

UNIVERSIDAD DEL VALLE DE GUATEMALA

Facultad de Ingeniería



Diseño de un sistema para la gestión y  
visualización de datos de clientes para el análisis  
estratégico de mercadeo

Trabajo de graduación presentado por  
Diana Marie Vásquez López  
para optar al grado académico de Licenciada en  
International Marketing and Business Analytics

Guatemala, 2025



UNIVERSIDAD DEL VALLE DE GUATEMALA

Facultad de Ingeniería



Diseño de un sistema para la gestión y  
visualización de datos de clientes para el análisis  
estratégico de mercadeo

Trabajo de graduación presentado por  
Diana Marie Vásquez López  
para optar al grado académico de Licenciada en  
International Marketing and Business Analytics

Guatemala, 2025

UNIVERSIDAD DEL VALLE DE GUATEMALA

Guatemala, Guatemala C.A.

ACTA DE EVALUACIÓN DE TRABAJO DE GRADUACIÓN

En la ciudad de Guatemala, a las 19:00 horas, a los dieciocho días del mes de junio del año dos mil veinticinco, se reunieron los infrascritos Licenciado Michael Finkel, Ingeniero Carlos Celada y MBA, José Carlos Cárcamo Acuña, para practicar al estudiante universitario.

Diana Marie Vásquez López

el EXAMEN DE GRADUACIÓN, previo a optar al grado de

Licenciada en International Marketing and Business Analytics

Procedimos a practicarlo conforme a los estatutos y reglamentos vigentes, y otorgamos al estudiante, de forma individual y secreta, las notas de aprobado, aprobado, aprobado

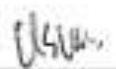
Por lo que lo declaramos Aprobado

(Aprobado - reprobado)

No habiendo nada más que hacer constar, salvo las observaciones que anotamos al pie, damos por terminado el acto a las 20:10 horas horas, en el mismo lugar y fecha.



Licenciado Michael Finkel



Ing. Carlos Celada



MBA. José Carlos Cárcamo  
Director del Departamento  
International Marketing.

Observaciones Hacemos ver que el trabajo se sugiere  
para Mención Honorífica.



# Índice

Índice.....	v
Lista de cuadros.....	vi
Lista de figuras.....	vi
Resumen.....	viii
I. Introducción.....	1
II. Justificación.....	2
III. Objetivos.....	4
IV. Marco teórico.....	5
V. Metodología.....	18
A. Fase 1 Diseño del esquema de recolección de datos.....	18
B. Fase 2 Análisis y exploración de datos.....	24
C. Fase 3 Creación del dashboard.....	26
D. Fase 4 Diseño de esquema para la interpretación y creación de perfiles y segmentos.....	30
VI. Análisis de la empresa.....	31
VII. Diseño del esquema de recolección de datos.....	34
VIII. Recolección de datos simulados.....	48
IX. Tablero para la visualización de datos.....	57
X. Metodología para la perfilación y segmentación.....	71
XI. Conclusiones.....	79
XII. Recomendaciones.....	80
XIII. Bibliografía.....	81
XIV. Anexos.....	83

## Lista de cuadros

Cuadro 1: Productos para la redención del programa de lealtad.....	45
--	----

## Lista de figuras

Figura 1. Participación de la venta monetaria por producto .....	39
Figura 2. Participación del volumen de venta por producto .....	39
Figura 3. Distribución del ticket promedio .....	40
Figura 4. <i>Marketing</i> 10 productos con mejor margen.....	41
Figura 5. Relación entre el volumen de venta y el margen por producto .....	42
Figura 6. Relación entre la venta monetaria y el margen por producto .....	43
Figura 7. Matriz de correlación de variables numéricas .....	48
Figura 8. Distribución de las edades de los usuarios .....	49
Figura 9. Distribución de clientes por sexo.....	49
Figura 10. Ventas totales por sucursal .....	50
Figura 11. Ventas mensuales .....	50
Figura 12. Relación entre el día del mes y la venta del día.....	51
Figura 13. Distribución de la frecuencia de órdenes por cliente.....	52
Figura 14. Resumen estadístico de la frecuencia de compra .....	52
Figura 15. Distribución de la recencia de compra .....	53
Figura 16. Resumen estadístico de la recencia de compra.....	53
Figura 17. Distribución del <i>ticket</i> por orden de compra.....	54
Figura 18: Resumen estadístico del monto por factura.....	54
Figura 19 <i>Marketing</i> 10 productos con más ventas .....	55
Figura 20. Productos más canjeados .....	56
Figura 21. <i>Dashboard</i> de resultados del programa de lealtad.....	58
Figura 22. <i>Dashboard</i> de perfil de clientes.....	61
Figura 23. Participación por sucursal por edad.....	62
Figura 24. Horarios de venta más frecuentes .....	63
Figura 25. <i>Dashboard</i> peso de venta y volumen de órdenes .....	65
Figura 26. <i>Dashboard</i> de segmentación .....	68
Figura 27. Propósito del manual .....	71
Figura 28. <i>Buyer persona</i> .....	72
Figura 29. Perfil .....	72
Figura 30. Objetivos.....	73
Figura 31. Motivaciones .....	73
Figura 32. Retos .....	74
Figura 33. Comportamiento de compra .....	74
Figura 34. Canales de comunicación .....	75

Figura 35. Canales de comunicación .....	75
Figura 36. Canales de comunicación .....	76
Figura 37. Gráficos interactivos.....	76
Figura 38. Segmentos por puntuación RFM .....	77
Figura 39. Interacción tablero de segmentos .....	77
Figura 40. cuadro de información de clientes .....	78

## Resumen

En este trabajo se propone el desarrollo del diseño de un sistema para la recolección, análisis y segmentación de datos de clientes, con el objetivo de optimizar las estrategias de *marketing* mediante el análisis de perfiles y comportamientos. Para la recolección de datos, se desarrollará una estrategia de fidelización estructurada que permitirá recopilar información relevante sobre los clientes, asimismo, ayudará a identificar patrones de consumo y a segmentar a los clientes según su comportamiento y características.

Dado que este proyecto se enfoca únicamente en el diseño, la base de datos utilizada para el análisis será simulada con inteligencia artificial. Esta base de datos permitirá modelar distintos escenarios de comportamiento del cliente y evaluar el funcionamiento del sistema de gestión de datos, sin necesidad de acceso a información transaccional real que se obtendría del plan de fidelización.

Con la información organizada y estructurada, se espera realizar un análisis detallado de los clientes utilizando herramientas de visualización de datos. Los resultados se presentarán mediante un *dashboard*, facilitando la interpretación de los datos y permitirá implementar soluciones más efectivas. Este enfoque ayudará a resolver problemas relacionados con la gestión de datos, permitiendo la identificación precisa del perfil de los clientes y facilitando la segmentación efectiva. Información recolectada. Esta visualización permitirá tomar decisiones estratégicas y operativas de manera ágil, ayudando a la empresa a tener un control más efectivo de sus recursos y esfuerzos de *marketing*.

Por último, se estructurará una metodología para que la empresa utilice la información obtenida de manera efectiva. Esto incluirá la creación de perfiles detallados de los clientes y la implementación de criterios claros para segmentar a los clientes, lo cual permitirá crear estrategias de *marketing* más enfocadas y alineadas con los intereses y necesidades de su público objetivo. Se espera que el sistema diseñado ayude a la empresa a optimizar sus esfuerzos de *marketing* y aumentar su eficiencia operativa.

# I. Introducción

La disponibilidad de información para analizar los datos de clientes de las empresas se ha convertido en un elemento clave en el éxito que buscan las empresas para mejorar su toma de decisiones estratégicas, operáticas y la optimización de campañas. En el presente trabajo se tiene como finalidad diseñar un sistema de recolección, segmentación y visualización de datos para Lulis, una empresa guatemalteca dedicada a la venta de galletas. El presente trabajo tiene el objetivo de crear una herramienta que les permita tener acceso a la información de sus clientes, crear perfiles y segmentos con las que puedan tomar decisiones informadas y generar estrategias de mercadeo. Se buscaba diseñar una estrategia junto con una herramienta que le permitiera a la empresa visualizar la información de forma resumida y clara.

El tema de este trabajo se delimita al diseño conceptual de un sistema que recoja, ordene y visualice datos de clientes. Para ello se diseñó un programa de lealtad, el cual se planteó utilizando las transacciones históricas de la empresa y un análisis de la competencia. Para evaluar los resultados, se usaron bases de datos simuladas de clientes, ventas y redenciones. No se contempla la implementación práctica del sistema, sino únicamente su diseño y validación teórica, enfocándose en demostrar su viabilidad y su aporte potencial a la estrategia comercial de la empresa. El proyecto abarca el análisis de variables demográficas, comportamiento de compra y participación en programas de fidelización, y establece las bases para futuras acciones de segmentación.

La metodología para el desarrollo de este trabajo incluyó la simulación de bases de datos mediante herramientas de inteligencia artificial y creación de código, el análisis exploratorio de datos (EDA) para detectar tendencias y patrones relevantes, la creación de tableros para la visualización ordenada de la información obtenida, y la aplicación del modelo de segmentación RFM (Recencia, Frecuencia y Valor Monetario) para la creación de segmentos. Además, se desarrolló una metodología para el uso de la información de los *dashboards* que facilitan la interpretación de los resultados y le permite a la empresa poder obtener el *buyer* persona de la marca, además de obtener los segmentos y el comportamiento de los mismos.

En este trabajo se destaca la creación de un sistema para la recolección de datos y su análisis que permite una visión más completa y segmentada de los clientes, facilita la toma de decisiones basada en datos y abre la oportunidad de diseñar estrategias personalizadas que aumenten la retención y el valor del cliente en el tiempo. Asimismo, con el análisis de resultados, se puede obtener la viabilidad de la implementación de un programa de lealtad, la rentabilidad del mismo y los beneficios que le trae a la empresa, no solo para la recolección de datos, sino que también para atraer clientes a canales propios y crear accionables para mantener y aumentar la lealtad de los consumidores.

## II. Justificación

En la actualidad, para una empresa es de suma importancia el tener conocimiento de sus clientes, principalmente para poder generar estrategias dirigidas a ellos y que éstas sean exitosas. Para esto, las empresas deben contar con herramientas y capacidad para hacer una recolección de datos, poder analizarlos, segmentarlos y visualizarlos, permitiéndoles tener un mejor entendimiento acerca de las preferencias y comportamientos de sus clientes, esto con la finalidad de poder anticiparse a las necesidades de los consumidores y también poder adaptarse a las tendencias del mercado.

Según un artículo publicado en el 2023 por el Instituto Tecnológico de Monterey, el doctor Luis Eduardo Falcón Morales, director de la Maestría en Inteligencia Artificial Aplicada, explica que en la actualidad los datos se generan a partir de prácticamente todo en distintos formatos; sin embargo, lo importante es saber cómo realizar la recolección de los datos mediante un proceso de búsqueda y medición a través de distintas fuentes, esto con la finalidad de obtener información que pueda generarle valor a la empresa. El doctor también señala la importancia de definir un objetivo y que la data recolectada sea útil para la resolución de este. Entre los usos más distintivos de la recolección de datos se encuentran la identificación de oportunidades de negocio, el análisis de datos estructurados y no estructurados para comprender su contexto, el almacenamiento de datos según las características del público objetivo y, sobre todo, el entendimiento profundo de los comportamientos de clientes, usuarios y leads, lo cual puede transformar los datos en estrategias accionables para las empresas.

La recolección de datos es un proceso esencial para garantizar decisiones basadas en información confiable y rica en contexto. Según Mazhar et al. (2021), en el artículo *Methods of Data Collection: A Fundamental Tool of Research*, publicado por Advanced Research Publications, este proceso permite evaluar hipótesis y tomar decisiones fundamentadas mediante métodos validados científicamente. Los datos recolectados pueden ser primarios, obtenidos por observación directa, entrevistas o cuestionarios; o secundarios, recopilados de publicaciones existentes, bases de datos organizadas y otras fuentes relevantes.

Uno de los principales problemas que las empresas pueden enfrentar a la hora de hacer la recolección de datos, es un bajo índice de respuestas a las encuestas o formularios. Aunque se sigan todas las mejores prácticas, como tener un objetivo bien definido, un borrador perfecto y una población objetivo claramente identificada, en algún punto, el éxito del proyecto depende de la disposición de los encuestados a participar (SurveyMonkey, 2024). Es aquí donde las estrategias de fidelización de los clientes resultan ser elemento clave en el proceso de recolección de datos, ya que uno de los principales obstáculos para obtener información personal de los consumidores es su reticencia a compartirla sin un incentivo claro. Según Smith y Murphy (2019), en el artículo *Consumer privacy and data sharing behavior*, los consumidores suelen ser reacios a proporcionar datos personales debido a preocupaciones sobre la privacidad y el uso indebido de su información. Por lo tanto, ofrecer incentivos, como descuentos, promociones exclusivas o beneficios adicionales, se ha convertido en una práctica común para motivar a los clientes a compartir sus datos de manera voluntaria y con confianza.

Luego de obtener los datos necesarios para responder la pregunta planteada es fundamental analizarlos, convirtiendo los datos a información clave para la toma de decisiones. La información obtenida debe ser procesada y presentada de manera que sea sencillo interpretar el mensaje para una correcta toma de decisiones informadas. Según Medina (2008, p.12), "escucha una pieza de información y, tres días después, recordarás el 10 % de ella. Añade una imagen y recordarás el 65 %". Otro estudio del Instituto Tecnológico de Massachusetts (MIT) indica que

el 90 % de la información transmitida a nuestro cerebro es de manera visual, identificando imágenes vistas en tan solo 13 milisegundos (p.3). Hoy en día, la visualización de datos se ha convertido en una de las principales herramientas para la toma de decisiones estratégicas dentro de una empresa. La visualización de datos resulta ser útil para limpieza y exploración de datos, detectar datos atípicos, identificar tendencias y grupos, detectar patrones, evaluar y presentar resultados. Esto es esencial para el análisis y extracción de los datos, verificando la calidad de la data y que ésta sea de utilidad para cumplir con el objetivo de la empresa. A medida que la estructura de los datos se vuelve más compleja, incrementa la necesidad de una visualización de datos eficiente.

Dentro de una empresa, es de suma importancia la identificación de sus consumidores para la segmentación de sus estrategias. La mejor forma de poder identificar a los consumidores es mediante estos procesos de recolección y visualización de datos, con la finalidad de conocer a sus clientes y tomar decisiones basadas en el perfil de estos. La definición clara de una segmentación de los clientes de la empresa puede generarle distintos beneficios tales como la creación de campañas personalizadas que resulten ser más efectivas, eficiencia a la hora de enfocar los recursos a los segmentos correctos y la satisfacción de los clientes al mejorar la experiencia de compra de los productos o servicios.

## III. Objetivos

### A. General

Diseñar un sistema integral de gestión de datos que abarque la recolección, análisis y visualización de la información de los clientes, con la intención de generar perfiles estratégicos y segmentaciones que permitan optimizar las estrategias de *marketing*.

### B. Específicos

1. Proponer un programa de fidelización estructurado para la recolección de datos, simulados con inteligencia artificial, con el propósito de identificar patrones de consumo y segmentar a los clientes para optimizar campañas dirigidas al público objetivo.
2. Diseñar una estructura de visualización de datos que facilite la interpretación rápida y eficiente de la información recolectada, asegurando que el sistema sea intuitivo y útil para optimizar las decisiones estratégicas y operativas de la empresa.
3. Estructurar una metodología para que la empresa pueda utilizar la información obtenida a través de la recolección de datos, estableciendo los criterios necesarios para construir perfiles detallados, y facilitar la segmentación efectiva para futuras estrategias de *marketing*.

## IV. Marco teórico

### A. Recolección de datos en *marketing* y *business analytics*

La recopilación de datos es un proceso metodológico para la obtención de información acerca de un tema en específico, en el caso del *marketing*, este proceso es útil para obtener información estructurada acerca de clientes, tendencias de consumo, resultados de campañas y transacciones. Es primordial el poder garantizar que los datos sean recopilados de manera legal y ética.

Según el artículo *7 métodos de recopilación de datos en el análisis empresarial* publicado por Harvard Business School, existen tres tipos de datos del consumidor:

- *First-party data*: estos son los datos de primera mano, son aquellos que la organización recopila directamente de los usuarios.
- *Second-party data*: son aquellos datos que se obtienen a través de otra empresa u organización acerca de los clientes de la misma, es decir, son los datos de primera mano de esa empresa.
- *Third-party data*: estos son los datos procedentes de fuentes externas, estos pueden haber sido vendidos por organizaciones que no poseen una conexión directa con su empresa o usuarios.

Los datos considerados los más valiosos son aquellos que se consiguen de primera mano, pues le permite a la empresa el poder tener información acerca del consumo y comportamiento de sus clientes, desde una fuente confiable.

Existe dos tipos de clasificación de datos:

- **Cualitativos**: aquellos datos o variables que no pueden ser medidas en términos numéricos, tales como categorías o características, estos datos tienen una naturaleza contextual.
- **Cuantitativos**: estos son aquellos que si pueden medirse y expresarse numéricamente. Para hacer la recolección de datos es importante primero definir ciertos factores:
  - Plantear la pregunta que se desea responder con los datos recopilados
  - De quién o de dónde se necesita recopilar la data
  - El plazo en el que se recopilarán los datos
  - El o los métodos que se utilizarán para la recolección de datos

La recopilación de datos es una parte muy importante en el éxito de una empresa, ya que esto les puede permitir el tener una mejor precisión, integridad y mayor relevancia en la data para su organización y cumplir los objetivos planteados. La información que se obtiene a través de los datos le permite a la empresa mantenerse informada y tomar acciones para la mejora de distintas estrategias.

Según el artículo *Recolección de datos: herramientas y plantillas*, publicado por HubsPot, existen distintos métodos para la recopilación de datos, algunos de estos:

1. Encuestas o cuestionarios: son cuestionarios físicos o digitales que ayudan a recopilar datos tanto cuantitativos como cualitativos. Estas pueden ser realizadas de forma físicas o enviadas a en línea, lo que permitiría tener un alcance amplio y al mismo tiempo es una opción más económica. Es importante tener claridad sobre

los objetivos de la investigación para hacer las preguntas correctas y obtener los resultados deseados. Existen 2 tipos de cuestionarios, se debe elegir el que mejor se adapte a las necesidades de la empresa:

- Cuestionario abierto: permite explorar de manera más profunda las perspectivas de los participantes sobre un tema y poder obtener información más detallada.
  - Cuestionario cerrado: este cuestionario proporciona respuestas más limitadas, esto debido a la formulación de las preguntas, además de que estas preguntas suelen ser de opción múltiple o respuestas simples de “sí/no”.
2. Seguimiento de transacciones: cada vez que un cliente realiza una compra, es posible darles seguimiento a esos datos, permitiéndole a la empresa el poder tomar decisiones al tener una mejor comprensión sobre el comportamiento de compra de sus clientes. La empresa puede obtener distintos *insights* que pueden ser utilizados para el desarrollo de estrategias de *marketing* más efectivas, además les permite comprender la fidelidad de sus clientes, *Top* productos más vendidos, frecuencias de compra y *ticket* promedio de compra.
  3. Entrevistas y grupos de discusión: este método es utilizado para obtener información mediante preguntas directas. Es realizado mediante una comunicación interpersonal en la que se obtiene las respuestas de los participantes de manera verbal acerca de un tema en específico. Las entrevistas pueden ser realizadas de manera presencial o por teléfono, sin embargo, se recomienda el poder ver las reacciones y comportamientos físicos de los participantes durante la entrevista, estas requieren un entrevistador y un informante. Para llevar a cabo una entrevista, es importante tener definida la información que se desea obtener y preparar las preguntas, además de escuchar atentamente, crear un ambiente cordial en el que los participantes puedan sentirse cómodos y que las preguntas que se realicen sean claras para facilitar la obtención de respuestas útiles.
  4. Observación: es una herramienta discreta y sencilla para la recolección de datos sin tener la necesidad de depender de un intermediario. Este método no es intrusivo y se enfoca en la evaluación del comportamiento de los participantes de manera continua. Para llevarla a cabo, se pueden registrar las observaciones en notas, grabaciones o plataformas digitales. Es una técnica muy utilizada, sin embargo, al ser un método bastante superficial, pueden dejarse fuera datos cruciales que pueden ser clave para el análisis. Es importante que la persona que realizará la observación evite posibles sesos o prejuicios personales, con la finalidad que la información y las decisiones que se tomen sean más acertadas.
  5. Formularios: es una fuente de datos que los usuarios proporcionan de manera voluntaria. Estos ayudan a la obtención de información como nombre, correo y edad, lo que resulta útil para la segmentación de mercado. En ocasiones se ofrecen descuentos, suscripciones o contenido relevante como método de incentivo para que los usuarios quieran llenar estos formularios.
  6. Monitoreo de redes sociales: las redes sociales pueden proporcionar una gran cantidad de datos importantes acerca del sector sobre la marca, la competencia y los potenciales clientes. Algunas de estas plataformas cuentan con herramientas de análisis de rendimiento, por las cuales se puede obtener información relevante.

## B. Creación y limpieza de bases de datos

### 1. Base de datos

Una base de datos es la recopilación organizada de información o datos estructurados, los cuales normalmente son almacenados de forma electrónica. Una base de datos es controlada la mayoría de las veces por un sistema de gestión de base de datos (OCI, 2020). Es una herramienta que ayuda a recompilar y organizar información. Estas pueden almacenar información acerca de personas, productos, pedidos, entre otros. Una base de datos computarizada es un contenedor de objetos, estas pueden tener más de un cuadro (Microsoft, 2024).

Una base de datos diseñada correctamente, permite el acceso a información actualizada y precisa. Según un artículo publicado por Microsoft llamado “*Conceptos básicos del diseño de una base de datos*”, existe una serie de pasos para el proceso del correcto diseño de una base de datos, los pasos son los siguientes:

- Determinar el propósito de las bases de datos: antes de diseñar una base de datos, es importante el determinar el propósito de la misma y hacerse preguntas como: ¿Para qué servirá esta base de datos? ¿Cómo se utilizará? ¿Quién es el usuario? El tener claridad en el propósito de la base de datos ayudará a enfocarse en los siguientes pasos.
- Buscar y organizar la información necesaria: es importante identificar la información que se quiere registrar en la base de datos. Se debe recopilar la información existente, como formularios o registros previos, y con esto identificar los elementos clave que la base de datos debería de almacenar.
- Dividir la información en cuadros: para este paso, se debe identificar los temas o entidades principales, estas se dividen en cuadros para evitar la redundancia y facilitar actualizaciones.
- Convertir los elementos de información en columnas: cada cuadro debe tener definida las columnas que contendrán la información específica sobre cada entidad. Cada columna del cuadro debe de representar un solo tipo de dato que esté relacionado con el tema del cuadro, no se deben combinar distintos tipos de datos en una misma columna.
- Especificar las claves principales: la clave o llave principal de un cuadro es una columna que se usa para identificar de manera única cada registro o fila dentro del cuadro. Este identificador es útil para asociar registros en otros cuadros
- Establecer las relaciones de cuadros: una vez organizada la información en distintos cuadros, es necesario establecer la relación entre ellas. Las relaciones se crean a través de claves o llaves externas, que son columnas en un cuadro que hacen referencia a la clave principal de otro cuadro. Al tener establecidas estas relaciones, es importante verificar que no se dupliquen los datos y que cada cuadro se conecte correctamente.
- Perfeccionar el diseño: luego de haber definido las cuadros, columnas y relaciones, es importante probar que el diseño de la base de datos sea funcional. Se debe comprobar si se pueden obtener los resultados deseados mediante los cuadros, de lo contrario, realizar los ajustes necesarios.

## 2. Tablas

Son objetos fundamentales de una base de datos, ya que en ellas se conserva la información o los datos. Dentro de una base de datos, un cuadro es similar en apariencia a una hoja de cálculos, ya que los datos son almacenados en filas y columnas. Cada fila se denomina un registro y cada columna, un campo. Un campo es un único elemento de información, un tipo de elemento que aparecerá en cada registro, estos deben tener un tipo de dato específico, como texto, número o fechas, con el fin de garantizar que los datos se manejen de manera correcta. Los cuadros se deben organizar de manera que se eviten las redundancias, es decir, que se almacenen de manera eficiente y que los datos no se repitan innecesariamente (Microsoft, 2024).

## 3. Consultas

Estas permiten realizar diversas funciones con los datos almacenados en los cuadros. Las consultas son utilizadas para recuperar datos específicos de los cuadros. Las consultas son útiles para tener vistas de los datos combinados cuando estos se encuentran distribuidos en diferentes cuadros. Al hacer consultas se tiene la opción de filtrar los datos, para que puedan obtenerse los registros que realmente se desean obtener.

- Herramientas para la consulta de datos: existen distintas herramientas para la consulta de datos, algunas de ellas:
  - SQL: se inventó en la década de 1970, el primer proveedor de este lenguaje fue “Relational Software” ahora conocido como “Oracle”. es un lenguaje de consulta estructurada, es un lenguaje de programación que permite almacenar y procesar información en una base de datos relacional. SQL consulta, manipula, y define los datos (Amazon Web Services, Inc., 2020).
  - R: presentada por primera vez en 1993, es un lenguaje de programación para la computación estadística y creación de gráficos, que está equipada para gestionar un gran volumen de datos. Ofrece una amplia variedad de técnicas estadísticas, además de ser una herramienta extensible, permitiendo agregar nuevas funciones y paquetes. Tiene bibliotecas para la manipulación y consultas de grandes conjuntos de datos. (R-project.org, s. f.)
  - Python: Guido Van Rossum creó Python en 1989 en el Centrum Wiskunde & Informática. Es un lenguaje de programación utilizado en las aplicaciones web, desarrollo de softwares, ciencia de datos y machine learning. Es un lenguaje ampliamente utilizado debido a su eficiencia y facilidad para aprender a utilizar, además de que se puede ejecutar en distintas plataformas (Amazon Web Services, Inc., 2022).

## 4. Limpieza de datos

Es el proceso que elimina los datos que no pertenecen a su conjunto de datos. Este proceso corrige y elimina datos dañados, con formato incorrecto, duplicados o incompletos dentro de la base de datos. Cuando se tiene una base de datos con múltiples cuadros y estas son combinadas, en muchas ocasiones los datos se duplican o etiquetan incorrectamente. Si los datos son incorrectos, los resultados no son confiables, afectando el resultado final del objetivo de la base de datos (Tableau, 2024). Para hacer una limpieza de datos es fundamental tener en consideración lo siguiente:

- Eliminar observaciones duplicadas o irrelevantes: es de suma importancia el eliminar los datos duplicados que suelen surgir al combinar conjuntos de datos de distintas fuentes, al igual que observaciones irrelevantes. De esta manera se puede mejorar la eficiencia y el rendimiento del análisis.
- Corregir los errores estructurales: estos incluyen errores de formato, nomenclatura incorrecta, errores tipográficos, que pueden llevar a categorías mal etiquetadas, por ello es importante estandarizar la información y corregir los errores.
- Filtrar los valores atípicos: los datos atípicos deben ser evaluados ya que pueden ser errores o datos válidos que confirmen alguna hipótesis.
- Gestionar los datos faltantes: cuando existen valores faltantes es importante lidiar con ellos, existen formas de poder manejarlos, la primera es el poder descartar las observaciones que tengan valores faltantes, sin embargo, al hacerlo se estaría perdiendo información. Otra opción es el ingresar los valores basándose en otras observaciones, pero, podría perderse la integridad de los datos al estar haciendo suposiciones.
- Validar y controlar la calidad: es primordial validar la calidad de los datos recopilados. Es necesario comprobar que los datos sean coherentes, sigan las reglas correctas y que sean útiles para confirmar o refutar la hipótesis planteada. La calidad de datos es importante para evitar conclusiones o decisiones errónea.

## C. Análisis y visualización de datos

### 1. Análisis de datos

Es un método de inspección, limpieza, transformación y modelado de datos, con la finalidad de obtener información útil, poder sacar conclusiones y apoyar la toma de decisiones. Es un proceso en el que se utilizan diferentes técnicas y herramientas para interpretar datos procedentes de distintas fuentes tanto estructuradas como no estructuradas (Crabtree & Nehme, 2024). Este proceso convierte los datos recopilados en información práctica. El análisis de datos pretende encontrar tendencias y resolver problemas mediante la información obtenida a través de los datos. Estos pueden formar procesos empresariales, mejorar y tener una eficiente toma de decisiones e impulsar el éxito de una empresa. El análisis de datos ayuda a las empresas a poder obtener una mejor visibilidad y conocimiento acerca de sus procesos, productos o servicios. Además, les proporciona información detallada acerca de la experiencia del cliente y distintos problemas que puedan darse en su recorrido de compra (Amazon Web Services, Inc., 2022).

Según el artículo *¿Qué es el análisis de datos?* publicado por Coursera en el 2023, existe un proceso o serie de pasos que seguir para un correcto análisis:

1. Identificar: encontrar la pregunta que se desea responder, el problema que se quiere resolver y qué se necesita para medirlo
2. Recopilar: se debe recopilar la información adecuada que ayudará a responder la pregunta identificada. La recopilación de datos puede proceder de distintas fuentes, tanto internas como el software de gestión de relaciones con los clientes (CRM), o de fuentes secundarias.
3. Limpiar: los datos deben pasar por el proceso de limpieza para poder ser analizados. Esto implica eliminación de datos anómalos, duplicados y

estatización de estructura de datos

4. Analizar los datos: al manipular los datos mediante distintas herramientas de análisis se empiezan a encontrar patrones, tendencias, correlaciones y variaciones que ayudan a obtener todo tipo hallazgos.
5. Interpretar: lo más importante del análisis de datos es la interpretación de los resultados obtenidos para ver en que medida las conclusiones encontradas pueden responder la pregunta inicial.

La interpretación de los datos proporciona información crucial para la toma de decisiones informadas dentro de las empresas. La toma de decisiones basada en datos, puede definirse como un proceso de toma de decisiones que se basen en hechos, métricas y datos, y no en intuición. De acuerdo con McKinsey Global Institute, una empresa de consultoría de gestión global, las organizaciones que utilizan datos de manera estratégica son más eficaces en atraer nuevos clientes, fidelizarlos y generar una rentabilidad superior al promedio (Coursera, 2023).

## 2. Visualización de datos

La visualización de datos es la representación gráfica de información y datos obtenidos. Es el uso de elementos visuales tales como gráficos de distintos tipos, mapas y cuadros, estos permiten poder tener la información de una manera más accesible y fácil de comprender, es decir que, se trasladan datos de alto volumen o numéricos a una forma visual que es más fácil de procesar. Las herramientas para visualizar datos ayudan a mejorar y automatizar el proceso de comunicación visual, permitiendo extraer información práctica. La visualización de datos dentro de una empresa es de gran importancia debido a que los datos pueden ser en muchas ocasiones difíciles de procesar y comprender. Es por ello que se deben preparar y presentar los datos de manera adecuada, de una forma visual para que aquellos responsables de la toma de decisiones, puedan identificar las relaciones entre los datos y detectar patrones o tendencias ocultas (Amazon Web Services, Inc., 2022).

Los ojos de las personas son atraídos por colores y patrones, dándoles la habilidad de identificar rápidamente la diferencia entre colores y figuras. La visualización de datos es una forma de arte visual que capta el interés de las personas y mantiene su atención en el mensaje. Cuando se ve un gráfico, de inmediato pueden identificarse las tendencias y los valores atípicos (Tableau, 2025).

Según Amazon Web Services, Inc., (2022), existen algunas ventajas de la visualización de datos:

- Toma de decisiones estratégicas: ayuda a la alta dirección a interpretar los datos rápidamente, identificar patrones y tendencias, y obtener información clave para tomar decisiones informadas.
- Mejora del servicio al cliente: permite identificar más puntualmente las necesidades y deseos de los clientes, de esta manera es más fácil mejorar el servicio y corregir deficiencia y reducir ineficiencias operativas
- Mayor compromiso con los empleados: facilita la comunicación de los resultados encontrados a través del análisis realizado, permitiendo trabajar en objetivos comunes, medir los progresos y mejorar las motivaciones de los equipos.

Asimismo, se menciona que existe un proceso para la visualización de datos:

- Definir los objetivos: establecer los objetivos que debe cumplir la visualización es importante, identificando las preguntas que el conjunto de

datos puede responder. Esto determinará el tipo de datos, el análisis y los elementos visuales necesarios.

- Recopilación de datos: se debe de identificar las fuentes de datos de las que se extraerá la información, ya sean fuentes internas, externas, datos históricos o cualquier conjunto de dato disponible para la realización del análisis.
- Limpieza de datos: eliminar datos redundantes, incoherentes o anómalos, realizar filtrados para que los datos se ajusten a los criterios de la pregunta de investigación.
- Elementos visuales: elegir los elementos visuales adecuados según la relación entre los datos, para la presentación del mensaje que se desea comunicar. Utilizar herramientas de visualización para generar los gráficos. Es necesario aplicar principios de diseño como el uso adecuado de colores, tipos y tamaños de letra, títulos explicativos que faciliten la comprensión de los datos.

