis de datos o desarrollo de gráficos en orden y de forma simultánea gracias a las 4 vistas que ofrece.
				

Tabla 5.2: Los 5 IDEs más utilizados para la Ciencia de Datos. [24] [28]

Librerías

Un científico de datos debe utilizar varias librerías según el lenguaje de programación en que esté desarrollando el proyecto [44]. Una librería en programación consiste en un conjunto de archivos, en forma de código o datos, que se importan y utilizan para desarrollar sistemas de forma más sencilla [13].

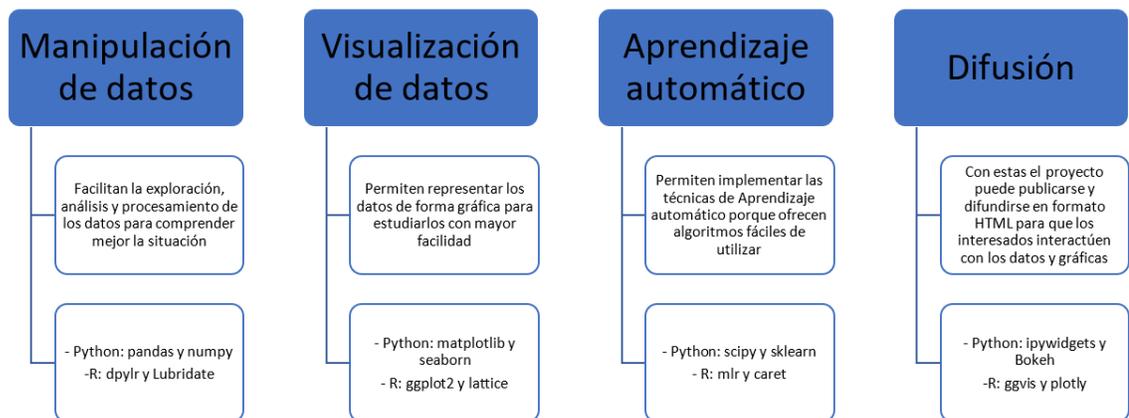


Figura 5.5: Las 7 Librerías recomendadas según el proceso y lenguaje de programación utilizado. [4] [6]

Herramienta sin programación

A continuación, se listan las herramientas más reconocidas para la Ciencia de Datos. Para ninguna de estas se requiere experiencia en programación o manejo de bases de datos, pero si es indispensable tener una visión clara de los procesos involucrados en la Ciencia de Datos y comprender la lógica de los mismos.

En todas estas herramientas lo más recomendable es utilizar datos previamente limpios, ya que la mayoría se enfoca en el análisis y la visualización de datos en entornos empresariales con el fin de identificar KPIs e información valiosa que sean de utilidad en la toma de decisiones. [10]

- Power BI Desktop

Esta herramienta para la inteligencia de negocios es reconocida por su interfaz amigable y fácil de utilizar. Permite limpiar los datos de fácilmente y los procesa para que las organizaciones vean la información de forma comprensible como gráficas o tablas. Permite recopilar la información desde hojas de cálculo hasta bases de datos y aplicaciones en servidores locales o en la nube. [26]

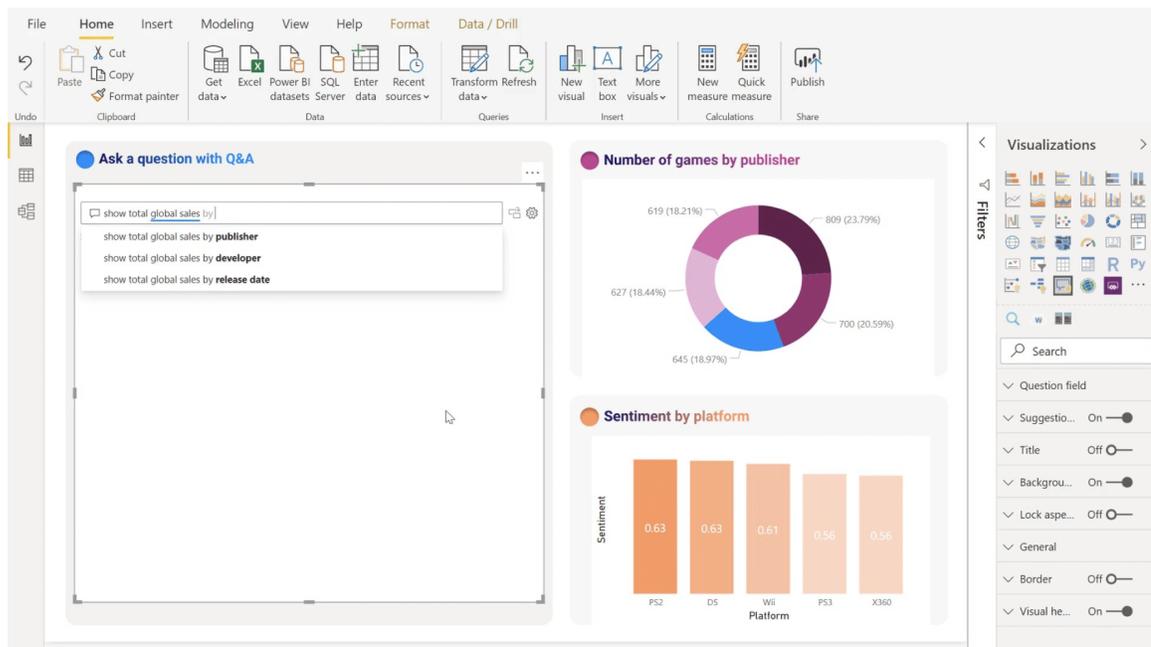


Figura 5.6: Ambiente de trabajo de Power BI Desktop [34]

- Tableau Public

Es otra herramienta para la exploración y visualización de datos, cuyos resultados se comparten públicamente. Permite la creación de visualizaciones intuitivas e interactivas en múltiples formatos. Cualquier persona puede crear o acceder a visualizaciones desde la plataforma web o desktop de forma gratuita, con el objetivo de aprender de otros usuarios mientras integra los datos en su vida diaria. [42]

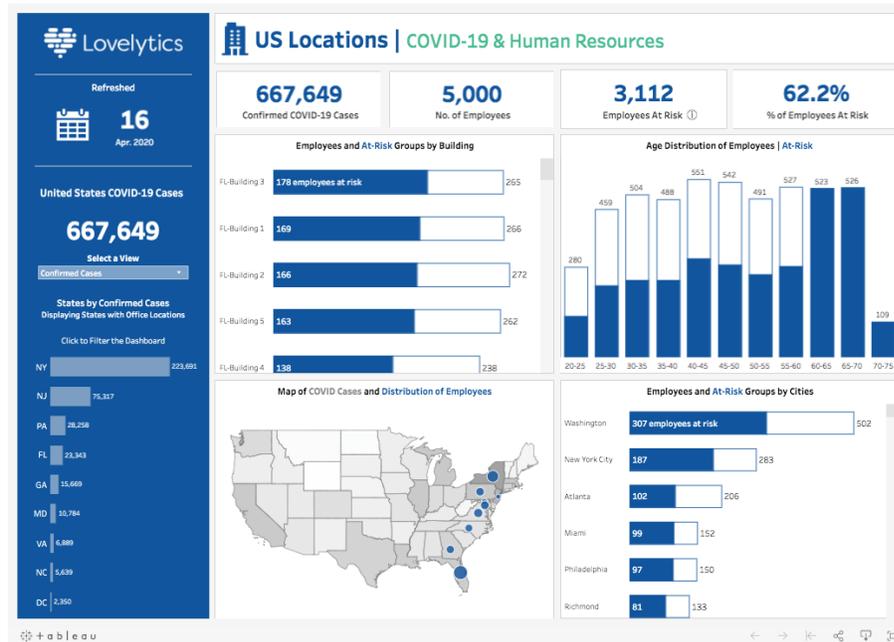


Figura 5.7: Ambiente de trabajo de Tableau Public [42]

■ Excel

Es un entorno basado en hojas de cálculo que permite ordenar los datos a partir de filas y columnas. Se caracteriza por su opción de aplicar cálculos y fórmulas a los datos de forma sencilla y ágil, así como la generación de gráficas. [34]



Figura 5.8: Ambiente de trabajo de Excel [34]

- **Infogr.am**

Es una herramienta para la visualización intuitiva e interactiva de datos o métricas. Presenta 6 opciones de plantillas con personalización limitada; Luego de seleccionar una, se cargan los datos, se agregan los comentarios y se comparte. La versión gratuita, que se accede desde cualquier navegador, permite realizar hasta 10 visualizaciones o infografías. [27]

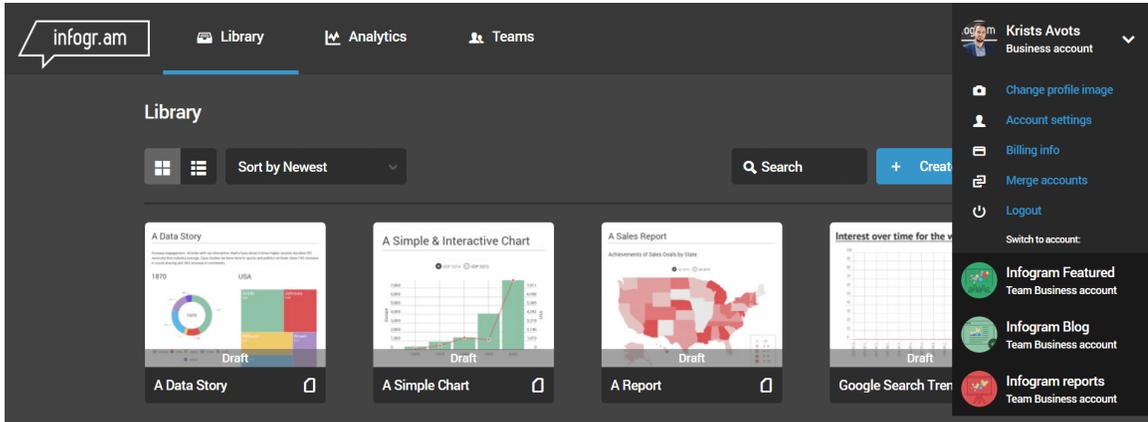


Figura 5.9: Ambiente de trabajo de Ingrofr.am [27]

- **Picktochart**

Ofrece la opción de elaborar infografías, presentaciones o reportes de los datos cargados por el usuario. Dichos datos deben subirse en archivos CSV para que sea posible generar gráficos simples o complejos. Permite personalizar la fuente del texto, colores, logos e íconos del diseño en general. [19]

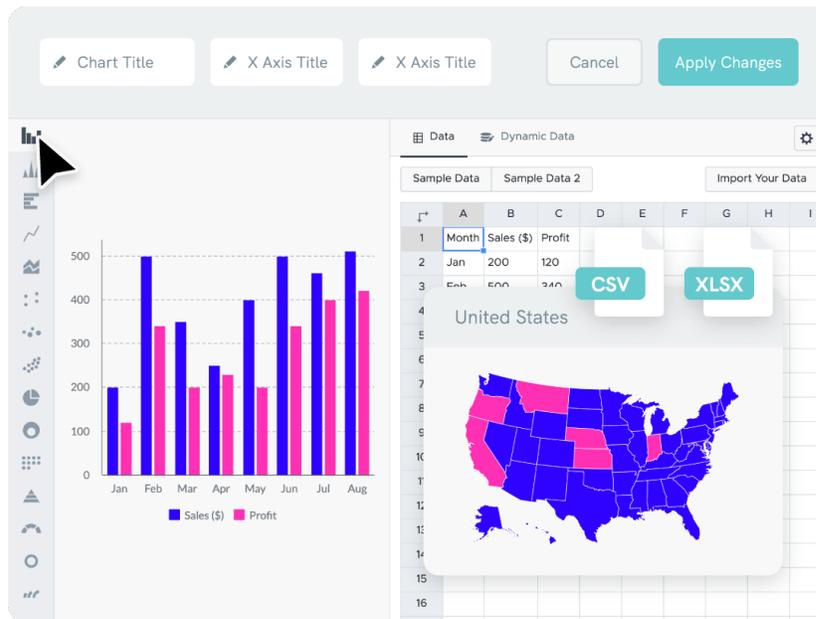


Figura 5.10: Ambiente de trabajo de Picktochart [19]

Power Bi Desktop	<ul style="list-style-type: none"> • Esta versión gratuita ofrece muchas funcionalidades • Para acceder a una versión más completa es necesario contar con un plan mensual de membresía (\$9.99- \$4,995) por cada usuario • La versión gratuita no permite el trabajo colaborativo 	
Tableau Public	<ul style="list-style-type: none"> • Acceso desde el ordenador y móvil • Acceso desde navegador o desktop • Es una plataforma para datos públicos por lo que cualquier persona puede ver los análisis publicados • El plan anual de la membresía para la versión pro, que incluye privacidad de datos, ronda entre los \$144 y \$840 	
Excel	<ul style="list-style-type: none"> • Incluido en el paquete de Microsoft 360 (\$6.99 al mes) • Versión gratuita en el navegador • Acceso desde el ordenador y móvil • Permite el trabajo colaborativo 	
Infogr.am	<ul style="list-style-type: none"> • Por cada proyecto (10 máximo) pueden crearse hasta 5 páginas • Acceso a todos los tipos de gráficos interactivos y 13 mapas • Control de privacidad restringido • La membresía mensual pro que permite descargar, compartir de forma privada, y acceder a plantillas e imágenes premium vale \$19. 	
Piktochart	<ul style="list-style-type: none"> • Las infografías realizadas no incluyen marca de agua • Se tienen acceso a todas las plantillas disponibles en la plataforma • Permite el trabajo colaborativo y la realización de hasta 5 proyectos • La versión pro requiere de una membresía mensual por usuario de \$14 	

Tabla 5.3: Tabla comparativa, elaborada por el autor, de las herramientas más reconocidas para la Ciencia de Datos que no requieren experiencia en programación.

5.5. Aprendiendo a caminar como científico de datos

5.5.1. Recopilación de los datos

Isaac González (2017) [23] describe que los datos son el activo más importante para una empresa y sin ellos no se puede realizar análisis que permitan tomar mejores decisiones comerciales.

Por ello es importante definir cómo encontrar y extraer los datos correctos para analizar y modelar. El primer paso es identificar las fuentes de información con las que se cuenta y dónde se encuentran estas. Lo ideal sería tener los datos agrupados en un ‘*Data warehouse*’, pero en la realidad las empresas tienen los datos almacenados en distintas unidades como bases de datos, sistemas de computación, hojas de cálculo, archivos y documentos impresos. [43]

Una vez obtenidos los datos, se prosigue a integrar la información de todas las fuentes en una estructura ordenada y adecuada para la herramienta a utilizar durante el análisis. [23]

5.5.2. Preparación de los datos

Este proceso dispone los datos en un formato adecuado para realizar los análisis y modelaciones posteriores.

Limpieza

Son operaciones para corregir y eliminar datos innecesarios o equivocados. Se debe validar que cada variable del conjunto de datos no sea información obsoleta o incorrecta. Los valores en cada variable deben ser alguno de la lista de valores esperados o aceptados, por ejemplo, el monto de venta debe ser una cadena de números y no letras. [37]

Transformación

Nisbet, Elder y Miner (2009) [37] indican que este proceso consiste en expresar las variables en la forma que se necesite. Es importante que tanto las variables numéricas como categóricas tengan una estructura que permita responder, a través de los algoritmos, las preguntas planteadas al inicio. Algunos ejemplos serían convertir los formatos de fecha, uniformizar categorías como de “Sí” y “No”, o cambiar la orientación de los datos.

Balanceo

El desbalance de datos es un problema recurrente al momento de trabajar con datos clasificados en grupos o clases. Consiste en la dominancia de un grupo, es decir que en el conjunto de datos se tienen más registros de una clase que de otras. [20] De acuerdo con Bagnato (2019) [5], existen varias soluciones para balancear la distribución de los datos. La primera, conocida como “sub-muestreo” consiste en eliminar registros de la clase dominante. La segunda solución es replicar nuevos registros de la clase minoritaria y se conoce como “sobre-muestreo”. Otra manera de contrarrestar el desbalance es aplicar las primeras dos soluciones de forma simultánea.

5.5.3. Exploración de los datos

La teoría establece que por lo general el proceso de comprensión y preparación es un ciclo en el que se regresa y va entre estos dos grandes procesos. Esto se debe a que es necesario aprender más sobre el conjunto de datos para mejorar su preparación y viceversa. [37]

Estadísticas descriptivas

Consiste en obtener las métricas para cada variable del conjunto de datos y evaluar la relación entre las variables. También involucra realizar gráficas que faciliten la comprensión de la información estadística y relaciones complejas entre variables. [16] Las estadísticas descriptivas son:

Estadística descriptiva	¿Qué es?
Media	Es el promedio de todas las observaciones.
Mediana	El valor central en la distribución de los datos de forma ascendente o descendente.
Moda	Valor más frecuente en un conjunto de valores.

Estadística descriptiva	¿Qué es?
Desviación estándar	Indica qué tan dispersos se encuentran los datos con relación a la media.
Mínimo	Consiste en el valor más pequeño de la variable en estudio.
Máximo	Consiste en el valor más grande de la variable en estudio.
Tablas de frecuencia	Ordenación en forma de tabla que contiene todos los datos en estudio y la frecuencia de cada uno.
Histogramas	Representación gráfica de una variable en forma de barras proporcionales a la frecuencia de los valores representados que permite visualizar la distribución de estos.

Tabla 5.4: Conceptos de las estadísticas descriptivas básicas. [33] [8]

Filtración

Se selecciona un subconjunto de datos, que cumplen con ciertos requisitos, para enfocarse en ellos durante el análisis. [3] Algunos ejemplos de filtros pueden ser:

- Registros específicos seleccionados.
- Columnas necesarias seleccionadas.
- Filtrar por categorías como género.
- Filtrar por rango de valores como edad, ubicación, etc.

Abstracción

Consiste en extraer la información relevante para facilitar la interpretación de los datos. La estructura de la base de datos se modifica con el fin de adaptarla o mejorarla. Por ejemplo, si se tiene una base de datos con los gastos mensuales de los clientes y esta tiene 12 registros (filas) para cada cliente, se puede crear una nueva base de datos con 12 columnas, es decir una para el gasto de cada mes y una sola fila por cliente. [32]

Según Lavrac, Keravnou, & Zupan (2000) [29], la abstracción puede ser de diversos tipos:

- Cualitativa: a partir de una expresión numérica se crea una expresión cualitativa, en este caso de clasificación. Por ejemplo, si el análisis se enfoca en clientes entre 20 y 30 años, se crea una nueva variable llamada “adultos jóvenes” y los clientes en dicho rango de edad se abstraen con 1 en dicha variable y 0 para los que no aplique.
- Generalización: casos específicos se asignan a una clase. Por ejemplo, si se analizan preferencias centroamericanas, la variable “Nacionalidad” se abstrae con 1 en una variable denominada “Centroamericana” si la persona tiene registrado alguno de los 5 países centroamericanos en la variable “Nacionalidad” y 0 en caso contrario.

Perfilamiento

Observar las distribuciones de los datos en cada variable del conjunto. Es importante descubrir la tendencia central de los datos, el valor promedio, la desviación estándar, los cuartiles, los potenciales datos atípicos, cantidad de datos faltantes, datos sospechosos o inútiles. [37]

5.5.4. Diseño del primer informe

Un informe de análisis de datos contiene datos cuantitativos y cualitativos para evaluar las estrategias y rendimiento de una empresa y formular recomendaciones comerciales. [56].

Es indispensable que sea accesible para todos los miembros del equipo. Debe contener los **KPIs**: importantes de la empresa para evaluar cuáles objetivos y metas se están alcanzando y en dónde es necesario optimizar las actividades. Presentar las tendencias o patrones descubiertos facilitará el proceso para innovar y destacar. [54].

Basándose en los resultados del análisis, la empresa puede ofrecer contenido, servicios o productos aún más valiosos para los clientes. Es vital realizar un informe de calidad ya que funcionará como una fuente confiable de información para toda la empresa, permite una mejor comprensión del negocio y mejora la productividad. [56] A continuación, se encuentra una tabla con los pasos recomendados para desarrollar un informe de análisis de datos de forma correcta:

#	Paso	Descripción
1	Elegir la audiencia y problema	<p>El reporte debe ajustarse a la experiencia, situación y conocimiento de la audiencia a la que va dirigido. Es importante responder las preguntas: ¿Quién leerá este reporte? ¿Qué le interesa saber a esas personas? ¿Qué pregunta quiero responder a nivel empresarial?</p> <p>Con las respuestas obtenidas se define el lenguaje y vocabulario a utilizar, el nivel de tecnicismo, las gráficas a incluir y qué tan profundo se explicará cada sección.</p>
2	Estructurar el reporte	<p>Planificar las secciones que se incluirán y qué temas cubrirán cada una de ellas. Por lo general la estructura de un reporte de análisis debe contar con:</p> <ul style="list-style-type: none">■ Introducción■ Cuerpo (datos, métodos, análisis y resultados)■ Conclusiones y recomendaciones

#	Paso	Descripción
3	Desarrollar el cuerpo	<p>Es la sección más importante y se recomienda incluir las siguientes subsecciones</p> <ul style="list-style-type: none"> ■ <u>Datos</u>: especificar la fuentes y herramientas utilizadas para obtener los datos analizados. Para que el lector comprenda mejor el análisis, debe explicarse cuáles datos fueron utilizados y por qué. ■ <u>Métodos</u>: describir los métodos utilizados para analizar los datos, el por qué se seleccionaron y cómo estos fueron aplicados. ■ <u>Análisis</u>: presentar cómo se analizaron los datos con los métodos detallados anteriormente. Deben incluirse cálculos, gráficas y análisis paso por paso. Cada uno de estos debe tener una descripción clara y fácil de comprender ■ <u>Resultados</u>: es la sección de mayor interés porque ofrece nueva información, respuestas y evidencia sobre la situación actual de la empresa. Cada resultado debe de explicarse apoyándose de valores numéricos e indicar que representa de manera global. Se finaliza con un resumen general. Los resultados pueden presentarse en forma de imágenes, gráficas, tablas o mapas, cada uno con un título y descripción breve.
4	Escribir la introducción	Explicar de forma breve lo que cubrirá el informe para que la audiencia sepa qué esperar. Debe incluir la pregunta planteada al inicio y respondida con el análisis, además del contexto y antecedentes del análisis.
5	Agregar conclusiones y recomendaciones	La conclusión debe ser breve y concisa, señalar los hallazgos más importantes descritos en el cuerpo del análisis. Es importante incluir lo que implican dichos hallazgos para la empresa y qué se recomienda hacer mantener o mejorar los resultados.

Tabla 5.5: Pasos recomendados para realizar un informe de análisis de datos

Al finalizar un reporte, los expertos [15] recomiendan leerlo nuevamente por completo para evaluarlo de forma objetiva. En este proceso es importante estar dispuesto a modificaciones que mejoren el informe y análisis en general. La fase de evaluación y edición debe repetirse hasta estar satisfecho con el resultado final.

6.1. Planteamiento del problema

Como se describe en las secciones anteriores, las PYMES en Guatemala han sido afectadas negativamente por factores externos e internos. En un país en desarrollo como Guatemala, el término “Ciencia de los datos” es poco conocido. La ignorancia respecto al tema evita que los empresarios aprovechen este campo de la computación en beneficio de su empresa para tomar mejores decisiones de producción, mercadeo y comercialización.

Antes de continuar, es importante plantear algunas preguntas como: ¿Los gerentes y dueños de las PYMES conocen el término Ciencia de los Datos? ¿Qué parte de las PYMES cuentan con un departamento especializado en la ciencia de los datos para elaborar análisis que les ayuden a tomar mejores decisiones? ¿Cuántas de las PYMES en Guatemala han tenido acercamiento con la programación en Python para la ciencia de los datos? ¿Las PYMES en el país realizan algún tipo de análisis para las ventas realizadas? ¿Los datos que las empresas registran de las ventas realizadas son útiles para elaborar un análisis desde el enfoque de la ciencia de los datos? ¿Dónde y cómo registran las PYMES los datos de las ventas? ¿Los empresarios conocen el beneficio de elaborar análisis sobre los datos de las ventas para potenciar las mismas a corto plazo? ¿Es interesante para los empresarios aprender una forma sencilla y gratuita de implementar la ciencia de los datos en su negocio para analizar grandes cantidades de datos y con los resultados detectar las debilidades y oportunidades de su negocio? ¿Es útil para las PYMES de Guatemala implementar la herramienta de la ciencia de datos?

Con la hipótesis de que las PYMES en Guatemala no han tenido acercamiento con la Ciencia de Datos, pero es de su interés conocer dicha ciencia para potenciar sus ventas y comprender mejor la situación de sus propias empresas. Se propuso la realización de un manual dirigido a las PYMES, interesadas en conocer los principios básicos de la Ciencia de Datos, el cual proveerá los conocimientos básicos de dicha ciencia. Será un manual de fácil comprensión que incluirá el procedimiento necesario para realizar un pre-procesamiento y análisis de datos desde un enfoque empresarial.

6.2. Tipo de investigación

Este trabajo de investigación tiene un enfoque descriptivo y cuantitativo. Para ello, se realizó una recopilación bibliográfica sobre la temática y una investigación de campo (encuestas y grupo focal). Con el fin de conocer de primera mano la situación de las PYMES en Guatemala respecto a la Ciencia de Datos, la primera fase de la investigación de campo consistió en realizar una encuesta hacia un grupo muestra de PYMES. Para contar con otra perspectiva de la misma situación, en la siguiente fase se llevó a cabo un grupo focal con estudiantes universitarios expertos en la Ciencia de Datos. Finalmente se realizó un taller introductorio a la Ciencia de Datos al grupo muestra de PYMES en el que se indagó más sobre su situación en dicho campo y se validaron las herramientas desarrolladas.

6.3. Alcance

El primer universo de esta investigación fueron las 44,882 pequeñas y medianas empresas activas en Guatemala. La muestra utilizada de este universo fue el 0.05 %, es decir 25 PYMES.

El otro universo fueron los 5,393 estudiantes de la Universidad del Valle de Guatemala. Se utilizó el 0.07% de dicho universo para realizar el grupo focal a 4 estudiantes con experiencia en Administración de empresas y Ciencia de Datos.

6.4. Técnica de investigación

Se realizó una encuesta con “[Google forms](#)” y se envió al grupo muestra de las PYMES. Cada empresa ingresó su información de contacto para llevar un registro de las 25 PYMES involucradas. Todos los datos obtenidos fueron manejados con total confidencialidad. La encuesta constó de preguntas abiertas y cerradas. A continuación, la batería de preguntas:

- ¿Conoce el término Ciencia de Datos?
- ¿Cómo empresario conoce el beneficio de elaborar análisis sobre los datos de sus compras y ventas para potenciar las mismas a corto plazo?
- ¿Ha tenido acercamiento con la programación en Python?
- ¿Qué herramienta considera que sería de utilidad en su empresa para incrementar las ventas y tomar mejores decisiones respecto al negocio?
- ¿En su empresa se realiza algún tipo de análisis para las ventas realizadas?
- ¿Le gustaría aprender una forma sencilla y gratuita de implementar la Ciencia de Datos en su empresa para analizar grandes cantidades de datos y con cuyos resultados puede detectar las debilidades y oportunidades de su negocio?
- ¿Qué datos registra sobre las ventas realizadas (nombre de cliente, edad del cliente, ubicación, monto, productos, hora, modo de pago, etc.)?
- ¿En dónde registra estos datos?

El objetivo de la encuesta era responder a preguntas como: ¿Qué datos registran las PYMES sobre las compras y ventas realizadas (nombre de cliente, edad de cliente, ubicación, monto, productos, hora, modo de pago, etc.)? Para definir las variables del conjunto de datos ejemplo para el tablero y manual.

Posteriormente se organizó un grupo focal con el fin de tener una visión más clara de las actitudes, necesidades, intereses, recursos y motivaciones del grupo objetivo desde la perspectiva de estudiantes universitarios. La actividad fue orientada hacia la resolución de algunas preguntas como: ¿Qué herramienta es útil para aumentar la competitividad empresarial en las PYMES? ¿Qué herramienta es de interés para las PYMES que desean incrementar su competitividad empresarial? y ¿Con qué herramientas tecnológicas cuentan las PYMES para tomar mejores decisiones? Esto con el fin de desarrollar un tablero apto para el nivel tecnológico de las PYMES.