Existen muchos elementos de visualización de datos, algunos de ellos:

- Cuadros: filas y columnas utilizadas para comparar variables. pueden mostrar un gran volumen de información de una manera estructurada. Debe ser utilizada solo cuando lo amerite, de lo contrario, tanta información puede abrumar al usuario y complicar la comprensión de la información obtenida.
- Gráficos circulares: los gráficos circulares muestran la distribución de un total en segmentos, ideal para ver proporciones.
- Gráficos de barras apiladas: permiten comparar categorías a lo largo de una dimensión, mostrando subcategorías apiladas en una sola barra, facilitando la comparación entre varias partes de un todo.
- Gráficos de líneas: los gráficos de líneas son útiles para mostrar tendencias a lo largo del tiempo.
- Gráficos de áreas: similares a los de líneas, pero rellenan el área debajo de la línea, destacando la magnitud del cambio, ideal para mostrar el volumen en comparación con las líneas de tiempo.
- Diagrama caja y bigotes: representa la distribución de los datos mostrando la mediana, cuartiles y valores atípicos. Es útil para identificar la dispersión y detectar valores anómalos en un conjunto de datos.
- Histogramas: muestran la distribución de un conjunto de datos en intervalos. Ayudan a visualizar la frecuencia con la que ocurren diferentes rangos de valores
- Gráficos de dispersión: permiten visualizar la relación entre dos variables numéricas. Son útiles para identificar correlaciones o patrones en los datos.
- Mapas de calor: representan datos a través de variaciones de color, donde los colores indican la intensidad de la variable en cada área

### 3. Análisis exploratorio

También conocido como EDA, es utilizado para analizar e investigar conjunto de datos y resumir sus características principales, acompañado de visualización de datos. Ayuda a determinar cuál es la mejor manera de manipular las bases de datos para obtener las respuestas que se necesitan para responder a la pregunta inicial, esto les facilita a los científicos de datos a descubrir patrones, encontrar anomalías o probar una hipótesis. El EDA tiene como objetivo ayudar a analizar los datos antes de hacer suposiciones, ayudando

a identificar errores, así como a comprender los patrones dentro de los datos y encontrar relaciones entre variables. El análisis exploratorio es utilizado para garantizar que los resultados que producen sean válidos y aplicables a los objetivos planteados, al igual que es una herramienta para validar que las preguntas hechas sean las correctas (Ibm, 2021).

#### 4. *Insights*

Es una verdad revelada o un descubrimiento ya existente, pero sin identificar. El término es utilizado en investigación de mercados, *marketing*, comunicación y en empresas para referirse a una idea reveladora que proporciona información clave para resolver un problema o pregunta (Alonso, 2025).

### D. Inteligencia artificial

La inteligencia artificial (IA) se resume en tecnologías que le permiten a las computadoras el poder realizar una serie de funciones a avanzadas, el poder simular inteligencia y capacidades humanas tales como ver, comprender, analizar, realizar conclusiones y recomendaciones, entre otras cosas. Es un campo de la ciencia enfocado en la creación de máquinas que tengan habilidades de razonamiento, aprendizaje y acciones parecidas a las de la inteligencia humana, o incluso que involucre datos que excedan la capacidad humana para analizarlos. La inteligencia artificial para el uso empresarial, es utilizada para el análisis de datos, generación de predicciones y previsiones, categorización de objetos, recomendaciones y más. Los sistemas de IA aprenden y mejoran mediante la exposición a grandes cantidades de datos, permitiendo identificar distritos patrones y relaciones. Tiene un proceso de aprendizaje que implica distintos algoritmos, estos son un conjunto de reglas o instrucciones que guían el análisis de la IA (Google, 2025).

Existen distintos tipos de inteligencia artificial, puede ser organizada de distintas maneras, esto según las etapas de desarrollo o las acciones que realcen. Algunos tipos de IA son:

- Máquinas reactivas: este tipo de IA solo reacciona ante diferentes estímulos basados en reglas preprogramadas. No usa memoria por lo que no puede aprender con datos nuevos.
- Memoria limitada: este tipo de IA usa la memoria para ir mejorando con el tiempo mediante el entrenamiento con datos nuevos, regularmente, mediante una red neuronal artificial.
- Teoría de la mente: no existe en la actualidad una IA con teoría de la mente, si embargo, este término hace referencia la inteligencia artificial que puede imitar la mente humana, con capacidades de decisiones similares a un humano, incluyendo reconocer emociones y reaccionar ante situaciones sociales.
- Autoconocimiento: un tipo de IA que aún no existe, pero se está estudiando y desarrollando el poder crear una máquina con la capacidad de tener conocimiento de su propia existencia y capacidades intelectuales y emocionales de un humano.
- Existen modelos para el entrenamiento de inteligencia artificial. La IA con memoria limitada, que es la más común, es aquella que mejora con el tiempo ya que se entrena con datos nuevos, utiliza algoritmos para entrenar datos y obtener resultados. Existen tres distintos tipos de modelos de aprendizaje automático:

- Aprendizaje supervisado: modelo de aprendizaje automático en el cual se asigna una entrada específica a un resultado determinado, esto a través de entrenamientos etiquetados. Este tipo de aprendizaje lo que hace es enviar variedad de información etiquetada de manera específica, esto con la finalidad de que la IA obtenga la capacidad de reconocer e identificar la información y etiquetarla de la misma manera.
- Aprendizaje no supervisado: es un modelo de aprendizaje automático, mediante el cual la IA aprende por medio de patrones en función de datos no etiquetados. En este modelo el algoritmo aprende de datos y los clasifica en grupos en función de diversos atributos.
- Aprendizaje por refuerzo: modelo por el cual un “agente” aprende a realizar las tareas correspondientes mediante distintas pruebas y errores hasta que se logren los resultados deseados. Este agente recibiría un refuerzo positivo en caso de que la tarea sea realizada de manera correcta, y un refuerzo negativo si fuese lo contrario.
- El uso de la inteligencia artificial puede ser beneficioso en muchos aspectos, por ejemplo:
  - Automatización: la IA puede ayudar a automatizar procesos y flujos de trabajo, asimismo puede trabajar de manera independiente y autónoma.
  - Reducción de errores humanos: al usar IA se eliminan los errores manuales a la hora de procesar datos, las estadísticas, ensamblaje en fabricación y muchas otras áreas, todo esto a través de la automatización y algoritmos que son entrenados para seguir los mismos procesos siempre.
  - Rápido y preciso: puede procesar información más rápida que la capacidad de un ser humano, esto por medio de la búsqueda de patrones y descubrimiento de relaciones entre datos.
  - Disponibilidad infinita: la IA no tiene limitaciones, no es necesario horarios establecidos o factores alternos que puedan frenar la labor de un ser humano. La IA puede llegar a trabajar continuamente en tareas asignadas.
  - Investigación y desarrollo acelerado: la inteligencia artificial tiene la capacidad de poder analizar una gran cantidad de datos con rapidez y puede acelerar avances en investigaciones y desarrollos.

## E. Marketing

La American *Marketing* Association (AMA, 2017) define el *marketing* como:

“*Marketing* is the activity, set of institutions, and processes for creating, communicating, delivering, and exchanging offerings that have value for customers, clients, partners, and society at large” ( párr. 1).

En español, esta definición se puede traducir como:

“El *marketing* es la actividad, el conjunto de instituciones, y los procesos para crear, comunicar, entregar e intercambiar ofertas que tienen valor para los compradores, clientes, socios y la sociedad en general” (traducción propia).

También conocido como mercadotecnia, es una metodología comercial que usa

información relevante sobre los productos o servicios de una empresa, asimismo, el contenido creado para influir en la toma de decisiones. El *marketing* sirve para distintas cosas, además de ayudar a incrementar las ventas, también es utilizado para aumentar la visibilidad y presencia del producto o servicio (Londoño & HubSpot, 2021). El *marketing* tiene distintas funciones dentro de una empresa, algunas de ellas:

- Dar a conocer la marca: cuando una empresa está empezando o se está introduciendo a un nuevo mercado, contar con una estrategia de *marketing* es fundamental, es importante que esta esté dirigida hacia el grupo objetivo de la marca, con la finalidad de que los usuarios se familiaricen con ella y con el tiempo puedan reconocerla fácilmente, aunque no hay probado los productos o servicios que se ofrecen.
- Presentar un nuevo producto: aunque la marca ya tenga una reputación y un posicionamiento sólido, siempre es importante crear campañas de *marketing* cuando es lanzado un nuevo producto, para que los usuarios puedan tener presente la marca en todo momento.
- Captar nuevos clientes: una campaña siempre necesita el poder atraer nuevos clientes, principalmente porque el mercado evoluciona y cambia constantemente. Un mercado que se encuentra en esta constante evolución tiende a cambiar de necesidades a medida que cambian las situaciones de vida, por lo que siempre busca alternativas para solventar sus necesidades. Es por ello que es necesario implementar estrategias de *marketing* que puedan informar y atraer a los consumidores a los que se les puede brindar una solución a sus necesidades.
- Fidelizar clientes: aunque los clientes ya estén comprando o consumiendo la marca, es necesario el poder centrarse en construir relaciones a largo plazo con los clientes de la empresa.

El *marketing* en la actualidad no está enfocado únicamente en la promoción de productos, es una estrategia fundamental para el éxito y crecimiento de una empresa. Para que el *marketing* dentro de una empresa sea efectivo, es importante establecer objetivos medibles, esto es esencial para dirigir las actividades de manera eficiente, y que además se alineen a las metas de la empresa. Los objetivos del *marketing* dependerán del objetivo principal de la empresa o de la campaña que se desee generar (Londoño & HubSpot, 2021). Algunos de estos objetivos son:

- Aumentar el reconocimiento de marca: en un mercado saturado, en el que los consumidores están expuestos a una gran variedad de opciones, es importante poder y saber destacarse. Este objetivo busca que los consumidores identifiquen la marca y la asocien con ciertos valores. Para lograr este objetivo usualmente se hace por medio de campañas de branding, uso de una variedad de canales de comunicación y una identidad visual distintiva.
- Generar leads y aumentar ventas: este objetivo se enfoca en atraer clientes potenciales para poder convertirlos en compradores, para esto es de suma importancia que la marca genere contenido relevante, que sus campañas estén bien segmentadas y que sus procesos de venta sean eficientes y prácticos.
- Mejorar la retención de clientes: la importancia de tener estrategias de retención de clientes radica en que para una empresa es mucho más rentable el retener a un cliente que captar nuevos. Para lograr esta retención es importante contar con un buen servicio al cliente, personalización, comunicación activa y crear programas de lealtad.
- Aumentar la participación del mercado: este objetivo se basa en el poder

ganar terreno frente a la competencia, para ello es importante que la empresa se encuentre en constante innovación, expansión, contar con precios competitivos y contar con productos y servicios que puedan generarle valor al mercado objetivo definido.

- Oportunidades de negocio: se basa en la exploración de nichos, con productos que atiendan las necesidades de un grupo específico de clientes, impulsando el crecimiento estratégico.
- Mejorar la imagen de marca: el objetivo es generar una mejor percepción que sea positiva en el público, esto se hace por medio de acciones como relaciones públicas, comunicación y contenido transparente.

## F. Estrategias de fidelización

La fidelización de clientes es un proceso que consiste en crear y mantener una relación duradera y positiva con los consumidores, es importante poder satisfacer sus necesidades y expectativas, generar lealtad y con esto asegurar frecuencia de compra y recomendaciones. Una estrategia de fidelización es la iniciativa de recompensas que la empresa le ofrece a sus clientes que realizan compras frecuentes. Este tipo de estrategias pretende ofrecer regalos, mercancía gratuita, recompensas o cupones. A pesar de todos estos beneficios que las empresas ofrecen, no todos los clientes se mantienen activos, lo que quiere decir que los usuarios no están recibiendo un valor agregado de parte de la empresa. Es importante que uno de los principales fines de estas estrategias sea el ofrecer un valor adicional a los clientes que se encuentran suscritos a estas membresías de fidelización (HubSpot, 2023).

Un artículo de Salesforce llamado *Fidelización de clientes*, indica que la fidelización de clientes, además de aumentar el valor y el número de ventas y el tiempo de retención de los clientes, puede traer otros beneficios para la empresa, algunos de ellos:

- Mantener información relevante en bases de datos, permitiendo conocer de mejor manera a los clientes, además de poder tener información acerca de sus comportamientos de compra y proyectar acciones en distintas áreas.
- Entre más información se obtenga, mejor será la planificación, por ende, los productos o servicios serán mejores y de esta manera se atraerán más clientes.
- Resulta más sencillo el poder sacar innovaciones para aquellos clientes que se tienen mapeados y se conocen sus gustos, preferencias y necesidades.
- Agrega un valor a la marca y crea un diferenciador de la competencia
- Proporciona el análisis de los resultados de las acciones de *marketing* y la venta continua.
- Permite anticiparse a cambios en el comportamiento del mercado.
- Estimula la interacción y creación de contenido positivo a través de las redes sociales.

## G. Segmentación de clientes

Uno de los aspectos más importantes en el *marketing* es comprender a quién se está dirigiendo. La segmentación de mercado se ha vuelto indispensable para cualquier estrategia de *marketing* efectiva. La segmentación le permite a las empresas el poder identificar y dirigirse a grupos específicos de consumidores con características y

necesidades similares, el identificar estos grupos puede asegurarle a la empresa que los mensajes que se transmitan sean más relevantes y atractivos para cada uno de ellos. Al no contar con una segmentación adecuada, las empresas pueden desperdiciar recursos en audiencias no interesadas, ocasionando que se pierda una valiosa oportunidad de conexión y conversión. Al aplicar este concepto, es posible diseñar campañas que sean más precisas y efectivas, de esta manera se atraen a los clientes adecuados y el nivel de satisfacción y lealtad aumentaría (HubSpot, 2024).

La segmentación es una estrategia utilizada para dividir el público de una marca o negocio en grupos más pequeños, que compartan ciertas características. Esta estrategia es crucial dentro de las empresas para conocer mejor su público objetivo y generar un mayor índice de éxito en las estrategias de *marketing* y ventas, ya que se conecta con los clientes ideales y los recursos fueron empleados de manera responsable (HubSpot, 2024).

## 1. Variables de la segmentación

Las variables de *marketing* permiten dividir a la audiencia en grupos basados en características como necesidades, creencias compartidas o comportamiento en línea. Los cuatro tipos principales de segmentación son demográfica, geográfica, psicográfica y de comportamiento (Mailchimp, 2023).

- Segmentación demográfica: permite identificar al público objetivo, un punto clave para la creación de perfiles de clientes. La segmentación basada en perfiles es un buen punto de partida si se desea conocer a los consumidores de la empresa. Algunas de las variables son edad, sexo, ingresos, etnia, ocupación, entre otras.
- Segmentación geográfica: ayuda a clasificar al público de la empresa según su ubicación, usando variables como ciudad, estado, país o zona horaria. Es útil para campañas dirigidas a un entorno en específico.
- Segmentación psicográfica: utiliza características psicológicas del consumidor, como actitudes, valores, estilo de vida e intereses. Es una variable más compleja, sin embargo, permite crear campañas más eficientes al entender las razones de compras de los clientes.
- Segmentación conductual: esta variable agrupa a los consumidores según su comportamiento de compra e interacción con la marca, además de la lealtad o satisfacción. Permite el poder anticipar necesidades o crear estrategias personalizadas, como enviar ofertas a usuarios pertenecientes al programa de lealtad.

## 2. *Buyer persona*

Un *buyer persona* es el perfil detallado del cliente ideal basado en los hechos reales y observaciones de los consumidores de la empresa. Es la representación de quién es, lo que necesita, los problemas a los que se enfrenta y cómo, como empresa, se le puede ayudar. Un *buyer* es una herramienta de segmentación que alinea las estrategias de *marketing*, ventas y servicio para conectar de una mejor manera con la audiencia. El desarrollar un *buyer* es fundamental para la creación de contenido, diseño de producto, seguimiento de leads y ventas, además de actividades relacionadas con la adquisición y retención de clientes (HubSpot, 2024).

Para hacer un *buyer persona* se debe de definir los siguiente:

- Nombre y apellido

- Datos demográficos
- Análisis de comportamientos y diferenciadores
- Objetivos y retos
- Comentarios y objeciones
- Puntos de contacto

## H. RFM

Las siglas RFM significan recencia, frecuencia y valor monetario, estas tres variables son parte de una técnica de análisis conductual de clientes, un modelo llamado RFM. Este modelo toma en cuenta los hábitos de compra de los consumidores y con esto ayuda a identificar a los principales clientes y separarlos por segmentos para generar mejores estrategias publicitarias. El modelo se basa en la asignación de puntuaciones en base las variables ya mencionadas.

- Recencia: esta variable hace referencia a cuán reciente o cuantos días transcurridos han pasado desde la última compra realizada por el cliente. Esta es una variable importante ya que los clientes que han realizado su compra en un periodo corto de tiempo tienen a la marca en mente y es más probable que vuelvan a realizar una compra. Las puntuaciones para esta variable varían dependiendo de los parámetros que cada empresa decida colocar ya que algunas empresas no tengan clientes que realicen sus pedidos y cortos periodos de tiempo debido al tipo de producto o servicio que ofrecen.
- Frecuencia: indica cada cuánto o cuántas veces un cliente realiza compras dentro de un periodo de tiempo. Al analizar la frecuencia de las transacciones de los clientes, se puede identificar a clientes habituales, medir la fidelidad de los mismo, entender cuáles son los clientes de alto valor y con esto poder generar estrategias de retención.
- Valor monetario: es lo que indica la cantidad de dinero que un cliente gasta en promedio en un periodo de tiempo en la compra de productos o servicios de la marca. Esta variable puede proporcionar distintas características sobre el comportamiento de compra de los consumidores. Asimismo, esta métrica ayuda a identificar los clientes más rentables, revelar oportunidades de alteraciones en los precios y además la creación de estrategias más personalizadas.

Este modelo permite crear segmentos y poder hacer campañas de *marketing* que resulten ser más eficaces. Este método permite conocer cuáles son los clientes más valiosos de la empresa y cuántos clientes han perdido a lo largo del tiempo. Además de identificar a los clientes que sustentan el negocio y poder crear estrategias de retención y fomentar la lealtad a la marca. Asimismo, no se encuentra información únicamente de los mejores clientes, sino que también se puede conocer cuál es la tasa de abandono, clientes en riesgo de abandonar la marca, entender como incrementar la tasa de conversión, entre otros. Cada segmento extraído de este análisis es usualmente atendido por separado según las necesidades y patrones de comportamiento.

## V. Metodología

### A. Fase 1 Diseño del esquema de recolección de datos

La primera etapa del desarrollo del presente trabajo consistió en una reunión inicial con los gerentes de la empresa, cuyo propósito fue establecer contacto directo con los tomadores de decisiones, conocer el contexto general del negocio y comprender a profundidad su funcionamiento operativo. Durante esta reunión se obtuvo una visión integral del modelo comercial, la estructura interna, los procesos de atención al cliente y, en particular, el manejo actual de la información. Esta aproximación permitió identificar tanto las fortalezas como las debilidades del sistema vigente, así como las principales necesidades relacionadas con la gestión y aprovechamiento de los datos para la generación de sus estrategias.

Como resultado de este diagnóstico, se determinó que la empresa no contaba con una base estructurada de datos sobre sus clientes, lo cual limitaba su capacidad para implementar estrategias de segmentación y comunicación efectivas. Debido a esto, se realizó una propuesta en la cual se desarrollaría un esquema de recolección de datos centrado en la implementación de un programa de fidelización, el cual se integraría en el canal de WhatsApp con el *chatbot* que actualmente manejan, ya que este es el medio por el cual obtienen una mayor rentabilidad, además de tomar en cuenta la practicidad de la plataforma, ya que no solo permite a la empresa realizar las personalizaciones necesarias para su plan de lealtad, sino que también, los clientes de la marca ya se encuentran acostumbrados esta plataforma. Se consideró que esta estrategia no solo facilitaría la obtención de datos mediante formularios, sino que, al ofrecer recompensas, incrementaría el interés y la disposición de los clientes para compartir su información.

El primer paso fue realizar un *benchmark* competitivo o referencias en el mercado para obtener una visión clara sobre las mejores prácticas actuales en la industria de alimentos. Se seleccionaron cuatro marcas reconocidas del sector de alimentos, para realizar el análisis se consideraron tres factores clave: la plataforma o medio por el cual gestionaban sus programas, ya sean aplicaciones móviles, web, entre otros, la modalidad del programa, es decir si era por acumulación de puntos, *cashback*, cupones, entre otros, y las reglas generales que regían cada programa. Este análisis proporcionó valiosos *insights* sobre cómo se posicionaban los programas de fidelización de algunas empresas que se encuentran en la industria de alimentos, permitiendo identificar oportunidades para el diseño y ejecución del propio programa propuesto.

Antes de desarrollar el diseño técnico del plan de fidelización, se establecieron los objetivos, general y específicos, que justificarían su implementación. Posteriormente, se procedió a la estructuración del programa de fidelización, antes de desarrollar dicha estructura, era importante tener el listado de productos que se utilizaría como incentivos en el programa de lealtad, es por ello que el primer paso fue analizar información transaccional histórica de la empresa y los márgenes de cada producto, con la finalidad de escoger los productos que además de incentivar a los consumidores a realizar compras recurrentes, pudiesen generar un beneficio y además ser rentables para la empresa. Para hacer el análisis de esta información, primero se plantearon las preguntas que se querían responder con los datos, se formularon 6 preguntas y con eso se empezó el análisis exploratorio.

La información de la empresa fue extraída mediante una consulta SQL (Anexo I) utilizando PostgreSQL, sistema que se conecta directamente con el software interno de la empresa. Para esto, se tuvo una breve reunión con el proveedor del sistema, esto para obtener los nombres de los cuadros y las variables que se encuentran en cada una de los cuadros, El *query* utilizado permitió obtener las siguientes variables: fecha, número de orden, tipo de producto, sucursal, producto, precio, cantidad de producto, valor total de la venta y canal de transacción. Para hacer esta extracción de datos se

hizo lo siguiente:

- Primero se usó un “SELECET” para empezar el código jalando las columnas que se necesitan consultar.
  - orden.id AS id\_orden: toma el campo id del cuadro pos\_order y lo muestra con el nombre id\_orden, que es el número único de cada orden.
  - templ.type AS tipo\_producto: toma el tipo de producto desde el cuadro 19producto\_template y devuelve la categoría de tipo de producto (service, 19producto, combo)
  - sucursal.name AS sucursal: proveniente de pos\_config, devuelve el nombre de la sucursal en la que se realizó la compra
  - 19rodu.full\_product\_name AS producto: la columna que lleva el nombre completo del producto que se vendió, guardado en pos\_order\_line.
  - 19rodu.price\_unit AS precio: es el precio unitario del producto en la línea de venta.
  - 19rodu.qty AS cantidad\_producto: cantidad de ese producto que se compró en esa orden.
  - 19rodu.price\_subtotal\_incl AS venta: es el subtotal con impuestos incluidos de esa línea de producto.
- Para obtener el canal por el cuál fue realizada la venta fue necesario utilizar la columna de método de pago y catalogarlo en una columna de “canal” con un nombre más fácil de entender. Para poder realizar esto se usó un “CASE” el cual revisa el contenido de metodo.name, que está guardado en un formato JSON y extrae el ‘en\_US’. Dependiendo de lo que venga ahí, asigna un valor de texto para la columna canal. En la empresa existen 7 canales de venta, y algunos otros utilizados ocasionalmente que son definidos como “canales extra”.
- Luego de esto se realiza otro “CASE” para obtener la diferenciación entre facturas y facturas anuladas que son “reembolsos”. Para esto de la cuadro account\_move se toma el tipo de movimiento (move\_type), si es “out\_invoice”, lo muestra como “Factura”. Si es “out\_refund”, lo muestra como “Refund”, cualquier otro valor, se muestra como “Otro”.
- En este sistema existen aproximadamente 765 cuadros. La información de las ventas se encuentra distribuida en distintas cuadros, por lo cual, es necesario combinarlas. Para esto se usaron algunos join, principalmente “INNER JOIN”, usado para mostrar únicamente las filas que cumplen con la condición de unión de ambas cuadros, es decir, se busca la columna coincidente en ambas cuadros y se extraen las filas que coinciden, y “LEFT JOIN”, el cual devuelve todas las filas de la cuadro izquierda y únicamente las filas que tienen coincidencia de la cuadro derecha. Para obtener la cuadro final junto a todas las consultas se realizó lo siguiente:
  - pos\_order AS orden: pos\_order es el cuadro principal, donde se almacena la información de la orden/venta completa en el punto de venta.
  - INNER JOIN pos\_order\_line AS 19rodu: se usa este cuadro para ver el detalle de los productos vendidos en esa orden. Orden.id = 19rodu.order\_id: esto hará que cada línea está relacionada con la orden por order\_id.
  - INNER JOIN pos\_session AS sesión: esta unión funciona para saber en qué sesión de venta fue realizada la orden.
  - INNER JOIN pos\_config AS sucursal: sirve para conocer en que sucursal fue realizada la orden.
  - INNER JOIN 19producto\_product AS 19producto y 19producto\_template AS templ: es para traer la información del producto que se vendió.

- INNER JOIN pos\_payment AS payment y pos\_payment\_method AS 20roduct: para saber los métodos de pago que se usó en la orden.
- LEFT JOIN account\_move AS account: esta unión es para asociar la orden con su comprobante contable. Se usa un LEFT JOIN para que en caso de no existir la factura, no se descarte la orden completa.
- Para este análisis se necesitó únicamente las transacciones realizadas por medio del bot de Whatsapp, por lo que dentro del código se decidió hacer el filtrado para obtener la información únicamente de este canal. Con un “WHERE” se hizo el filtrado para obtener de la columna “canal” aquellas transacciones realizadas con el bot.

Una vez obtenida la base de datos, se exportó a un archivo en formato Excel para su posterior limpieza y análisis. En esta etapa, se aplicaron una serie de filtros y modificaciones para construir el cuadro final que sería utilizada en el análisis. En primer lugar, se eliminaron los productos de temporada para concentrarse únicamente en aquellos que forman parte del menú fijo de la empresa. Luego de esto, se filtró la columna “tipo de producto”, excluyendo servicios adicionales como los envíos logísticos, y conservando únicamente las categorías “20roducto” y “combo”.

Para resguardar la confidencialidad de la información, se realizaron ajustes en los precios y nombres de los productos. Esta modificación se gestionó a través de una hoja adicional en Excel, donde en la columna A se registraron los nombres originales de los productos y en la columna B se colocaron los nombres modificados que serían utilizados en el análisis posterior. Para hacer la sustitución de los nombres de los productos en el cuadro original se usó la siguiente fórmula:

=BUSCARV(G2, nombres!\$A\$2:\$B\$23, 2, FALSO)

De esta manera se pudo sustituir automáticamente los nombres originales por los nuevos en el cuadro original. Después de haber cambiado los nombres de los productos, fue necesario hacer una alteración en los precios, esta alteración fue realizada con un porcentaje distinto, ya sea positivo o negativo, la cual fue realizada por categorías, a cada categoría se le asignó un porcentaje de variación de precio.

Para la realización del análisis de productos se hizo un cuadro con márgenes, los cuales se asumieron para cada uno de los productos, todos dentro de un rango proporcionado por la empresa, esto para continuar con la confidencialidad de la información de la empresa. Ya con el cuadro de transacciones, llamada “DATA\_BOT1” y la cuadro “margenesbot”, se empezó con dicho análisis. Antes de empezar con el código, primero se plantearon los objetivos de hacer ese análisis, con la finalidad de conocer la exploración necesaria para obtener la información relevante para el planteamiento de los productos que se pondrían en el plan de fidelización.

Este análisis fue hecho en la herramienta de “Jupyter Notebook”, es una aplicación web que permite ejecutar código de Python, herramienta utilizada para hacer consultas, exploraciones, visualizaciones, entre otras cosas, el código de la minería de datos se encuentra en el Anexo II. Ya con ambos cuadros, se empezó a realizar el código necesario para obtener información presentada en distintas visualizaciones.

Lo que se hizo fue lo siguiente:

- Importar la librería “pandas”, una librería que sirve para el análisis y manipulación de datos es útil para leer, limpiar, transformar y analizar los datos de manera eficiente.
- Se usó pandas para leer un archivo CSV guardado en una carpeta desde la computadora.
- Se imprimió el encabezado del primer cuadro para poder verificar que los datos se

hayan procesado correctamente, además de tener presente los nombres de cada columna.

- Se verificó que no hubiera datos nulos en las columnas.
- Se agruparon los datos por producto para poder obtener el volumen por producto y las ventas por producto. Luego, se calculó la participación por producto.
- Se creó el primer gráfico, un *treemap* en el que se observa la participación por producto en volumen de venta. En la etiqueta de datos se muestra el nombre del producto, la cantidad vendida y el porcentaje de participación de cada producto.
- El segundo gráfico, de igual manera un *treemap* en el que se observa la participación por producto en ventas monetarias. En la etiqueta de datos se muestra el nombre del producto, la venta en quetzales y el porcentaje de participación de cada producto.
- El tercer gráfico fue un histograma para analizar la distribución del *ticket* promedio de compra, esto para entender parte del comportamiento de compra actual de los consumidores de la empresa.
- Para tener un análisis de márgenes de los productos se cargó la segunda cuadro llamada “margenesbot”.
- El cuarto gráfico fue un análisis del *marketing* 10 de productos con mejor margen, para esto se hizo un código para ordenar los márgenes de manera descendente y se escogen los 10 primeros, con esto se realizó un gráfico de barras horizontales.
- Se combinaron los cuadros de transacciones y márgenes utilizando la columna producto, con la finalidad de obtener un análisis completo y entender que productos califican para el plan de fidelización.
- El quinto gráfico fue un gráfico de dispersión, utilizando líneas verticales y horizontales creando cuadrantes para obtener una mejor interpretación. En este gráfico se analizó el volumen de ventas vs el margen porcentual por cada producto.
- El sexto gráfico realizado, también fue un gráfico de dispersión, utilizando líneas verticales y horizontales creando cuadrantes. En este gráfico se analizaron las ventas vs el margen porcentual por cada producto.

Con el análisis hecho y la interpretación de cada uno de los gráficos, se realizó nuevamente una reunión con los gerentes de la empresa. El objetivo de dicha reunión fue el poder presentar los hallazgos que se obtuvieron, y basado en esta información, se definieron los criterios para poder seleccionar los productos que serían incluidos como opciones de canje dentro del plan de fidelización. Durante la reunión se acordó que la elección de los productos estaría basada en el volumen de ventas, el valor monetario en el *ticket* promedio y los márgenes de cada producto. Con esto en mente, se estableció que productos con un alto volumen de ventas, que no necesariamente tienen los mejores márgenes, podrían formar parte del programa para incentivar la participación de los consumidores. Asimismo, se consideró la inclusión de productos con márgenes elevados, pero con bajo nivel de rotación, como una estrategia para darles mayor visibilidad.

Luego de haber definido los criterios y tener el listado de productos que se utilizarían, fue necesario entender cómo se asignarían los puntos necesarios para canjear cada producto, esto basado en un concepto que los gerentes llaman “valor percibido por punto”. Este valor es calculado dividiendo el precio normal de venta entre los puntos requeridos para realizar el canje, esto quiere decir que, este valor es el beneficio económico que los clientes perciben por cada punto acumulado. Con el objetivo de asegurar un equilibrio entre la rentabilidad y este beneficio percibido, se definió que este valor debería ser menor o igual que 10 %. Esta asignación de puntos se realizó de una manera más estratégica, siempre tomando en cuenta el valor máximo de valor percibido. Es así que, los productos con alta demanda, aunque no tuvieran el mejor margen o un precio tan elevado, se les asignó una mayor cantidad de puntos, esto debido a que, al tratarse de productos altamente valorados

por los clientes, estos productos resultan ser un incentivo en la participación y acumulación de una mayor cantidad de puntos. Asimismo, los productos con márgenes altos y precios más bajos, pero de poca rotación, se les asignó una menor cantidad de puntos, con el objetivo de fomentar su consumo. Esta asignación de puntos respondió a un equilibrio entre la popularidad de productos, rentabilidad y la percepción del valor, de esta manera se optimizaría el impacto del plan de fidelización.