Finalmente, para conocer la opinión de las PYMES respecto al tablero y manual básico desarrollado, se llevó a cabo un taller introductorio a la Ciencia de Datos, con 7 de las PYMES participantes de la encuesta. Se plantearon preguntas que pretendían descubrir la utilidad que dichos materiales representan para las PYMES, así como los comentarios y sugerencias para futuros proyectos.

7.1. Encuesta

7.1.1. Convocatoria

La convocatoria a las PYMES se realizó enviando un boletín digital (Ver [Ilustración 7](#)) a través de diversos canales. Entre ellos la unidad de egresados de la Universidad del Valle, Centro de Recursos para Emprendedores en Acción de la Universidad del Valle, la Comercializadora del CampusTec, el Centro Municipal de Emprendimiento (CME) y la Secretaría Nacional de Ciencia y Tecnología (Senacyt).

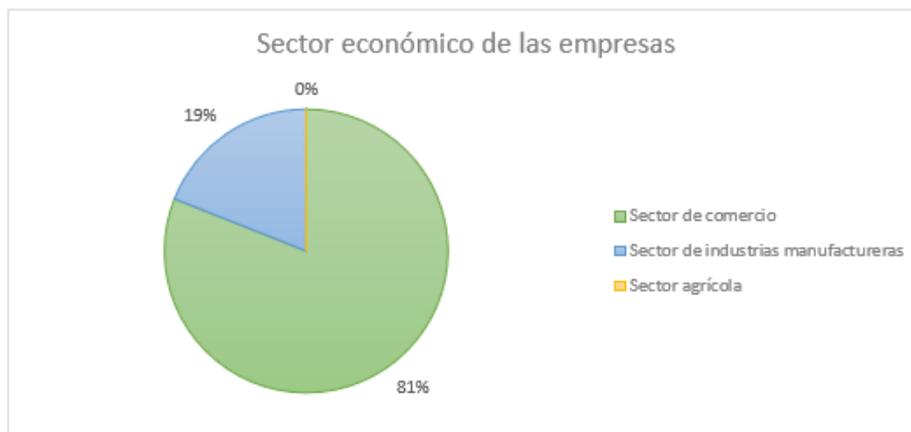
Se sabe que la convocatoria llegó a más de 50 PYMES y se esperaba obtener respuesta de mínimo 17 PYMES. Fueron 25 PYMES las que respondieron, además de ponerse a disposición para recibir el taller introductorio a la Ciencia de Datos.

7.1.2. Instrumento de investigación

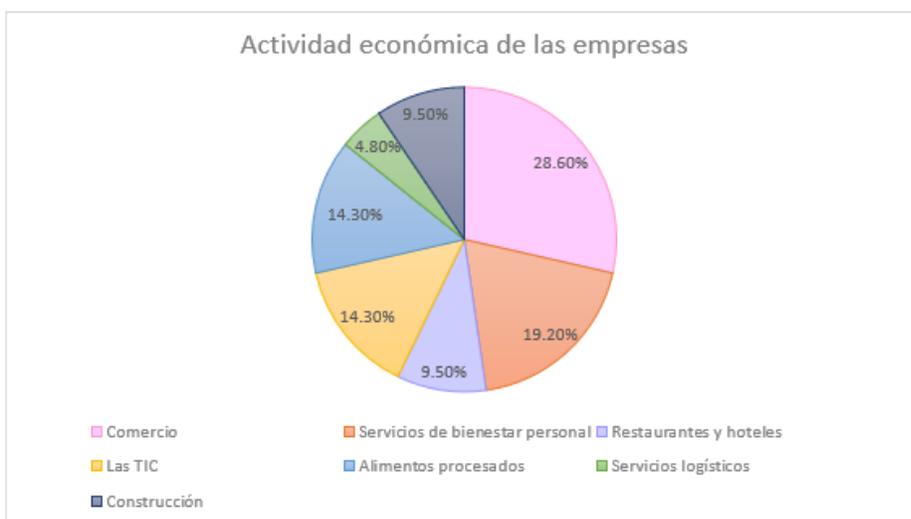
Enlace a la encuesta que se compartió con la muestra: <https://bit.ly/EncuestaCienciaDeDatos>

7.1.3. Resultados

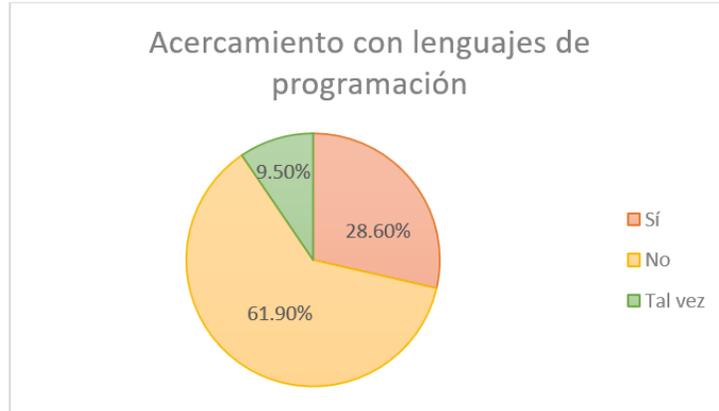
Los resultados de la pregunta No. 1 demuestran que el 81 % de las PYMES que respondieron a la encuesta pertenecen a los sectores de Comercio y 19 % a las Industria Manufacturera. La participación de empresas agrícolas fue nula en esta encuesta.



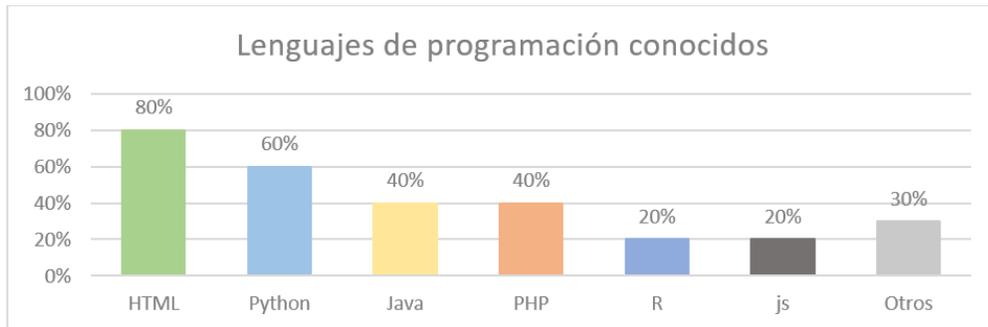
Respecto a la actividad económica de las PYMES participantes, el 28.6% se dedican a actividades comerciales y el 19.2% a servicios de bienestar personal como belleza y ejercicio. Un 14.3% de los participantes indicaron dedicarse a actividades relacionadas con las TICs y otro 14.3% a la producción alimentos procesados. El 9.5% de las PYMES indican dedicarse a la construcción, otro 9.5% a actividades de hotelería y restaurantes. El 4.8% restante se dedican a ofrecer servicios logísticos.



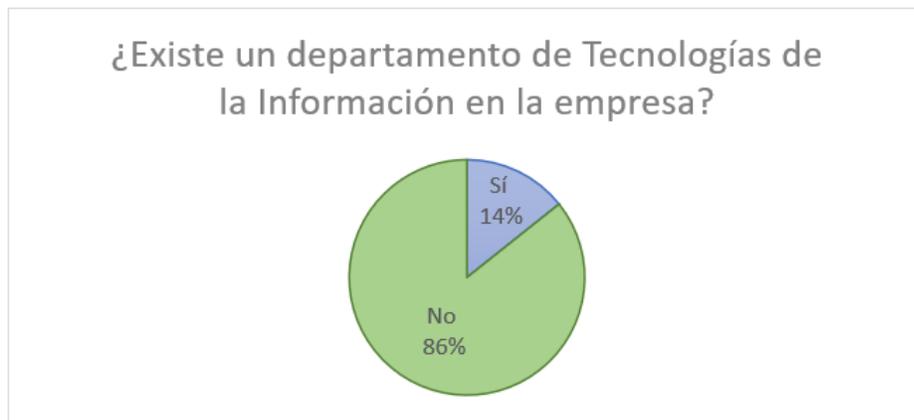
Con la pregunta No. 3 se descubrió que el 28.6% de las PYMES han tenido algún tipo de experiencia con lenguajes de programación. La experiencia con lenguajes de programación es nula en el 61.9% de las empresas y 9.5% no están seguros.



El 80 % de las PYMES, que han tenido cierta experiencia con algún lenguaje de programación, indicaron conocer HTML. El 60 % conoce Python y tanto Java como PHP son conocidos por el 40 % de las PYMES. JavaScript (js) y R son lenguajes conocidos por el 20 % respectivamente. El 30 % de las PYMES indicaron conocer otros lenguajes de programación, entre ellos C++, Fortran, Odoos y .Net.



Los datos indican que un 86 % de las PYMES no cuentan con un departamento de Tecnologías de la Información (IT).



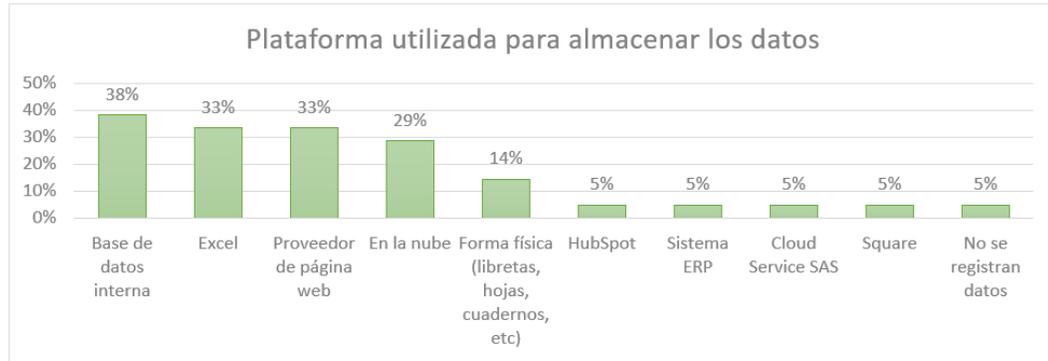
Varias (71 %) PYMES sí realizan algún tipo de análisis para las ventas realizadas. El 19 % de las empresas encuestadas no planifica dentro de sus actividades algún tipo de análisis sobre las ventas realizadas y el 10 % no están segura de hacerlo o no. Las empresas que afirmaron analizar sus ventas señalaron hacerlo a través de indicadores como KPIs y estadísticas de redes sociales. Otras indicaron utilizar sistemas de terceros como Square App, mynube y CRM.



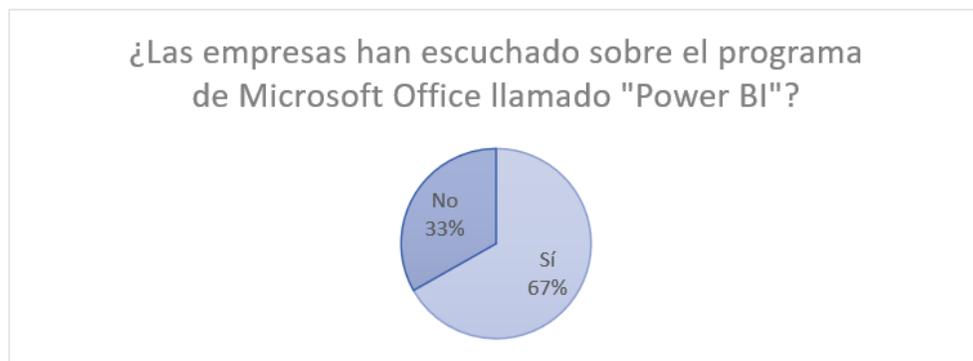
Para las ventas realizadas, el 95 % de las PYMES indicaron registrar el monto de la venta y 91 % los productos vendidos. El 86 % toma en cuenta tanto el nombre del cliente como el modo de pago seleccionado por el mismo. El 48 % registra la ubicación del cliente y 38 % la hora de la venta. Tanto la edad del cliente, como los datos de contacto y NIT son datos registrados por el 14 % de las PYMES. El género, cómo el cliente llegó a la empresa y cuáles son los datos de envío es información registrada únicamente por el 5 % de las PYMES. El 5 % de las empresas proyectan no registrar datos sobre las ventas realizadas.



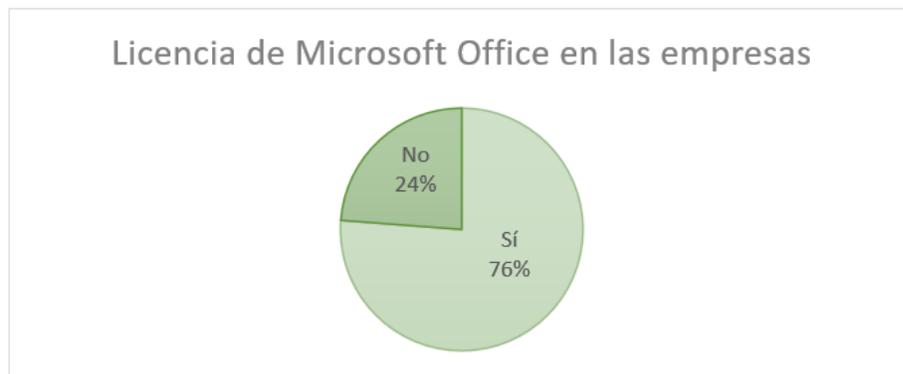
El 38 % de las PYMES almacenan los datos en una base de datos interna, mientras que un 33 % lo hace con la aplicación Excel y otro 33 % en el espacio habilitado de su página web. El 29 % indicaron llevar los registros en la nube y 14 % de forma física. Las PYMES indicaron otras plataformas de almacenamiento como HubSpot, sistemas ERP, Cloud Service SAS y Square con una concurrencia del 5 % cada uno. Mientras que un 5 % de los participantes no registran datos.



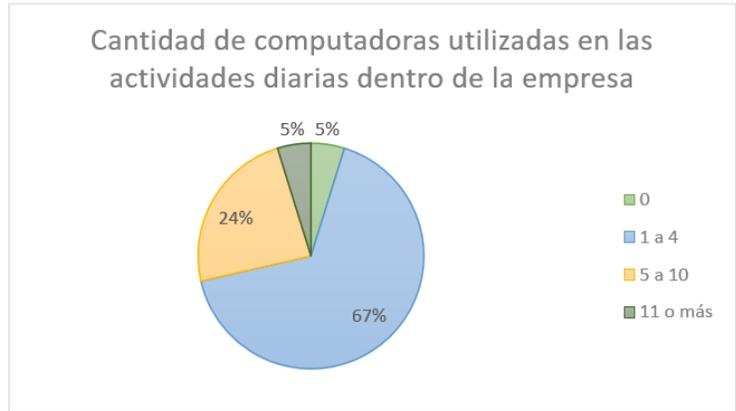
Se observó que el 33 % de las PYMES han escuchado sobre “Power BI”, un programa desarrollado por Microsoft Office.



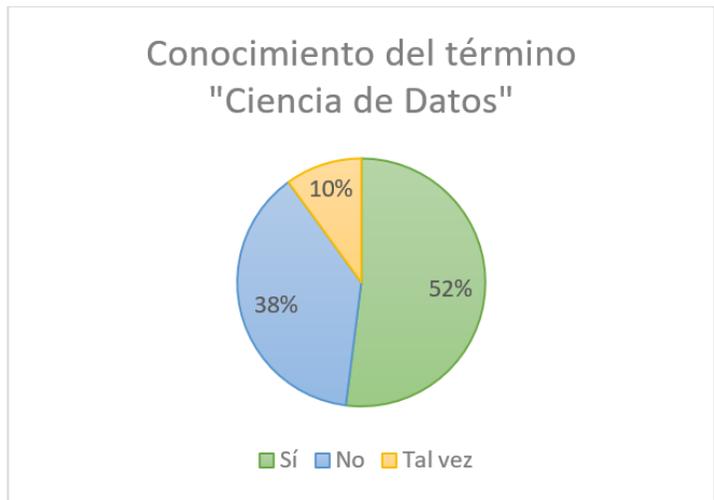
La mayoría (76 %) de las PYMES cuentan con una licencia de Microsoft Office vigente en su organización.



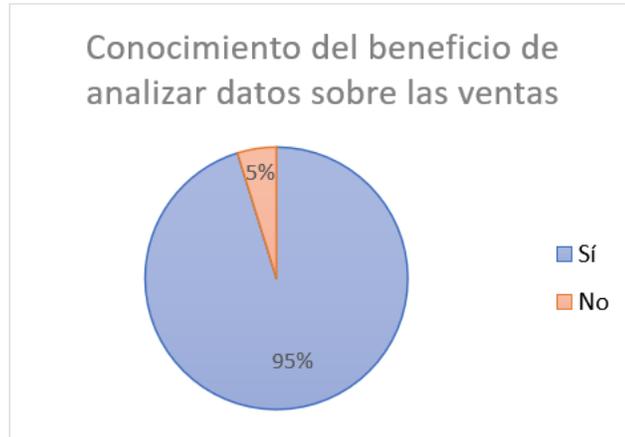
Respecto a la cantidad de computadoras utilizadas en las actividades diarias dentro de las empresas, el 67 % de las PYMES indicaron tener entre 1 y 4 computadoras. El 24 % utiliza entre 5 y 10 computadoras, otro 4.8 % más de 11 computadoras y únicamente el 5 % indicaron no utilizar computadora en sus actividades diarias.



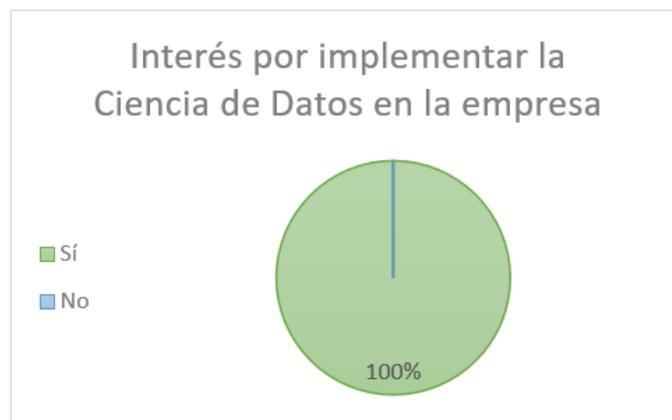
El 52% de las PYMES conocen de alguna manera el término “Ciencia de Datos”. Sin embargo, al profundizar en la pregunta el 84% de los participantes indicaron no implementar la “Ciencia de Datos” en sus empresas. Esto confirma la hipótesis de que la mayoría de las PYMES no han tenido acercamiento con la Ciencia de Datos.



El 95% de las PYMES participantes conocen los beneficios que trae el análisis de datos sobre las ventas.



Con la última pregunta: ¿Le gustaría aprender una forma sencilla y gratuita de implementar la Ciencia de Datos en su empresa para analizar grandes cantidades de datos y con cuyos resultados puede detectar las debilidades y oportunidades de su negocio?, fue posible confirmar la hipótesis. El 100 % de las PYMES están interesadas en implementar la Ciencia de Datos en su empresa para obtener los beneficios de esta.



7.2. Grupo focal

7.2.1. Alcance e hipótesis

El objetivo de este grupo focal era conocer, desde la perspectiva de estudiantes universitarios en el campo de la Ciencia de Datos, la situación de las PYMES en dicho campo y si realmente un manual sería útil para las PYMES en la implementación de este.

La información brindada por los estudiantes permitió conocer qué herramientas para la Ciencia de Datos, en el ámbito empresarial, les enseñan a utilizar en los cursos universitarios. Esto con el fin de poder conocer nuevas herramientas, programas en tendencia y la dificultad que representan para personas no ingenieras en Ciencias de la Computación.

7.2.2. Información general

Fecha de realización:	30 de julio de 2022
Horario:	9:00am – 10:30am
Modalidad:	Virtual, vía zoom
Cantidad de participantes:	4

7.2.3. Participantes

Nombre	Carrera	Año
Apolo Alvarado Calderón	Ingeniería en Ciencia de la Administración	4to
Joseline Jerline Gómez Rodríguez	Ingeniería en Ciencia de la Administración	4to
María Mercedes Retolaza Reyna	Ingeniería en Ciencias de la Computación y TI	5to
Kevin Rolando Romero Lorenzana	MBA+	Maestría

7.2.4. Responsables

Nombre	Responsabilidad	Carrera	Año
Paula Camila González Ortega	Moderadora	Ingeniería en Ciencias de la Computación y TI	5to
María Inés Vásquez Figueroa	Co-moderadora	Ingeniería en Ciencias de la Computación y TI	5to

7.2.5. Convocatoria

La convocatoria se realizó a través del nodo “Comunidad UVG” del correo institucional. Se solicitó el apoyo de estudiantes, de 4to y 5to año con experiencia en administración o ciencia de datos, para participar en el grupo focal a cambio de horas beca.

7.2.6. Hallazgos**

PREGUNTAS	RESULTADO
¿Cuál es su PYMES nacional favorita? ¿Por qué?	Al decir empresa nacional piensan automáticamente en grandes empresas. Identificar una PYME nacional les tomó un poco tiempo.
¿Sabes qué es la ciencia de datos?	Cada participante dio su definición de ciencia de datos, todas fueron correctas y se notó que sabían sobre el tema.
¿Cuál es su experiencia con la ciencia de datos?	Tuvieron su primera experiencia con la Ciencia de Datos en el curso de “Data Mining” en la UVG.
¿Qué opinan y cuál ha sido su experiencia con Power BI?	Para usar Power BI no es necesario tener experiencia en programación por ello es el favorito de todos.

PREGUNTAS	RESULTADO
¿Conoce los beneficios de la ciencia de datos?	Los datos son el activo más poderoso de cualquier empresa por lo que analizarlos permite encontrar nuevas estrategias de marketing y administración, así como identificar que las metas internas de la empresa se estén cumpliendo según los KPIs.
¿Conoce alguna empresa que aplique ciencia de datos y cómo lo aplica?	Todas las empresas mencionadas fueron grandes o macroempresas como Tigo, McDonald' s y la Cervecería Centroamericana. La posible razón es que estas si cuentan con la capacidad económica para invertir en la Ciencia de Datos.
¿Qué herramienta es útil para aumentar la competitividad empresarial en las PYMES?	Fue una pregunta difícil de responder para todos. Luego de la primera intervención, ya los demás participaron como lo venían haciendo en las preguntas anteriores (dinámicamente)
¿Qué información (qué desean consultar) es de interés para las PYMES que desean incrementar su competitividad empresarial? ¿En dónde la registran?	- Todos coinciden que las PYMES utilizan Excel para almacenar sus datos. - Se infirió que al inicio las PYMES no conocen la importancia de almacenar estos datos y el análisis de estos. Eso estanca su proceso de crecimiento y reduce su competitividad empresarial.
¿Con qué herramientas tecnológicas cuentan las PYMES para tomar mejores decisiones?	Tener una estructura sólida en internet le facilita la recopilación de datos a las empresas.
¿Cree usted que un manual básico de ciencia de datos sería útil y de interés para las PYMES que desean implementar esta metodología?	- Todos confirman que un manual básico de la Ciencia de Datos será de interés para las PYMES. - Resaltan la importancia de mencionar la utilidad y beneficios de este método para atraer a más empresas.
¿Si existiera un manual básico sobre la ciencia de datos dirigido a las PYMES, qué procesos y conceptos de esta metodología deberían de incluirse?	Recalaron que sea un manual muy básico y amigable, que enseñe qué es, la importancia y utilidades de la Ciencia de Datos y no un proceso de análisis específico para cada modelo de negocio.

**Tabla de respuesta en [Tabla 5](#) de anexos

7.3. Taller introductorio

7.3.1. Información general

Fecha de realización:	19 de octubre de 2022
Horario:	4:30pm – 5:30pm
Modalidad:	Virtual, vía zoom

7.3.2. Convocatoria

La convocatoria se realizó vía correo electrónico a las 25 PYMES muestra. Se envió la invitación de una reunión por zoom para introducirlos a la Ciencia de Datos y tablero desarrollado, pero únicamente 7 PYMES confirmaron su asistencia.

7.3.3. Hallazgos

PREGUNTAS	RESULTADO
¿Creen que le suma valor a su empresa?	Definitivamente sí. La Ciencia de Datos en Guatemala es un campo en “pañales” y lo que en este momento se necesita es evangelizar a las empresas sobre el tema, dar a conocer qué es, para qué sirve y su importancia.
¿Lo usarían? ¿Sí/no y por qué?	Lo usarían si alguien les explica previamente cómo usar el tablero y porqué usarlo. Las cooperativas que apoyan en el tema de exportaciones a las empresas del interior serían un aliado perfecto para educar a las PYMES sobre la Ciencia de Datos y este tablero con el fin de que ellas identifiquen a cuáles países exportan más productos, qué producto es el más exportado, etc.
¿Utilizan otra plataforma que haga algo similar?	Algunas empresas utilizan Power BI, pero son muy pocas por el alto precio de este. Lo más cercano a la Ciencia de Datos que la mayoría de las PYMES utilizan son las redes sociales y KPIs. Algunas utilizan otras plataformas como square app, Odoos y Mynube, pero no tienen un proceso de limpieza, exploración y análisis de datos establecido además de lo que dichas aplicaciones les ofrecen.
¿Les gustaría algo más personalizado?	Sí. Si este fuera un modelo de negocio, un servicio de consultoría sería lo más valioso que se podría ofrecer en conjunto con este tablero. Sería una buena estrategia de negocio tener la opción de que las empresas adquieran un tablero desarrollado específicamente para sus necesidades.
Comentario y sugerencias	El término “Ciencia de Datos” puede ser intimidante para muchas PYMES, lo mejor es explicarles el tema con lenguaje no técnico y desde cero.

8.1. ¿Por qué un manual de Ciencia de los Datos?

De acuerdo con Duhalt Krauss (1968) [14] un manual es una herramienta que transmite de forma ordenada y sistemática los conocimientos, experiencias y términos relacionados a un tema en específico.

Un manual de Ciencia de Datos es útil en las organizaciones porque permiten la normalización de las operaciones, reducir el desperdicio de tiempo, multiplicar la cultura organizacional en esta ciencia e incrementar o consolidar la posición competitiva. Contar con un manual de este método reduce la dependencia de contar con un experto en el tema como parte del equipo de trabajo y hace posible la introducción al análisis de datos básico. Sin embargo, esta dependencia no desaparece por completo. Un manual no cuenta con todos los conocimientos del experto, quien por lo general es el encargado de elaborar dicho material. [2]

8.2. Contenido del manual de Ciencia de Datos

Siguiendo las sugerencias obtenidas durante el grupo focal con los estudiantes, se planificó el contenido a abordar en el manual de usuario. Se estableció que el propósito principal de este sería orientar al lector sobre los conceptos básicos sobre la Ciencia de Datos desde una perspectiva empresarial, en la que el conocimiento previo sobre el método sea casi nulo. Así como un guía de uso del tablero en el que prepararán y analizarán los datos de su negocio.

Secciones por incluirse en el manual de usuario:

1. Portada
2. Índice
3. Introducción

4. Glosario
5. Ciencia de Datos en el ámbito empresarial
 - a) ¿Qué es?
 - b) Importancia
 - c) Habilidades necesarias
 - d) Ejemplos de aplicación
 - e) Plataformas para su aplicación sin necesidad de programación
6. Guía de uso del tablero
 - a) Requerimientos
 - b) Limpieza
 - c) Exploración
 - d) Modelos
7. Referencias

8.3. Manual

La introducción del manual describe de forma breve el contenido que el usuario encontrará en el manual con el fin de llamar su atención e introducirlo al mundo de la Ciencia de Datos. Para el glosario se seleccionaron los conceptos básicos relacionados a la Ciencia de Datos, especialmente a los procesos de limpieza y preparación, así como algunos otros términos que se abordan en las siguientes secciones del manual. Los conceptos fueron desarrollados de la manera más sencilla posible para que cualquier persona, sin importar su experiencia en el campo de la computación, comprenda su significado y por lo tanto el resto del manual sea mucho más fácil de asimilar.