A continuación, se procedió a establecer la plataforma en la que se estaría implementando el programa y estructurar el funcionamiento del mismo, definiéndose que operaría bajo un sistema de puntos, por cada quetzal gastado, el cliente recibiría un punto. Asimismo, se establecieron las reglas para el canje de los productos, tales como el monto mínimo de compra y el uso de los puntos. Se explicó que los puntos son canjeables por medio de la plataforma, puede ser por entregas a domicilio o para recoger en tienda, además de que no pueden ser canjeados parcialmente para obtener algún descuento.

Después de entender el funcionamiento del plan de lealtad, se procedió a definir las variables necesarias que debían de solicitarse en el formulario de registro del cliente. Se identificó la información esencial que se debía obtener, el objetivo principal era recopilar datos relevantes para construir un perfil de cliente de la marca, además de generar los grupos de segmentación necesarios para entender el comportamiento de los consumidores.

Dentro de la estructuración del programa se decidió plantear un diseño general para un plan de comunicación, el cual estaría enfocado en maximizar la efectividad del programa. Este plan se dividió en tres etapas claramente definidas: concienciación, activación y retención. En la etapa de concienciación, el objetivo era informar y captar el interés inicial del cliente, en la etapa de activación, se buscó incentivar el registro inmediato mediante acciones concretas y promociones especiales, y en la etapa de retención, se implementaron estrategias de comunicación personalizadas para mantener activo y fidelizado al cliente a largo plazo. Luego de esto se definieron los canales principales de comunicación como redes sociales u puntos de venta físicos, y se propuso lo que se haría en cada uno de ellos para comunicar tanto el lanzamiento como el seguimiento del plan de fidelización.

Debido a las limitaciones de tiempo y la imposibilidad de implementar el programa de fidelización de manera real durante esta etapa del proyecto, fue necesario llevar a cabo una simulación de datos utilizando técnicas avanzadas de Inteligencia Artificial (IA). Para asegurar la calidad y utilidad de los datos generados, primero se definieron claramente todas las variables necesarias que permitieran realizar análisis posteriores precisos y alineados con los objetivos planteados en la investigación.

Las variables esenciales para el posterior análisis se decidieron dividir en 3 bases de datos distintas, las cuales representan diferentes aspectos tanto de los clientes como de sus comportamientos de compra. La primera base de datos, denominada “Clientes”, tuvo como propósito recopilar información clave para generar perfiles precisos de los consumidores. Las variables definidas fueron que se preguntarían durante el formulario de registro, las variables son nombre y apellido, fecha de cumpleaños, correo electrónico, género y la red social por la cual se enteraron del programa de fidelización, el número de teléfono no es necesario, ya que, al realizar sus compras por medio de la plataforma de WhatsApp, ya se cuenta con esa información. Estas variables fueron elegidas, ya que permiten segmentar adecuadamente a los clientes y establecer estrategias personalizadas según sus características demográficas y preferencias de comunicación.

La segunda base de datos generada fue la que llevaba información sobre las “Redenciones”

realizadas por los clientes dentro del programa. El objetivo de esta base fue registrar todas las transacciones relacionadas con el canje de productos utilizando los puntos acumulados. Las variables elegidas para este cuadro fueron:

- Número del cliente: se usó como una llave primaria para hacer la relación entre tablas y poder obtener información de las redenciones de cada cliente.
- Fecha de la redención: fecha en la que se realizó el canje de los puntos
- Sucursal donde se efectuó el canje
- Producto canjeado.

La información en esta base es fundamental para entender el comportamiento de los canjes que realizan los clientes, si los productos que se colocan son efectivos e incentivan a la recompra o si debiesen ser cambiados, identificar preferencias por sucursales y forma parte del cálculo del costo de adquisición de clientes y el desempeño del programa en general.

Por último, la tercera base de datos simulada fue la de “Transacciones”, fue diseñada para simular el comportamiento histórico y transaccional, detallando las compras realizadas por los clientes en la empresa. Cada una de las variables incluidas fue seleccionada y definida con el propósito de poder tener información valiosa para analizar y poder entender comportamientos de compra y generar grupos de segmentación, además de contar con las variables necesarias para poder relacionar todos los cuadros. Las variables para esta base de datos fueron:

- Fecha: día específico en que el cliente realizó la compra, esta variable es funcional para un análisis temporal y de estacionalidad.
- Día de la semana: indica el día en que se realizó la transacción, será utilizada para un análisis igualmente de temporalidad y conocer los días con una mayor recurrencia de compra.
- Número de orden: código único que identifica cada transacción individualmente.
- Teléfono del cliente: número telefónico registrado del cliente, usado como identificador para enlazar con las otras bases de datos.
- Sucursal: punto de venta específico donde se realizó la transacción, es importante para medir desempeño por ubicación y las zonas geográficas con oportunidades de mejoras o de aperturas, además es útil para entender la zona geográfica en la que se desenvuelven los consumidores.
- Producto: nombre específico del producto adquirido, clave para análisis de demanda y preferencias del consumidor.
- Precio: precio unitario del producto en el momento de la compra, sirve para calcular rentabilidad y evaluar sensibilidad a precios.
- Cantidad vendida: es el número de unidades adquiridas por producto en cada transacción, es para medir volumen venta y patrones de compra.
- Venta total: monto total en dinero, es el resultado de multiplicar el precio por la cantidad vendida en cada transacción, es necesaria para medir los ingresos, el *ticket* promedio, efectividad de la estrategia entre otros.
- Hora de compra: hora en la que se realizó la compra, ayuda a identificar horarios con mayores ventas.
- *Rating*: es la calificación otorgada a cada producto por el cliente tras la compra, se representa en una escala numérica, para evaluar la satisfacción con el producto.
- Comentario: espacio para que los clientes puedan colocar su opinión o retroalimentación adicional sobre su experiencia, es para análisis cualitativos y

gestión de la calidad.

- Categoría: cada producto se clasifica internamente en categorías, esto es para tener un orden y también para el análisis de ventas por tipo o grupo de productos y poder generar estrategias específicas de *marketing*.
- Método de entrega: refleja si la compra fue entregada por domicilio o retirada directamente por el cliente, relevante para un análisis operativo o de comportamiento o preferencias de los clientes.

El objetivo de simular la información del programa de fidelización al público es entender los resultados que se obtendrían, como se interpretarían y como se usarían para la generación de estrategias, además de comprender como se recolectaría la data y la estructuración de la información generada por medio de este sistema. Dado que el proyecto se limita a la propuesta de un diseño de este sistema, sin llegar a la implementación, se optó por simular los datos que se obtendrían una vez estuviera en funcionamiento. Para poder realizar esta implementación, se utilizó la herramienta de Inteligencia Artificial (IA) como apoyo en la generación del código que se utilizó para crear la base de datos simulada. La elección de esta herramienta fue debido a su capacidad de interpretar instrucciones detalladas, interpretar el contexto y generar soluciones automatizadas y personalizables en tiempo real. La herramienta permite optimizar el proceso de construcción del código, además de facilitar la incorporación de múltiples criterios al proceso de simulación, como la distribución estadística de variables, la lógica de comportamiento de los clientes, y la consistencia entre columnas relacionadas. De esta manera se pudo garantizar que los cuadros simulados fueran representativas de escenarios realistas, permitiendo generar bases de datos que mantuvieran coherencia con los objetivos del estudio.

Fue necesario redactar un prompt con todas las especificaciones requeridas, la cantidad de cuadros y columnas que cada una debería llevar, la relación entre columnas y cuadros, distribución de frecuencias, entre otros. En el prompt se solicitó que la simulación fuese creada por medio de un código para Python, dado que este lenguaje ofrece un amplio acceso a bibliotecas diseñadas para la creación de datos ficticios, en este caso la biblioteca principal que se utilizó fue “Faker”. El generar el código para Python permite un control detallado sobre la lógica de generación, distribución estadística y coherencia entre variables, esto resulta ser relevante debido a que una simulación generada directamente por IA recae en una distribución normal de los datos por lo que no es significativo, es por eso que un código de Python puede permitir ajustar esta normalidad.

El código obtenido permitió simular información de clientes, transacciones diarias durante 1 año, y redenciones, respetando patrones de comportamiento y características realistas que se proporcionaron basado en la información dada por la empresa, esto con el objetivo de poner a prueba el funcionamiento del plan de fidelización desde una perspectiva analítica. El prompt enviado a la inteligencia artificial y la conversación para obtener el código de simulación se pueden encontrar en el Anexo III, Anexo IV y Anexo V. El código para obtener la simulación se encuentra en el Anexo VI.

## B. Fase 2 Análisis y exploración de datos

Al tener la estructura para la implementación del plan de fidelización, por el cual se realizaría la recolección de datos, y las 3 cuadros generadas con el código proporcionado por la inteligencia artificial, se procedió al análisis de la información que se obtuvo de la simulación. Se realizó un análisis exploratorio, con la finalidad de obtener una vista previa de los datos, tanto para la verificación de que los datos simulados puedan tener cierta relevancia y no actuaran únicamente con normalidad, como para entender la información y hacer la estructuración del *dashboard* que se presentaría posteriormente. En el análisis se pretendía comprender distintos patrones de

comportamiento de los clientes, además de obtener hallazgos que sirvieran como base para el desarrollo de una herramienta significativa que ayudara a la empresa a tomar decisiones estratégicas basadas en datos. Para poder realizar este análisis exploratorio, primero se plantearon preguntas generales sobre información que se quisiera obtener.

Para este análisis se volvió a utilizar la plataforma de Jupyter, con código de Python (Anexo VII). El procedimiento técnico se desarrolló de la siguiente manera:

- Se importaron las librerías necesarias para el análisis (pandas para manipulación de datos, numpy para funciones estadísticas, y matplotlib y seaborn para visualización).
- Se leyeron tres archivos .csv que contenían las tres bases simuladas: base\_clientes, base\_transacciones y base\_redenciones.
- Para poder realizar ciertas visualizaciones fue necesario extraer la hora desde la columna de hora de compra, y se creó una columna que convierte el día de la semana a formato numérico para análisis posteriores.
- Luego se utilizó “isnull().sum()”, para corroborar de que las columnas relevantes no tuvieran valores nulos que pudieran impactar en el análisis.
- Para calcular la edad de los clientes se hizo uso de la columna de fecha de nacimiento, se calculó la edad restando esa fecha de la fecha máxima de la base de datos, dividiendo los días resultantes entre 365.
- Se combinó el cuadro de transacciones con la edad de los clientes, utilizando el número de teléfono como clave.
- El primer gráfico realizado fue una matriz de correlación. Para esto, se seleccionaron las variables relevantes para análisis de correlación: precio, cantidad vendida, venta total, edad, hora y día de la semana. Luego se generó visualizada mediante un *heatmap*.
- El segundo gráfico fue un histograma para observar la distribución de edades de los clientes.
- El tercer gráfico realizado fue un gráfico de barras para poder visualizar la distribución de la cantidad de clientes por sexo, usando un diferenciador de color para su mejor comprensión.
- Luego, se creó un gráfico de barras para observar los ingresos generados por cada sucursal, para ver cuáles son las sucursales que le generan más ventas a a empresa.
- Se creó una nueva columna año\_mes para agrupar las ventas por mes, y se generó un gráfico de líneas para observar su evolución de las ventas en el tiempo.
- El quinto gráfico que se hizo es un gráfico de dispersión para poder analizar la relación entre el día del mes y el gasto total, con esto se pudo identificar concentraciones o caídas de ventas según los días del mes.
- Luego se realizó un histograma para analizar el comportamiento del *ticket* promedio por factura y se realizó un resumen estadístico para entender mejor cómo se comportan los datos por percentiles.
- Para el séptimo gráfico se agrupó la base de datos por número de teléfono y número de orden para poder obtener la frecuencia de compra, es decir, la cantidad de veces que ha comprado cada cliente, además, también se realizó un resumen estadístico para analizar los percentiles.
- Se calculó la cantidad de días desde la última compra de cada cliente hasta la fecha más reciente de la base de datos, con esto se sacó la recencia, para saber la cantidad de días que pasaron desde la última compra de cada cliente.
- Luego se hizo un gráfico de barras para comparar las sucursales por ingresos monetarios. Para esto se agruparon las transacciones por producto, calculando la venta total.
- Por último, se agruparon los productos en la base de datos de redenciones y se hizo un gráfico de barras que indica que productos se han canjeado más veces.

Al finalizar la exploración de datos, se realizó el respectivo análisis e interpretación de cada uno de los gráficos para entender la existencia de posibles patrones y comportamientos de compra,

además de encontrar la información necesaria para la perfilación y creación de segmentos. Con esta información se procedió a la creación y elaboración de los tableros que mostrarían la información en visualizaciones de datos resumidos para facilitar la toma de decisiones en la empresa.

### C. Fase 3 Creación del *dashboard*

Luego de haber analizado los datos en Python y haber obtenido información relevante, se planificó la estructura del *dashboard* que presentaría los resultados y cumpliría con los objetivos principales del proyecto: evaluar el desempeño del plan de lealtad y facilitar la perfilación y segmentación de los clientes. Para esto, lo primero que se hizo fue la planificación del *dashboard*, donde se definieron tres vistas principales: Resultados, Perfil de clientes y Segmentación. Para cada página se planteó primero su objetivo específico y las preguntas clave que se pretendían responder mediante las visualizaciones. Esta planificación fue fundamental para asegurar que cada sección del *dashboard* tuviera una función clara y contribuyera a la toma de decisiones estratégicas.

Luego de haber definido los objetivos, se procedió con el diseño visual del *dashboard*. Se eligió una paleta de colores utilizando la página web Colors (Anexo VIII). El enlace de la paleta generada se copió en la plataforma Bibb, la cual permite crear temas personalizados para Power BI (Anexo IX). El tema fue descargado y posteriormente cargado en Power BI, asegurando así un diseño visual consistente en todo el *dashboard*. Ya con el tema cargado, se organizó el orden de las visualizaciones y se definió qué tipo de gráficos se utilizaría en cada página, esto con el objetivo de tener una estructura lógica que facilitara el análisis de los datos.

Es importante mencionar que una vez concluido el primer diseño completo del *dashboard*, con todo y los gráficos y medidas, este fue presentado a la gerencia de la empresa, quienes realizaron observaciones y sugerencias de mejora, las cuales fueron tomadas en cuenta, con estas se realizó la versión final de los tableros que se presentarían.

Para iniciar con la construcción del *dashboard*, se subieron las tres bases de datos simuladas: base\_transacciones, base\_clientes y base\_redenciones. Luego de haber subido los tres cuadros se realizaron las relaciones entre ellas, mientras se avanzaba en la elaboración de los tableros, nuevos cuadros iban surgiendo, las cuales, de igual manera se configuraron las relaciones entre ellas, esto para garantizar la correcta interacción de toda la información en Power BI (Anexo X).

La primera página que se desarrolló fue la de Resultados. Para esta sección, debido a las preguntas planteadas y lo que se quería responder, se realizaron dos simulaciones adicionales utilizando inteligencia artificial, estas simulaciones fueron:

- Cuadro meta de venta: para obtener este cuadro, lo primero que se hizo fue crear una tabla dinámica con las variables de fecha, sucursal y venta real. Esta información se le dio a la IA mediante un prompt (Anexo XI), con esto se obtuvo el cuadro con las metas de ventas mensuales por sucursal.
- Cuadro de gastos de *marketing*: para este cuadro lo que se hizo fue solicitarle a la IA la generación de gastos mensuales por red social, utilizando otro prompt (Anexo XII), obteniendo un cuadro con las variables de fecha, red social y monto de inversión.

Con estas simulaciones listas, se procedió a crear las visualizaciones correspondientes para la vista de resultados:

- Venta real: se hizo una suma de la variable venta total.

- Meta de venta: se hizo una suma de la variable meta venta.
- Avance: se hizo una medida calculada dividiendo la venta real entre la meta de venta para obtener el porcentaje de cumplimiento sobre la venta.
- Clientes con redención: conteo de clientes que realizaron al menos una redención.
- Tasa de retención: se creó una columna de periodo, con esto calcularon los clientes por periodo y luego los clientes retenidos, con estas medidas se calculó la tasa de retención dividiendo los clientes retenidos entre los clientes por periodo
- Retorno sobre la inversión (ROI): para este gráfico primero se creó un cuadro con el producto, precio y margen y de ahí se ha calculado el costo del producto. Con esto se hizo una medida que sumaba del total del costo de cada uno de los productos redimidos. Adicional, se ha calculado el total de lo invertido en *marketing* convertido a quetzales, lo que en total daba el costo de inversión. Luego se hizo el cálculo del ROI con la siguiente fórmula:

$$(VENTAS TOTAL - INVERSIÓN) / INVERSIÓN$$

- Ventas vs Meta en el tiempo: gráfico de líneas comparando venta real contra meta mensual.
- Venta vs Meta por sucursal: gráfico de barras comparando venta real y meta por cada sucursal.
- Usuarios registrados vs usuarios con compra: para este gráfico se generó un cuadro que hacía un recuento de los clientes que aparecían en la base de datos de clientes y de los clientes que aparecen en la base de transacciones, luego se categorizaron y se hizo un gráfico de barras comparativo
- *Ticket* promedio por tipo de cliente: se hizo un gráfico de barras comparando el *ticket* promedio entre clientes que realizaron redenciones y los que no.
- Sucursales con más redenciones: gráfico de barras mostrando el conteo de canjes por sucursal.
- Producto más redimido: gráfico de barras con el conteo de redenciones por producto.

Luego de haber completado la vista de resultados, se desarrolló el tablero de “Perfil de clientes”, el cual tenía como objetivo analizar las características demográficas y de comportamiento de los clientes que participaron en el programa de fidelización. Las medidas y visualizaciones realizadas para esta página fueron las siguientes:

- Edad promedio: para calcularla, se generó primero una columna nueva que calcula la edad de cada cliente a partir de su fecha de nacimiento. Una vez obtenida esta columna, se calculó el promedio general de edad de todos los clientes.
- Género más frecuente: se colocó una tarjeta que muestra el género que más se repite dentro de la base de clientes. Para esto se creó una medida que identifica y devuelve el valor con mayor frecuencia en la columna de género.
- Sucursal más frecuentada: se construyó una medida que identifica cuál es la sucursal que aparece más veces en la base de transacciones. Esta información se presentó en una tarjeta.
- *Ticket* promedio: se calculó haciendo el promedio del monto de venta por cada orden registrada.
- Frecuencia promedio: se creó una medida que cuenta cuántas veces aparece cada número de teléfono en la base de transacciones, y se calculó el promedio de estas repeticiones para conocer la frecuencia general.
- Proporción por género: se utilizó un objeto visual llamado Infographic Designer, el cual permite representar visualmente la proporción entre hombres y mujeres mediante íconos o figuras.

- Distribución de edad: se elaboró un gráfico de barras que muestra cómo se distribuyen las edades dentro de la base de clientes.
- Participación por sucursal según rango de edad: se creó primero una columna calculada en la base de clientes que clasifica a los clientes en rangos de edad de cinco en cinco años. Luego, se utilizó un gráfico de barras apiladas para visualizar qué sucursales tienen mayor cantidad de clientes y cuál es la participación de cada rango etario en ellas.
- Proporción por canal: se diseñó un gráfico que muestra qué porcentaje de clientes llegó a través de cada canal de captación, es decir, cómo se enteraron del programa de lealtad.
- Método de entrega: se elaboró un gráfico comparativo para analizar la distribución de los métodos de entrega seleccionados por los clientes.
- Horarios más frecuentes: se realizó un mapa de calor en forma de matriz, utilizando formato condicional. Para esto, se transformó la columna de hora de compra en números enteros y se creó una medida que calcula el promedio de compras por hora y por día. En la matriz, se ubicaron los días de la semana en las columnas, las horas en las filas y en los valores se colocó el promedio de compra, permitiendo identificar los picos de actividad.
- Ventas por categoría y producto: se usó una visualización jerárquica que permite explorar qué categorías de productos generan más ventas y dentro de ellas, cuáles son los productos más relevantes en términos de volumen.
- Lo que piensan los clientes: se generó una nube de palabras (wordcloud) a partir de la columna de comentarios. Para ello, se excluyeron palabras irrelevantes como conectores o términos comunes que no aportaban al análisis.

Como complemento del tablero de perfil, se construyó una segunda vista que permite analizar el peso que representa en las ventas cada perfil o segmento seleccionado. Esta página fue diseñada para actuar como un tablero interactivo, permitiendo aplicar filtros y visualizar los resultados asociados a cada subgrupo de clientes.

Las visualizaciones incluidas fueron las siguientes :

- Ventas totales: suma total de ventas realizadas en el periodo analizado.
- Volumen de órdenes: suma total de las órdenes emitidas por los clientes.
- Peso en la venta: se utilizó una matriz donde se colocaron los rangos de edad en las columnas, mientras que las filas fueron divididas por género y franja horaria. En los valores se mostró el porcentaje de participación en ventas sobre el total.
- Peso en volumen: matriz con la misma estructura que la anterior, pero en los valores se colocó la cantidad de órdenes en porcentaje sobre el total general.
- Ventas por sucursal: gráfico de barras que muestra la suma de las ventas por sucursal. Esta visualización también cumple la función de filtro para identificar qué segmentos contribuyen más en cada punto de venta.
- Frecuencia de compra: gráfico de barras que muestra cuántos clientes han comprado cierta cantidad de veces. Esta visualización se utiliza como filtro para evaluar qué segmentos tienen mayor impacto en ventas y volumen.
- Rangos de *ticket* por orden: se creó una columna que clasifica los montos de *ticket* en rangos de Q30 en Q30. Esto permite visualizar, a través de un gráfico, cuáles son los montos más comunes por orden. Al igual que las anteriores, esta visualización también funciona como filtro para el resto de los gráficos.

La tercera vista desarrollada se enfocó en la segmentación de clientes utilizando el modelo

RFM (Recencia, Frecuencia, Valor Monetario), con el fin de clasificar a los clientes según su comportamiento de compra y facilitar la toma de decisiones basadas en valor de cliente.

- Pasos para la construcción del modelo RFM:
- Se creó un cuadro donde se calcularon las tres variables clave:
- Recencia: número de días desde la última compra del cliente.
- Frecuencia: número de órdenes realizadas por el cliente durante el periodo analizado.
- Valor monetario: monto total gastado por el cliente en el mismo periodo.
- Luego se consolidó esta información en un nuevo cuadro que contenía el número de teléfono del cliente junto con sus medidas de recencia, frecuencia y valor monetario.
- Posteriormente, se definieron los rangos o punteos para cada una de las variables, basados en los resultados obtenidos durante el análisis exploratorio. Las puntuaciones iban del 1 al 5, siendo 1 el menor y 5 el mayor.
- Se creó una nueva columna que concatenaba los tres puntajes (R, F, M) para obtener el score completo de cada cliente.
- Se definieron siete segmentos basados en las combinaciones más relevantes entre las 123 posibles. A cada combinación se le asignó un segmento utilizando criterios estratégicos. El cuadro con estas combinaciones y los nombres de los segmentos fue cargada en Power BI y relacionada con los demás cuadros.

Las visualizaciones desarrolladas para este tablero fueron las siguientes:

- Total de clientes: visualización que muestra cuántos clientes fueron clasificados.
- Recencia promedio: medida que indica, en promedio, hace cuántos días los clientes realizaron su última compra.
- Frecuencia promedio: promedio de órdenes realizadas por cliente.
- Valor monetario promedio: promedio del gasto total por cliente en el periodo.
- Participación por segmento: *treemap* donde se muestra el porcentaje de clientes en cada uno de los siete segmentos.
- Conversión de segmentos por trimestre: gráfico de cintas que representa cómo los clientes migraron entre segmentos en distintos trimestres, utilizando el número de teléfono como identificador.
- Ventas por segmento en el tiempo: gráfico de líneas que permite visualizar la evolución de las ventas por grupo de clientes.
- *Marketing* 5 productos más vendidos: gráfico de barras que muestra los productos con mayor rotación.
- Participación por segmento por sucursal: gráfico de barras 100 % apiladas para ver el peso de cada segmento dentro de cada sucursal.
- Gráfico de dispersión por métricas RFM: se crearon dos cuadros con filtros por variable (R, F o M), y dos medidas que calculan el promedio de las variables seleccionadas. En el gráfico de dispersión se colocaron estas medidas en los ejes X y Y, permitiendo visualizar distintos cruces según el criterio elegido.
- Cuadro de información de clientes: cuadro que muestra los datos generales de los clientes, su puntuación RFM y el segmento asignado. Este cuadro sirve para examinar en detalle la composición de cada grupo filtrado.

Luego de haber concluido con la generación de los tableros, se colocaron distintos filtros y botones que ayudarían a la interacción y personalización del mismo, además se agregó una carátula para su presentación. Al finalizar, el resultado fue presentado a la gerencia, quienes aprobaron cada una de las páginas, visualizaciones y paletas de colores.

Al tener aprobado el *dashboard*, se realizó un análisis por cada una de las páginas realizadas, se interpretó cada uno de los gráficos con la finalidad de entender la información obtenida de la simulación, estas interpretaciones pueden ser utilizadas para comprender cada gráfico y poder obtener valiosos *insights* al momento de hacer la implementación del programa de lealtad.

## D. Fase 4 Diseño de esquema para la interpretación y creación de perfiles y segmentos

Luego de haber realizado la interpretación de los datos del *dashboard* se elaboró una presentación que explicaba cuáles eran las visualizaciones del *dashboard* que debían utilizarse para la construcción de los perfiles. Se seleccionaron aquellas gráficas que contenían información relevante para describir a los clientes, como la distribución de edad, la proporción de género, los canales de captación, el método de entrega preferido, la frecuencia de compra, el *ticket* promedio, los horarios más comunes de compra y la sucursal más frecuentada. Esta selección fue clave para asegurar que los perfiles estuvieran contruidos sobre variables concretas y cuantificables.

Se hizo una explicación y sugerencias sobre como llenar y realizar cada uno de los componentes que debe de llevar un *buyer persona* detallando los gráficos y las posibles inferencias que se podían realizar en cada uno de ellos. Los componentes para realizar el perfil fueron: Perfil, Objetivos, Motivaciones, Retos, Comportamiento de compra y Canales de comunicación.

Posteriormente, se estableció la metodología para evaluar el peso en ventas y el peso en volumen de transacciones de cada perfil de cliente. Para ello, se utilizaron las vistas complementarias del *dashboard*, aplicando filtros cruzados sobre rangos de edad, sexo, franja horaria, método de entrega y sucursal. De esta manera, se pudo observar qué combinaciones de características tenían mayor participación en el total de ventas y en la cantidad de órdenes generadas. El objetivo de este análisis fue identificar qué segmentos representaban un mayor aporte económico para la empresa y cuáles podían ser considerados prioritarios para futuras estrategias de fidelización.

Luego se realizaron las páginas en donde se muestra cada segmento y las combinaciones de puntuaciones RFM que lo conforman, para entender cómo se eligió cada uno de los segmentos. Después se colocó la indicación de que el *dashbaord* de segmentación contiene gráficos que ayudan ver el peso de cada segmento y la información encontrada para obtener el segmento que se desea analizar o con el que se quiere implementar alguna campaña. Por último, se muestra el cuadro con la información de los clientes, y se muestra como descargarla para poder realizar la implementación de estrategias.

## VI. Análisis de la empresa

### A. Contexto actual de la empresa

Lulis es una empresa guatemalteca que inició sus operaciones en el año 2019 como *dark kitchen*. Actualmente, la empresa posee sucursales ubicadas en la Ciudad de Guatemala. Lulis se especializa en la producción y venta de galletas artesanales y otros productos horneados. Se ha posicionado como una marca reconocida en el segmento de repostería *prémium*. Uno de sus principales diferenciadores es el ofrecer productos recién horneados, además de contar con ingredientes de alta calidad y cuidadosamente seleccionados. Asimismo, se ha caracterizado por sus productos de temporada, los cuales salen una vez al mes aproximadamente. Los productos de la empresa brindan una experiencia indulgente a sus clientes.

Un pilar fundamental para la empresa es su enfoque en la experiencia del cliente. La marca se asegura de que el servicio que se presta en cada punto de contacto sea sobresaliente, garantizando que los clientes se lleven productos de alta calidad, junto con un servicio satisfactorio. Ambos factores son las principales razones de que los clientes de la empresa resulten ser tan fieles a la marca, ya que se genera un sentimiento de confianza, y que sus compras se relacionan a una buena experiencia.

La marca cuenta con una estrategia que combina la distribución directa, en la que la venta es directa de la empresa hacia el consumidor final, e indirecta corta, la cual consiste en la venta de los productos por medio de un intermediario hacia el consumidor final. En la venta directa, la empresa cuenta con canales físicos y digitales, sus puntos de venta se encuentran ubicados en zonas estratégicas de la Ciudad de Guatemala, y en digital cuentan con atención por medio de las distintas redes sociales. Para la distribución indirecta cuentan con intermediarios tales como Uber Eats y Pedidos Ya, plataformas de entrega de productos a domicilio, son plataformas que conectan a los usuarios con los restaurantes u otros comercios.

La empresa es joven en el mercado guatemalteco, pues no lleva tantos años operando en comparación con otras marcas de repostería establecidas. Sin embargo, ha logrado posicionarse rápidamente gracias a la calidad de sus productos y su enfoque en la experiencia del cliente. Actualmente, la empresa se encuentra en una etapa de crecimiento, además de estar en constante expansión, explorando nuevas oportunidades para abrir más puntos de venta y llegar a una mayor cantidad de sectores en el país. Como parte de este crecimiento, Lulis se mantiene en un proceso de innovación y adaptación a las tendencias del mercado, buscando actualizar su portafolio, mejorar su servicio y optimizar su presencia digital para seguir consolidando su marca y atraer a más clientes.

### B. Manejo de datos

En la actualidad, la empresa cuenta con un sistema de facturación llamado Odo, un sistema creado para que las empresas puedan gestionar y centralizar su información, cuenta con distintos módulos y aplicaciones para las distintas áreas de la empresa, desde este sistema se manejan módulos de venta, contabilidad, inventarios, punto de venta, entre otros. El sistema permite crear reportes personalizados dentro de la plataforma con las variables que se deseen, estos reportes, si se desea, se pueden descargar en formatos CSV o en Excel, sin embargo, para fines de análisis, los reportes no poseen un formato funcional, es por ello que para descargar reportes con datos en tiempo real, se debe primero hacer una conexión entre el sistema y un sistema llamado PostgreSQL, el cual gestiona y almacena bases de datos relacionales con

código abierto, por medio de lenguaje de SQL se pueden realizar consultas de datos y reportes específicos, a grandes conjuntos de datos.

Otro inconveniente que la empresa posee en el manejo de sus datos es la falta de soporte en el sistema, por lo cual, no ha sido implementado y alimentado de la manera correcta en los primeros años de uso, esto debido a que no contaban con una persona encargada de manejar y administrar correctamente la información que se ingresaba al sistema y la falta de un proceso generalizado para el ingreso de la información al mismo. Es por ello que los datos históricos no son relevantes y al ser analizados darían como resultados información de poco valor. Otro suceso que afectó a la información de la empresa fue el cambio de proveedores del sistema, por lo que mucha información se perdió. A raíz de esta situación, la empresa buscó a una persona que se encargara del sistema, lo que ha permitido lograr un orden en su funcionamiento y mejorar su operatividad. Ahora, la información y los datos que arroja el sistema son correctos y eficientes, lo que facilita su análisis y contribuye a una toma de decisiones más informada.

### C. Toma de decisiones estratégicas de mercadeo actual

En términos de *marketing*, la marca tiene una fuerte presencia en redes sociales, como Instagram, Facebook y TikTok, donde comparte contenido atractivo que resalta la calidad de sus productos. El contenido de las redes sociales es realizado a través de una agencia de publicidad que la empresa ha contratado, la agencia se encarga del manejo de pautas en redes como Instagram y Facebook, además de la creación de contenido en estas plataformas. La empresa también implementa promociones y campañas especiales para impulsar ventas y fidelizar a su audiencia; sin embargo, estas promociones son realizadas con objetivos ambiguos, lo que ha dificultado la medición de su efectividad y el impacto real en el comportamiento del consumidor.

Hasta hace poco, ellos no trabajaban con decisiones basadas en datos, lo que limitaba la capacidad de la empresa para optimizar sus estrategias de *marketing* y comprender a profundidad el desempeño de sus acciones comerciales. Actualmente, la empresa se encuentra en un proceso de construcción de una cultura de decisiones basadas en datos, con el objetivo de utilizar información estructurada y análisis de datos para tomar decisiones más efectivas. Este cambio permitirá mejorar la segmentación de clientes, afinar las campañas publicitarias y optimizar las promociones para que generen un mayor impacto en las ventas y la fidelización de los consumidores.