La sección 5 del manual, es decir “La Ciencia de Datos en el ámbito empresarial” buscan sumergir a las PYMES en el mundo científico desde la perspectiva de negocios, la conocida por ellos. En esta parte del manual se comparte a grandes rasgos en qué consiste la implementación de la Ciencia de Datos dentro de una empresa, así como los beneficios que esta trae, la importancia de esta en los distintos departamentos de un negocio y las habilidades necesarias para llevar a cabo un análisis científico de datos.

Se enfatiza que no es indispensable contar con conocimientos en el campo de la computación para iniciar en el mundo de la Ciencia de Datos. Esto con el fin de motivar a los empresarios a implementar dicha ciencia sin importar el área de su formación profesional. Se describen las aplicaciones de dicha ciencia, en empresas reconocidas como Coca-Cola, Netflix y Metlife, para demostrarle a las PYMES algunos de los beneficios que pueden obtener al implementar la Ciencia de Datos en sus organizaciones y la expansión de este método a nivel mundial. Para finalizar se muestra comparten 4 plataformas que permiten ejecutar procesos de la Ciencia de Datos a nivel empresarial sin la necesidad de tener experiencia en informática. En la sección 6 se comparte el enlace para acceder al tablero desde cualquier navegador y una breve explicación lo que podrán hacer en él. Se continua con el listado de requerimientos para su uso en temas de tecnología e información, especialmente el archivo .csv con los datos a analizar que debe contar mínimo con las siguientes columnas:

- No. de factura o transacción
- Fecha de la venta o compra
- Monto de la venta o compra
- Producto o servicio vendido o adquirido
- Cantidad de productos vendidos/adquiridos

En esa misma sección se explica cómo transformar un archivo .xlsx a .csv desde Excel, ya que el tipo de archivos más comunes a nivel empresarial son .xlsx. Una vez cubierta dicha información se comienza con la guía, paso a paso, para el uso del tablero. Cada paso cuenta con una imagen que oriente al usuario todo lo posible desde la carga de archivos, hasta la limpieza, exploración y uso de modelos. En la parte de exploración se explica cómo descargar los reportes HTML y se incluye un enlace a [youtube](#) (Recurso 6 en anexos) con una demostración de cómo analizarlos, ya que dicho proceso se explica mejor de manera audiovisual. El manual de usuario puede ser visualizado en este [enlace](#) (Recurso 2 en anexos)

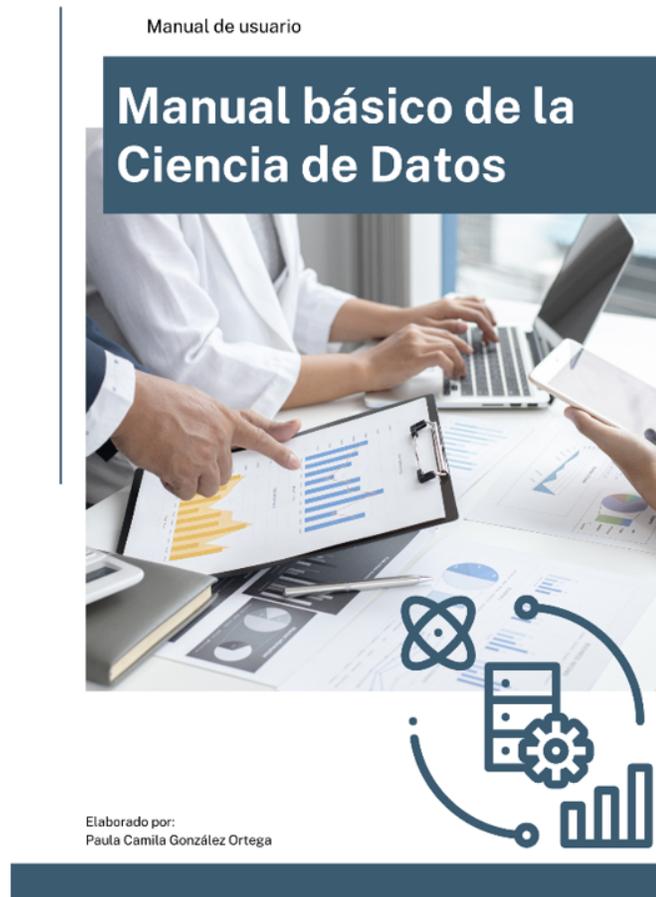


Figura 8.1: Portada del manual de usuario elaborado para las PYMES

8.4. Tablero

El primer paso fue buscar conjuntos de datos en internet para utilizarlos como base. Estos fueron modificados de tal manera que contaran con las variables que la mayoría de las PYMES indicó registrar de sus ventas y compras. De este proceso se obtuvieron [2 conjuntos de datos](#) (Recurso 3 en anexos), cada uno de un giro de negocio distinto: pastelería y floristería.

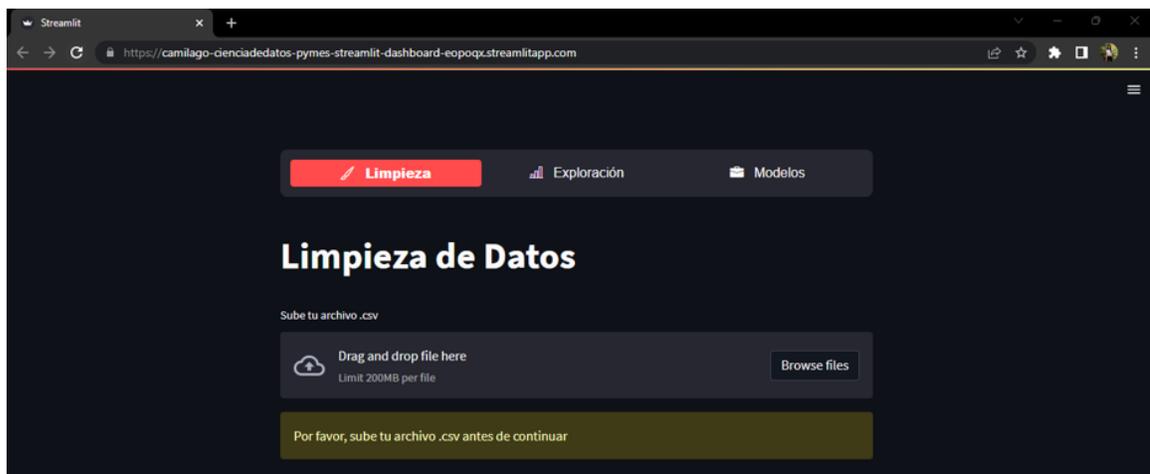
El siguiente paso consistió en la elaboración del tablero con el que las PYMES incursionarán en la Ciencia de Datos. Se utilizó la librería Streamlit en conjunto con muchas otras, con el fin de crear un tipo de caja negra en la que los gerentes de empresas carguen el conjunto de datos, visualicen una primera exploración, seleccionen los procesos de limpieza a ejecutar, generen los

reportes que permitan la exploración y análisis de los datos y finalmente utilicen alguno de los modelos de aprendizaje automático.

Se seleccionó Streamlit para el desarrollo de tablero por la facilidad de compartir el recurso. A diferencia de programas como Power BI y Tableau, el tablero desarrollado se puede compartir de forma gratuita a través de un link sin necesidad que el usuario final instale o configure algo en su computadora.

Si bien Power BI tiene una interfaz bastante intuitiva, su licencia al mes es de \$9.90 y con la información recopilada en este trabajo era imposible asumir que todas las empresas estarían dispuestas a invertir en ello. Además, la versión gratuita restringe compartir, editar o consumir tableros o informes desarrollados por otros usuarios, así como la descarga de reportes interactivos (los cuales si fueron generados utilizando la librería Lux).

La opción de utilizar Tableau fue descartada, ya que su interfaz no es fácil de utilizar y la mayoría de PYMES indicaron no tener expertos en computación en su equipo de trabajo. Además, con la versión gratuita los análisis realizados por las PYMES serían públicos y eso podría representar un peligro para su estabilidad.



Se crearon diferentes funciones para la limpieza de los conjuntos de datos. Las opciones habilitadas son:

- Eliminar caracteres especiales
- Cambiar mayúsculas a minúsculas
- Convertir en mayúscula la primera letra de cada palabra
- Eliminar filas duplicadas
- Eliminar valores nulos
- Cambiar los tipos de datos

¡Limpieza de datos!

Selecciona las técnicas de limpieza que deseas realizar:

- Eliminar caracteres especiales
- Cambiar mayúsculas a minúsculas
- Convertir en mayúscula la primera letra de cada palabra
- Eliminar filas duplicadas
- Eliminar valores nulos
- Cambiar los tipos de datos

Posteriormente se utilizaron la librería Lux y PandasProfiling para generar reportes interactivos y estáticos, respectivamente. Estos se elaboran con los datos limpios obtenidos en el proceso anterior. Con dichos reportes se obtiene información valiosa sobre los negocios analizados. Estas librerías fueron seleccionadas por generar una gran variedad de gráficas, entre ellas la de los productos más vendidos, ventas por día o mes, comparación de ventas entre años distintos, correlación entre las variables, etc. y al descargarlas en formato HTML facilita su visualización en las empresas.

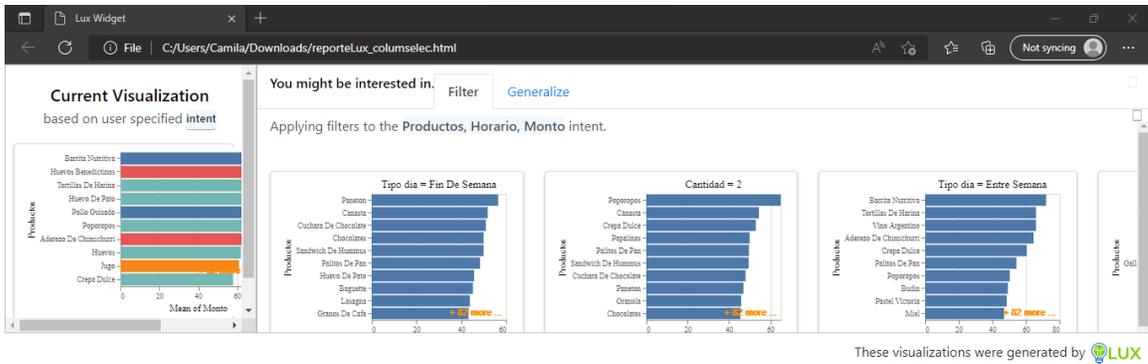


Figura 8.2: Reporte interactivo elaborado con los datos ejemplos

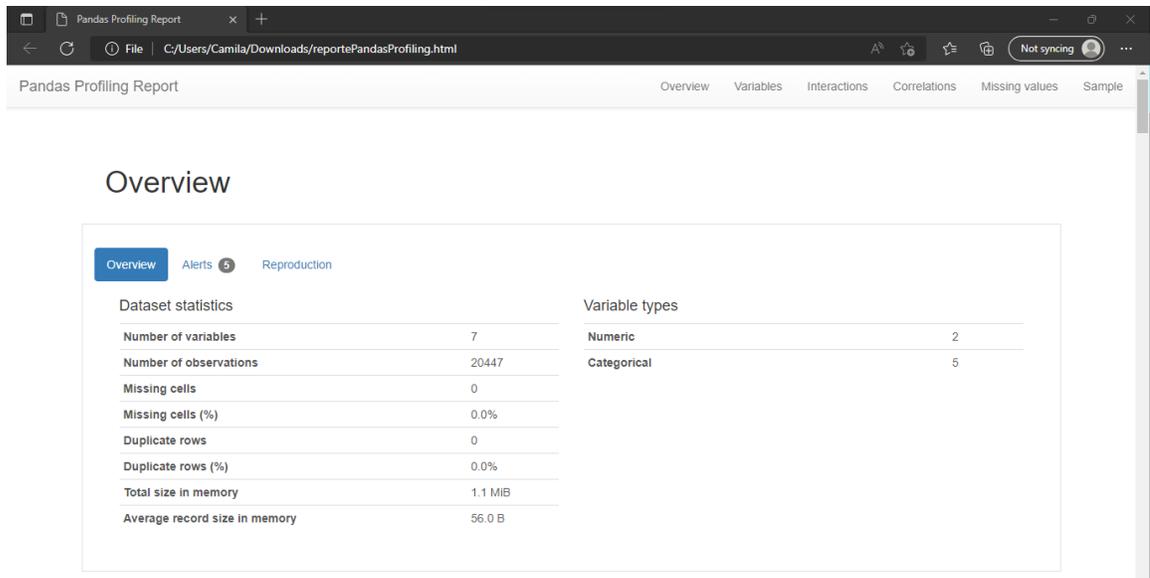


Figura 8.3: Reporte estático elaborado con los datos ejemplos

Finalmente se implementó una sección extra de aprendizaje automático utilizando como base los conjuntos de datos elaborados en el primer paso y los intereses de las PYMES. Para el conjunto de datos de la pastelería se implementó un modelo de asociación utilizando el Algoritmo Apriori y ciertas reglas de asociación de la librería mlxtend. Esto con el fin de ofrecer a los interesados un ejemplo de cómo descubrir la combinación más frecuente de productos comprados por los clientes. El ejemplo del modelo de regresión lineal se realizó con los datos de la floristería y la librería sklearn. Este le permite al usuario ingresar el año, mes o día de la semana del que desea predecir la demanda.

Es importante mencionar que todo el proceso de preparación y entrenamiento de los modelos es invisible para el usuario. Luego de cargar el conjunto de datos limpios, se procede a categorizar el campo de productos y a extraer el mes, día y año de la fecha para posteriormente categorizarlos. Una vez terminado este proceso, se evalúa el modelo que el usuario desea ejecutar para realizar el respectivo entrenamiento y mostrar los resultados al usuario.