### D. Identificación de necesidades de la empresa

Actualmente, la empresa enfrenta desafíos en su estrategia de mercadeo, principalmente debido a la ausencia de un plan estructurado, con decisiones fundamentadas, que permita una gestión eficiente de sus canales de comunicación y adquisición de clientes. Si bien esta estrategia ha permitido mantener presencia digital, su impacto es limitado, ya que no poseen información valiosa de los resultados para poder actuar con estos, además debido a la falta de un enfoque basado en el conocimiento de los segmentos de clientes, las estrategias en ocasiones, las estrategias no están bien dirigidas, lo que resulta en una falta de relevancia y visibilidad. Esto puede llevar a una escasa aceptación por parte de los involucrados, lo que a su vez provoca pérdidas o la ineficacia de los objetivos planteados.

Actualmente la empresa no posee los datos suficientes para poder hacer la perfilación de sus clientes, y a pesar de contar con la información para conocer los comportamientos, intereses y hábitos de consumo de los clientes, no han realizado los análisis para poder realizar segmentaciones efectivas, lo que impide la optimización de las estrategias de *marketing*. Se

emiten campañas o promociones que en ocasiones no resultan ser efectivas. Asimismo, un reto de la empresa es que depende en gran medida de plataformas de distribución indirecta, lo que limita su capacidad de control sobre la experiencia del consumidor y los datos generados a lo largo del proceso de compra de los mismos. Esta dependencia del uso de estas plataformas reduce la rentabilidad y dificulta la creación de estrategias de fidelización personalizadas que puedan fortalecer la relación con los clientes.

La empresa no cuenta con un sistema estructurado para la recolección de los datos necesarios para realizar distintas segmentaciones. Asimismo, necesitan encontrar una manera de gestionar y visualizar estos datos para la toma de decisiones en el área de mercadeo. Sus estrategias deben de ir enfocadas en atraer a consumidores a los canales propios de la empresa, siendo estos los 9 puntos de venta físicos, el chatbot de WhatsApp y redes sociales como Instagram y Facebook, donde también se toman pedidos, esto debido a la rentabilidad que se obtiene al hacer uso de sus canales.

## VII. Diseño del esquema de recolección de datos

Luego de conocer las necesidades actuales de la empresa, se planteó una estrategia con el objetivo de responder y apoyar al crecimiento estratégico de la empresa. Dicha estrategia permite el fortalecimiento de las relaciones que la empresa tiene con sus clientes, el poder conocer mejor el perfil y comportamiento de los clientes, mejorar las decisiones estratégicas de mercadeo y optimizar la efectividad de sus estrategias. La empresa no cuenta con la suficiente información para el análisis y comprensión de los perfiles de sus consumidores y el comportamiento de cada uno de ellos, es por ello que, para lograr el desarrollo de la estrategia, es importante contar con datos relevantes y actualizados sobre los consumidores y sus datos transaccionales, con ellos, se espera obtener *insights* accionables para crear estrategias más personalizadas y efectivas.

La estrategia consiste en el desarrollo de un plan de fidelización, que además de ser un incentivo para la recompra y generación de lealtad de los clientes, es un medio por el cual se puede realizar la recolección de los datos necesarios para el análisis propuesto. Por medio de este sistema de recolección se obtiene información clave como la frecuencia de compra, el *ticket* promedio, productos más vendidos, edad, género, número de teléfono, entre otros. Dicha información es útil para la construcción de perfiles y segmentos que serán utilizados para la creación de estrategias que puedan generar valor a los clientes basado en sus preferencias, además de tomar decisiones fundamentadas en datos que contribuyan a la eficiencia de la empresa.

### A. Elección de un programa de fidelización

La elección de implementar un plan de fidelización como la estrategia para la recolección de datos responde a diversas necesidades específicas que se han identificado dentro de la empresa. En primer lugar, Lulis tiene una clara necesidad de recopilar datos de sus clientes, tanto personales como transaccionales, con el fin de construir perfiles, segmentar adecuadamente y tomar decisiones basadas en información real. Sin embargo, la recolección de datos puede resultar complicada si no se ofrece un beneficio tangible a los clientes a cambio de su información. En este sentido, un programa de fidelización representa una solución efectiva, ya que permite incentivar la participación de los consumidores a través de recompensas, mientras se obtiene la data necesaria para fortalecer la toma de decisiones.

Además, se identificó la necesidad de migrar progresivamente a los clientes que hoy consumen a través de plataformas externas como PedidosYa y Uber Eats hacia los canales propios de la marca. Estas plataformas, aunque útiles para ampliar el alcance, afectan directamente la rentabilidad del negocio debido a las altas comisiones que implican. A través de un plan de fidelización gestionado desde canales propios (como una app o sitio web), se puede ofrecer un valor añadido al cliente y motivarlo a comprar directamente, aumentando así el margen de ganancia para la empresa.

Otra razón clave detrás de esta decisión es que los clientes de Lulis ya presentan altos niveles de fidelidad, según lo expresado por los propios dueños y personal, quienes mencionan que incluso han aprendido a reconocer a varios clientes frecuentes. Esta observación demuestra un gran potencial para profundizar el vínculo con estos consumidores, pero también plantea la necesidad de conocerlos mejor, entender quiénes son, qué los motiva a comprar y qué otros segmentos podrían estar creciendo en torno a la marca. Con esta información, no solo se podrá recompensar de manera más justa y estratégica a los clientes leales, sino también diseñar acciones de *marketing* más personalizadas y efectivas, alineadas a las preferencias y comportamientos reales del consumidor.

En resumen, el plan de fidelización fue elegido por su capacidad para responder a múltiples necesidades de forma integrada: obtener datos valiosos, mejorar la rentabilidad canalizando las ventas a medios propios, aprovechar la lealtad ya existente en la base de clientes, y construir una estrategia más precisa y centrada en el consumidor.

## B. Referencia de mercado

Se realizó un análisis de la competencia con el objetivo de identificar estrategias similares dentro de la industria de alimentos. Dentro de la búsqueda, no se hallaron empresas que formen parte de la competencia directa. Sin embargo, dentro de la industria de alimentos y bebidas, principalmente en el sector de restaurantes y comida rápida, sí se encontraron programas de fidelización que buscan incentivar la recompra y recompensar a sus clientes más fieles. Algunos de los lugares encontrados que manejan este tipo de iniciativas son:

- Starbucks: forma parte de una cadena internacional de cafeterías que ofrece una amplia variedad de bebidas y productos de panadería. Tienen un fuerte enfoque en la experiencia y personalización hacia los clientes. Starbucks inició sus operaciones de retail en Guatemala en el 2011 y desde entonces empezó su crecimiento por la región. (*Starbucks Celebra La Inauguración de Su Tienda 1,500 En Latinoamérica Con La Primera Cafetería En Antigua, Guatemala - Starbucks Historias*, 2022). El 25 de octubre del 2024 Starbucks anunció el lanzamiento de su programa de lealtad en Guatemala, por medio de este programa los clientes de la marca obtienen distintos beneficios, promociones especiales, acceso prioritario a algunos productos y eventos exclusivos. (*Starbucks Lanza En Guatemala Su Programa de Lealtad, Starbucks® Rewards - Starbucks Historias*, 2024)
  - Plataforma: aplicación móvil disponible en Android, también puede ser usada en la web.
  - Modalidad del programa: luego de descargar la aplicación se llena el formulario de registro que solicita información como: nombre, apellido, apodo, correo electrónico, fecha de cumpleaños, teléfono, contraseña. Es un programa de puntos, a los que ellos llaman “estrellas”, al pagar con efectivo, tarjeta de crédito o débito 1 estrella equivale a Q1, mientras que con saldo de una Starbucks card se acumulan 2 estrella por cada Q1. Los incentivos son dados en forma de cupón, el beneficio producto dependerá de la cantidad de estrellas y el nivel al que se pertenece. Existen 2 niveles de beneficios en Starbucks Rewards, esto se determina por la cantidad de estrellas acumuladas. El primer nivel, “Nivel Green”, primer nivel obtenido al crear la cuenta, en este nivel se tiene acceso a cambiar 3 distintos cupones como 1 shot de café extra mensual, 1 café americano por cada compra de una bolsa de grano de café y la bebida de cumpleaños, adicional ofrecen una bebida por cada 1200 estrellas acumuladas. El segundo nivel, “Nivel Gold”, se obtiene al haber acumulado 1600 estrellas en el periodo de un año. Para mantener este nivel es necesario acumular 1600 estrellas anuales, los beneficios de este nivel son los mismos que el Nivel Green más opción de cambio de grano a grano premium, rellena el vaso de café o té del día.
  - Reglamento: Las estrellas tienen vigencia de 1 año a partir del último día del mes en el que se acumularon. Las compras pueden ser pagas dentro de la aplicación o en el punto de venta físico y se acumulan las estrellas al escanear el código QR. No cuentan con servicio a domicilio, las compras realizadas dentro de la aplicación deben ser recogidas en tienda. El crear una cuenta no tiene ningún costo, se deben tener mínimo 18 años para poder crear una cuenta, para poder crear la cuenta es de carácter obligatorio hacer el llenado del formulario de registro, se puede tener una cuenta por correo

electrónico.

- McDonalds: es una cadena de comida rápida orifinaria de Estados Unidos. La primera sucursal en Guatemala llegó en 1974. Para el 2022 ya tenían 100 restaurantes alrededor de la región. En el 2021 Mcdonalds lanzó su programa de lealtad, un programa en donde se acumulan puntos y se redimen por productos (*McDonald's Mesoamérica, 2025*).
  - Plataforma: aplicación móvil disponible para Android y iOS. (Anexo x)
  - Modalidad del programa: la modalidad del plan de lealtad de acumulación de puntos, cada quetzal gastado dentro de la aplicación es igual a un punto. Hay un catálogo de productos disponibles para canjear por los puntos, en total son 16 productos incluyendo productos de desayuno y almuerzo. Hay 5 categoría de puntos, la más baja es por productos como antojos o acompañamientos y la más alta es el canje de una de sus hamburguesas más populares. Se pueden hacer canjes de productos desde los 165 puntos por productos como derretidos, pastel de manzana, entre otros, con valores de entre los Q15 a Q20. Los productos con 230 puntos requeridos tienen precios de entre Q22 a Q25. Luego la categoría de 310 puntos se encuentran productos de precios entre Q24 a Q30. En la categoría de 530 puntos hay productos de precios entre Q30 y Q35. Por último el producto canjeable por 630 puntos tiene un precio de Q43.
  - Reglamento: para poder hacer redención de los puntos es necesario ser mayor de edad. Los usuarios pueden hacer el canje por la cantidad de puntos que desee, dependiendo de la disponibilidad de puntos acumulados, además de que el monto mínimo de compra debe de ser para los desayunos Q40 y para el almuerzo y cena Q50. Los puntos pueden ser redimidos únicamente por medio de la aplicación de envíos a domicilio. Los puntos no pueden ser canjeados por dinero en efectivo ni pueden ser transferidos (Mcdonalds, s.f).
  
- Subway: es la cadena de sándwiches más grande del mundo. Inició sus operaciones en Guatemala en septiembre de 1996. En la actualidad, Subway cuenta con más de 70 sucursales en el país. La aplicación de Subway para realizar pedidos incluye un apartado de "lealtad", la aplicación fue lanzada en el año 2021 para Guatemala. (Softonic, 2025)
  - Plataforma: aplicación móvil disponible en Android y iOS.
  - Modalidad del programa: el programa se llama "subway Card". Para acceder al programa de lealtad primero debe hacerse un registro, para el cual se deben llenar datos como: nombre completo, email, DPI, teléfono, contraseña, sexo. Los puntos son acumulados automáticamente si se hace el pedido a domicilio por medio de la aplicación, o si se realizó la compra en un punto de venta físico se muestra la "Subway card" con un código de barras y se acumulan los puntos. La acumulación de puntos varía según el nivel en el que se clasifique el cliente, nivel que depende del consumo registrado. Los puntos se acumulan dependiendo de su nivel, se mide puntos por dólar en este caso. Existen cinco niveles dentro del programa: Regular, Bronce, Plata, Oro y Platino. Los clientes que acumulan entre 25 y 49 puntos en ese periodo son clasificados como Bronce y obtienen 1.25 puntos por cada dólar gastado. Quienes alcanzan entre 50 y 124 puntos acceden al nivel Plata, acumulando 1.50 puntos por dólar. Los clientes con entre 125 y 324 puntos entran en el nivel Oro, con una tasa de 1.75 puntos por dólar. Finalmente, aquellos que superan los 325 puntos en el semestre acceden al nivel Platino, recibiendo 2 puntos por cada dólar de compra. Los clientes que no alcanzan los 25 puntos permanecen en el nivel Regular, acumulando 1 punto por cada dólar. Hay 16 productos disponibles en el catálogo para

poder canjear, hay productos que pueden ser canjeados desde los 15 puntos hasta con 110 puntos, los precios de estos productos varían entre Q8 a Q42. El primer ingreso se otorgan 70 puntos por bienvenida.

- Reglamento: se puede hacer redención de los puntos únicamente personas mayores de 18 años. Solo pueden participar consumidores finales; es decir, personas que adquieren productos para su consumo personal, quedan excluidas las compras corporativas. Las compras realizadas con promociones o cupones de temporada no aplican para la acumulación de puntos. Los puntos son intransferibles y pueden ser canjeados únicamente por los productos puestos en el catálogo de redenciones. Los puntos tienen una vigencia de únicamente 6 meses, para subir de nivel los puntos necesarios deben ser acumulados dentro de este periodo de tiempo. La tarjeta puede ser utilizada únicamente 3 veces al día.
- TrefraGroup: es una cadena de restaurantes en Guatemala. El primer restaurante se inauguró bajo el nombre de “Trefratelli” en 1995, a partir de entonces empezó su crecimiento, abriendo sucursales en distintas zonas del país. Con este crecimiento se expandieron y diversificaron los estilos de comida. Para 2025 el portafolio de marcas del grupo incluía Trefratelli, Frisco Grill, JK Ming y La Cantina. La empresa lanzó su programa de lealtad llamado Trefra Club, el cuál puede ser utilizado en cualquiera de sus restaurantes (Trefra Group, 2025).
  - Plataforma: aplicación móvil disponible en Android e iOS, también en la página web.
  - Modalidad del programa: el programa se llama “**Trefra Club**”. Para participar, el cliente debe registrarse a través del sitio web o la app, proporcionando datos como nombre, correo electrónico, contraseña, DPI, número de teléfono, fecha de nacimiento. Los usuarios reciben un **cashback del 5 %** en consumos mayores a Q50 en cualquiera de los restaurantes del grupo. El cashback acumulado puede canjearse por productos dentro del menú, se acumula en la cuenta del cliente y puede visualizarse desde la aplicación o el sitio web. El monto acumulado puede ser utilizado como forma de pago en compras futuras dentro de cualquiera de los restaurantes del grupo, tanto en consumo en el lugar como en pedidos para llevar. Por realizar el registro se otorga un beneficio de Q100 cashback.
  - Reglamento: está dirigido exclusivamente a personas mayores de 18 años y a consumidores finales, quedando excluidas las compras corporativas. Para que una compra sea válida en la acumulación de beneficios, esta debe ser igual o superior a Q50. El cashback obtenido tiene una vigencia de seis meses a partir de su acumulación, y tanto este como los demás beneficios del programa son estrictamente personales e intransferibles. Además, no pueden redimirse en promociones bancarias tipo 2x1, no pueden ser canjeadas por efectivo y no pueden ser transferidos ni reembolsados.

Luego de haber realizado la exploración de programas de lealtad en la industria de alimentos se ha observado una tendencia de utilización de incentivos, en su mayoría, por medio de aplicaciones que requieren el llenado de un formulario de registro para poder acceder a los beneficios que proporciona la empresa, además de fomentar la lealtad a la marca e incentivar la recompra. Sin embargo, existen varias restricciones para el uso de estos programas, entre ellas se encuentra el requisito de ser mayor de edad, el requerimiento de un consumo mínimo para acumular puntos o cashback, la intransferibilidad de los beneficios, la exclusión de compras corporativas o promociones, así como la vigencia limitada de los beneficios acumulados.

## C. Objetivos del plan de fidelización

Implementar un plan de fidelización basado en puntos para la recolección de información sobre los clientes, con la finalidad de mejorar la toma de decisiones de *marketing* y fortalecer la relación entre los clientes y la marca.

- a. Recolectar y analizar datos detallados sobre patrones de consumo y preferencias de los clientes, utilizando un sistema estructurado de recompensas por puntos.
- b. Incrementar la migración de clientes desde plataformas hacia los canales propios, utilizando incentivos atractivos que fomenten las compras directas, aumentando la preferencia hacia los canales directos de la empresa.
- c. Mejorar la segmentación de clientes mediante la generación de un modelo de análisis conductual, con la finalidad optimizar campañas de *marketing* que incrementen la recompra, eleven el gasto promedio por cliente y fortalezcan la relación con la marca.
- d. Desarrollar un panel de visualización de datos que muestre métricas clave del programa de fidelización, ayude a generar el perfil de los clientes, facilite el análisis de tendencias y optimice la toma de decisiones.

## D. Análisis de datos transaccionales

Utilizando una base histórica de datos transaccionales de la empresa y una base de márgenes, se realizó un análisis de los resultados que se obtuvieron a partir de la exploración y procesamiento de la data mediante un código de programación en Python, el código completo se encuentra adjunto en el Anexo X. Para este análisis se usaron distintas visualizaciones gráficas que permitieron procesar la información de manera más eficiente, permitiendo generar *insights* que ayudarían con la selección de los productos a utilizar en el programa de fidelización. El análisis se orientó a responder las siguientes preguntas:

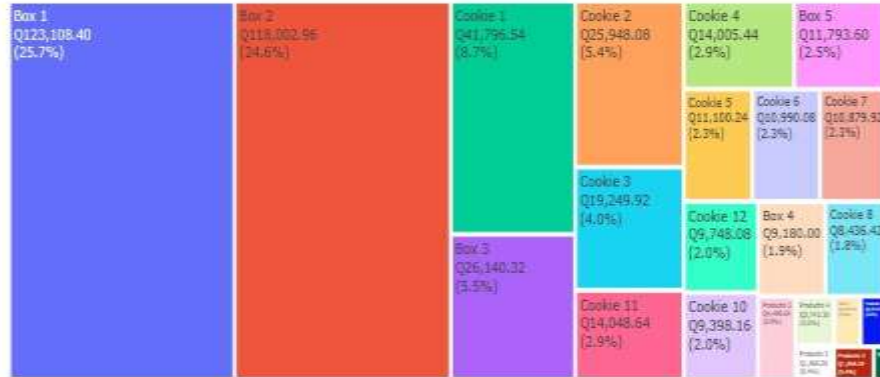
- ¿Cuál es la participación de cada producto en el volumen total de ventas?
- ¿Cuál es la participación de cada producto en el ingreso total por ventas?
- ¿Cómo se distribuye el *ticket* promedio de compra en los clientes?
- ¿Cuáles son los 10 productos con mayor margen de rentabilidad?
- ¿Cuál es la relación entre el volumen de ventas y el margen de rentabilidad por producto?
- ¿Cómo se relacionan los ingresos totales por ventas con el margen de rentabilidad por producto?

### 1. Resultados

- a. En este gráfico se observa la proporción que cada producto aporta al total de ingresos generados por ventas en valor monetario, se pretende entender la contribución económica que aporta cada producto. En el gráfico puede verse que los productos con una mayor participación en términos de ingreso son “Box 1” y “Box 2”, generando Q123,108.40, 25.7 % de participación y Q118,002.96, 24.6 % de participación respectivamente, las cuales representan más del 50 % de los ingresos de la empresa. La “Cookie 1” es el tercer producto con mayor ingreso con Q41,796.54 con una participación del 8.7 %, al tener un precio bajo

esto indica que la rotación del producto es alta.

**Figura 1. Participación de la venta monetaria por producto**



Fuente: Elaboración propia

- b. Se muestra la participación del volumen de venta por producto. Lo que se observa en el gráfico es que la “Cookie 1” es el producto con el mayor volumen de ventas, con una participación del 22 % con un total de 3,110 unidades vendidas. El segundo producto con un mayor volumen de ventas es la “Cookie 2” con participación del 13 %. Se observa que el producto con mayor ingreso monetario (figura 1) es el tercer producto con mayor volumen de ventas, lo que indica que los ingresos de este producto se deben a una combinación entre el precio y la rotación del mismo. Por otro lado, se puede ver que el producto “Box 2” que tiene la segunda mayor participación en ventas (figura 1) tiene únicamente un 5 % de participación en volumen, lo que puede indicar que las ventas se deben al precio elevado del producto.

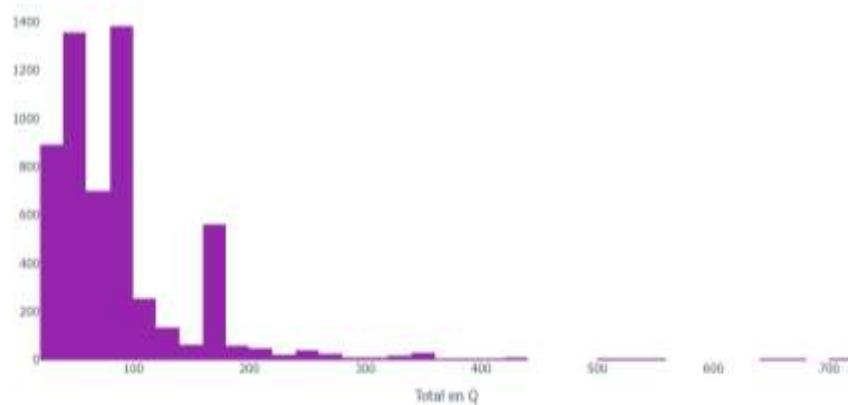
**Figura 2. Participación del volumen de venta por producto**



Fuente: Elaboración propia

- c. En el histograma se observa una clara cola hacia la derecha, es decir, el gráfico posee un sesgo positivo, lo que indica que los valores se encuentran concentrados en los valores más bajos, en este caso el gráfico indica que en promedio los clientes por cada orden de compran, gastan alrededor de entre Q25 y Q100, con una probabilidad mayor que el *ticket* tenga valores entre los Q50 y Q90. Asimismo, se observa que la cola hacia la derecha es larga, indicando que, aunque en menor proporción, hay facturas que alcanzan montos por encima de los Q200, y algunos hasta Q700. Este comportamiento sugiere la existencia de un segmento reducido que hace compras de mayor volumen o de productos de *ticket* promedio elevado.

**Figura 3. Distribución del ticket promedio**

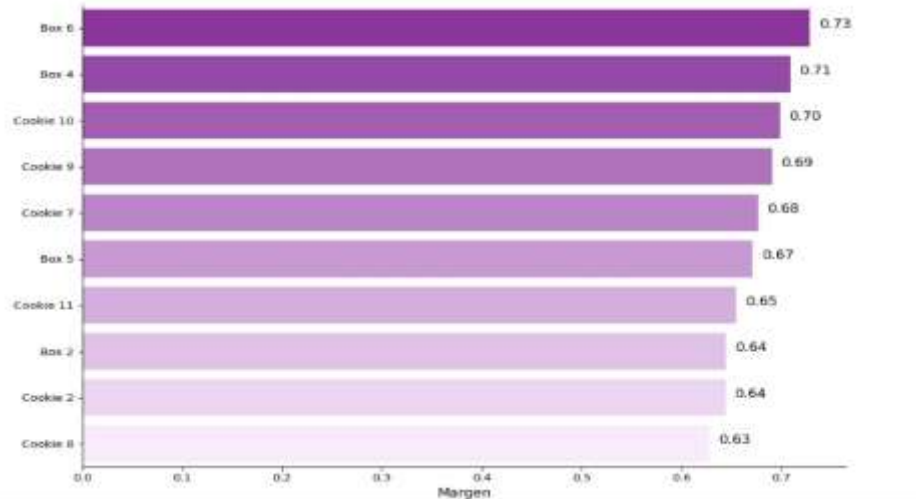


Fuente: Elaboración propia

- d. En el gráfico se ve el *marketing* 10 de los productos con mejores márgenes de todo el catálogo. El margen se encuentra en decimales, lo que quiere decir que los valores más cercanos a 1 son los márgenes más altos. El producto con un mayor margen es “Box 6” con un margen de 73 %, seguido de “Box 4” con un margen de 71 %. En el *marketing* 10 de los márgenes más altos, el menor margen es de 63 % de la “Cookie 8”. Se observa que tanto los productos de la categoría de cookies como la categoría de boxes se encuentran entre los más altos, lo que sugiera que el margen no depende directamente del tipo de producto, también depende de la estrategia de precios aplicada y el manejo de los costos de cada producto.

**Figura 4. Marketing 10 productos con mejor margen**

Fuente: Elaboración propia



- e. En el gráfico se observa la relación entre el volumen de ventas y el margen, por producto. Se observa que la mayoría de los productos tienen una rotación promedio de entre 22 a 2000 unidades vendidas, sin embargo se observa que la “Cookie 1” es un dato atípico ya que tiene en total 3110 unidades vendidas.

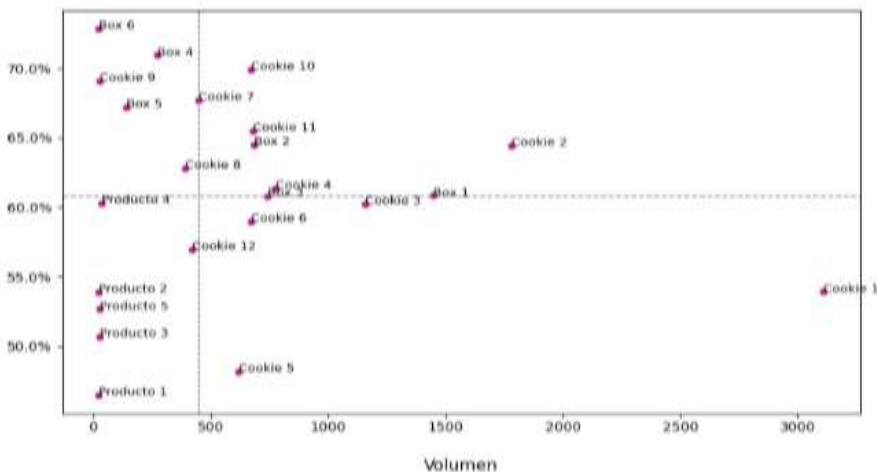
En el cuadrante superior izquierdo pueden observarse aquellos productos con un alto margen, pero baja rotación, en este cuadrante se encuentran los productos: “Box 6”, “Box 4”, “Cookie 9”, “Box 5”, “Cookie 7” y “Cookie 8”. Estos productos tienen un margen arriba del promedio, pero de los volúmenes de venta más bajos, esto puede representar una oportunidad para la empresa, al ser productos con buenos márgenes pueden usarse en promociones o como incentivos para darles una mayor visibilidad.

En el cuadrante superior derecho se encuentran aquellos productos con margen alto y una alta rotación, son los productos de mayor valor para la empresa, además que son excelentes para poner como incentivos ya que son productos populares que pueden colocarse para atraer a usuarios. La mayoría de productos se concentran en este cuadrante, algunos de los productos con mayor rotación en este cuadrante son: “Cookie 2”, “Box 1”, “Cookie 11”, “Cookie 10”, entre otros.

Para el cuadrante inferior izquierdo, se observa que se encuentran todos los productos que pertenecen a la categoría de “Varios”, son todos aquellos productos extras que ha sacado la marca para diversificar su catálogo de productos y entrar a competir en otras categorías. Sin embargo, estos productos no son productos que aporten significativamente a la empresa ya que los márgenes son mucho más bajos y tienen poca rotación.

Por último, en el cuadrante inferior derecho, el cual muestra productos de bajo margen pero de alto volumen de ventas. En este cuadrante se encuentra la “Cookie 1” el producto más vendido, pero con un margen menor al promedio.

Figura 5. Relación entre el volumen de venta y el margen por producto



Fuente: Elaboración propia

- f. En este gráfico de dispersión, donde se observa la relación entre los ingresos monetarios y el margen de cada producto. De manera general, se observa una alta concentración en de productos que generaron ingresos hasta de aproximadamente Q30,000. Sin embargo, en esta distribución se observa una cola larga hacia la derecha, el siguiente dato que se encuentra arriba de los Q40,00 es la “Cookie 1”, y con ingresos de hasta Q120,00 se observan 2 productos, “Box 2” y “Box 3”.

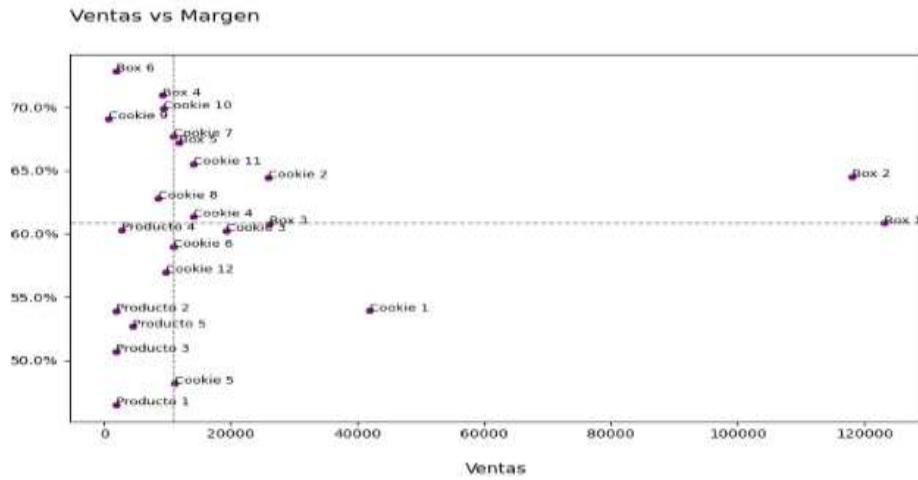
En el cuadrante superior izquierda se observan productos con márgenes arriba del promedio, pero con ingresos bajos, en este cuadrante se observan nuevamente productos como “Box 6”, “Box 4”, “Cookie 8”, entre otros. Esto puede deberse también a que alguno de los productos ha sido lanzado recientemente, o no se le ha dado la visibilidad suficiente, se deben evaluar estos factores para entender si es conveniente continuar con dichos productos en el menú.

En el cuadrante superior derecho se encuentran productos con márgenes altos y ventas monetarias significativas, se encuentran los productos que representan un 50 % de los ingresos, haciéndolos los productos importantes para la empresa, ya que poseen un buen desempeño tanto en la rentabilidad como en los ingresos.

En el cuadrante inferior izquierdo se vuelven a mostrar los productos de la categoría “Varios”, es importante prestar atención a estos productos ya que al ser poco rentables es poco atractivo invertir en esfuerzos de *marketing* para el aumento de su rotación e ingresos, es importante evaluar la relevancia de estos dentro del menú.

En el cuadrante inferior derecho, donde las ventas son significativos pero la rentabilidad es baja, se encuentra el producto más popular de la marca, la “Cookie 1”, producto con una rotación muy alta y una venta monetaria relevante, sin embargo con una baja rentabilidad, esto podría deberse a la falta de revisión de costos del producto. Asimismo se encuentran productos como “Cookie 3” y “Cookie 5”, con una venta menos relevante.

Figura 6. Relación entre la venta monetaria y el margen por producto



Fuente: Elaboración propia

El análisis exploratorio de los datos transaccionales históricos de la empresa, permitió identificar patrones clave en el comportamiento de ventas y rentabilidad de los productos. Se obtuvo información para la toma de decisiones estratégicas sobre los productos que irían dentro del catálogo de productos canjeables para el programa de fidelización. Dentro de la información analizada se encontró que un grupo reducido de productos, como “Box 1”, “Box 2” y “Cookie 1”, concentra la mayor parte de los ingresos, aunque no todos presentan altos márgenes de rentabilidad, sin embargo estos pueden ser productos que funcionen como incentivos para la participación, asignándole una mayor cantidad de puntos para que puedan ser canjeados. Asimismo, se identificaron productos con márgenes elevados pero baja rotación, los cuales podrían ser usados como productos canjeables con la finalidad de darles una mayor visibilidad y aumentar su rotación. La segmentación por cuadrantes en los gráficos facilitó la visualización de productos estratégicos, aquellos con alto margen y alta rotación, que son ideales para incluir en el programa de fidelización.

## E. Estructuración del plan de fidelización

### 1. Planteamiento del plan de fidelización

Para el desarrollo del plan de fidelización, fue necesario establecer la plataforma en la que se desarrollaría, para elegir el canal por el cual se desarrollaría esta estrategia se tomaron en cuenta factores como la fácil interacción con la plataforma, la simplicidad con el uso de la herramienta, personalización del programa de fidelización, y posibles integraciones con el sistema actual de la empresa. Tomando esos factores en cuenta, se decidió, junto a la gerencia de la empresa, que la plataforma adecuada para la futura implementación del programa sería la misma plataforma que se utiliza actualmente para el manejo del chatbot de Whatsapp. El sistema funcionará iniciando una conversación con el chatbot, que enviará un enlace a la página de pedidos habitual. En esta página, los clientes podrán visualizar sus puntos acumulados, los productos canjeables y la cantidad de puntos requerida para cada uno.

Es posible establecer ciertas reglas para la realización de pedidos, acumulación de puntos y canjes de productos. Dentro de la plataforma es posible realizar estos ajustes, con la finalidad de que el programa sea rentable para la empresa. Para poder realizar un pedido, el monto mínimo de compra es de Q50, asegurando la compra de mínimo 2 productos, incentivando un volumen de compra sostenible para la empresa. Los clientes acumulan un punto por cada quetzal gastados, es decir, 1 punto es igual a Q1. Los puntos son canjeables únicamente por los productos previamente establecidos, no pueden ser canjeados por otro producto del menú ni tampoco usados parcialmente para obtener un descuento. Los puntos acumulados tienen una validez de 12 meses para fomentar la recurrencia de compra. Se regalarán 50 puntos a cada usuario por el primer registro al programa. Durante períodos clave, como días festivos, se ofrecen puntos dobles para promover las ventas. Estas reglas son diseñadas para equilibrar los incentivos al cliente con los objetivos comerciales, además facilitan la captura de datos de compra.

### 2. Puntos de contacto

El programa de fidelización se estructura en torno a diversos puntos de contacto diseñados para maximizar la inscripción y participación de los clientes. Las historias de Instagram y Facebook, con enlaces directos al chatbot, captan a los seguidores mediante contenido visual atractivo. La generación de “push notifications”, mensajes enviados por medio de whatsapp a números que han escrito anteriormente, invitando a participar en el programa de fidelización a clientes previos, utilizando un arte atractivo y un mensaje que haga un llamado a la acción. Una sección destacada en el sitio web ofrece que redirige a los usuarios directamente a la página principal de la plataforma para hacer el registro inicial, usando un diseño intuitivo y llamativo con la finalidad de fomentar el registro de los usuarios. En los puntos de venta físicos, contar con QR que dirijan a la página de registro, además de usar las tablets en tiendas para permitirles a los usuarios realizar sus registros en el instante si lo desean, esto permiten a los clientes inscribirse al momento, y así integrar la experiencia presencial con el programa digital. Estos puntos de contacto, estratégicamente diversificados, aseguran una incorporación accesible y efectiva, promoviendo el registro y uso del programa.

### 3. Productos y sistema de puntos

Los productos disponibles para la redención fueron elegidos a partir del análisis de las transacciones históricas de la empresa junto con los márgenes de cada uno de los productos. Se eligieron productos de alta demanda, para incentivar a los consumidores a participar, productos con excelente margen, pero con menor volumen de ventas, con la finalidad de promover su rotación y darle mayor visibilidad. La asignación de puntos para cada producto se realizó tomando en cuenta 3 factores: el margen, la rotación del producto y el valor percibido por el cliente, este último es un porcentaje que se obtiene al dividir el precio entre la cantidad de puntos asignada, por ejemplo, un producto de Q25 podría requerir 250 puntos, reflejando un equilibrio entre atractivo para el cliente y rentabilidad. Tomando en cuenta esas variables, se asignaron los puntos a cada producto. Los productos elegidos junto con los puntos para realizar el canje se encuentran en la Cuadro 1.

**Cuadro 1. Cuadro de productos para la redención del programa de lealtad**

Producto	Puntos	Precio	Valor percibido	Margen
Cookie 1	400	Q23.8	0.05	60 %
Cookie 2	320	Q23.8	0.07	67 %
Box 3	480	Q43.3	0.09	69 %
Cookie 10	380	Q23.8	0.06	65 %
Cookie 9	430	Q34.7	0.08	62 %

Fuente: Elaboración propia

### 4. Datos para el registro

Para desarrollar un análisis efectivo sobre los perfiles de los clientes de la marca, la parte más importante es poder contar con la información necesaria y relevante. Para recopilar datos estratégicos que impulsen decisiones basadas en datos y campañas personalizadas, el chatbot del programa de fidelización solicita la siguiente información durante el registro, diseñada para ser intuitiva y recopilada con el consentimiento explícito del cliente:

- Nombre y apellido: con una entrada de texto para identificar al cliente individualmente, permitiendo comunicaciones personalizadas y perfiles únicos.
- Fecha de cumpleaños: con formato de fecha (“yyyy-mm-dd”) para conocer las edades más comunes que compran recurrentemente en la empresa. Asimismo, se podrán diseñar campañas dirigidas, como promociones de cortesía en cumpleaños, fomentando la recurrencia.
- Correo electrónico: entrada de texto para establecer un canal de comunicación directa, permitiendo notificaciones personalizadas.
- Sexo: selección múltiple (masculino, femenino) para optimizar la segmentación demográfica, apoyando campañas específicas y predicción de preferencias. Esta parte del formulario se dejará como opcional.
- Canal de descubrimiento del programa: selección múltiple (Instagram, Facebook, TikTok, WhatsApp, recomendación personal, tienda física) para evaluar la eficacia de los puntos de contacto y optimizar estrategias de *marketing*.

La información de registro será necesaria para la perfilación de clientes y generación de segmentos, con la finalidad de generar estrategias enfocadas para que estas resulten ser más efectivas y con mejores resultados

## 5. Estrategia general de comunicación

Se busca el planteamiento de los lineamientos generales para el desarrollo de la estrategia de comunicación del programa de fidelización, la cual busca maximizar la captación de participantes, fomentar un engagement continuo y promover interacciones en canales propios mediante una campaña multicanal estructurada en tres fases:

- **Concienciación:** durante la primera fase el objetivo principal es comunicar la existencia del programa de fidelización de la marca. Durante esta fase es importante destacar la propuesta de valor del programa, resaltando los beneficios que se ofrecen para poder atraer a los clientes potenciales. En esta etapa se genera interés, asegurando también que los usuarios comprenden los beneficios y el propósito de el programa que se ofrece. Para esta fase, los mensajes que se comunican son claros, resaltan la facilidad del uso del programa, desde el registro hasta el canje de los productos, además de comunicar las recompensas disponibles para generar curiosidad y deseo por conocer más de la marca y su plan de lealtad.
- **Activación:** para esta fase, lo importante es la conversión de leads, es decir que, el interés generado en la primera fase debe convertirse en una acción, esto se hace incentivando a los usuarios a hacer su registro en la plataforma y comenzar a formar parte del plan de lealtad. El objetivo principal de la etapa es maximizar la inscripción de participantes, obtener la mayor conversión posible. Durante esta etapa, el contenido que debe generarse es un contenido de sentido de urgencia y valor añadido, usando incentivos y promociones temporales.
- **Retención:** es importante recordarle a los usuarios que pertenecen al programa de fidelización de la marca y motivarlos a completar los puntos necesarios para el canje de productos. Por ello, esta etapa está enfocada en mantener el engagement a largo plazo, incentivando a los clientes a comprar de manera recurrente. El objetivo de esta fase es aumentar la participación continua en el programa, transformando a los clientes en clientes habituales. Para esta etapa es importante tener una comunicación más personalizada y constante, en donde se ofrezcan recompensas exclusivas y recordatorios de los beneficios.

Los canales seleccionados para la comunicación de del programa de lealtad son redes sociales, puntos de venta físicos y WhatsApp. Los canales se integran estratégicamente para ofrecer una experiencia visualmente atractiva, accesible y coherente, diseñada para maximizar la captación de participantes y la recopilación de datos valiosos. Estos canales fueron elegidos por su capacidad para conectar con los clientes en múltiples puntos de contacto, aprovechar el atractivo visual de los productos de la empresa y facilitar interacciones directas que promuevan el registro y la participación activa.

- **Redes sociales:** los canales principales del programa serán Instagram, Facebook y TikTok, creando contenido orgánico. Para mantener la constancia y relevancia, se debe cumplir con una planificación de tres contenidos semanales relacionados con el programa, acompañados de historias donde se muestran los productos destacados y clientes disfrutando de los beneficios.

Además, quincenalmente se compartirán videos testimoniales de clientes satisfechos o mostrando el funcionamiento de la plataforma. Para los primeros meses de lanzamiento, la estrategia es generar una campaña de publicidad pagada en redes, con la finalidad de llegar a más personas. En días especiales se harán dinámicas como puntos dobles, o envío gratis para los miembros del plan de fidelización, promociones que siempre serán anunciadas mediante contenido orgánico por las redes sociales.

- Punto de venta: los materiales visuales diseñados son importantes, específicamente para reforzar la identidad de la marca. Esto incluye carteles visibles en mostradores y displays estratégicamente ubicados en áreas de espera, todos ellos con códigos QR que dirigen directamente al chatbot de WhatsApp para facilitar el registro inmediato. El personal de ventas recibirá capacitación para explicar claramente los beneficios del programa, integrando la experiencia presencial y digital, de esa manera es posible adaptarse a los distintos comportamientos de compra de los clientes.
- WhatsApp: es un canal clave tanto para la inscripción inicial como para mantener un engagement continuo con los clientes registrados. Este canal permite enviar mensajes personalizados con actualizaciones relevantes del saldo de puntos, ofertas exclusivas y saludos personalizados en ocasiones especiales como cumpleaños, siempre con un tono cercano y amigable, tal y como la personalidad de la marca. Las comunicaciones incluirán ejemplos claros que motiven a la compra. La frecuencia y estilo de estos mensajes estarán cuidadosamente planificados para evitar saturar al usuario, manteniendo siempre una comunicación relevante y fortaleciendo continuamente la relación entre el cliente y la marca.

## VIII. Recolección de datos simulados

### A. Simulación

De la simulación se obtuvieron tres cuadros:

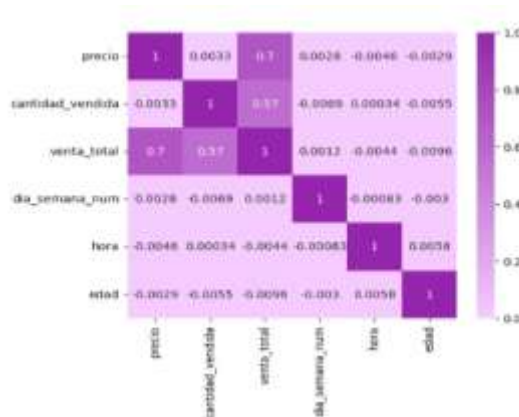
- Base de clientes: este cuadro contiene información demográfica y de contacto de los clientes que participan en el programa de fidelización.
- Base de transacciones: registra cada compra realizada por los clientes en el canal del bot de WhatsApp, reflejando aspectos como fecha, monto, producto, entre otros datos contextuales.
- Base de redenciones: Este cuadro refleja cuándo y dónde se hicieron los canjes, así como qué producto fue redimido. Representa la cantidad de veces que un cliente utilizó sus puntos acumulados para realizar canjes.

A partir de estas tres cuadros simuladas se extrajo la información necesaria para llevar a cabo el análisis exploratorio de datos. Cada una de ellas aportó variables clave que fueron utilizadas para comenzar a entender el comportamiento general de los clientes, las compras realizadas y el uso del sistema de canje. Esta información sirvió como base para iniciar con el procesamiento y visualización de los datos en la siguiente etapa del estudio.

### B. Análisis exploratorio

1. La matriz de correlación muestra la relación lineal que hay entre variables. En este gráfico se observa que la variable de precio y venta tienen una fuerte relación positiva, con un 0.71, indicando que entre mayor sea el precio, mayor será la venta. Asimismo, se observa una relación positiva entre la cantidad vendida y la venta total con un 0.57, lo que indica que las unidades vendidas es un factor relevante que influye en las ventas totales. Se observa también una relación negativa, aunque casi nula entre el precio del producto y la cantidad vendida, esto quiere decir que el precio del producto no es un factor determinante en el volumen venta del mismo. En la matriz también se pueden ver variables temporales y demográficas, que no presentan una relación significativa con las variables de comportamiento de compra.

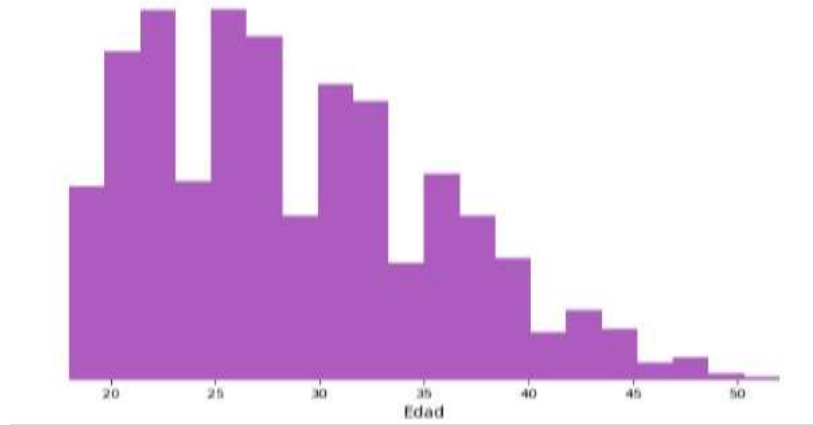
Figura 7. Matriz de correlación de variables numéricas



Fuente: Elaboración propia.

2. En este histograma se observa la distribución de las edades de los clientes. Como resultado se obtuvo un gráfico con cola hacia la derecha, es decir que la mayoría de los datos se encuentran concentrados en valores bajos, en este caso, la mayoría de los clientes se encuentra en rangos de edad entre los 20 y los 30 años. Sin embargo, se observa una distribución de edades de un mínimo aproximado de 18 años y un máximo de 55.

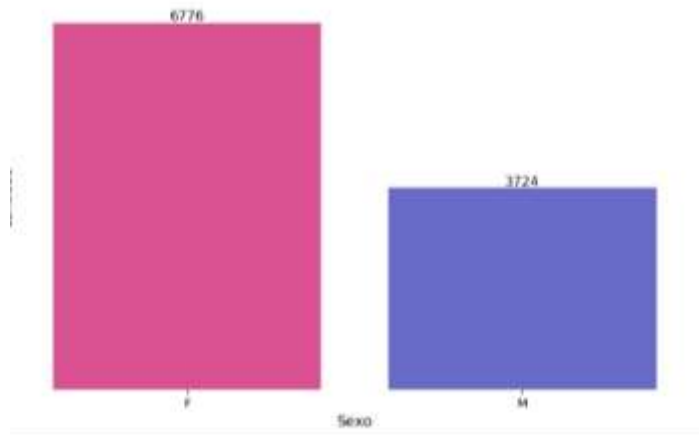
**Figura 8. Distribución de las edades de los usuarios**



Fuente: Elaboración propia

3. El siguiente gráfico de barras se muestra la cantidad de clientes por sexo, siendo la "F" de "Femenino" y "M" de "Masculino". La cantidad de clientes que es del sexo femenino, con 6776 clientes, resulta ser casi el doble que la de sexo masculino, con 3724. Indicando que los clientes de la marca en promedio son más los que tienen sexo femenino.

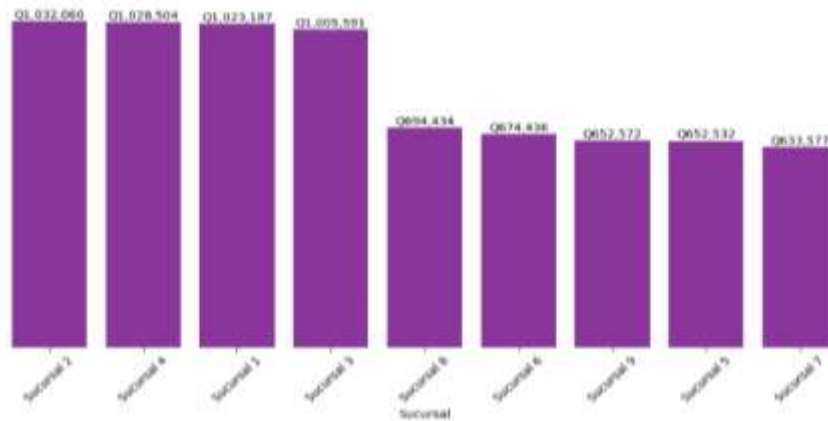
**Figura 9. Distribución de clientes por sexo**



Fuente: Elaboración propia.

4. El siguiente gráfico es un gráfico de barras, donde se observan las ventas totales por sucursal. El gráfico muestra que existe 4 tiendas con mayores ingresos monetarios, “Sucursal 2”, “Sucursal 4”, “Sucursal 1” y “Sucursal 3”, las cuatro sucursales superaron el millón de quetzales. La sucursal con más ingresos es “Sucursal 2” con Q1,032,060, mientras que la sucursal con menos ingresos fue “Sucursal 6” con Q633,577.

**Figura 10. Ventas totales por sucursal**



Fuente: Elaboración propia

5. Este gráfico de líneas muestra las ventas mensuales en el transcurso de un año. Se observa una tendencia positiva durante el periodo analizado. El mes con menos ventas fue septiembre del 2025, mientras que el mes con mejores ingresos fue marzo del 2026, con ventas que superan los Q7,000,000.

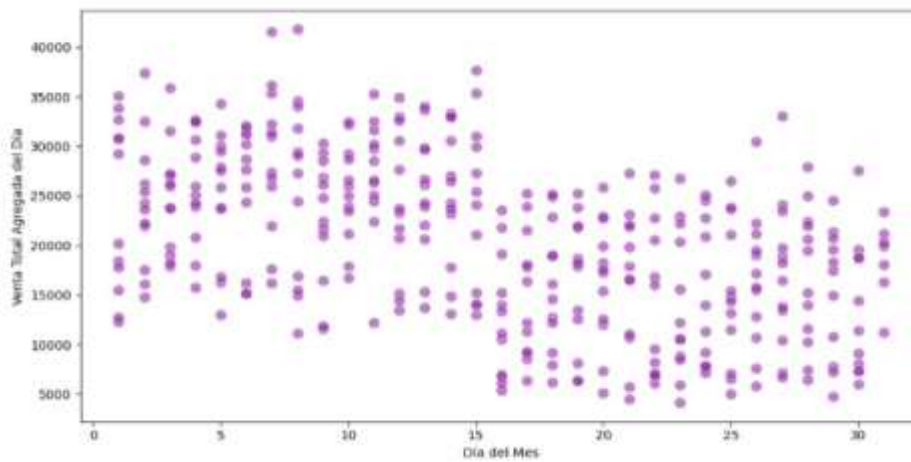
**Figura 11. Ventas mensuales**



Fuente: Elaboración propia

6. EL gráfico muestra la relación que hay entre la variable del día del mes vs las ventas realizadas. La relación de ambas variables es una relación negativa no lineal, lo que quiere decir que a medida que el día del mes aumenta, las ventas disminuyen. Esta inflexión se muestra a partir de la segunda quincena del mes, indicando que en promedio, hay un decrecimiento de las ventas las últimas dos semanas del mes. En el gráfico también puede verse una gran dispersión de los datos, indicando variabilidad significativa entre las ventas de cada día.

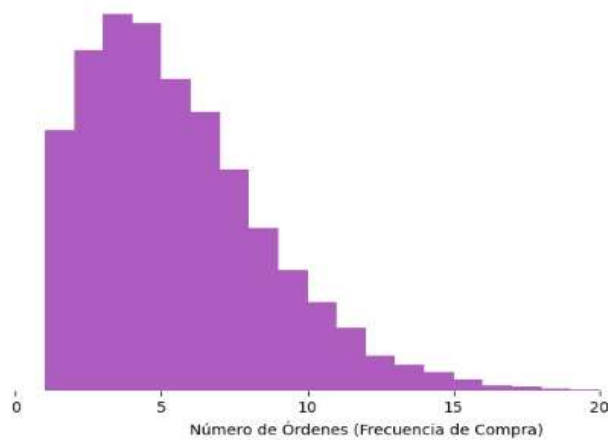
**Figura 12. Relación entre el día del mes y la venta del día**



Fuente: Elaboración propia

7. El siguiente gráfico es un histograma que muestra la distribución de la frecuencia de compra de los clientes. Se muestra una cola hacia la derecha, el histograma tiene un sesgo positivo, indicando que la mayoría de los datos se encuentra en valores bajos, lo que quiere decir que en promedio los clientes tienen una frecuencia de compra en un rango de entre 1 a 5 veces en el periodo de tiempo analizado. Sin embargo, se observa que algunos de los clientes han llegado a comprar hasta un máximo aproximado de 20 veces dentro de ese periodo.

**Figura 13. Distribución de la frecuencia de órdenes por cliente**



Fuente: Elaboración propia

8. En este resumen estadístico se describe la cantidad de veces que un cliente ha realizado una compra. En total hay 9817 clientes que han realizado al menos una compra. En promedio cada cliente ha realizado un aproximado de 5.09 compras, redondeado a 5 compras por cliente. La mediana es de 5, indicando que al menos la mitad de los clientes ha realizado 5 compras en este periodo de tiempo. En los percentiles se ve que el 20 % de los clientes realizó como máximo 2 compras, el 40 % compró hasta 4 veces y el 80 % no superó las 8 compras. Esto indica que la mayoría de los clientes presenta una frecuencia baja o moderada, mientras que solo un pequeño grupo, que es el 20 % restante, muestra un comportamiento de compra más frecuente.

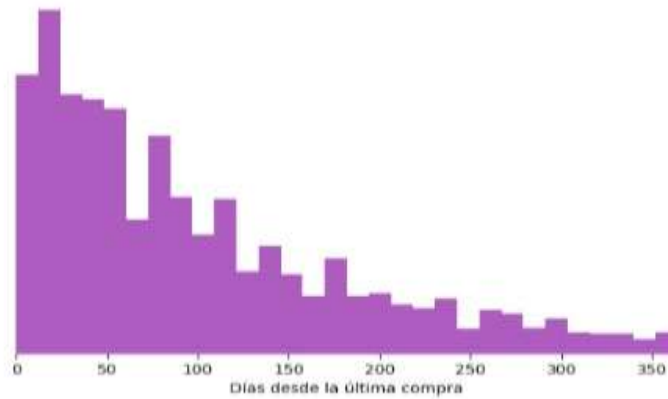
**Figura 14. Resumen estadístico de la frecuencia de compra**

```
=== Resumen de la Frecuencia de Compra ===
count      9817.000000
mean       5.093206
std        3.128785
min        1.000000
20%        2.000000
40%        4.000000
50%        5.000000
60%        5.000000
80%        8.000000
max        23.000000
Name: frecuencia_compra, dtype: float64
```

Fuente: Elaboración propia

9. En la distribución de la recencia se vuelve a observar un sesgo positivo, que indica una concentración de datos en los valores más bajos, en este caso, esto significa que hay una alta concentración de compras en un rango entre los 0 hasta los 75 días aproximadamente. Sin embargo, también se observan clientes que su última compra fue hace más de 300 días.

**Figura 15. Distribución de la recencia de compra**



Fuente: Elaboración propia

10. En este resumen estadístico se describe la cantidad de días transcurridos desde la última compra de cada cliente. En promedio, han pasado aproximadamente 100 días desde la última compra, aunque este valor varía ampliamente, con una desviación estándar de 85.79 días. La mediana es de 75 días, lo que indica que al menos la mitad de los clientes realizó su última compra hace 75 días o menos. Los percentiles indican que, el 20 % de los clientes compró por última vez hace 25 días o menos, el 40 % dentro de los últimos 53 días y el 80 % hace 171 días o menos.

**Figura 16. Resumen estadístico de la recencia de compra**

```

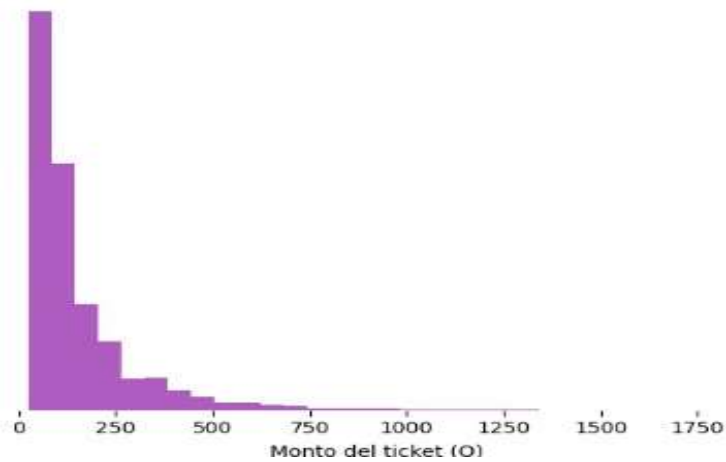
=== Resumen de la Recencia (días) ===
count      9817.000000
mean       99.851890
std        85.794885
min         0.000000
20%        25.000000
40%        53.000000
50%        75.000000
60%        96.000000
80%       171.000000
max       364.000000
Name: recencia dias, dtype: float64

```

Fuente: Elaboración propia

11. En este histograma se muestra la distribución del monto por factura, en el que se observa una cola hacia la derecha, en este caso se ve una mayor concentración en valores bajos, es decir que el *ticket* promedio se encuentra en un rango aproximado entre Q50 a Q100. Sin embargo, se observa que a lo largo del periodo analizado hubo compras con montos de hasta un aproximado de Q1,750.

Figura 17. Distribución del *ticket* por orden de compra



Fuente: Elaboración propia

12. En este resumen estadístico se describe el monto total de cada factura registrada en el periodo analizado. Se cuenta con un total de 50,000 transacciones. El monto promedio por factura es de aproximadamente Q147.93, pero con una alta variabilidad, ya que tiene una desviación estándar de Q156.60. La mediana es de Q95.96, lo que indica que al menos la mitad de las facturas fueron por montos iguales o inferiores a ese valor. En cuanto a los percentiles, el 20 % de las facturas tiene un valor igual o menor a Q43.96, el 40 % es menor o igual a Q80.39, y el 80 % no supera los Q213.

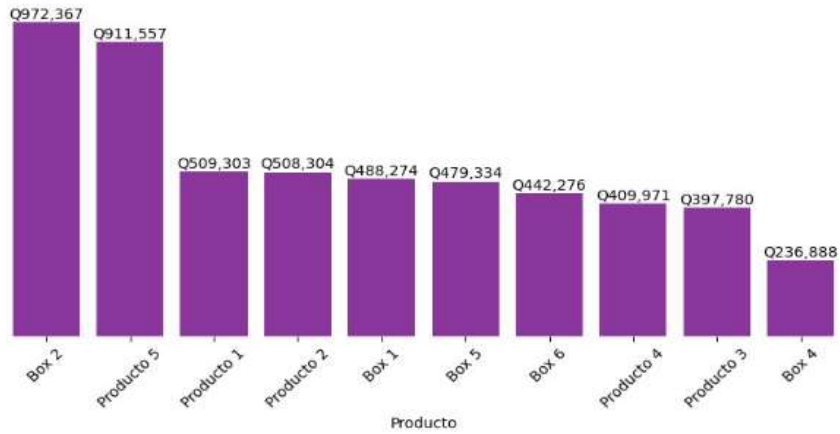
Figura 18: Resumen estadístico del monto por factura

```
Resumen estadístico del monto por factura
count      50000.000000
mean       147.937976
std        156.607430
min         23.800000
20%        43.960000
40%        80.396000
50%        95.960000
60%       113.670000
80%       213.010000
max       1822.760000
Name: venta_total, dtype: float64
```

Fuente: Elaboración propia

13. Este gráfico de barras muestra el *marketing 10* de productos que generan un mayor ingreso a la empresa. Se observa que 2 de los productos sobresalen son mayores ingresos económicos, estos productos son “Box 2” y “Producto 5”, con ingresos de Q972,367 y Q911,557 respectivamente, siendo los principales generadores de ventas. A estos les siguen “Producto 1”, “Producto 2” y “Box 1”, cuyos ingresos se encuentran por encima de los Q450,000. Los productos del *marketing 10* que menos ingresos generan son “Producto 4” con ventas de Q397,780 y por último la “Box 3” con ventas de Q236,888.

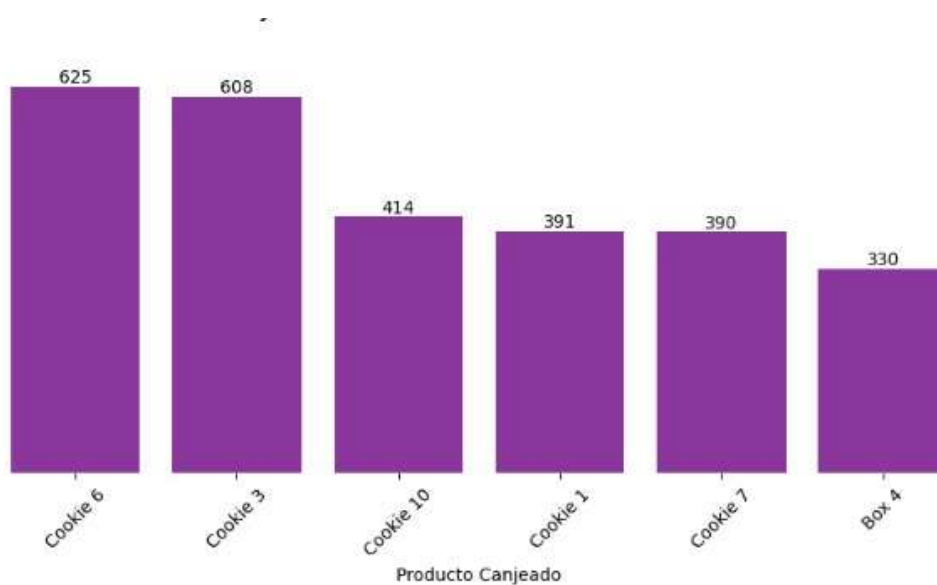
**Figura 19 Marketing 10 productos con más ventas**



Fuente: Elaboración propia

14. En este gráfico de barras se presentan los productos con mayor cantidad de redenciones. Del listado de productos puestos para que los clientes pudieran hacer el canje de sus puntos el producto más canjeado fue la “Cookie 6” con 625 redenciones, seguida de la “Cookie 3” con 608 redenciones, luego “Cookie 10” con 414, “Cookie 1” con 381 y por último “Box 4” con 330 redenciones.

**Figura 20. Productos más canjeados**



Fuente: Elaboración propia.

Este análisis exploratorio fue de utilidad para poder comprender el perfil de los clientes, su comportamiento de compra y la dinámica de ventas. Se evidenció una fuerte relación entre el precio y las ventas monetarias, aunque el precio no influye significativamente en el volumen de unidades vendidas. Asimismo, se encontró que el público de la marca se concentra principalmente en mujeres jóvenes de entre 20 y 30 años. Las sucursales presentan diferencias notables en desempeño, destacando cuatro puntos de venta con ingresos superiores al millón de quetzales. Además, las ventas mensuales muestran una tendencia positiva a lo largo del tiempo, aunque se observó una caída en las ventas en la segunda mitad de cada mes. La mayoría de los clientes ha realizado entre 1 y 5 compras en el periodo analizado, con una recencia promedio de 100 días, y con un *ticket* promedio que se concentra entre Q50 y Q100, aunque se observa que existen facturas con valores significativamente más altos.

## IX. Tablero para la visualización de datos

Como parte del análisis de resultados, se diseñó un *dashboard* o tablero interactivo que tenía como finalidad el poder facilitar la interpretación de los datos obtenidos del plan de lealtad, en este caso, de los datos que se obtuvieron de la simulación, y con esto se busca que la empresa pueda tomar decisiones estratégicas basadas en datos. Esta herramienta permite visualizar de forma clara y dinámica la información más relevante del plan de fidelización, ayudando a identificar patrones y evaluar el desempeño del programa, además de que permite comprender mejor el comportamiento de los clientes, quienes son y con esto generar segmentos que ayuden a crear estrategias y campañas mejor enfocadas.

Este *dashboard* fue construido en Power BI, basado en los cuadros simulados y estructurado en distintas páginas o vistas, cada página tiene un enfoque distinto y busca responder preguntas clave del proyecto. El proceso de creación incluyó la preparación previa de los datos, modelado, carga en la herramienta de visualización y diseño de filtros, formulación de medidas y columnas necesarias para las distintas visualizaciones que se deseaban obtener.

### A. Resultados del programa de Lealtad

Visualizar el impacto general del plan de fidelización en términos de comparación de ventas reales vs la meta, redenciones de los productos canjeables, y retorno sobre la inversión (ROI), con la finalidad de evaluar si la estrategia planteada fue efectiva.

Se plantearon las preguntas que se buscaban responder con las visualizaciones dentro de esta página del tablero, las preguntas planteadas:

- ¿Cuál fue la venta total durante el periodo analizado?
- ¿Se cumplió la meta previamente establecida?
- ¿Cuál fue el retorno sobre la inversión?
- ¿Los clientes con redenciones en promedio gastan más que los que no han hecho canje?
- ¿Todos los usuarios registrados han realizado compras?
- ¿Qué porcentaje de los clientes se retuvieron durante el periodo analizado?
- ¿Cuáles fueron los productos con más redenciones?
- ¿En qué sucursales se hizo la mayor cantidad de redenciones?