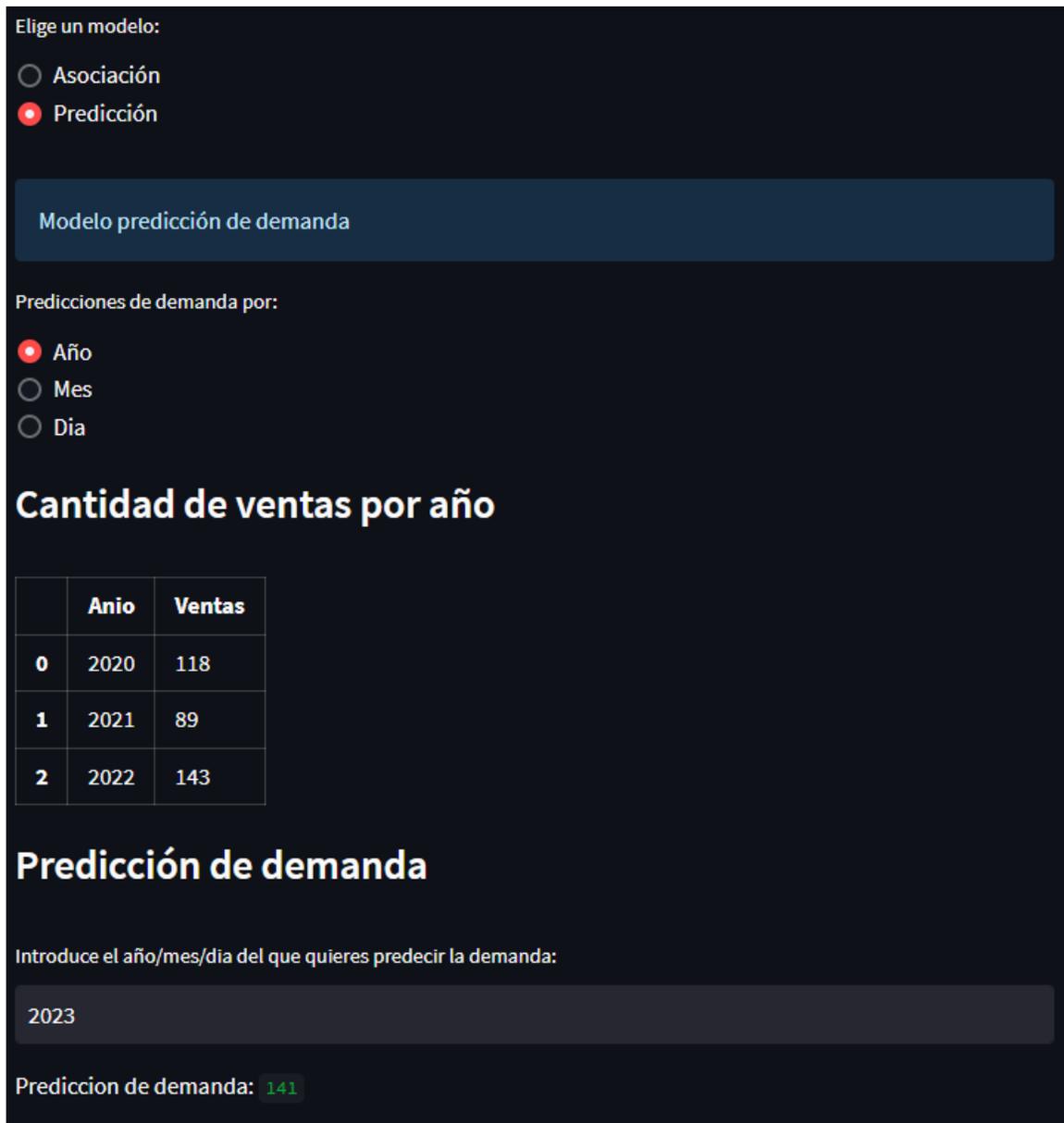


Figura 8.4: Modelo de regresión lineal para predecir la demanda del año ingresado

Para el despliegue del tablero se utilizó la plataforma gratuita de Share Streamlit asociada con Github. Actualmente el tablero se encuentra disponible en este [enlace](#) (Recurso 4 en anexos). Para ver más detalles de cada proceso y el código desarrollado se recomienda leer el [manual técnico](#) (Recurso 5 en anexos) y visitar el [repositorio de Github](#) (Recurso 6 en anexos).

- A través de la encuesta realizada se confirmó que el 84 % de las PYMES no han tenido acercamiento con la Ciencia de Datos, pero el 100 % están interesadas en conocer dicha ciencia para potenciar sus ventas y comprender mejor la situación de sus propias empresas.
- En el taller introductorio las PYMES confirmaron que los materiales desarrollados sí suman valor a su empresa y serán herramientas utilizadas en su negocio para tomar mejores decisiones, comprender mejor la situación de la empresa y encontrar tanto debilidades como fortalezas.
- Los métodos aplicados para el manejo y predicción de datos fueron explicados a las PYMES de forma general y sin términos técnicos porque el 86 % de las PYMES no cuentan con un departamento de TI y el 62 % no tienen experiencia con lenguajes de programación.
- Las PYMES aprobaron el manual básico y accesible sobre los principios fundamentales de la Ciencia de Datos que se creó en este proyecto.
- Si se elaboró un conjunto de datos ejemplo para que las PYMES lo replicaran con los datos de su empresa, pero fue necesario reducir el número de variables porque las PYMES registran pocas variables de sus ventas.
- El conjunto de datos ejemplo no contaba con las variables suficientes para elaborar una muestra, valiosa y de calidad, del proceso de exploración y análisis de datos.
- El tablero desarrollado, enfocado en la limpieza y exploración de Datos, es accesible para las PYMES desde cualquier navegador sin necesidad de instalar nada en sus computadoras.
- Las PYMES que implementan la Ciencia de Datos en su estrategia de negocio, llevan una ventaja competitiva por encima de aquellas que no lo hacen. Accenture (2016) lo confirma al indicar que se registran incrementos en su productividad laboral de un 40 %, se mejoran los planes de mercado y ventas y de acuerdo con Harris (2013) se reportan altas ganancias de sus inversiones y mayor satisfacción de sus clientes.
- Las PYMES que participaron en el taller introductorio indicaron que dicho taller es indispensable para saber cómo utilizar el tablero desarrollado, ya que antes de implementar nuevas herramientas o procesos prefieren contar con servicios tipo consultoría.
- Las PYMES mostraron interés en el taller y material sobre la Ciencia de Datos, pero la mayoría no asistió a la actividad por diversas razones.

- Además de un tablero y manual, se recomienda planificar un programa de talleres que eduque a las PYMES sobre la Ciencia de Datos desde cero porque, así como existen empresarios que conocen del término hay otros que nunca lo han escuchado.
- Se recomienda crear alianzas con cooperativas que trabajan con PYMES del interior para exportar sus productos. A través de ellas se podría difundir el concepto e importancia de la Ciencia de Datos en las empresas y las herramientas desarrolladas en este proyecto.
- Para futuras investigaciones se sugiere definir grupos muestra más grandes para reducir los márgenes de error e incrementar la confidencialidad de las respuestas.
- Ya que el 76 % de las PYMES cuentan con una licencia de Microsoft Office vigente en su organización, se recomienda enseñarle a las PYMES cómo aprovechar las herramientas de Excel para realizar análisis de datos.
- A los investigadores se les recomienda enfocarse en una industria específica con el fin de poder diseñar un tablero específicamente para las necesidades de las PYMES de esa industria y no tan generalizado como el de este proyecto.
- A las PYMES se les recomienda obtener más información sobre sus clientes (edad, ubicación, identidad de género, etc.), ya sea a través de los registros de ventas o redes sociales. Esto permitirá elaborar un análisis de datos de calidad para potenciar sus ventas y compras, mejorar las decisiones a nivel empresarial y ofrecer una visión clara de la situación del negocio.
- Para futuras investigaciones se recomienda crear un conjunto de datos ejemplo con más variables que el utilizado en este proyecto. Esto con el fin de enseñarle a las PYMES qué variables registrar de sus ventas y compras para poder elaborar análisis de datos valiosos y de calidad.
- Se recomienda investigar más sobre la disposición de las PYMES para invertir tiempo y dinero para el aprendizaje e implementación de la Ciencia de Datos. Ya que pueden estar interesadas en el tema, pero no dispuestas a invertir.
- A los próximos investigadores se les sugiere desarrollar soluciones con programas como Power BI o Tableau, los cuales ofrecen ya herramientas preparadas para el análisis y exploración de datos.

-
- [1] Accenture: *AI and the Future of Growth*. En *AI and the Future of Growth*, 2016.
- [2] Alvarez Torres, Martin: *Manual para elaborar manuales de políticas y procedimientos*. Panorama Editorial, México, 1996.
- [3] ArcGIS Pro: *Filtrar datos en la tabla*, 2020 [Online]. <https://pro.arcgis.com/es/pro-app/2.9/help/data/tables/filter-data-in-the-table.htm>.
- [4] Arias, Emilio: *Las 20 mejores librerías para Data Scientists en R*, 2019 [Online]. <https://todobi.com/las-20-mejores-librerias-para-data/>.
- [5] Bagnato, Juan: *Aprende Machine Learning*, 2019 [Online]. <https://www.aprendemachinelearning.com/clasificacion-con-datos-desbalanceados/>.
- [6] Bagnato, Juan: *10 Librerías Python para Data Science y Machine Learning*, 2020 [Online]. <https://blogvisionarios.com/articulos-ia/10-librerias-python-para-data-science-y-machine-learning/>.
- [7] Bowne-Anderson, Hugo: *What Data Scientists Really Do, According to 35 Data Scientists*, 15 de Febrero, 2018. <https://www.sudeep.co/data-science/2018/02/09/Understanding-the-Data-Science-Lifecycle.html>.
- [8] Cabral, Angélica, Francisco Cimé, Román Cob y Elisa Zapata: *Probabilidad y Estadística*. Colegio de Estudios Científicos y Tecnológicos del Estado de Campeche, México, 2021.
- [9] Cambridge Spark: *Why Economists should embrace Data Science*, 20 de Abril, 2022. <https://www.cambridgespark.com/info/why-economists-should-embrace-data-science>.
- [10] Costa, Claire: *Las Mejores Herramientas De Ciencia De Datos Para Los Científicos De Datos*, 2021 [Online]. <https://www.datasource.ai/es/data-science-articles/las-mejores-herramientas-de-ciencia-de-datos-para-los-cientificos-de-datos>.
- [11] Data Flair: *Data Science for Business – 7 Major Implementations of Data Science in Businesses*, 2022. <https://data-flair.training/blogs/data-science-for-business/>.
- [12] Data Science: *¿Qué es la ciencia de los datos?*, 2020. <https://datascience.eu/es/aprendizaje-automatico/que-es-la-ciencia-de-los-datos/>.
- [13] devCamp: *Qué es una librería en programación*, 2020 [Online]. <https://devcamp.es/que-es-libreria-programacion/>.

- [14] Duhalt Krauss, Miguel: *Los manuales de procedimientos en las oficinas públicas*. UNAM [Facultad de Ciencias Políticas y Sociales], México, 1968.
- [15] Fender, Jessica: *Writing a Good Data Analysis Report: 7 Steps*, 2022 [Online]. <https://www.modernanalyst.com/Resources/Articles/tabid/115/ID/6070/Writing-a-Good-Data-Analysis-Report-7-Steps.aspx>.
- [16] Fernández Fernández, Santiago, José María Cordero Sánchez y Alejandro Córdoba Largo: *Estadística descriptiva*. ESIC Editorial, Madrid, 2002.
- [17] Foote, Keith: *A Brief History of Data Science*, 16 de Octubre, 2021. <https://www.dataversity.net/brief-history-data-science/#>.
- [18] García, Fernan: *RStudio, IDE para programar con R. Instalación y primeros pasos*, 2010 [Online]. <https://www.arsys.es/blog/rstudio>.
- [19] García, Geraldine: *Precios*, 2015 [Online]. <https://blogs.iadb.org/conocimiento-abierto/es/6-herramientas-gratuitas-para-analizar-y-visualizar-datos/>.
- [20] García, Joaquín: *Comparativa de técnicas de balanceo de datos. Aplicación a un caso real para la predicción de fuga de clientes*, 2021 [Online]. https://digibuo.uniovi.es/dspace/bitstream/handle/10651/60629/TFM_Joaqu%C3%ADnGarc%C3%ADaAbad.pdf?sequence=4.
- [21] García-Maroto, M.: *Data Science (Ciencia de Datos): Aclaración de conceptos básicos*, 1 de Septiembre, 2021. <https://isdfundacion.org/2021/09/01/data-science-ciencia-de-datos-aclaracion-de-conceptos-basicos/>.
- [22] Gasull: *World of Data - Conceptos Básico*, 22 de Diciembre, 2020. <https://netmind.net/es/world-of-data-conceptos-basicos/>.
- [23] González, Isaac: *Big Data para CEOs y Directores de Marketing: Como dominar Big Data Analytics en 5 semanas para directivos Versión Kindle*. Data Science 4 Business, España, 2017.
- [24] González, Ligdi: *IDE para Machine Learning*, 2018 [Online]. <https://aprendeia.com/ide-para-machine-learning-con-python/>.
- [25] Harris, H: *THE DATA PRODUCTS VENN DIAGRAM*, 2013. <https://www.datacommunitydc.org/blog/2013/09/the-data-products-venn-diagram>.
- [26] Hyman, Jack: *Microsoft Power BI For Dummies*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2022.
- [27] Infogram: *Precios*, 2022 [Online]. <https://infogram.com/es/>.
- [28] JetBrains s.r.o: *Herramientas de JetBrains para ciencia de datos y Big Data*, 2019 [Online]. <https://www.jetbrains.com/es-es/data-tools/>.
- [29] Lavrac, Nada, Elpida Keravnou y Blaz Zupan: *Intelligent data analysis in medicine*. University of Ljubljana, Faculty of Computer and Information Sciences, Slovenia, 2000.
- [30] Lemus-Delgado, Daniel y Ricardo Pérez Navarro: *Ciencia de datos y estudios globales: aportaciones y desafíos metodológicos*. Colombia Internacional, 1(102):41–62, 2020. <https://doi.org/10.7440/colombiaint102.2020.03>.
- [31] León Cermeño, Dayan Roselyn de: *Análisis del uso de Tecnologías de la Información y Comunicación*. Tesis de Doctorado, Universidad San Carlos de Guatemala, Guatemala, noviembre 2017.
- [32] Martín, Francisco: *Operaciones con Bases de Datos Ofimáticas y Corporativas*. RA-MA, Madrid, 2007.