La vista y los gráficos realizados en esta página del *dashboard* son los siguientes:

Figura 21. Dashboard de resultados del programa de lealtad



Fuente: Elaboración propia.

KPI's:

- Venta total: se registró un total Q7.4 millones de ventas en un periodo de 1 año.
- Meta de venta: la meta de venta para el primer año de del programa de lealtad es de Q7.32 millones.
- Avance: el avance de la venta real sobre la meta de ventas es de 101.05 %, indicando que, para el final del periodo analizado, las ventas reales habían superado la meta propuesta.
- Clientes con redención: el total de clientes que han realizado al menos 1 canje

de sus puntos por un producto es de 2312 clientes.

- Tasa de retención: el porcentaje de clientes que se retuvo en el periodo analizado es de 25.09 %, es decir que el

Visualizaciones:

- A. Retorno sobre la inversión: la tasa de ROI mensual durante el primer año se mantuvo siempre positiva y significativamente alta, indicando que los ingresos obtenidos son mucho mayores que los gastos invertidos. El primer mes de lanzamiento se observa el ROI más bajo, sin embargo, este sigue siendo de un 627.41 %, lo que puede deberse a los gastos invertidos en campañas de lanzamiento. Julio es el mes con el ROI más alto dentro de este periodo el ROI fue de 1844.37 %, puede deberse a una baja inversión en mercadeo y altos ingresos. Los siguientes meses el ROI se mantiene por encima del 1000 %. En diciembre hubo una ligera caída en la tasa de ROI llegando a 869.94 % puede deberse a la temporada. Los últimos meses analizados tienen una tendencia a la baja, pero siempre con valores positivos.
- B. Ventas vs Meta en el tiempo: en el gráfico de líneas se observa a detalle la comparación mensual entre las ventas reales y la meta propuesta. La venta real estuvo similar o por encima en casi cada mes, exceptuando el mes de septiembre, donde no se llegó a cumplir la meta de ventas. En el gráfico se observa un comportamiento positivo, y en constante crecimiento, a pesar de tener meses con algunas bajas, las ventas se han mantenido por encima de los Q500,000, llegando a su punto máximo en marzo del 2026 con, ventas aproximadas de Q700,000.
- C. Venta vs Meta por sucursal: en este gráfico de barras se muestra una comparativa entre las ventas reales acumuladas por sucursal vs la meta de venta. En total hay 9 sucursales, de las cuales 6 estuvieron por encima de su meta, estas sucursales fueron: “Sucursal 2”, “Sucursal 4”, “Sucursal 1”, “Sucursal 6”, “Sucursal 5”, “Sucursal 7”. Por otro lado, las sucursales que no cumplieron con su meta son: “Sucursal 3” donde el faltante fue de un aproximado de Q8,000, “Sucursal 8” donde faltó un aproximado de Q24,000, por último, la “Sucursal 9”, donde hizo falta Q6,000 aproximadamente.
- D. Usuarios registrados vs usuarios con compra: este gráfico de barras muestra la comparación entre los usuarios que llenaron el formulario de registro y los usuarios que si efectuaron al menos una compra. En este caso se observa que la cantidad de personas registradas es de 10,500, mientras que la cantidad de personas que realizaron mínimo una compra es de 9,817, es decir que el 93.49% de los clientes registrados hicieron al menos una compra.
- E. *Ticket* promedio: este gráfico pretende mostrar la comparación del *ticket* promedio dependiendo si son clientes que han hecho redenciones o no. En este se observa que el *ticket* promedio de las personas que si han llegado a canjear sus puntos es mayor que el de las personas sin redenciones, siendo un *ticket* promedio de Q187 vs uno de Q143. Esto indica que en promedio las personas que ya han realizado redenciones gastan más que las personas que aún no han hecho canje.
- F. Sucursales con más redenciones: el gráfico de barras muestra la cantidad de redenciones que se han realizado por cada sucursal. En el gráfico se muestran las 2 sucursales con mayor cantidad de redenciones, la “Sucursal 1” con una cantidad de 415 redenciones y “Sucursal 2” con 399 redenciones.
- G. Producto más redimido: se muestra la cantidad de redenciones que se han hecho por producto. En el gráfico se observa que los productos con una mayor cantidad de redenciones fueron: “Cookie 6” con 625 redenciones y “Cookie 3” con 608 redenciones.

Luego de haber analizado la primera página de resultados del tablero, se puede decir

que, el impacto del primer año del lanzamiento del programa de lealtad tuvo un impacto positivo. Las ventas totales superaron ligeramente la meta anual, lo que refleja una planificación adecuada y una buena respuesta del mercado. Las ventas totales superaron ligeramente la meta anual, lo que refleja una planificación adecuada y una buena respuesta del mercado. Además de que el ROI mensual consistentemente alto sugiere que la inversión en el programa fue altamente rentable, incluso en la fase de lanzamiento. La conversión de clientes registrados a los que realmente compraron también es elevada siendo de un 93.49 %. En general se observa que la estrategia del programa de lealtad generó buenos resultados en términos financieros para la empresa, además de cumplir con el objetivo de recolección de datos.

## B. Perfil de clientes

Entender las características demográficas y de comportamiento de los clientes que participaron en el plan de lealtad, con el fin de construir un perfil del consumidor de la marca. Permitted to identify consumption patterns, segment clients in a more precise way and optimize loyalty strategies.

Para la generación de esta página del tablero nuevamente se realizó el planteamiento de las preguntas que se deseaban responder con las visualizaciones a realizar. Las preguntas planteadas en esta página iban enfocadas en entender el perfil del cliente y el comportamiento de estos. Estas preguntas son las siguientes:

- ¿Cuál es la frecuencia promedio de compra? ¿Hay una diferencia significativa entre hombres y mujeres?
- ¿Cuál es la edad promedio de los clientes
- ¿Cuál es la distribución de las edades de los clientes?
- ¿Hay una diferencia significativa entre hombres y mujeres?
- ¿Qué edades compran más en cada sucursal?
- ¿Cuál es la proporción de canales más usados por los clientes?
- ¿Qué método de entrega es más frecuente?
- ¿Qué días u horarios compran más?
- ¿Cuál es la venta por producto?
- ¿Qué dicen los clientes de la empresa?

La vista de la página de “Perfil de clientes” es la siguiente:



porcentaje de participación de 64.53 %, mientras que la participación del género “Masculino” es de 35.47 %.

- B. Distribución de edades: en este gráfico se muestra una distribución con cola hacia la derecha, es decir que tiene un sesgo positivo, indicando que la mayoría de los valores se concentra en valores bajos. En este caso se puede ver que las edades de los clientes se concentran en un rango de entre 18 a 30 años, sin embargo, se tiene registro de clientes con edades que alcanzan los 50 años, pero en una menor proporción.
- C. Participación por sucursal por rango de edad: en este gráfico de participación por sucursal (Figura 20) muestra la participación en volumen de órdenes de compra, por sucursal y categorizado por rangos de clientes. La sucursal que tiene un mayor volumen de compra es “Sucursal 1”, seguida de la “Sucursal 2”, acumulando un total de 4,769 y 4,751 órdenes respectivamente. Mientras que las sucursales con una menor cantidad de números de orden son la “Sucursal 7” con 2,874 órdenes. Este gráfico se categoriza también por rangos de edad, se muestra que en todas las sucursales el rango de edad que más predomina es el de “24-29”, seguido del rango “18-23”. Un claro ejemplo es la “Sucursal 1” en la cual, la cantidad de órdenes generadas por el rango de “24-29” es de 1,496, mientras que el rango de “18-23” fueron 1,304 órdenes.

**Figura 23. Participación por sucursal por edad**



Fuente: Elaboración propia.

- D. Proporción por canal: en el gráfico de proporción por canal, además de entender el medio por el cual se enteraron del programa de lealtad y cuál es el canal en el que mejor está funcionando la publicidad, se buscó comprender el uso y preferencia de las aplicaciones o redes de los consumidores de la marca. En este caso, se obtuvo que la mayoría de los clientes, con un 39.41 %, se enteró del programa de lealtad a través de “Instagram”, seguido de “Recomendaciones”, siendo de un 25.47 %, luego se encuentra “Facebook” con una participación del 15.44 %, “TikTok” con 14.72 % y de último se tiene a “Whatsapp” con un 4.96 %
- E. Método de entrega: el gráfico presenta las preferencias de los clientes entre el envío a domicilio, pagando un extra por envío, vs las personas que realizan sus pedidos para recoger en la sucursal. En el gráfico se obtuvo que la mayoría de las órdenes realizadas fueron de envío a domicilio, con un total de 22,174 órdenes. Por otro lado, el total de números de orden con método “Pickup” es de 11,877, quiere decir que en su mayoría los clientes

prefieren pagar un cargo adicional por el envío para que el producto llegue hacia ellos.

- F. Horarios más frecuentes: este gráfico (Figura 21) es un *Heatmap*, en el que se refleja el promedio de compras realizadas en un día de la semana en un horario específico. En el gráfico se observa una tendencia de compras en horas entre las 14:00 y 18:00 horas, con una mayor probabilidad de compra en día de jueves a domingo. El día y la hora con el mayor promedio de compras realizadas es el día “Domingo” a las “17:00” con compras promedio de 29.98. Por otro lado, se puede decir que los horarios con menos compras son de 9:00 a 11:00, siendo el día “Lunes” a las “8:00” el horario con menos compras, con un promedio de 1.97.

**Figura 24. Horarios de venta más frecuentes**



Fuente: Elaboración propia.

- G. Ventas por categoría y producto: en este esquema jerárquico se observan las ventas totales que ha tenido cada categoría, luego cada producto perteneciente a la categoría. En este caso se observa que la categoría con mayor volumen de ventas es “Cookies” con cantidades vendidas de 55,866, el producto con mayor rotación de esta categoría es la “Cookies 12” con 4,850 unidades vendidas. La segunda categoría que genera más ingresos es “Boxes” con una cantidad vendida de 27,456, dentro de la categoría, el producto con mejores ventas es “Box 5” con 4,806 unidades vendidas. Por último, está la categoría de “Varios”, con un volumen de venta de 23,296, y su producto con más ventas es “Producto 4” con una cantidad de 4,709.
- H. Lo que piensan los clientes: se realizó un *wordcloud* para poder analizar las palabras más frecuentes entre los comentarios que dejan los usuarios. En este caso las palabras con más menciones son “brownies” se puede inferir que ha sido un producto popular, o que ha recibido diversos problemas de presentación y calidad por lo que está siendo mencionado recurrentemente, Acá se observan palabras como “frías”, “secos”, entre otros, que pueden ayudar a la empresa las oportunidades que tienen en temas de calidad. Asimismo, se observan comentarios como “deliciosos”, “ricos”, “suaves”, que indican que el producto está siendo, en su mayoría, aceptado por los consumidores.

En el análisis realizado de los datos demográficos y de comportamiento de los clientes, se

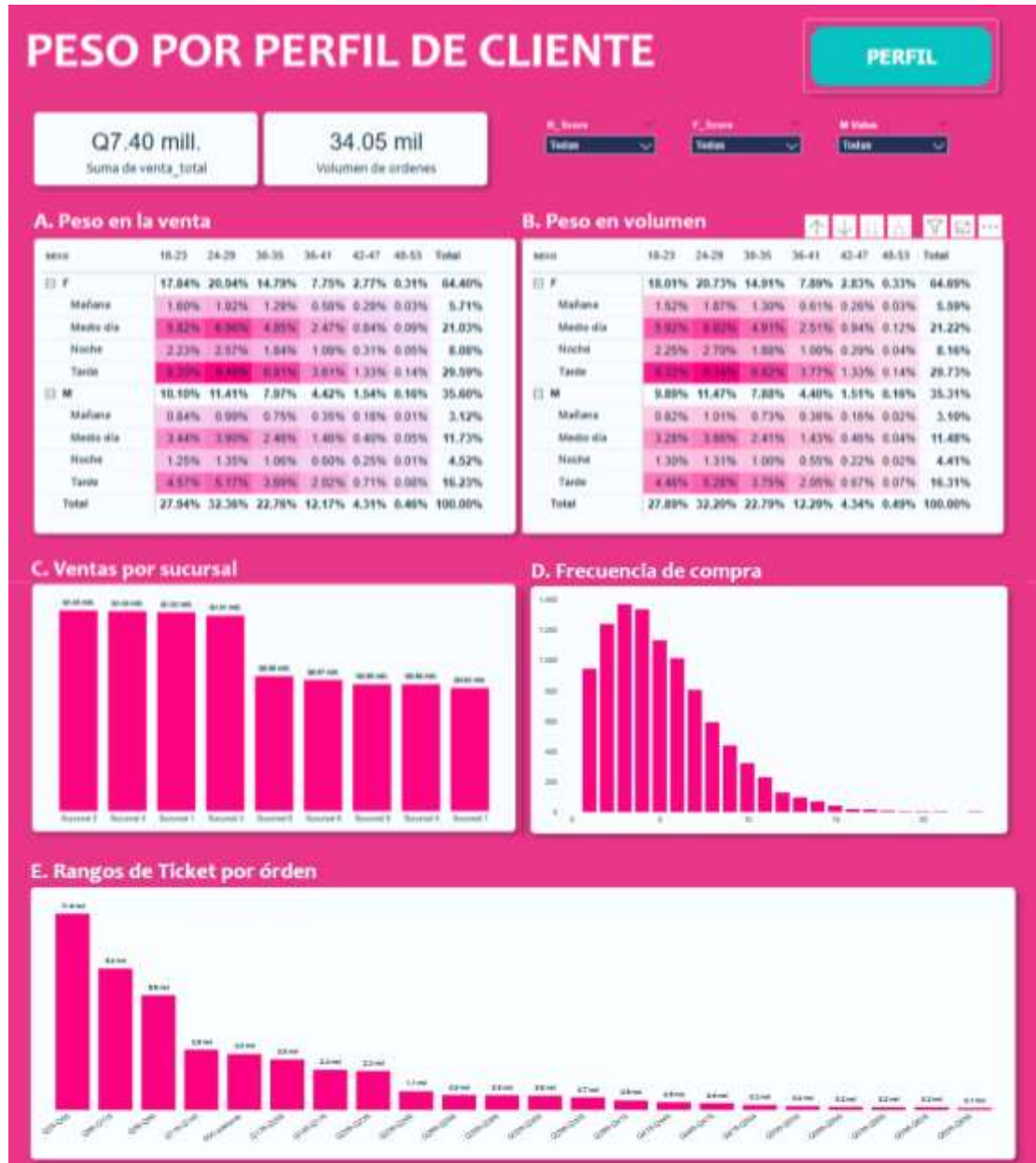
obtuvo que el perfil del cliente de la marca es joven, mayoritariamente femenino, con una edad promedio de 28 años, y con una fuerte preferencia por realizar compras en la “Sucursal 1”. Estos clientes presentan un *ticket* promedio de Q147 y una frecuencia de compra aproximada de 5 veces al año. En cuanto su preferencia en redes es un cliente que usa Instagram y que es influenciado por las recomendaciones que recibe. Además, es un cliente con preferencia a ordenar por servicio a domicilio, indicándola importancia de contar con una buena logística. Asimismo, el cliente tiene una tendencia de realizar sus compras en fines de semana en horarios de la tarde.

## C. Peso del segmento en ventas

Evaluar el peso que tienen los diferentes segmentos de clientes, construidos a partir de las visualizaciones, en el total de ventas y volumen de órdenes, para entender en que segmentos es importante invertir esfuerzos de mercadeo.

Esta página se usó como respaldo para la página de perfil, para además de entender el promedio de clientes que consumen la marca, entender el peso que representan en las ventas y en el volumen de las facturas. Para eso se realizaron matrices es segmentadas por género y franja horaria, con esto se podría identificar los grupos con mayor impacto económico y en volumen de órdenes. Asimismo, se hicieron gráficos complementarios sobre ventas por sucursal, frecuencia de compra y *ticket* promedio, con la finalidad de ayudar a definir los segmentos clave y patrones de comportamiento. En conjunto, el análisis permite visualizar de forma clara qué segmentos aportan más valor al negocio.

Figura 25. Dashboard peso de venta y volumen de órdenes



Fuente: Elaboración propia.

**KPI's:**

- Venta total: la venta total es de aproximadamente Q7.4 millones.
- Volumen de órdenes total: el total de transacciones realizadas en el primer año es de un aproximado de 34.05 mil órdenes.

**Visualizaciones:**

- A. Peso de la venta: en esta matriz se muestran el peso de las ventas totales que representa cada segmento, es decir, las ventas según su rango de edad, sexo y franja horaria. En la matriz se observa que el sexo “Femenino” contribuye con más del 64% de las ventas mientras que el “Masculino” tiene un peso de más

del 35%. En la franja horario “Tarde” tiene un peso en las ventas de 45.82 %. Asimismo, se obtiene que el rango de edad que representa un mayor porcentaje de las ventas es “24-29” con un peso de 32.35 %. La combinación de estas tres variables tiene un peso de 9.4 9 %

- B. Peso en volumen: la matriz demuestra el peso en la venta, de igual manera que la matriz A, según sexo, franja horaria y rangos de edad. De igual manera se obtuvo que el sexo que representa un mayor peso es “Femenino” con más del 64%, en la “Tarde” con un peso de 46.04 % y el rango de edad con más peso es “24-29” con 32.2%. La combinación de las tres variables es el segmento con mayor representatividad siendo de 9.34 %
- C. Ventas por sucursal: el gráfico muestra que la sucursal con la mayor cantidad de ventas es “Sucursal 2” con ventas de Q1,032,060, seguido de “Sucursal 4” con ventas de Q1,028,504. La sucursal que poseen un menor ingreso monetario es “Sucursal 7” con ingresos de Q622,577.
- D. Frecuencia de compra: el gráfico representa la frecuencia de compras por cliente, es decir que cuantos clientes han comprado cierta cantidad de veces. La mayoría de los clientes (1371) han comprado 3 veces a lo largo del periodo analizado. En promedio la mayoría de los clientes ha realizado entre 3 y 5 compras. A partir de las 6 compras se empieza a observar una distribución sesgada hacia la derecha, indicando valores mucho más bajos. El máximo de compras fue de 23 compras.
- E. Rangos de *Ticket* por orden: el gráfico de barras muestra la cantidad de órdenes que hay dentro de los distintos rangos de *ticket* promedio por orden. La mayoría de las órdenes (11,351) se encuentran en un rango de entre Q25 y Q55 gastados por orden, seguido del rango Q86 a Q115 con 8,178 órdenes en este rango. En tercer lugar, se tiene el rango de Q56 y Q85 con 6,633 órdenes.

## D. Segmentación RFM

Visualizar los segmentos resultantes del modelo RFM (Recencia, Frecuencia, Monetario), con el fin de entender las diferencias entre grupos de clientes y facilitar acciones personalizadas según su valor y comportamiento.

Para el realizar el modelo RFM fue necesario entender cuáles son los segmentos que se querían entender y con que combinación de RFM score se clasificaría cada segmento. Se eligieron 7 segmentos, estos son:

### 1. VIP

Son clientes con alta recencia, alta frecuencia y alto valor monetario. Este segmento agrupa a los clientes que tienen un alto desempeño en las tres métricas del modelo RFM, con puntajes de 4 o 5 en recencia, frecuencia y valor monetario. Estas combinaciones, como 555 (máximo en todas las métricas) o 544 (recencia y valor altos con frecuencia ligeramente menor), representan una compra continua, compras regulares y un gasto significativo. Se optó por tomar este segmento con las calificaciones más altas ya que representa a los clientes más rentables y leales. Este es un segmento con un alto valor para la empresa, por lo que sería importante la implementación de estrategias de retención.

### 2. Leales

Son clientes con alta frecuencia y buen valor monetario, aunque no necesariamente con compras recientes. Este segmento incluye a clientes con puntuaciones variables (3-5), las combinaciones para este segmento varían entre estos valores ya que, por ejemplo, una puntuación de 543 (recencia, frecuencia y valor altos) indican una relación reciente con la marca, con compras regulares y gastos relevantes.

Este segmento fue definido para poder identificar a los clientes que poseen una fidelidad histórica y una contribución económica que los convierten en activos valiosos para la empresa. La elección de separar este segmento del “VIP” se debe a que, los clientes leales tienen un buen historial, sin embargo, tienen al menos una métrica promedio, por lo que no es tan valioso como un VIP.

### 3. Prometedores

Este segmento posee clientes con buena recencia, baja frecuencia, y buen valor monetario. Este segmento incluye clientes con recencia media-alta (3-5), frecuencia baja (1-3) y valor monetario medio-alto (3-5), un ejemplo de esto es la combinación 315 (recencia alta, frecuencia baja, valor alto). El segmento fue definido para poder identificar a clientes que han realizado compras recientes y significativas, lo que sugiere un alto potencial de conversión hacia el segmento de “Leales”. Estas combinaciones reflejan que los clientes tienen un interés inicial fuerte, por ello se clasifica como un cliente que tiene potencial de convertirse en un cliente leal.

### 4. No podemos perder

En este segmento los clientes tienen un alto valor y frecuencia histórica, pero con baja recencia. Este segmento es crucial ya que agrupa a clientes con recencia baja (1-2), pero alta frecuencia (3-5) y valor monetario elevado (3-5), como las combinaciones 145 (recencia baja, frecuencia y valor altos), indicando que eran clientes históricamente valiosos, que compraban varias veces, con *tickets* promedio altos, pero llevan mucho tiempo sin volver a realizar una compra. Para este tipo de segmento es importante priorizar esfuerzos intensivos de retención para recuperar su compromiso con la marca.

### 5. Dormidos

Estos son clientes con un valor monetario medio y frecuencia media o baja, que no han comprado en mucho tiempo. Para este segmento se usaron puntuaciones en las que la recencia es baja (1-2), frecuencia es baja a media (1-3) y valor monetario medio (1-3). Son clientes que han interactuado con la marca en el pasado, pero cuya inactividad reciente y baja regularidad sugieren una desconexión con la marca. Este segmento tiene un potencial de activación ya que tienen un valor monetario medio.

### 6. Ocasionales

Estos clientes se clasifican así debido que tienen una baja frecuencia y valor monetario, pero con recencia media o alta. Las combinaciones de puntuaciones de las métricas de este segmento son de una recencia media-alta (4-5), frecuencia baja (1-3) y valor monetario bajo (1-3), como las combinaciones 411 (alta recencia, frecuencia y valor bajos). El segmento se definió de esta manera ya que estos pueden ser clientes nuevos, o clientes que compran muy pocas veces cada cierto tiempo.

### 7. Perdidos

En este segmento los clientes tienen una baja recencia, baja frecuencia y bajo valor monetario. El segmento tiene la combinación de los puntajes más bajos, principalmente todos aquellos que tienen de recencia (1), la frecuencia variable (1-5) al igual que el valor monetario. Las combinaciones reflejan una interacción mínima, con compras no recientes, esporádicas y de bajo impacto económico, indicando una probable desconexión total con la marca.

Para la generación de las visualizaciones del *dashboard* se plantearon las siguientes preguntas:

- ¿Cuál es el total de clientes?



#### KPI's:

- Total clientes: el total de los clientes registrados es de 10,500
- Recencia promedio: en promedio han pasado 100 días desde la última compra de los clientes, puede que este promedio se vea afectado por la alta variabilidad de los datos.
- Frecuencia promedio: en promedio los clientes han comprado al menos 5 veces durante el periodo analizado.
- Valor monetario promedio: en promedio, cada cliente ha gastado un monto Q753.48 a lo largo del primer año del programa de fidelidad.

#### Visualizaciones:

- A. Proporción de clientes por segmento: el gráfico es un *treemap* gráfico que muestra la proporción de clientes por cada segmento. En este caso se observa que el segmento con mayor participación es el segmento “VIP” con un 21.39%, indicando que la mayoría de los clientes de la marca son muy leales, con una continua compra y un *ticket* promedio alto. En segundo lugar se encuentra el segmento “No podemos perder” con una participación del 20.66%, seguido del segmento de “Prometedores” con 20.32%, luego está el segmento de “Leales” con 16.39%. Por último, se tienen los segmentos “Dormidos”, “Perdidos” y “Ocasionales” tienen un porcentaje de participación menor al 10%
- B. Conversión de segmentos por trimestre: en este gráfico de cintas se muestra la conversión trimestral de segmentos, con el objetivo de entender si hubo cambios en la cantidad de clientes por segmento en cada trimestre. Se muestra que segmentos como “prometedores” o “no podemos perder” terminan en segmentos como “perdidos” al no atender a estos clientes. Los segmentos “Leales” y “VIP” se mantienen estables, sin embargo, si se observan ligeros decrecimientos. En los demás segmentos se observa estabilidad, indicando que hay oportunidades de fidelizarlos con alguna estrategia.
- C. Ventas por segmento en el tiempo: en este gráfico de líneas se observa que el segmento “VIP” mantuvo ventas estables hasta enero del 2026, donde sus ventas incrementaron significativamente, llegando a ser de Q442,907, siendo el segmento con mayores ingresos monetarios. El segmento de “Leales” había mantenido ventas estables hasta diciembre del 2025, mes en el que sus ventas incrementaron, llegando a ser Q313,192, sin embargo, estas decrecieron hasta llegar a su punto más bajo con Q81,191. El segundo segmento con mayores ingresos fue del segmento “Prometedores”, el cual sus ventas se vieron incrementadas en diciembre del 2025, y en marzo del 2026 llegaron a ser de Q184,565. Se observa que el segmento “Perdidos” dejan de percibirse ingresos a partir de octubre del 2025, “No podemos perder” ya no hubo ventas en diciembre del 2025 y “Dormidos” a partir de febrero del 2026.
- D. *Marketing* 5 productos más vendidos: se observa la comparación entre los 5 productos con mayor rotación. En primer lugar, se encuentra el producto “Cookie 12” con un total de 4850 unidades vendidas. En segundo lugar está el producto “Cookie 11” con un total de 4,840 unidades. Se puede ver que los 5 productos más vendidos superan las 4,700 unidades vendidas.
- E. Proporción del segmento por sucursal: se quiere entender la participación de cada segmento en cada una de las sucursales. En este caso el segmento que predomina es “VIP” en todas las sucursales, superando el 34% de participación en las 9 sucursales. Se observa que la participación de todos los segmentos es similar en cada una de las sucursales.
- F. Gráfico de dispersión por métricas RFM: este gráfico presenta la relación entre las variables RFM, el gráfico permite filtrar tanto el eje y como el eje x con las distintas variables RFM. En la combinación Recency vs Frequency,

se evidencia que los clientes VIP y Leales se agrupan en las zonas con mayor frecuencia de compra y recencia alta, lo que confirma su compromiso activo con la marca. Por el contrario, los segmentos Dormidos y Perdidos se posicionan en áreas de baja frecuencia y baja recencia, reflejando inactividad prolongada. En la relación de Recency vs Monetary, los clientes VIP nuevamente se puede ver que han realizado compras recientes con alto valor monetario, mientras que segmentos como Prometedores muestran compras recientes, pero con menor gasto, lo cual representa una oportunidad de crecimiento. En cambio, clientes como No podemos perder poseen un alto valor monetario histórico, pero menor recencia, lo que los ubica en una zona de alerta por posible abandono. Finalmente, la combinación Frequency vs Monetary permite observar que los clientes más rentables son aquellos que compran con mayor frecuencia y además gastan más, mientras que segmentos como Ocasionales se mantienen en niveles bajos en ambas métricas.

- G. Cuadro de información de clientes: en este cuadro se encuentra la información de los clientes, en caso de querer descargar el cuadro y poder generar estrategias usando los datos de cada cliente.

## X. Metodología para la perfilación y segmentación

El *dashboard* elaborado debe ser correctamente interpretado para poder generar el perfil del cliente de la marca y los distintos segmentos que tienen. Para esto se elaboró una presentación en la cual se estructuró el proceso para la perfilación y segmentación de clientes utilizando la información contenida en el *dashboard*. Primero, se identificaron las visualizaciones clave dentro del tablero que permitían construir los perfiles de clientes, analizando variables demográficas y de comportamiento como edad, género, canal de captación, frecuencia de compra, *ticket* promedio y método de entrega. Con esto, se evaluó también el peso de cada perfil en las ventas totales y en la cantidad de transacciones realizadas, para determinar cuáles segmentos representaban una mayor contribución al negocio. Posteriormente, se explicó la generación de la segmentación basada en el modelo RFM (Recencia, Frecuencia y Valor Monetario), explicando las variables a tomar en cuenta para la creación del modelo, la asignación de puntajes y la clasificación de clientes en segmentos estratégicos.

El contenido de la presentación:

- a. Se realizó una introducción del propósito de la generación del manual.

**Figura 27. Propósito del manual**



Fuente: Elaboración propia.

- b. Se definieron las características de un “*Buyer persona*”

Figura 28. Buyer persona

## Buyer persona



Incluye:

- Datos demográficos
- Comportamiento de compra
- Objetivos y motivaciones
- Preferencias de comunicación

Fuente: Elaboración propia.

- c. Perfil: se explicó qué gráficos contenían la información demográfica necesaria para hacer el perfil del cliente.

Figura 29. Perfil



Fuente: Elaboración propia.

- d. Objetivos: los objetivos requieren un análisis cualitativo, sin embargo, con los datos obtenidos del *dashboard* se pueden realizar algunas inferencias. Se explica el gráfico que se puede utilizar (G. Ventas por categoría y producto) para hacer las deducciones sobre este paso de la creación del perfil.

Figura 30. Objetivos



Fuente: Elaboración propia.

- e. Motivaciones: al igual que los objetivos, es una variable que requiere un análisis cualitativo, pero, por medio de algunos de los gráficos del *dashboard* pueden sacarse conclusiones sobre las motivaciones de los clientes. Algunos de los gráficos son: D. Proporción por canal, E. Método de entrega, H. Lo que piensan los clientes.

Figura 31. Motivaciones



Fuente: Elaboración propia.

- f. Retos: esta es una variable también cualitativa. Las inferencias pueden realizarse utilizando gráficos como E. Método de entrega, H. Lo que piensan los clientes.

Figura 32. Retos



Se debe inferir en los retos que el buyer pueda presentar basándose en los comentario y algunos comportamientos de compra recurrente. Por ejemplo: en este caso se usó la preferencia del uso del servicio a domicilio como un reto por los atrasos en servicios logísticos.

Fuente: Elaboración propia.

- g. Comportamiento de compra: esta información puede ser extraída de los distintos gráficos que se muestran en el tablero, en ellos se proporciona información como, preferencias de métodos de entrega, horarios de compra más frecuente, *ticket* promedio y frecuencia de compra promedio.

Figura 33. Comportamiento de compra



Fuente: Elaboración propia.

- h. Canales de comunicación: se usa el gráfico de proporción de canales para obtener los

canales de comunicación preferidos de los clientes.

**Figura 34. Canales de comunicación**



En el gráfico D. se observan el peso de cada canal

Fuente: Elaboración propia.

- i. Creación del *Buyer persona*: con la información analizada se hace una propuesta de lo que debería de ser el *Buyer persona* de la empresa.

**Figura 35. Canales de comunicación**



Fuente: Elaboración propia.

- j. Matriz de peso del subsegmento por venta y volumen: con el objetivo de entender el peso y participación que el *buyer* elegido tiene dentro de las ventas y el volumen de transacciones se explica la matriz con mapa de calor para identificar el segmento con una mayor participación.

Figura 36. Canales de comunicación

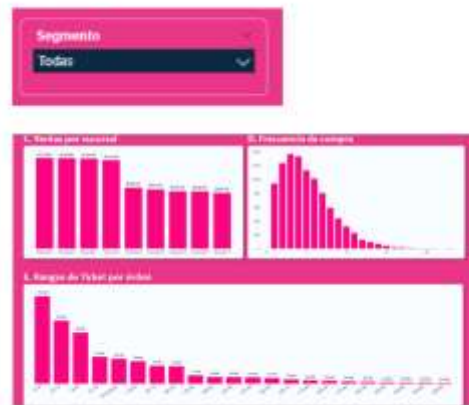


Las tarjetas de datos indican la venta total y el volumen de órdenes total. Los Heatmap están diseñados para resalta el valor más alto, en este caso ayudará a identificar la combinación de características que generan el segmento con mayor peso.

Fuente: Elaboración propia.

- k. Gráficos interactivos: para poder tener mejor filtrado el segmento elegido y ver su peso en las ventas y en el volumen de órdenes, se explica la realización de gráficos interactivos que muestran las tendencias y ayudan a filtrar.

Figura 37. Gráficos interactivos



Los gráficos C, Ventas por sucursal, D, Frecuencia de compra y E, Rango de ticket por orden. Estos son gráficos puestos con el objetivo de observar tendencias y con esto poder filtrar hasta llegar al segmento deseado.