- [33] Martínez Bencardino, Ciro: *Estadística básica aplicada*. Ecoe Ediciones, Bogotá, 2019.
- [34] Microsoft: *Excel*, 2022 [Online]. <https://www.microsoft.com/en-us/microsoft-365/p/excel/cfq7ttc0hr4r?activetab=pivot:overviewtab>.
- [35] Ministerio de Economía: *Sistema Nacional de Información MIPYME Guatemala*, 2015 [Online]. http://www.mineco.gob.gt/sites/default/files/Comunicacion%20Social/sistema_nacional_de_informacion_mipyme_guatemala_ano_base_2015.pdf.
- [36] Mora, J: *El futuro de la Ciencia de los Datos*, 2010. <https://www.javeriana.edu.co/documents/12847/10949798/El+futuro+de+la+ciencia+de+datos/842b726f-b258-4d69-bdbf-7078c7523e42>.
- [37] Nisbet, Robert, John Elder y Gary Miner: *Handbook of Statistical Analysis and Data Mining Applications*. Academic Press, Massachusetts, Estados Unidos, 2009.
- [38] OCSE-MIPYME: *Sistema Nacional de Información MIPYME Guatemala, año base 2015*, 2017 [Online]. http://www.mineco.gob.gt/sites/default/files/Comunicacion%20Social/sistema_nacional_de_informacion_mipyme_guatemala_ano_base_2015.pdf.
- [39] Olarte Ramos, Julian: *Los orígenes del Científico de Datos*, 2021. <https://datos.org.co/los-orignes-del-cientifico-de-datos/>.
- [40] Oracle: *¿Qué es la ciencia de datos?*, 2020. <https://www.oracle.com/mx/data-science/what-is-data-science/>.
- [41] Pazos, E.: *La situación de la ciencia en Guatemala, con datos*, 2017. <https://nomada.gt/cotidianidad/la-situacion-de-la-ciencia-en-guatemala-con-datos/>.
- [42] Pensertrust: *Evaluación Tableau Software*, 2022 [Online]. https://pensertrust.com/conozca-mas-sobre-tableau-software/?gclid=CjwKCAjwwL6aBhB1EiwADycBINK5KJqqW6szBiuSDH81HTT_k-2HfEw1bGGHPz2zabNz2cQ-Jt04BRoCZEGQAvD_BwE.
- [43] Provost, Foster y Tom Fawcett: *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media, Inc., Estados Unidos, 2013.
- [44] Solver Machine Learning: *6 Librerías imprescindibles de Python para Machine Learning*, 2021 [Online]. <https://iasolver.es/6-librerias-de-python-para-machine-learning/>.
- [45] Stedman, C: *What is data science? The ultimate guide*, 2020. <https://www.techtarget.com/searchenterpriseai/definition/data-science>.
- [46] Sudeep: *Understanding the Data Science Lifecycle*, 02 de Febrero, 2018. <https://www.sudeep.co/data-science/2018/02/09/Understanding-the-Data-Science-Lifecycle.html>.
- [47] SYNEX Westcon-Comstor: *CIENCIA DE DATOS: ¿QUÉ ES Y POR QUÉ DEBERÍAN INVERTIR LAS EMPRESAS?*, 20 de Octubre, 2021. <https://digital.la.synnex.com/ciencia-de-datos-que-es-y-por-que-deberian-invertir-las-empresas>.
- [48] Tableau: *5 ejemplos reales de inteligencia de negocios en acción*, 2022. <https://www.tableau.com/es-mx/learn/articles/business-intelligence-examples>.
- [49] Tierney, B: *Data Science Is Multidisciplinary*, 2012. <https://oralytics.com/2012/06/13/data-science-is-multidisciplinary/>.
- [50] Tukey, J. W.: *The future of data analysis*. The annals of mathematical statistics, 33(1):1-67, 1962.

- [51] Universidad Católica de Murcia: *¿Qué necesitas para ser Científico de Datos o Data Scientist?*, 2022 [Online]. https://keepcoding.io/blog/que-se-necesita-para-ser-cientifico-de-datos/#4_habilidades_o_requisitos_que_necesitas.
- [52] Upadhyaya, M.: *Impact of Big Data on Analytics*, 2014. https://www.slideshare.net/slideshow/embed_code/36866068.
- [53] Viceministerio de desarrollo de la microempresa, pequeña y mediana empresa: *Informe de Situación y Evolución del Sector MIPYME de Guatemala 2015-2017*, 2018 [Online]. https://www.mineco.gob.gt/sites/default/files/MIPYMES/informedesituacion_y_evolu_delsector_mipymedeguatemala2015-2017.pdf.
- [54] Weedmark, David: *How to Write a Data Report*, 2020 [Online]. <https://smallbusiness.chron.com/write-data-report-61330.html>.
- [55] World Bank Enterprise Survey: *Informe de Evaluación Anual del Plan Estratégico Institucional 2017-2023*, 17 de Agosto, 2020 [Online]. https://www.enterprisesurveys.org/content/dam/enterprisesurveys/documents/covid/country-profile-Guatemala_English.pdf.
- [56] Zaric, Stefana: *How to Write Data Analysis Reports in 9 Easy Steps*, 2021 [Online]. <https://databox.com/data-analysis-report>.

12.1. Ilustraciones

Ilustración 1

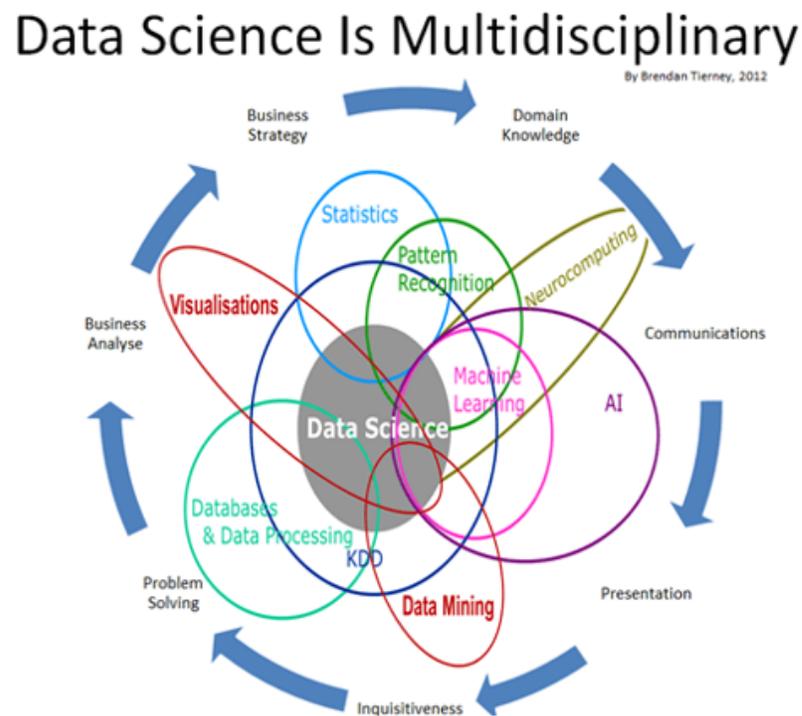


Figura 12.1: Disciplinas y habilidades involucradas en la Ciencia de Datos. [\[49\]](#)

12.1.1. Ilustración 2

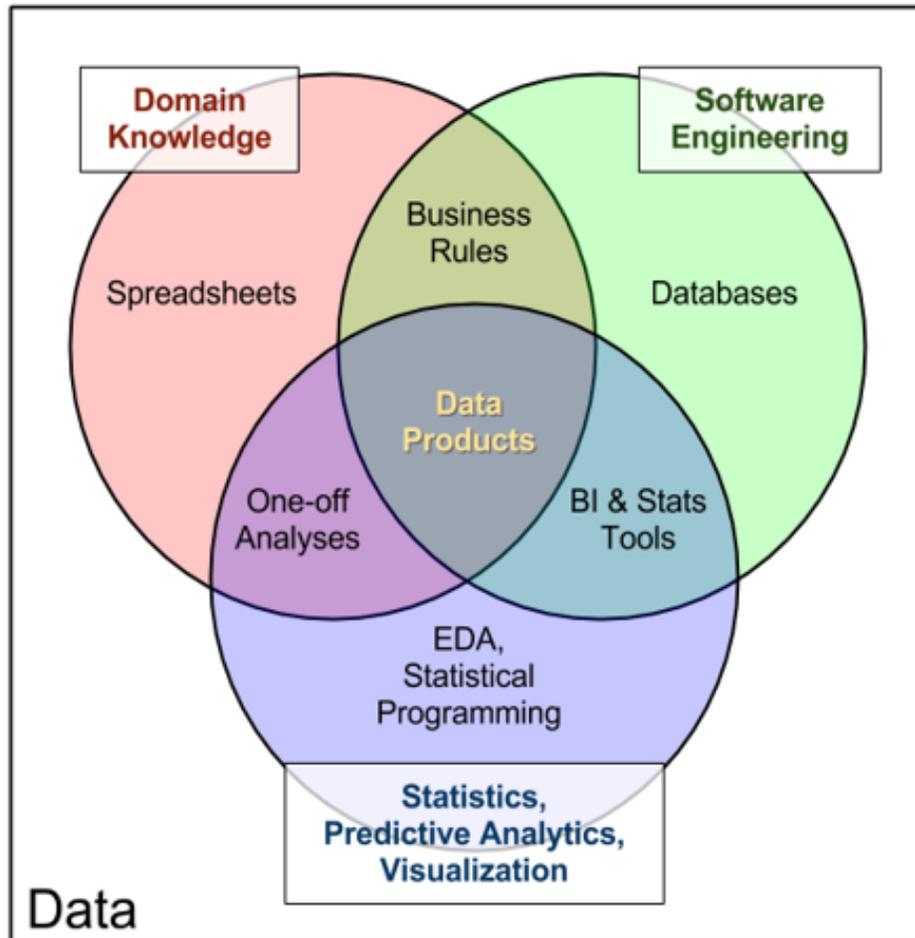


Figura 12.2: Disciplinas necesarias para la aplicación de la Ciencia de Datos. [25]

12.1.2. Ilustración 3



Figura 12.3: Cinco etapas del ciclo de vida de la Ciencia de Datos: captar, mantener, procesar, analizar y comunicar. [12]

12.1.3. Ilustración 4

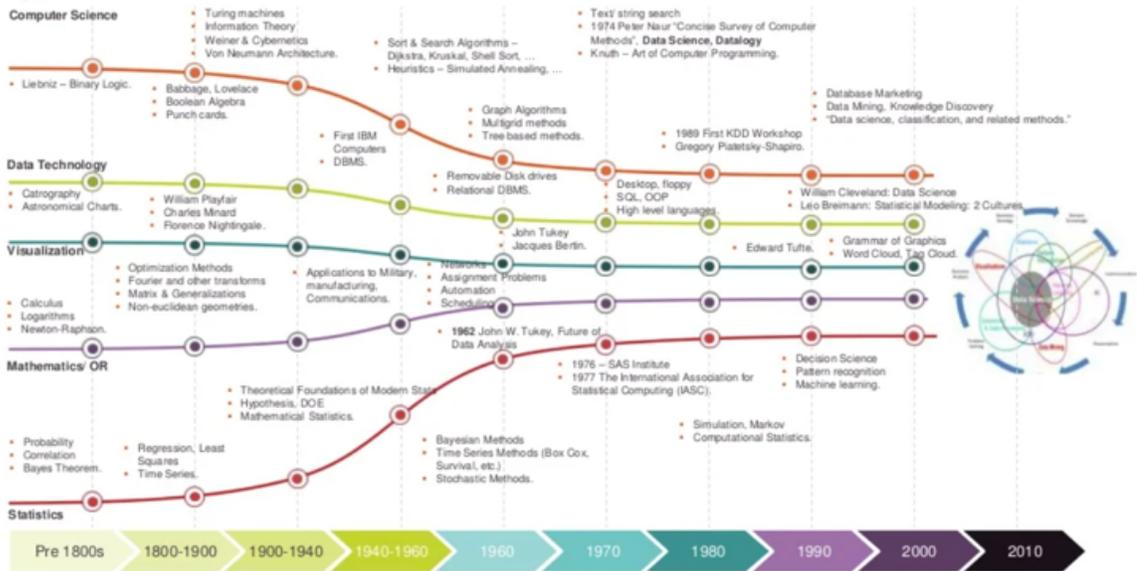


Figura 12.4: Línea del tiempo de la Ciencia de Datos. [52]

12.1.4. Ilustración 5

	Economics	Data Science
Research focus	Focus on understanding causal mechanisms of a system, e.g. 'What are the main factors that increase the credit risk of commercial loans?'	Focus on making predictions within a system, e.g. 'What is the best model to predict the credit risk of commercial loans?'
Toolkit	Stata, Matlab, Excel & VBA, SAS, SPSS, R	Python, R, database systems, big data frameworks (e.g. Spark), version control software
Model validation	Quasi-random experiments: e.g. regression discontinuity, instrumental variables	Cross validation techniques, e.g. k-fold cross validation
Data collection and processing	Mostly from spreadsheets, e.g. from international and public bodies	APIs, web scraping, ETL pipelines, etc.

Figura 12.5: Diferencias entre la economía y la Ciencia de Datos. [9]

12.1.5. Ilustración 6



Figura 12.6: Los procesos en la Ciencia de Datos. [\[46\]](#)

12.1.6. Ilustración 7

SUMÉRGETE EN LA CIENCIA DE DATOS

Si eres una pequeña o mediana empresa ayúdame a responder esta corta encuesta para mi proyecto de graduación:

<https://bit.ly/EncuestaCienciaDeDatos>

Un grupo selecto de los empresarios participantes recibirán un manual y taller básico sobre la Ciencia de Datos para potencializar su negocio a través de esta metodología

UVG UNIVERSIDAD DEL VALLE DE GUATEMALA

ING FACULTAD DE INGENIERÍA

Figura 12.7: Boletín digital utilizado en la convocatoria de PYMES para la encuesta.

12.1.7. Ilustración 8

Situación de las PYMES en la Ciencia de Datos

La presente encuesta es parte de los instrumentos investigativos de mi proyecto de graduación para obtener el título de ingeniera en Ciencias de la Computación y Tecnologías de la Información en la Universidad del Valle de Guatemala. Todos los datos utilizados serán confidenciales y únicamente utilizados con fines educativos.