Seleccionando la barra del gráfico que se desea filtrar se filtran directamente las matrices, si se desea más de un filtro al mismo tiempo puede seleccionarse la otra columna presionando "ctrl".

Fuente: Elaboración propia.

- l. Definición de segmentos y variaciones en la puntuación: cada segmento propuesto se le asigno una combinación de puntuaciones RFM, estas puntuaciones son asignas en

base a criterio, pueden ser modificadas, al igual que los nombres de los segmentos.

**Figura 38. Segmentos por puntuación RFM**



Fuente: Elaboración propia.

- m. Interacción en el tablero de segmentación: el tablero de segmentación permite observar los distintos segmentos y sus comportamientos.

**Figura 39. Interacción tablero de segmentos**



Fuente: Elaboración propia.

- n. Cuadro con información de clientes: se creó una cuadro con la finalidad de extraer la

información de los clientes en un archivo para poder ejecutar las estrategias.

**Figura 40. cuadro de información de clientes**

### Tabla con información de clientes



Luego de filtrar y elegir el segmento en el que la empresa desea enfocar su estrategia, se puede descargar la tabla con la información de cada cliente y poder implementar las estrategias elegidas.

Fuente: Elaboración propia.

## XI. Conclusiones

Se diseñó un sistema integral de gestión de datos que abarca desde la recolección, análisis y visualización de la información de clientes. Este sistema permitió implementar modelos de segmentación y técnicas de perfilación como el desarrollo de *buyer personas*. Este sistema brinda una base para que la empresa pueda construir estrategias de *marketing* mejor dirigidas. La creación de perfiles estratégicos a partir de la información estructurada permite que las futuras campañas estén basadas en datos concretos, ajustándose al comportamiento y preferencias de los clientes.

Se planteó la estructura de un programa de fidelización, para ello se utilizaron datos históricos de la empresa, además de un análisis de competencia dentro de la industria de alimentos. El análisis de referencia del mercado permitió comprender las mejores prácticas utilizadas en otros programas de lealtad, esto fue útil como referencia para diseñar un esquema adaptado a las necesidades específicas de la empresa. Se utilizó inteligencia artificial para la generación de bases de datos ficticias adaptadas al contexto de la empresa, las cuales permitieron visualizar cómo se vería el desempeño del programa en términos de ventas, órdenes y comportamiento del cliente, en caso de ser implementado.

Se desarrollaron tableros de visualización de datos en *Power BI* que permitieron estructurar de forma clara y dinámica los datos obtenidos. Estos tableros facilitaron la interpretación de la información, proporcionando *insights* sobre el perfil de los clientes, su comportamiento de compra y sus patrones de consumo. Asimismo, se desarrolló un modelo de segmentación RFM, el cual permitió agrupar a los clientes en segmentos estratégicos según la combinación del puntaje de cada métrica: valor monetario, frecuencia de compra y recencia, y así ayudar a la empresa a definir de manera más precisa tanto sus segmentos de mercado como sus *buyer personas*, permitiéndoles construir estrategias diferenciadas para cada grupo.

Se creó una presentación para guiar a la empresa, la cual explica de manera práctica qué visualizaciones del tablero deben ser utilizadas para construir los *buyer personas* y para generar los segmentos de clientes. Por medio de esta herramienta el equipo gerencial y de *marketing* pueden aprovechar de forma autónoma la información disponible en los tableros para la toma de decisiones futuras.

## XII. Recomendaciones

Se recomienda que el siguiente paso, antes de lanzar oficialmente el programa de lealtad, sea realizar una proyección financiera, tomando en cuenta tanto los ingresos esperados como los gastos e inversiones que se deberán hacer para llevar a cabo el proyecto. Esta proyección será útil para analizar la rentabilidad estimada del programa y ajustar la estrategia si fuera necesario para garantizar su viabilidad económica.

Es fundamental que las acciones de *marketing* no se realicen de manera generalizada, por eso se recomienda generar estrategias dirigidas a segmentos específicos identificados mediante el modelo RFM y la construcción de *buyer personas*. Por ejemplo: se recomienda diseñar una campaña de reactivación para el segmento “No podemos perder”, ofreciendo beneficios exclusivos dentro de los canales propios, una campaña que genere un sentido de urgencia con mensajes como: “¡Solo por hoy!”. Además de comunicar un mensaje como: “Te extrañamos”, para reactivar el vínculo emocional que el cliente ya había establecido con la marca

Se recomienda invertir los esfuerzos de *marketing* en segmentos de alto impacto, priorizando la inversión de recursos y campañas en los segmentos que representen un peso significativo tanto en las ventas como en el volumen de transacciones, basándose en los análisis realizados a través del *dashboard*. De esta manera, poder maximizar el retorno de la inversión enfocándose en los perfiles de clientes que generan mayor valor para el negocio.

Se sugiere actualizar los análisis de segmentación y perfilación de forma periódica, utilizando la nueva información recolectada a través del programa de fidelización una vez implementado. Esto con la finalidad de mantener estrategias relevantes y adaptadas a la evolución del mercado y las preferencias de los consumidores.

### XIII. Bibliografía

- Alonso, M. (2025, 13 febrero). Insights: qué son y cómo aplicarlos a tu proyecto [2025] • Asana. *Asana*. <https://asana.com/es/resources/insights>
- Amazon Web Services, Inc. (2020, 24 noviembre). *¿Qué es SQL?* - AWS. <https://aws.amazon.com/es/what-is/sql/#:~:text=El%20lenguaje%20de%20consulta%20de,espec%C3%ADficos%20de%20una%20cuadro%20SQL>.
- Amazon Web Services, Inc. (2022a). *¿Qué es el análisis de datos? - Explicación del análisis de datos* - AWS. <https://aws.amazon.com/es/what-is/data-analytics/>
- Amazon Web Services, Inc. (2022b). *¿Qué es la visualización de datos? - Explicación de la visualización de datos* - AWS. <https://aws.amazon.com/es/what-is/data-visualization/>
- Amazon Web Services, Inc. (2022c). *¿Qué es Python? - Explicación del lenguaje Python* - AWS. <https://aws.amazon.com/es/what-is/python/>
- Crabtree, M. C., & Nehme, A. N. (2024, 26 abril). *¿Qué es el análisis de datos?* DataCamp. <https://www.datacamp.com/es/blog/what-is-data-analysis-expert-guide>
- Flores, N. (2023, 21 enero). Recolección de datos: qué es, ventajas y consejos para usarlos. *Tecnológico de Monterrey*. <https://blog.maestriasydiplomados.tec.mx/recoleccion-de-datos-que-es-ventajas-y-consejos-para-usarlos>
- Google. (2025). *¿Qué es la inteligencia artificial o IA?* | Google Cloud. Google Cloud. <https://cloud.google.com/learn/what-is-artificial-intelligence?hl=es-419>
- HubSpot. (2023, 4 abril). Qué es la fidelización del cliente y 7 estrategias para lograrlo. *HubSpot*. <https://blog.hubspot.es/service/guia-fidelidad-cliente#que-es>
- HubSpot. (2024a, junio 12). Segmentación de mercado: cómo hacerla + ejemplos vigentes. *HubSpot*. <https://blog.hubspot.es/marketing/segmentacion-mercado>
- HubSpot. (2024b, julio 24). Recolección de datos: herramientas y 4 plantillas gratuitas. *HubSpot*. <https://blog.hubspot.es/marketing/recoleccion-de-datos>
- HubSpot. (2024c, diciembre 20). Qué es un *buyer* persona y cómo crearlo (con plantillas gratis). *HubSpot*. <https://blog.hubspot.es/marketing/que-son-buyer-personas#que-es>
- Ibm. (2021, 4 octubre). Exploratory Data Analysis. *IBM*. <https://www.ibm.com/think/marketingics/exploratory-data-analysis>
- Ibm. (2025, 9 enero). Visualización de datos. *IBM*. <https://www.ibm.com/mx-es/marketingics/data-visualization>
- Londoño, P. L. & HubSpot. (2021, December 21). Introducción al *marketing*: qué es, para qué sirve, tipos y guía. HubSpot. <https://blog.hubspot.es/marketing/introduccion-al-marketing>
- Mailchimp. (2023, 13 febrero). *4 variables de segmentación de mercado* | Mailchimp. <https://mailchimp.com/es/resources/what-are-segmentation-variables/>
- Mazhar, S. A. (2021). Methods of Data Collection: A Fundamental Tool of Research. *JOURNAL OF INTEGRATED COMMUNITY HEALTH*, 10(01), 6-10. <https://doi.org/10.24321/2319.9113.202101>
- McDonald's Mesoamérica. (2025). Nuestra Compañía - Historia. [Mcdonaldsmesoamerica.com](https://mcdonaldsmesoamerica.com).

- <https://mcdonaldsmesoamerica.com/compania#Historia>
- Medina, J. (2009). *Brain rules: 12 Principles for Surviving and Thriving at Work, Home, and School*. Microsoft. (2024a). *Conceptos básicos del diseño de una base de datos - Soporte técnico de Microsoft*. Microsoft Support. <https://support.microsoft.com/es-es/marketingic/conceptos-b%C3%A1sicos-del-dise%C3%B1o-de-una-base-de-datos-eb2159cf-1e30-401a-8084-bd4f9c9ca1f5#bmnormalization>
- Microsoft. (2024b). *Conceptos básicos del diseño de una base de datos - Soporte técnico de Microsoft*. Microsoft Support. <https://support.microsoft.com/es-es/marketingic/conceptos-b%C3%A1sicos-del-dise%C3%B1o-de-una-base-de-datos-eb2159cf-1e30-401a-8084-bd4f9c9ca1f5#bmnormalization>
- Microsoft. (2024c). *Conceptos básicos sobre bases de datos - Soporte técnico de Microsoft*. Microsoft Support. <https://support.microsoft.com/es-es/marketingic/conceptos-b%C3%A1sicos-sobre-bases-de-datos-a849ac16-07c7-4a31-99483c8c94a7c204#:~:text=Una%20base%20de%20datos%20es,programa%20de%20procesamiento%20de%20texto>
- OCI. (2020, 24 noviembre). *What Is a Database?* Oracle Cloud Infrastructure (OCI). <https://www.oracle.com/mx/database/what-is-database/>
- R-project.org. (s. f.). *R: What is R?* R Project. <https://www.r-project.org/about.html>
- Smith, J., & Murphy, K. (2019). *Consumer privacy and data sharing behavior*. Journal of Marketing Research.
- Softonic. (2025, febrero 23). *Subway Guatemala: El Programa de Fidelización Digital*. <https://subway-guatemala.softonic.com/android>
- SurveyMonkey. (2024, 10 abril). *Incentivos: mejora tu índice de respuestas | SurveyMonkey*. <https://es.surveymonkey.com/learn/survey-best-practices/using-survey-incentives-to-improve-response-rates/>
- Starbucks lanza en Guatemala su programa de lealtad, Starbucks® Rewards - Starbucks Historias. (2024, October 25). Starbucks Historias. <https://historias.starbucks.com/press/2024/starbucks-lanza-en-guatemala-su-programa-de-lealtad-starbucks-rewards/>
- Tableau. (2024). *Guide to Data Cleaning: Definition, benefits, components, and how to clean your data*. <https://www.tableau.com/learn/articles/what-is-data-cleaning>
- Tableau. (2025). *Guía de visualización de datos para principiantes: definición, ejemplos y recursos de aprendizaje*. <https://www.tableau.com/es-mx/learn/articles/data-visualization>
- Trefra Club. (2025). *Preguntas frecuentes*. <https://club.trefragroup.com/preguntas-frecuentes>

## XIV. Anexos

Anexo I: Código SQL para la extracción de datos transaccionales de la empresa.

```
SELECT
orden.id AS id_orden,
linea.id AS id_producto_orden,
templ.type AS tipo_producto,
sucursal.name AS sucursal,
linea.full_product_name AS producto,
linea.price_unit AS precio,
linea.qty AS cantidad_producto,
linea.price_subtotal_incl AS venta,
CASE
  WHEN metodo.name::json->>'en_US' = 'Visa Link' THEN 'Digital'
  WHEN metodo.name::json->>'en_US' = 'Pagina Web' THEN 'Digital'
  WHEN metodo.name::json->>'en_US' = 'TRF BI'
    OR metodo.name::json->>'en_US' = 'POS VISA'
    OR UPPER(metodo.name::json->>'en_US') LIKE '%EFECTIVO%' THEN 'Punto de Venta'
  WHEN metodo.name::json->>'en_US' = 'UBER' THEN 'UBER'
  WHEN metodo.name::json->>'en_US' = 'PEDIDOS YA' THEN 'Pedidos Ya'
  WHEN metodo.name::json->>'en_US' = 'BOT' THEN 'BOT'
  ELSE 'Canal Extra'
END AS canal,
CASE
  WHEN account.move_type = 'out_invoice' THEN 'Factura'
  WHEN account.move_type = 'out_refund' THEN 'Refund'
  ELSE 'Otro'
END AS move_type
FROM pos_order AS orden
INNER JOIN pos_order_line AS linea ON orden.id = linea.order_id
INNER JOIN pos_session AS sesion ON orden.session_id = sesion.id
INNER JOIN pos_config AS sucursal ON sucursal.id = sesion.config_id
INNER JOIN product_product AS product ON linea.product_id = product.id
INNER JOIN product_template AS templ ON product.product_tmpl_id = templ.id
INNER JOIN pos_payment AS payment ON orden.id = payment.pos_order_id
INNER JOIN pos_payment_method AS metodo ON payment.payment_method_id = metodo.id
LEFT JOIN account_move AS account ON orden.account_move = account.id
WHERE account.move_type IN ('out_invoice', 'out_refund')
AND metodo.name::json->>'en_US' = 'BOT';
```

## Anexo II: código de Python para el análisis exploratorio de los datos transaccionales

```
#importar datos transaccionales

import pandas as pd
df = pd.read_csv("C:/Users/Usuario/OneDrive - lulisgt/Documentos/TESIS/DATA_BOT1.csv", encoding='latin1')
print(df.head())

#participación en ventas y volumen
import plotly.express as px

df = df.dropna(subset=["producto", "cantidad_producto", "venta"])

#Agrupar datos por producto
df_volume = df.groupby("producto", as_index=False).agg({"cantidad_producto": "sum"})
df_sales = df.groupby("producto", as_index=False).agg({"venta": "sum"})

#calcular el share
df_volume["share_volumen"] = df_volume["cantidad_producto"] / df_volume["cantidad_producto"].sum() * 100
df_sales["share_ventas"] = df_sales["venta"] / df_sales["venta"].sum() * 100

#treemap volumen
fig_volume = px.treemap(
    df_volume,
    path=["producto"],
    values="cantidad_producto",
    title="Participación del Volumen de Venta por producto",
    labels={"cantidad_producto": "Cantidad Vendida"},
    hover_data={"cantidad_producto": True, "share_volumen": True}
)

fig_volume.update_traces(textinfo="label+value+percent entry")
fig_volume.update_layout(
    title=dict(
        x=0.01,          # Alineado a la izquierda
        xanchor='left',
        yanchor='marketing',
        y=0.85,         # Más bajo = más "pad"
        font=dict(size=20)
    )
)

#treemap venta
fig_sales = px.treemap(

    df_sales,
    path=["producto"],
    values="venta",
    title="Participación de la Venta por Producto",
    labels={"venta": "Total Vendido"},
    hover_data={"venta": True, "share_ventas": True}
)
```

```

fig_sales.update_traces(
    textinfo="label+text+percent entry",
    texttemplate="% {label}<br>Q% {value:,.2f}<br>(<math>{percentEntry:.1\%}</math>)"
)

fig_sales.update_layout(
    title=dict(
        x=0.01,
        xanchor='left',
        yanchor='marketing',
        y=0.85,
        font=dict(size=20)
    )
)

fig_volume.show()
fig_sales.show()

# Histograma de ticket promedio

df_ticket = df.groupby("id_orden", as_index=False).agg({"venta": "sum"})

fig_ticket = px.histogram(
    df_ticket,
    x="venta",
    title="Distribución del Ticket Promedio",
    labels={"venta": "Total en Q"},
    nbins=50
)

fig_ticket.update_traces(marker=dict(color='#9525ad'))

fig_ticket.update_layout(
    plot_bgcolor='rgba(0,0,0,0)',
    paper_bgcolor='rgba(0,0,0,0)',
    yaxis_title='',

    title={
        'x': 0,          # Alineado a la izquierda (0 = izquierda, 0.5 = centro, 1 = derecha)
        'xanchor': 'left', # Anclamos desde la izquierda
        'font': {'size': 16} # Opcional: ajustar tamaño del texto
    }
)

fig_ticket.show()

#importar cuadro margenes
file_path = r"C:\Users\Usuario\OneDrive - lulisgt\Documentos\TESIS\margenesbot.xlsx"
df2 = pd.read_excel(file_path)

#marketing 10 margenes

```

```

import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from matplotlib.colors import to_rgb

#marketing 10 productos descendente
marketing_10_productos = df2.sort_values(by='Margen', ascending=False).head(10)

#color
base_color = np.array(to_rgb("#9525ad"))
white = np.array([1, 1, 1])
n = len(marketing_10_productos)
colors = [tuple(base_color + (white - base_color) * (i / (n - 1))) for i in range(n)]

#gráfico de barras mejor margen
plt.figure(figsize=(10, 8))
sns.barplot(
    x='Margen',
    y='producto',
    data=marketing_10_productos,
    palette=colors,
    hue='producto',
    dodge=False
)

for index, value in enumerate(marketing_10_productos['Margen']):
    plt.text(value + 0.01, index, f'{value:.2f}', fontsize=12)

plt.title('Marketing 10 Productos con Mejor Margen', fontsize=14, loc='left', pad=30)
plt.xlabel('Margen', fontsize=12)
plt.ylabel("", fontsize=14)
plt.legend([], [], frameon=False) # Ocultar leyenda
sns.despine()
plt.tight_layout()

plt.show()

#combinar ambas cuadros

df.columns = df.columns.str.strip()
df2.columns = df2.columns.str.strip()

print("Columnas de df:", df.columns)
print("Columnas de df2:", df2.columns)

#combinación usando la columna de producto
df_combined = pd.merge(df[['producto', 'cantidad_producto', 'venta']],
                      df2[['producto', 'Margen']],
                      left_on='producto', right_on='producto',
                      how='inner')

#agrupar por producto
df_combined = df_combined.groupby('producto').agg(
    volumen=('cantidad_producto', 'sum'),
    venta=('venta', 'sum'),

```

```

    margen=('Margen', 'first')
).reset_index()

df_combined['margen'] = df_combined['margen'] * 100

#distribuciones

import matplotlib.ticker as mtick

df_combined = df_combined.sort_values(by='margen', ascending=True)

x1_centro = df_combined['volumen'].median()
x2_centro = df_combined['venta'].median()
y_centro = df_combined['margen'].median()

#distribución volumen vs margen
plt.figure(figsize=(10, 6))
sns.scatterplot(x='volumen', y='margen', data=df_combined, color='#C71585')

for index, row in df_combined.iterrows():
    plt.text(row['volumen'] + 0.1, row['margen'], row['producto'], fontsize=9)

plt.title('Volumen vs Margen', fontsize=14, loc='left', pad=30)
plt.xlabel('Volumen', fontsize=12)
plt.gca().xaxis.labelpad = 20
plt.ylabel("")

plt.grid(False)

plt.axhline(y=y_centro, color='gray', linestyle='--', linewidth=1)
plt.axvline(x=x1_centro, color='gray', linestyle='--', linewidth=1)

plt.gca().yaxis.set_major_formatter(mtick.PercentFormatter())

plt.show()

#distribución ventas vs margen
plt.figure(figsize=(10, 6))
sns.scatterplot(x='venta', y='margen', data=df_combined, color='purple')

for index, row in df_combined.iterrows():
    plt.text(row['venta'] + 0.1, row['margen'], row['producto'], fontsize=9)

plt.title('Ventas vs Margen', fontsize=14, loc='left', pad=30)
plt.xlabel('Ventas', fontsize=12)
plt.gca().xaxis.labelpad = 20
plt.ylabel("")
plt.grid(False)

plt.axhline(y=y_centro, color='gray', linestyle='--', linewidth=1)
plt.axvline(x=x2_centro, color='gray', linestyle='--', linewidth=1)

plt.gca().yaxis.set_major_formatter(mtick.PercentFormatter())

plt.show()

```

### Anexo III: Prompt redactado a la IA para obtención de la simulación de bases de datos.

“Eres un científico de datos experto en python y simulaciones de bases de datos

----

Quiero hacer una simulación de varios cuadros utilizando faker para simular las ventas de una tienda de galletas, con un listado de productos ["Producto 1", "Producto 2", etc...], con 9 sucursales con un listado de ["Sucursal 1", "Sucursal 2", etc...]

Quiero simular las ventas en un rango de 6 meses, luego yo realizaré un análisis de segmentación de los clientes, por lo que los clientes deben de poder repetirse y tener comportamientos de compra variados. Esta segmentación la haré utilizando un modelo RFM, por lo que debes de usar perfiles de comportamiento variado y de comportamiento fiel o semifiel a algunas tiendas. Por ejemplo, un cliente puede ser totalmente fiel a una sola tienda y variado en productos, otro fiel en un par de productos, pero variado en tiendas, otros que compraron casualmente, etc. Define varias opciones de estos perfiles y simula su comportamiento. Utiliza diversas distribuciones de probabilidad. Además, asume que tienes un listado de promedio de compras mensuales, *ticket* de compra promedio y monto total mensual por cada tienda y trata de respetarlo hasta cierto punto que podría variar un poco.

Asimismo, la base de datos debería ser capaz de simular más de un producto en la misma factura para poder tener variedad, esto debe ser realizado basado en el número de orden, es decir que el número de orden debe poder repetirse hasta 5 veces, junto con otros aspectos como la sucursal y fecha, esto quiere decir que una persona en una misma factura compró hasta 5 productos diferentes.

Otra cosa que debe variar es la cantidad de productos que llevan por orden de compra.

La segmentación será por medio del número de teléfono, así que simula los nombres, teléfonos, correos, edades, sexos de los clientes. La edad debe estar distribuida de manera que una mayor concentración de clientes se encuentre en edades de 20 a 40 años aproximadamente. Asimismo, necesito la proporción de sexo sea mayor la femenina que masculino.

Se crearán 3 bases de datos, una transaccional donde se encuentra el registro de las compras en el tiempo, otra base de datos de clientes donde se almacena toda la información de ellos y una de redenciones, en donde se muestren los productos canjeados.

La primera base de datos debe llevar las siguientes variables: fecha, día de la semana (que haya un mayor consumo de jueves a domingo), número de orden, número de teléfono del cliente, sucursal (que haya sucursales con mayor consumo que otras), producto, precio, cantidad vendida, venta total, hora de compra (que hay una mayor cantidad de compras en la tarde), rating por producto (con una variación de 1 a 5, repitiéndose más los ratings de 4 y 5), una columna de comentarios (se te proporcionará la lista de posibles comentarios) y categoría que se te proporcionará en lista, método de entrega (debes elegir entre Delivery o Pickup, quiero que la distribución de esto sea 65% delivery y 35% pickup) la segunda base de datos que es la de clientes deberá llevar: nombre, número del cliente, fecha de cumpleaños, correo electrónico, sexo, red social de dónde vienen o donde se enteraron del plan de fidelización (agrega Instagram, Facebook, tiktok o recomendación de manera que la distribución de esto sea mayor mente Instagram, seguido de recomendaciones, luego tiktok y por último facebook) la tercera base de datos llevará: número de teléfono, fecha, sucursal (con variaciones visibles por cada sucursal) y producto canjeado (con variaciones visibles por cada producto), recuerda que solo pueden canjear estos productos después de haber gastado cierta cantidad de dinero, cada quetzal gastado es 1 punto acumulado, el listado de lo productos junto con la cantidad de puntos necesarios para canjear es el siguiente: Cookie 1, Cookie 3, Box 4, Cookie 6, Cookie 10, Cookie 7.

-----

Debes retornar el código para lograr la simulación y la segmentación.

Haz todas las preguntas necesarias para poder tener totalmente claro el panorama y necesidades del proyecto.”

#### Anexo IV: Respuesta de la IA al prompt inicial.

“¡Perfecto! Para poder elaborar un código que cumpla exactamente con tus necesidades y preparar el escenario de simulación de forma óptima, necesito aclarar algunos detalles. Por favor, ayúdame respondiendo las siguientes preguntas:

1. Rango de 6 meses
  - ¿Tienes un rango de fechas específico en mente (por ejemplo, de enero a junio del 2023)?
  - ¿O prefieres que se simule en un período más genérico (por ejemplo, desde el mes actual hacia atrás 6 meses)?
2. Volumen de datos
  - ¿Cuántos registros aproximados deseas en la base transaccional (número de ventas totales)?
  - ¿Cuántos clientes únicos te gustaría simular?
3. Listados promedio de compras mensuales, *ticket* promedio y monto total mensual por sucursal
  - ¿Tienes algún valor de referencia? Por ejemplo, “Sucursal 1” vende en promedio 500 *tickets* al mes con un monto total de 20,000 quetzales, “Sucursal 2” vende 300 *tickets* con un monto total de 12,000 quetzales, etc.
  - Si no los tienes muy definidos, podemos asumir ciertas distribuciones y luego ajustarlas.
4. Perfil de sucursales
  - Mencionaste que algunas sucursales deberían tener más consumo que otras. ¿Podrías darnos un orden aproximado de mayor a menor?
  - Por ejemplo, “Sucursal 1” (mayor venta), “Sucursal 2”, “Sucursal 3”, etc., hasta “Sucursal 9” (menor venta).
5. Productos
  - ¿Tienes ya la lista definitiva de productos (ej. “Producto 1”, “Producto 2”, ..., “Producto X”) o prefieres que generemos un listado de nombres de galletas más “realistas”?
  - ¿Puedes compartir la categoría (por ejemplo, “Chocolate”, “Vainilla”, “Rellenas”, “Especiales”, etc.) para cada producto si ya las tienes definidas?
  - ¿Deseas que el precio de cada producto sea fijo o que varíe un poco?
6. Comportamiento de los clientes (perfiles)
  - ¿Tienes algún número aproximado de perfiles o grupos que quieras simular? (p. ej., 4 o 5 grupos como “Totalmente fiel a 1 sucursal”, “Varía en producto pero no en sucursal”, “Comprador casual”, etc.)
  - Además, ¿hay algún porcentaje aproximado que quieras asignar a cada comportamiento?
7. Cantidad de productos por orden
  - Hasta 5 productos distintos en la misma factura. ¿Deseas alguna distribución especial? Por ejemplo, 60% de las ventas con 1-2 productos, 30% con 3 productos, 10% con 4 o 5 productos.

- ¿Hay algún *marketinge* en la cantidad total de un mismo producto en la misma orden? (por ejemplo, máximo 10 unidades del mismo producto en una sola compra).
8. Datos de los clientes
    - ¿Te interesa algún patrón de nombres (por región, por ejemplo) o simplemente nombres latinoamericanos genéricos está bien?
    - ¿Tienes algún rango de edad específico aparte de la concentración entre 20 y 40 años? (por ejemplo, ¿queremos un 70% entre 20 y 40, 20% entre 41 y 60, y 10% < 20 o > 60?).
    - Para la proporción de sexo, ¿qué porcentaje aproximado consideras femenino vs. masculino (y si quieres agregar no binario, etc.)?
  9. Distribución de redes sociales
    - Mencionaste que la mayoría proviene de Instagram, luego recomendación, luego TikTok y por último Facebook. ¿Podrías dar un rango aproximado en %? (p. ej., 40% Instagram, 30% recomendación, 20% TikTok, 10% Facebook).
  10. Comentarios y categorías
    - ¿Podrías compartir la lista de posibles comentarios que quieres que se incluyan?
    - ¿Tienes la lista de categorías de producto? (Mencionaste que “se te proporcionará la lista de categorías” — ¿está ya definida o prefieres que se cree algo genérico?)
  11. Método de entrega
    - Ya indicaste que sea 65% Delivery y 35% Pickup. ¿Hay alguna diferencia de precio o condición para cada uno o solo lo manejamos como una variable categórica?
  12. Ratings
    - Se repetirán más los ratings de 4 y 5. ¿Deseas un aproximado como 40% calificación 5, 30% calificación 4, 15% calificación 3, 10% calificación 2, 5% calificación 1 (por dar un ejemplo)?
    - prefieres otra proporción.
  13. Redenciones
    - Productos a canjear: Cookie 1, Cookie 3, Box 4, Cookie 6, Cookie 10, Cookie 7.
    - ¿Cuántos puntos se requieren para cada uno?
    - ¿Deseas que la base de redenciones sea considerablemente menor que la base total de ventas? (Por ejemplo, ~10% de los clientes canjean algo en esos 6 meses).
    - ¿Tienes un límite de canjes por cliente en esos 6 meses?
  14. Representación de los datos
    - ¿En qué formato deseas que se guarden los datos resultantes? (DataFrames en pandas, CSVs, o inserts SQL para una base de datos específica, etc.).
    - ¿Tienes alguna preferencia de librerías adicionales para usar, además de Faker (por ejemplo, NumPy, pandas, SQLAlchemy, etc.)?
  15. Output del código
    - ¿Deseas que el script te genere directamente los 3 CSV (transacciones, clientes y redenciones)?
    - ¿Quieres también un notebook/Python script que realice la segmentación RFM y muestre algunos ejemplos de resultados (agrupación de clientes)?

16. Cualquier otra restricción o requerimiento especial

- ¿Hay alguna otra regla de negocio que debamos considerar en la simulación?
- ¿Algún requerimiento legal o de formato específico (por ejemplo, enmascarar números de teléfono, correos, etc.) o podemos simular libremente?

Con estas respuestas podré generar un código completo en Python que simule los tres conjuntos de datos (transaccional, clientes y redenciones) con la lógica de comportamiento, distribución de probabilidades y perfiles RFM que necesitas. ¡Gracias de antemano!”

Anexo V: Prompt con respuestas para la IA.

1. El rango es de 1 de abril al 30 de septiembre
2. Se desea un aproximado de 25,000 filas. Se quieren 3500 clientes
3. asúmelo, siempre teniendo en cuenta que algunas sucursales venden más que otras. Sí hay picos de venta, principalmente de viernes a domingo, si hay días con más ventas en festividades como día de la madre, día del padre, o así
4. Si hay 4 tiendas *marketing* sellers y 5 tiendas que van entre 2 medias y 3 un poco más bajas
5. en realidad hay 24 productos, dividido en 3 categorías, los nombres de los productos son los siguientes: Categoría "Boxes" ( Box 1, Box 2, Box 3, Box 4, Box 5, Box 6), categoría "Cookies" (Cookie 1, Cookie 10, Cookie 11, Cookie 12, Cookie 2, Cookie 3, Cookie 4, Cookie 5, Cookie 6, Cookie 7, Cookie 8, Cookie 9) , categoría "Varios" (Producto 1, Producto 2, Producto 3, Producto 4, Producto 5)
6. Si tienen precios específicos asignados, box 1: Q92, box2: Q177.65, box 3: Q43.3, box 4: Q43.3, box 5: Q84.5, box 6: 84.5, cookie 1 a cookie 6: Q 23.8, cookie 7 a cookie 9: Q34.7, Cookie 10: Q23.8, Cookie 11 y Cookie 12: Q29.2, producto 1 y producto 2: Q92, producto 3 y producto 4: Q73.6, producto 5: Q164.6
7. Asume la cantidad de segmentos, ese modelado lo haré luego con la data transaccional. Asume los porcentajes
8. Que sea así, un máximo de 5 productos por orden y una distribución de 60% de las ventas con 1-2 productos, 30% con 3 productos, 10% con 4 o 5 productos.
9. Asume los nombres, pero que los número de teléfono sean con código de Guatemala (+502). asume las edades, pero quiero un mayor peso en edades jóvenes. Para la distribución de sexo debe ser un 65% femenino (F) y 35% masculino (M).
10. para las redes sociales quiero que sea un 45% instagram, 20% tiktok, 25% recomendación y 10% facebook
11. los comentarios no tiene que ser en todas las ordenes, es 1 comentario por número de orden, pero puede ser solo en el 20% de las órdenes en donde coloques comentarios, serán comentarios random de una lista que yo te daré, se dividen por comentarios positivos y negativos, el 70% de los comentarios deben venir de los positivos y 30% de los negativos: a continuación, te doy la lista de comentarios:

Comentarios Positivos:

"¡Están deliciosas! Muy ricas, me encantaron." "Las galletas son muy buenas, las más ricas." "¡Qué deliciosos brownies! ¡Los mejores!" "Todo está muy bueno, ¡muy rico!" "Las galletas son frescas y crujientes, ¡deliciosas!" "¡Están buenísimos! Me encantaron." "Muy ricas, las mejores galletas." "¡Increíble! Todo está delicioso." "Los brownies están muy buenos, ¡muy suaves!" "¡Las mejores galletas! Muy ricas." "Todo está muy rico, me encanta." "¡Qué buenas! Los mejores brownies." "Deliciosos, muy suaves y ricos." "Las galletas están frescas, ¡muy ricas!" "¡Todo está muy sabroso, muy recomendado!" "¡Todo riquísimo! Los brownies son geniales." "Muy buena calidad, todo estaba

delicioso." "Las galletas son las mejores, ¡están deliciosas!" "Los brownies son muy buenos, ¡super ricos!" "¡Increíble! Me encantaron las galletas."