Agradezco de antemano su colaboración.

Cualquier duda, comentario o consulta quedo a las órdenes:
Paula Camila González Ortega
gon18398@uvg.edu.gt

gon18398@uvg.edu.gt [Cambiar de cuenta](#) 

*Obligatorio

Correo *

Tu dirección de correo electrónico

[Siguiente](#) [Borrar formulario](#)

Figura 12.8: Primera sección de la encuesta elaborada

12.2. Tablas

12.2.1. Tabla 2

DEPARTAMENTOS	TAMAÑO DE EMPRESA					
	MICRO	PEQUEÑA	MEDIANA	GRANDE	TOTAL GENERAL	%
Guatemala	133,220	27,081	3,338	1,455	165,094	44,29
El Progreso	4,931	179	21	2	5,133	1,38
Sacatepéquez	12,785	732	44	13	13,574	3,64
Chimaltenango	10,594	448	32	6	11,080	2,97
Escuintla	13,317	959	68	27	14,371	3,86
Santa Rosa	8,097	257	14	2	8,370	2,25
Sololá	8,277	159	8	3	8,447	2,27
Totonicapán	6,425	105	15	1	6,546	1,76
Quetzaltenango	19,138	1,410	136	35	20,719	5,56
Suchitepequez	10,499	523	29	4	11,055	2,97
Retalhuleu	6,420	306	23	6	6,755	1,81
San Marcos	12,118	488	27	6	12,639	3,39
Huehuetenango	11,079	498	30	7	11,614	3,12
Quiché	10,399	302	20	3	10,724	2,88
Baja Verapaz	5,167	127	6	3	5,303	1,42
Alta Verapaz	11,181	477	39	2	11,699	3,14
Petén	9,213	448	24	3	9,688	2,60
Izabal	8,053	642	56	7	8,758	2,35
Zacapa	6,711	373	28	6	7,118	1,91
Chiquimula	7,934	417	29	7	8,387	2,25
Jalapa	5,474	153	15	0	5,642	1,51
Jutiapa	9,720	314	25	4	10,063	2,70
TOTAL GENERAL	330,752	36,398	4,027	1,602	372,779	100%

Tabla 12.1: Número de empresas registradas y activas clasificadas por tamaño y departamento. [35]

12.2.2. Tabla 3

Empresa	Importaciones	Exportaciones
Micro	2.7 %	10.5 %
Pequeña	6.3 %	23.7 %
Mediana	10.5 %	15.9 %
Grande	79.3 %	49.5 %
Sin clasificación	1.2 %	0.4 %
TOTAL	100 %	100 %

Tabla 12.2: Aporte al monto total de importaciones y exportaciones por tamaño de empresa. [38]

12.2.3. Tabla 4

<i>Actividad</i>	<i>Total MIPYMES</i>	<i>Aporte total</i>
<i>Agricultura, ganadería, caza, silvicultura, pesca, explotación de minas y canteras</i>	9,123	1.90%
<i>Industrias manufactureras</i>	37,384	7.79%
<i>Transporte, almacenamiento y comunicaciones</i>	44,258	9.22%
<i>Servicios privados</i>	95,243	19.84%
<i>Comercio al por mayor y al por menor; reparación de vehículos automotores, motocicletas, efectos personales y enseres domésticos</i>	140,709	29.32%
<i>Actividades inmobiliarias, empresariales y de alquiler</i>	153,222	31.93%
TOTAL	479,939	100%

Tabla 12.3: Distribución porcentual por actividad económica de las MIPYYES activas en 2017. [53]

12.2.4. Tabla 5

PREGUNTAS	RESULTADO
¿Cuál es su PYMES nacional favorita? ¿Por qué?	<ul style="list-style-type: none"> - La panadería de mi colonia - Carbón y leña, un restaurante de hamburguesas en Cayibel de Ciudad Cayalá. - OSI, una empresa de tecnología que ofrece la aplicación y servicio de delivery a empresas que lo necesiten. Es muy útil y segura - Panadería Santo Domingo en Carretera a El Salvador. Era muy pequeña, pero ha hido creciendo gradualmente y ahora ya tienen más sucursales en la ciudad.
¿Sabes qué es la ciencia de datos?	<ul style="list-style-type: none"> - Para mí “Data Science” es el análisis de información. - Es la recopilación de varios datos, ya sean geográficos o demográficos, que ayudan a complementar información. - Es la ciencia que obtiene información (que no se sabía que tenía) a partir de datos. - Lo asocio a programación y estadística.

PREGUNTAS	RESULTADO
¿Cuál es su experiencia con la ciencia de datos?	<ul style="list-style-type: none"> - El curso de “Data Mining” en la U. - Utilicé el programa llamado R específicamente para el análisis de datos. - Las dos herramientas que he utilizado en el banco es Tableau y Power BI. - He usado un programa que se llama Sage.
¿Qué opinan y cuál ha sido su experiencia con Power BI?	<ul style="list-style-type: none"> - La diferencia principal entre Power BI y Tableau es que Power BI tiene una licencia pagada por usuario bastante elevada. Una opción viable que te da las mismas herramientas, aunque puede que sea menos amigable, es Tableau que es de fuente abierta. - Power BI es mucho más visual. - Power BI es bastante útil y fácil de usar, pero para usarse tiene que ser una empresa bien constituida. - A mí en lo personal no me gusta la programación y Power BI es super amigable porque básicamente solo tenés que meter los datos que quieras a usar en Excel y ya luego filtras lo que quieras. - Uno puede trabajar con la cuenta gratis de Power BI, pero no se pueden publicar o compartir los tableros.
¿Conoce los beneficios de la ciencia de datos?	<ul style="list-style-type: none"> - La facilidad de reducir los datos en lo que necesitamos como empresa. - Los datos es el activo más poderoso de cualquier empresa y analizar ayuda a encontrar nuevas estrategias de marketing y administración. Permite identificar si nuestras metas internas se están cumpliendo con base a algunos KPIs.
¿Conoce alguna empresa que aplique ciencia de datos y cómo lo aplica?	<ul style="list-style-type: none"> - En Tigo tienen un departamento de Ciencia de Datos. No sé a cierta exacta cómo lo aplican pero seguro analizan los datos de sus clientes para mantenerlos e incluso atraer a más. - Imagino que restaurantes grandes como McDonald’s, lo han de usar. - Martinexsa, una empresa guatemalteca, se dedica a hacer análisis de datos a otras empresas. Uno de sus clientes fue McDonald’s y el proyecto fue evaluar el comportamiento de las personas que tenían una tarjeta de crédito para atraerlas. - La Cervecería Centroamericana trabaja con “<i>Data Science</i>”. Hacen análisis de la cultura o costumbres para ver el potencial de los productos. Cualquier empresa que use la Ciencia de Datos podrá ver el potencial que tiene, si su posicionamiento es correcto y si su propuesta es valiosa.

PREGUNTAS	RESULTADO
<p>¿Qué herramienta es útil para aumentar la competitividad empresarial en las PYMES?</p>	<ul style="list-style-type: none"> - Los restaurantes ven cuándo los clientes compran más y menos. Ofrecen promociones en los días que menos compran para generar más flujo de personas. Creo que es una ventaja competitiva porque si no sabes esos datos no lo podés implementar. - Yo me he dado cuenta de que las ventas pueden ser estacionales o cíclicas, entonces las empresas que ya tienen bastantes años en el mercado pueden tomar los datos para predecir en qué momento volverán a tener una venta fuerte. Un ejemplo de empresa con ventas estacionales es una librería porque no vende la misma cantidad de útiles en enero que en mayo. El análisis le permite a la librería prepararse mejor o definir una estrategia para aumentar el porcentaje de ventas en los meses de febrero a diciembre.
<p>¿Qué información (qué desean consultar) es de interés para las PYMES que desean incrementar su competitividad empresarial? ¿En dónde la registran?</p>	<ul style="list-style-type: none"> - Por temas de capital y conocimiento imagino que muchas tienen la información en Excel. Creo que la información útil es ventas, personas que entran a la tienda con intención de venta y los “leads” (clientes en están proceso). - Cuando empiezan las PYMES no le ponen importancia a eso o lo guardan de manera arcaica. Lo que más les interesa creería que es la tendencia del grupo. - Lo recomendable es que almacenen todos los datos posibles para poder identificar nuevas estrategias, proveedores, etc. Las opciones económicas para almacenar datos son Access y SQL aunque son herramientas difíciles de utilizar si no se tiene conocimiento en bases de datos. - Dudo que las PYMES tengan una estructura de base de datos, pero seguro si tienen algún Excel. Algo que útil es que digitalicen todos sus canales de ventas para saber más sobre sus ventas.
<p>¿Cree usted que las PYMES en general conocen y aplican la ciencia de los datos?</p>	<ul style="list-style-type: none"> - Muy pocas tienen ese asesoramiento inicial para establecer ciertas métricas necesarias para el crecimiento. Por lo que no se toman el tiempo de recopilar toda la información de los clientes, fechas de ventas y efectos de promociones en las ventas. - No lo aplican. La empresa se llama “Dressy” sí lo hizo y creció muchísimo. Identificaron los rangos de edad y ubicaciones de la mayoría de sus clientes para implementar envíos gratis a dichas zonas y tarifas accesibles a áreas cercanas. - Las PYMES no lo utilizan debido a la cultura. Aquí nos lanzamos a hacer las cosas sin investigación y si analizamos la curva de cada cuánto se cierran las empresas pequeñas nos damos cuenta de que la mayoría caen antes de los 4 años.

PREGUNTAS	RESULTADO
<p>¿Con qué herramientas tecnológicas cuentan las PYMES para tomar mejores decisiones?</p>	<ul style="list-style-type: none"> - “Business Analytics de Facebook” porque da estadísticas de Facebook, Instagram y WhatsApp. Con cuentas empresariales de WhatsApp y agregan sus catálogos, y pueden ver cuántas veces y cuántas personas ingresaron a su catálogo. Estas herramientas son buenas si la empresa está bien constituida en redes sociales. Si la empresa tiene un sitio web es recomendable usar “Google Analytics” porque de forma automatizada muestra cuántas personas ingresaron al sitio, cuánto tiempo permanecieron navegando, etc. - Para publicaciones de Facebook promocionadas se puede usar “Pixel”. En base a la publicación indica cuántas personas entraron a la página web y cuántas compraron. - Todas las herramientas de CRM o tipo “Sendinblue” que te ayudan a hacer campañas por email y al mismo tiempo te deja ver tu público y su comportamiento con el correo, son bastante útiles, amigables y accesibles.
<p>¿Cree usted que un manual básico de ciencia de datos sería útil y de interés para las PYMES que desean implementar esta metodología?</p>	<ul style="list-style-type: none"> - Si, pero debe incluir las formas de implementar el método en una empresa. Muchas empresas comienzan a ver los múltiples beneficios de la tecnología entonces si tomarían en cuenta un manual así. El manual debe ser amigable con ellos utilizando un lenguaje simple. - A mí al ser una PYME me encantaría tener este manual. Ahora creo que hay un problema porque hay muchas personas que no saben qué es esto. Hay que complementar el manual con información de para qué sirve todo esto para que la gente esté interesada. - Podría ser un buen modelo de negocio porque algo que necesitas cuando eres una PYME y no puedes tener a un experto dentro de tu empresa por costos, es un servicio de asesorías o capacitaciones. Que este manual sea la entrada a una asesoría un poco más especializada.
<p>¿Si existiera un manual básico sobre la ciencia de datos dirigido a las PYMES, qué procesos y conceptos de esta metodología deberían de incluirse?</p>	<ul style="list-style-type: none"> - ¿Qué es Ciencia de Datos? ¿Qué aplicaciones son más amigables? y ¿qué conocimientos debería de tener para utilizar esas aplicaciones? o ¿dónde puedo yo adquirir esos conocimientos? - Aparte de los beneficios y utilidad de la Ciencia de Datos me gustaría encontrar qué puedo lograr. - Ver cómo analizar los datos para poder obtener es información sobre su audiencia y alcance.

Tabla 12.4: Respuestas de los estudiantes de administración y ciencia de datos, durante el focus group realizado.

12.3. Recursos

12.3.1. Recurso 1

Vídeo explicativo de cómo analizar los reportes HTML generados en el tablero.
<https://youtu.be/JXkbaqp3kZc>

12.3.2. Recurso 2

Manual de usuario del tablero de la Ciencia de Datos. https://www.canva.com/design/DAFOYst0pRY/o2xBcXN6NFafnowWCve-Mg/view?utm_content=DAFOYst0pRY&utm_campaign=designshare&utm_medium=link&utm_source=publishsharelink

12.3.3. Recurso 3

Bases de datos utilizadas como ejemplos en el manual y las cuáles se elaboraron con base a la información que las PYMES indicaron recolectar de sus ventas y compras.
https://drive.google.com/file/d/1KldxBaW9Xg24cfJjqHP_vJmIhE060xHz/view?usp=sharing

12.3.4. Recurso 4

Tablero desarrollado en Streamlit para que las PYMES limpien y exploren los datos deseados de su empresa en formato .csv. <https://camilago-cienciadedatos-pymes-streamlit-dashboard-streamlitapp.com/>

12.3.5. Recurso 5

Manual técnico del tablero de la Ciencia de Datos. https://www.canva.com/design/DAFOYq0trsg/rG90ZNO6cuLrj1WsNxw/view?utm_content=DAFOYq0trsg&utm_campaign=designshare&utm_medium=link&utm_source=publishsharelink

12.3.6. Recurso 6

Repositorio de Github con el proyecto del tablero. https://github.com/CamilaGO/CienciaDeDatos_PYMES

Data warehouse: Consiste en un almacén de datos electrónico utilizado por una organización. Su función es guardar cantidades grandes de datos de forma segura, fiable, fácil de recuperar y administrar. [16](#)

Google forms: Una herramienta de Google que permite crear y publicar formularios útiles para encuestas, exámenes, asistencias a cursos, etc. y ver los resultados de este en forma gráfica. [23](#)

KPIs: Key Performance Indicators, por sus siglas en inglés, son los indicadores o valores cuantitativos que una empresa mide, compara y monitorea para conocer el desempeño de los procesos y así mejorar las estrategias del negocio. [19](#)