Comentarios Negativos:

"Las galletas venían frías, no me gustaron." "El brownie estaba un poco seco, no me convenció." "Las galletas estaban duras, no me gustaron mucho." "El sabor estaba bien, pero venían fríos." "Me esperaba algo más suave, pero estaban muy duros." "Las galletas se quebraron, no eran lo que esperaba." "Los brownies no estaban frescos, muy duros." "Las galletas no tenían mucho sabor, no me convencieron." "El brownie estaba frío y seco, no me gustó." "Las galletas venían rotas, no estaban bien." "El sabor es bueno, pero estaban frías." "Me decepcionaron, no eran lo que esperaba." "Las galletas se rompieron muy rápido, no eran frescas." "El brownie estaba muy seco, esperaba algo más suave." "No me gustaron, estaban muy duras y frías." "Las galletas estaban muy secas, no me gustaron." "Los brownies no estaban frescos, no los volveré a comprar." "Me esperaba algo más suave, no estaban tan ricos." "Estaban frías y un poco duras, no me convencieron." "El brownie estaba quemado, no me gustó."

quítale las tildes y los signos de exclamación

11. Manéjalo como variable categórica

12. Asúmelo

13. Estos son los productos con los puntos: Cookie 1: 400, Cookie 3: 320, Box 4: 480, Cookie 6 : 320, Cookie 10: 380, Cookie 7: 430

Asume el porcentaje de la cuadro de redenciones. No hay límite de canjes

14. Guárdalos en CSV. puedes usar cualquier librería

15. Si quiero que se generen los CSV. No hagas la segmentación, ese será parte de mi trabajo luego del análisis

16. Simula libremente

#### Anexo VI: Código para la simulación.

```
#####
# SIMULACIÓN DE BASE DE DATOS
# Ventas de Tienda de Galletas
#####
import numpy as np
import pandas as pd
import random
from faker import Faker
from datetime import datetime, timedelta, date
from collections import defaultdict
# Semilla para reproducibilidad (opcional)
np.random.seed(43)
random.seed(43)
fake = Faker('es_ES') # Faker en español (puedes usar 'en_US' o el local que prefieras)
# -----
# 1. CONFIGURACIONES Y PARÁMETROS GENERALES
# -----
# Rango de fechas: 6 meses comenzando el 1 de abril
start_date = date(2025, 4, 1) # 1 de abril (año arbitrario 2025)
end_date = date(2026, 3, 31)
# Cantidad total de clientes a simular
NUM_CLIENTES = 10500
# Cantidad de líneas transaccionales aproximadas
```

```

TARGET_NUM_LINES = 50_000
# Cantidad máxima de productos por orden (factura)
MAX_PRODUCTOS_POR_ORDEN = 5
# Definimos las 9 sucursales
sucursales = [
    "Sucursal 1", "Sucursal 2", "Sucursal 3",
    "Sucursal 4", "Sucursal 5", "Sucursal 6",
    "Sucursal 7", "Sucursal 8", "Sucursal 9"
]
# Suponemos que hay 4 sucursales "marketing sellers" y 5 sucursales "normales"
# Ajustamos un factor de probabilidad para asignar órdenes
marketing_sellers = ["Sucursal 1", "Sucursal 2", "Sucursal 3", "Sucursal 4"]
normal_sellers = ["Sucursal 5", "Sucursal 6", "Sucursal 7", "Sucursal 8", "Sucursal 9"]
# Probabilidad de que una orden vaya a una tienda "marketing" vs "normal"
# (Ajusta según la proporción que desees)
prob_marketing_sellers = 0.60 # 60% de las órdenes en 4 tiendas marketing
prob_normal_sellers = 0.40
# Creamos la distribución de probabilidad para todas las sucursales
prob_sucursales = {}
# Dividimos el peso entre las sucursales marketing y normales
peso_por_marketing = prob_marketing_sellers / len(marketing_sellers)
peso_por_normal = prob_normal_sellers / len(normal_sellers)
for suc in marketing_sellers:
    prob_sucursales[suc] = peso_por_marketing
for suc in normal_sellers:
    prob_sucursales[suc] = peso_por_normal
# Definimos la lista de productos con sus precios fijos
# (moneda en Quetzales 'Q')
# Nota: Se incluye la nomenclatura que diste: Box 1..6, Cookie 1..12, Producto 1..5
# Verifica que la ortografía coincide con la que desees.
productos_precios = {
    "Box 1": 92.0,
    "Box 2": 177.65,
    "Box 3": 43.3,
    "Box 4": 43.3,
    "Box 5": 84.5,
    "Box 6": 84.5,
    "Cookie 1": 23.8,
    "Cookie 2": 23.8,
    "Cookie 3": 23.8,
    "Cookie 4": 23.8,
    "Cookie 5": 23.8,
    "Cookie 6": 23.8,
    "Cookie 7": 34.7,
    "Cookie 8": 34.7,
    "Cookie 9": 34.7,
    "Cookie 10": 23.8,
    "Cookie 11": 29.2,
    "Cookie 12": 29.2,
    "Producto 1": 92.0,
    "Producto 2": 92.0,
    "Producto 3": 73.6,
    "Producto 4": 73.6,
    "Producto 5": 164.6
}

```

```

lista_productos = list(productos_precios.keys())
# Para la hora de compra, definimos un rango de 8am a 10pm, por ejemplo
hora_inicio = 8
hora_fin = 20
# -----t-----
# 2. CREACIÓN DE BASE DE CLIENTES
# -----
# Queremos 10,500 clientes con: nombre, teléfono, fecha cumpleaños, correo, sexo
# Edades de 18 a 55
def generar_telefono_guatemala():
    """
    Genera un número de teléfono guatemalteco con código de país +502
    seguido de 8 dígitos.
    Los 8 dígitos pueden empezar con 2, 3, 4, 5, 6 o 7,
    según las series más comunes en Guatemala.
    """
    first_digit = random.choice([2, 3, 4, 5, 6, 7])
    phone_digits = [str(first_digit)] + [str(random.randint(0, 9)) for _ in range(7)]
    phone_local = ".join(phone_digits)
    # prepend +502
    return f"+502{phone_local}"
def generar_edad_sesgada():
    """
    Genera una edad entre 18 y 55 con mayor concentración entre 18 y 35 años,
    usando una distribución beta sesgada hacia edades más jóvenes.
    """
    edad_beta = np.random.beta(1.5, 3.5) # más peso hacia edades bajas
    edad = int(18 + edad_beta * (55 - 18))
    return edad
redes_sociales = ["instagram", "facebook", "tiktok", "recomendacion", "whatsapp"]
prob_redes = [0.40, 0.15, 0.15, 0.25, 0.05] # suma=1.0
def escoger_red_social():
    """
    Elige la red social/origen de fidelización con una distribución específica.
    Por ejemplo: instagram 30%, facebook 25%, tiktok 15%, recomendacion 30%
    """
    return np.random.choice(redes_sociales, p=prob_redes)
def generar_cliente():
    """
    Genera un registro de cliente:
    - nombre
    - teléfono (GT)
    - fecha_cumple (18 a 55 años)
    - email
    - sexo
    - red_social (origen fidelización)
    """
    nombre = fake.name()
    telefono = generar_telefono_guatemala()
    # Edad con sesgo hacia los jóvenes
    edad = generar_edad_sesgada()
    hoy = datetime.now().date()
    birth_year = hoy.year - edad
    birth_month = random.randint(1, 12)
    birth_day = min(random.randint(1, 28), 28) # para evitar errores con días

```

```

fecha_cumple = date(birth_year, birth_month, birth_day)
correo = fake.email()
sexo = random.choices(["F", "M"], weights=[0.65, 0.35])[0]
red_social = escoger_red_social()
return [nombre, telefono, fecha_cumple, correo, sexo, red_social]
clientes_data = []
telefono_set = set()
while len(clientes_data) < NUM_CLIENTES:
    c = generar_cliente()
    # Evitamos duplicados de teléfono
    if c[1] not in telefono_set:
        clientes_data.append(c)
        telefono_set.add(c[1])
df_clientes = pd.DataFrame(
    clientes_data,
    columns=["nombre", "telefono", "fecha_cumple", "email", "sexo", "red_social"]
)
# -----
# 3. DEFINICIÓN DE PERFILES DE COMPORTAMIENTO
# -----
# Vamos a clasificar clientes en 4 tipos (tú puedes ajustar las proporciones):
# A) Fieles a 1 sucursal (30%)
# B) Fieles a 2-3 productos (20%)
# C) Comportamiento variado (50%)
#
# También podemos incluir la idea de "binomial negativa" vs "poisson"
# para número de órdenes en el periodo. Simplificaremos con un approach
# para la frecuencia esperada de compras.
num_clientes = len(df_clientes)
indices = df_clientes.index.tolist()
random.shuffle(indices) # mezclamos índices para asignar aleatoriamente
n_fiel_sucursal = int(num_clientes * 0.30)
n_fiel_productos = int(num_clientes * 0.20)
# El resto será variado
n_variado = num_clientes - n_fiel_sucursal - n_fiel_productos
# Asignamos grupos
grupo_sucursal = indices[:n_fiel_sucursal]
grupo_productos = indices[n_fiel_sucursal:n_fiel_sucursal+n_fiel_productos]
grupo_variado = indices[n_fiel_sucursal+n_fiel_productos:]
# Para cada grupo, definimos un "perfil":
# - Grupo fiel a 1 sucursal -> elegimos la sucursal preferida
# - Grupo fiel a 2-3 productos -> elegimos los productos preferidos
# - Grupo variado -> sin preferencia fuerte
perfil_cliente = {}
for i in grupo_sucursal:
    sucursal_preferida = random.choice(sucursales) # 1 sucursal
    perfil_cliente[i] = {
        "tipo": "fiel_sucursal",
        "sucursal_preferida": sucursal_preferida
    }
for i in grupo_productos:
    # Escogemos 2 o 3 productos preferidos
    n_prefs = random.choice([2, 3])
    prods_preferidos = random.sample(lista_productos, n_prefs)
    perfil_cliente[i] = {

```

```

        "tipo": "fiel_productos",
        "productos_preferidos": prods_preferidos
    }
for i in grupo_variado:
    perfil_cliente[i] = {
        "tipo": "variado"
    }
# -----
# 4. GENERACIÓN DE ÓRDENES Y LÍNEAS DE DETALLE
# -----
# Aproximadamente 25,000 órdenes para llegar cerca de 50,000 líneas
# (dado que cada orden puede tener 1-5 productos).
NUM_ORDENES = 50_000

# Para obtener un ID de orden único, usaremos un contador
orden_id_counter = 100000 # un número base cualquiera
# Vamos a almacenar las líneas transaccionales en esta lista
transacciones_data = []
# -----
# 1. Definimos la distribución de horas (pico desde las 2pm)
# -----
hour_weights_dict = {
    8: 0.2,
    9: 0.3,
    10: 0.4,
    11: 0.5,
    12: 0.7,
    13: 0.8,
    14: 1.8, # mayor peso a partir de 2pm
    15: 2.0,
    16: 2.0,
    17: 2.6,
    18: 2.8,
    19: 1.1,
    20: 0.9,
}
hours = sorted(hour_weights_dict.keys())
hour_weights = [hour_weights_dict[h] for h in hours]
sum_hw = sum(hour_weights)
hour_probs = [w / sum_hw for w in hour_weights]
# Para simular la distribución de días (con mayor peso viernes a domingo),
# definimos probabilidad aproximada de día de la semana (lunes=0.8, martes=0.9, ...),
# y también algunos picos festivos (ej. día de la madre = 10 mayo, día del padre = 17 junio, etc.)
day_weights = {
    0: 0.7, # Lunes (weekday() = 0)
    1: 0.8, # Martes
    2: 1.2, # Miércoles
    3: 1.45, # Jueves
    4: 2.3, # Viernes
    5: 2.1, # Sábado
    6: 2.0 # Domingo
}
# Días festivos (ejemplo) con multiplicador adicional
festive_days = {

```

```

(5, 10): 2.0, # 10 de mayo (Día de la madre en varios países latinos)
(6, 17): 1.8 # 17 de junio (ej. día del padre en algunos lugares; ajusta a tu realidad)
}
def escoger_hora_compra():
    """
    Elige una hora (entre 8 y 22) con mayor probabilidad en la tarde
    y un minuto aleatorio de 0 a 59.
    """
    h = np.random.choice(hours, p=hour_probs)
    m = random.randint(0, 59)
    return h, m
def random_date_in_range():
    """
    Devuelve una fecha (datetime.date) entre start_date y end_date
    con mayor peso en fines de semana, días festivos,
    y día del mes (más compras 1..15).
    """
    while True:
        delta = end_date - start_date
        random_days = random.randint(0, delta.days)
        d = start_date + timedelta(days=random_days)
        # Peso por día de la semana
        wday = d.weekday()
        base_weight = day_weights.get(wday, 1.0)
        # Peso festivo
        fest_key = (d.month, d.day)
        festive_weight = festive_days.get(fest_key, 1.0)
        # Peso por día del mes
        if 1 <= d.day <= 15:
            day_of_month_weight = 1.5 # Ajusta a tus preferencias
        else:
            day_of_month_weight = 0.7
        final_weight = base_weight * festive_weight * day_of_month_weight
        if random.uniform(0, 2) < final_weight:
            return d
def numero_lineas_orden():
    """
    Retorna cuántos productos tendrá la orden,
    con la siguiente distribución:
    - 60% de probabilidad => 1 o 2 productos
    - 30% => 3 productos
    - 10% => 4 o 5 productos
    """
    r = random.random()
    if r < 0.5:
        # 60% de probabilidad de que sea 1 o 2
        return random.choice([1, 2])
    elif r < 0.9:
        # del 0.6 al 0.9 => 30% para 3
        return 3
    else:
        # del 0.9 al 1 => 10% para 4 o 5
        return random.choice([4, 5])
# Creamos la distribución de probabilidad para todas las sucursales
prob_sucursales = {}

```

```

peso_por_marketing = prob_marketing_sellers / len(marketing_sellers)
peso_por_normal = prob_normal_sellers / len(normal_sellers)
for suc in marketing_sellers:
    prob_sucursales[suc] = peso_por_marketing
for suc in normal_sellers:
    prob_sucursales[suc] = peso_por_normal
# Función auxiliar para determinar si el cliente compra en su sucursal preferida
def escoger_sucursal(idx_cliente):
    """Elige la sucursal según el perfil de fidelidad."""
    tipo = perfil_cliente[idx_cliente]["tipo"]
    if tipo == "fiel_sucursal":
        return perfil_cliente[idx_cliente]["sucursal_preferida"]
    else:
        # Caso no sea fiel a sucursal -> elegimos sucursal según la distribución general
        r = random.random()
        cumulative = 0.0
        for suc in prob_sucursales:
            cumulative += prob_sucursales[suc]
            if r < cumulative:
                return suc
        # fallback
        return random.choice(sucursales)
# Función auxiliar para escoger un producto según el perfil
def escoger_productos(idx_cliente, n_productos):
    """Devuelve una lista de n_productos (strings) con la lógica de preferencia."""
    tipo = perfil_cliente[idx_cliente]["tipo"]
    if tipo == "fiel_productos":
        # Escogemos principalmente de sus productos preferidos
        prods_pref = perfil_cliente[idx_cliente]["productos_preferidos"]
        # Damos alta probabilidad a los preferidos, pero permitimos cierto mix:
        productos_seleccionados = []
        for _ in range(n_productos):
            if random.random() < 0.7: # 70% chances de que escoja uno de sus preferidos
                productos_seleccionados.append(random.choice(prods_pref))
            else:
                productos_seleccionados.append(random.choice(lista_productos))
        return productos_seleccionados
    else:
        # "fiel_sucursal" o "variado" -> escoge productos sin preferencia especial
        # (puedes añadir matices si deseas)
        return random.choices(lista_productos, k=n_productos)
# Para la cantidad vendida de cada producto, utilizamos Poisson (lambda=2, por ej.)
# y limitamos a un máximo de 10, para no exagerar.
def cantidad_vendida_poisson(lam=2):
    q = np.random.poisson(lam)
    if q < 1:
        q = 1
    elif q > 10:
        q = 10
    return q
# Para la frecuencia de órdenes por cliente en 6 meses,
# podemos usar una Binomial Negativa (o Poisson). Para simplificar,
# usaremos una Poisson con media ~ 10 compras en 6 meses (arbitrario).
def estimar_ordenes_por_cliente():
    # Ejemplo: Poisson con lambda=10

```

```

    return np.random.poisson(10)
# 1) Distribución "normal"
rating_choices_normal = [1, 2, 3, 4, 5]
rating_probs_normal = [0.05, 0.10, 0.15, 0.30, 0.40] # 70% en 4 o 5
# 2) Distribución "mejorada"
rating_choices_better = [1, 2, 3, 4, 5]
rating_probs_better = [0.02, 0.03, 0.10, 0.35, 0.50] # 85% en 4 o 5
# Definimos cuántos productos tendrán la distribución "mejorada"
k_mejorados = 10
productos_mejor_rating = random.sample(lista_productos, k_mejorados)
def obtener_rating_para_producto(prod):
    """Devuelve un rating usando la dist. mejorada si 'prod' está en productos_mejor_rating."""
    if prod in productos_mejor_rating:
        return np.random.choice(rating_choices_better, p=rating_probs_better)
    else:
        return np.random.choice(rating_choices_normal, p=rating_probs_normal)
# Listas de comentarios
comentarios_positivos = [
    "Estan deliciosas. Muy ricas, me encantaron.",
    "Las galletas son muy buenas, las mas ricas.",
    "Que deliciosos brownies. Los mejores",
    "Todo esta muy bueno, muy rico",
    "Las galletas son frescas y crujientes, deliciosas",
    "Estan buenisimos. Me encantaron.",
    "Muy ricas, las mejores galletas.",
    "Increible. Todo esta delicioso.",
    "Los brownies están muy buenos, muy suaves",
    "Las mejores galletas. Muy ricas.",
    "Todo esta muy rico, me encanta.",
    "Que buenas. Los mejores brownies",
    "Deliciosos, muy suaves y ricos.",
    "Las galletas estan frescas, muy ricas",
    "Todo esta muy sabroso, muy recomendado",
    "Todo riquisimo. Los brownies son geniales.",
    "Muy buena calidad, todo estaba delicioso.",
    "Las galletas son las mejores, estan deliciosas",
    "Los brownies son muy buenos, super ricos",
    "Increible. Me encantaron las galletas."
]
comentarios_negativos = [
    "Las galletas venian frias, no me gustaron.",
    "El brownie estaba un poco seco, no me convencio.",
    "Las galletas estaban duras, no me gustaron mucho.",
    "El sabor estaba bien, pero venian frios.",
    "Me esperaba algo mas suave, pero estaban muy duros.",
    "Las galletas se quebraron, no eran lo que esperaba.",
    "Los brownies no estaban frescos, muy duros.",
    "Las galletas no tenían mucho sabor.",
    "El brownie estaba frio y seco, no me gusto.",
    "Las galletas venian rotas, no estaban bien.",
    "El sabor es bueno, pero estaban frias.",
    "Me decepcionaron, no eran lo que esperaba.",
    "Las galletas venian crudas.",
    "El brownie estaba muy seco, esperaba algo mas suave.",
    "Estaban crudas."
]

```

```

    "Las galletas estaban muy secas.",
    "Estaban frias y un poco duras",
    "El brownie estaba quemado"
]
# Probabilidad de que una orden tenga comentario
prob_comentario = 0.20 # 20%
# De las órdenes que sí llevan comentario: 70% positivos, 30% negativos
prob_positivo = 0.60
# 6. MÉTODO DE ENTREGA
# Quieres 65% Delivery y 35% Pickup
opciones_entrega = ["Delivery", "Pickup"]
prob_entrega = [0.65, 0.35]
def obtener_metodo_entrega():
    """Retorna 'Delivery' (65%) o 'Pickup' (35%)."""
    return np.random.choice(opciones_entrega, p=prob_entrega)
# 7. CATEGORÍA DE PRODUCTOS
# Boxes: Box 1, Box 2, Box 3, Box 4, Box 5, Box 6
# Cookies: Cookie 1..12
# Varios: Producto 1..5
def obtener_categoria_producto(prod):
    if prod.startswith("Box"):
        return "Boxes"
    elif prod.startswith("Cookie"):
        return "Cookies"
    else:
        # Todo lo demás (Producto 1..5)
        return "Varios"
# -----
# Generamos 25k órdenes. Para cada orden, elegimos un cliente al azar
# (pero con una prob. ponderada en base a cuántas órdenes se espera que tengan).
# -----
# Primero estimamos cuántas órdenes se espera por cliente.
ordenes_esperadas = [estimar_ordenes_por_cliente() for _ in range(num_clientes)]
total_ordenes_esperadas = sum(ordenes_esperadas)

# Normalizamos para sacar probabilidad de que "toque" a cada cliente
ordenes_prob_cliente = [o / total_ordenes_esperadas for o in ordenes_esperadas]
all_orders = [] # almacenará (id_cliente, ...)

for i in range(NUM_ORDENES):
    # Elegimos un cliente según la prob
    idx_cliente = np.random.choice(df_clientes.index, p=ordenes_prob_cliente)
    all_orders.append(idx_cliente)
# Para cada orden, definimos (una sola vez) si hay comentario y cuál es
order_comments = {}
order_delivery = {}
import datetime
# Fecha de inicio de la simulación
start_date = datetime.date(2025, 4, 1)
# Tasa de crecimiento mensual (por ejemplo, 3%)
monthly_growth = 0.03
# Ahora, para cada orden, generamos de 1 a MAX_PRODUCTOS_POR_ORDEN lineas
# y las agregamos a la base transaccional
for idx_cliente in all_orders:
    orden_id_counter += 1

```

```

# Decidir si esta orden tendrá comentario
if random.random() < probab_comentario:
    # 70% de que sea positivo, 30% negativo
    if random.random() < probab_positivo:
        comentario = random.choice(comentarios_positivos)
    else:
        comentario = random.choice(comentarios_negativos)
else:
    comentario = None
# Decidir método de entrega
metodo_entrega = obtener_metodo_entrega()
order_comments[orden_id_counter] = comentario
order_delivery[orden_id_counter] = metodo_entrega
num_lineas = numero_lineas_orden()
# Escogemos la fecha y la hora
fecha_orden = random_date_in_range()
# Hora
hora_compra, minuto_compra = escoger_hora_compra()
# Sucursal
suc = escoger_sucursal(idx_cliente)
# Productos de la orden
prods = escoger_productos(idx_cliente, num_lineas)
# 1) Escoge la fecha de la orden:
fecha_orden = random_date_in_range()
# 2) Calcula meses transcurridos con el nombre correcto de variable:
months_since_start = (
    (fecha_orden.year - start_date.year) * 12
    + (fecha_orden.month - start_date.month)
)

# 3) Define tu tasa de crecimiento mensual
monthly_growth = 0.03
# 4) Calcula el multiplicador de tendencia
trend_multiplier = (1 + monthly_growth) ** months_since_start
for p in prods:
    # Cantidad vendida
    cant = cantidad_vendida_poisson(lam=2)
    # Asignamos el precio base del producto
    precio_unit = productos_precios[p] # Precio base directamente sin descuentos
    # Total de la venta para este producto
    total_line = round(precio_unit * cant * trend_multiplier, 2)
    # Rating
    rating = obtener_rating_para_producto(p)
    # Agregamos registro
# Categoría
categoria = obtener_categoria_producto(p)
transacciones_data.append([
    fecha_orden,
    fecha_orden.strftime("%A"), # día de la semana
    orden_id_counter,
    df_clientes.loc[idx_cliente, "telefono"],
    suc,
    p,
    precio_unit,

```

```

        cant,
        total_line,
        f"{hora_compra:02d}:{minuto_compra:02d}",
        rating,
        comentario,
        categoria,# Comentario (puede ser None)
        metodo_entrega, # Nuevo: método de entrega
            # Nuevo: categoría del producto
    ])
df_transacciones = pd.DataFrame(
    transacciones_data,
    columns=[
        "fecha",
        "dia_semana",
        "numero_orden",
        "telefono_cliente",
        "sucursal",
        "producto",
        "precio",
        "cantidad_vendida",
        "venta_total",
        "hora_compra",
        "rating",
        "comentario",
        "categoria",
        "metodo_entrega",
    ]
)
# Es probable que tengamos más o menos de 50,000 filas. Ajustamos si necesitamos
if len(df_transacciones) > TARGET_NUM_LINES:
    df_transacciones = df_transacciones.sample(n=TARGET_NUM_LINES, random_state=42).reset_index(drop=True)
elif len(df_transacciones) < TARGET_NUM_LINES:
    # Si fuera importante llegar a EXACTO, podríamos regenerar algunas,
    # pero para el ejemplo nos quedamos así.
    df_transacciones.reset_index(drop=True, inplace=True)
# Ordenamos por fecha y número de orden, solo para tener cierto orden lógico
df_transacciones.sort_values(by=["fecha", "numero_orden"], inplace=True)
df_transacciones.reset_index(drop=True, inplace=True)
# -----
# 5. GENERAR LA BASE DE CANJES (Redenciones de Puntos)
# -----
# Cada Q1 gastado = 1 punto.
# Los clientes pueden canjear:
# Cookie 1 (400 puntos)
# Cookie 3 (320 puntos)
# Box 4 (480 puntos)
# Cookie 6 (320 puntos)
# Cookie 10 (380 puntos)
# Cookie 7 (430 puntos)
# Solo se canjea si el cliente tiene puntos suficientes.
redeemable_products = {
    "Cookie 1": 400,
    "Cookie 3": 320,
    "Box 4": 480,
    "Cookie 6": 320,

```

```

"Cookie 10": 380,
"Cookie 7": 430
}

# Para simular canjes:
# 1) Agrupamos transacciones por cliente, en orden cronológico.
# 2) Acumulamos puntos y, después de cada compra, existe una probabilidad
# de que el cliente decida canjear si tiene suficientes puntos para algún producto.
df_transacciones_sorted = df_transacciones.sort_values(
    by=["telefono_cliente", "fecha", "numero_orden"]
).reset_index(drop=True)
redenciones_data = []
def intentar_canjear_puntos(telefono_cliente, fecha_c, sucursal, puntos_acumulados):
    """
    Dada la cantidad de puntos_acumulados, intentamos canjear con cierta probabilidad.
    Retorna (producto_canjeado, nuevos_puntos).
    Si no se canjea, retorna (None, puntos_acumulados).
    """
    # Probabilidad fija de 10% de intentar un canje tras cada compra.
    if random.random() < 0.10:
        # Verificar qué productos puede costear
        affordables = [p for p, cost in redeemable_products.items() if cost <= puntos_acumulados]
        if not affordables:
            return (None, puntos_acumulados)
        # Escoge uno al azar
        chosen = random.choice(affordables)
        cost_chosen = redeemable_products[chosen]
        new_points = puntos_acumulados - cost_chosen
        # Registramos el canje
        redenciones_data.append([
            telefono_cliente,
            fecha_c,
            sucursal,
            chosen
        ])
        return (chosen, new_points)
    else:
        return (None, puntos_acumulados)
# Iteramos por cliente
grouped = df_transacciones_sorted.groupby("telefono_cliente")
for tel, group in grouped:
    puntos = 0
    for i, row in group.iterrows():
        # Suma de puntos = total gastado en esa línea
        puntos += row["venta_total"]
        # Intentar canje
        _producto_canjeado, puntos = intentar_canjear_puntos(
            tel, row["fecha"], row["sucursal"], puntos
        )
# Creamos la DataFrame de redenciones
df_redenciones = pd.DataFrame(
    redenciones_data,
    columns=["numero_cliente", "fecha", "sucursal", "producto_canjeado"]
)
# Ordenamos por fecha

```

```

df_redenciones.sort_values(by=["numero_cliente", "fecha"], inplace=True)
df_redenciones.reset_index(drop=True, inplace=True)
print("Cantidad de registros de canjes generados:", len(df_redenciones))
#!/usr/bin/env python
# coding: utf-8
# -----
# 5. RESULTADOS
# -----
print("Cantidad de clientes generados:", len(df_clientes))
print("Cantidad de registros transaccionales:", len(df_transacciones))
# Veamos algunos ejemplos
print("\n--- EJEMPLOS DE LA BASE DE CLIENTES ---")
print(df_clientes.head(5))
print("\n--- EJEMPLOS DE LA BASE TRANSACCIONAL ---")
print(df_transacciones.head(10))
# Si deseas, guarda a CSV
df_clientes.to_csv("C:/Users/Usuario/OneDrive - lulisgt/Documentos/TESIS/base_clientes.csv", index=False)
df_transacciones.to_csv("C:/Users/Usuario/OneDrive - lulisgt/Documentos/TESIS/base_transacciones.csv",
index=False)
df_redenciones.to_csv("C:/Users/Usuario/OneDrive -lulisgt/Documentos/TESIS/base_redenciones.csv", index=False)

```

## Anexo VII: Código para la exploración de datos

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df_clientes = pd.read_csv("C:/Users/Usuario/OneDrive - lulisgt/Documentos/TESIS/base_clientes.csv")
df_trans = pd.read_csv("C:/Users/Usuario/OneDrive - lulisgt/Documentos/TESIS/base_transacciones.csv")
df_red = pd.read_csv("C:/Users/Usuario/OneDrive - lulisgt/Documentos/TESIS/base_redenciones.csv")
df_trans["fecha"] = pd.to_datetime(df_trans["fecha"], errors="coerce")
# extraer hora
df_trans['hora'] = df_trans['hora_compra'].str.split(':').str[0].astype(int)
# dia de la semana en número
map_day = {
    'Monday': 0, 'Tuesday': 1, 'Wednesday': 2, 'Thursday': 3,
    'Friday': 4, 'Saturday': 5, 'Sunday': 6
}
df_trans['dia_semana_num'] = df_trans['dia_semana'].map(map_day)
#fecha
df_trans['fecha'] = pd.to_datetime(df_trans['fecha'], errors='coerce')
#Sacar la edad de clientes
df_clientes['fecha_cumple'] = pd.to_datetime(df_clientes['fecha_cumple'], errors='coerce')
fecha_ref = df_trans['fecha'].max() # Tomamos la última fecha de transacciones como referencia
df_clientes['edad'] = ((fecha_ref - df_clientes['fecha_cumple']).dt.days // 365)
# unir edad a la cuadro
df_combinado = pd.merge(
    df_trans,
    df_clientes[['telefono', 'edad']],
    left_on='telefono_cliente',
    right_on='telefono',
    how='left'
)
# extraer hora
df_trans['hora'] = df_trans['hora_compra'].str.split(':').str[0].astype(int)
# dia de la semana en número
map_day = {
    'Monday': 0, 'Tuesday': 1, 'Wednesday': 2, 'Thursday': 3,
    'Friday': 4, 'Saturday': 5, 'Sunday': 6
}
df_trans['dia_semana_num'] = df_trans['dia_semana'].map(map_day)
#fecha
df_trans['fecha'] = pd.to_datetime(df_trans['fecha'], errors='coerce')
#Sacar la edad de clientes
df_clientes['fecha_cumple'] = pd.to_datetime(df_clientes['fecha_cumple'], errors='coerce')
fecha_ref = df_trans['fecha'].max() # Tomamos la última fecha de transacciones como referencia
df_clientes['edad'] = ((fecha_ref - df_clientes['fecha_cumple']).dt.days // 365)
# unir edad a la cuadro
df_combinado = pd.merge(
    df_trans,
    df_clientes[['telefono', 'edad']],
    left_on='telefono_cliente',
    right_on='telefono',
    how='left'
```

```

)
from matplotlib.colors import LinearSegmentedColormap
#variables
df_corr = df_combinado[[
    'precio',
    'cantidad_vendida',
    'venta_total',
    'dia_semana_num',
    'hora',
    'edad'
]].copy()
df_corr.dropna(inplace=True)
# corplot
corr_matrix = df_corr.corr()
single_color_cmap = LinearSegmentedColormap.from_list(
    "rosado_suave",
    ["#f6cbff", "#9525ad"],
    N=256
)

# Crear el heatmap
plt.figure(figsize=(6, 5))
sns.heatmap(
    corr_matrix,
    annot=True,
    cmap=single_color_cmap,
    vmin=0,
    vmax=1
)
plt.title("Matriz de Correlación", loc = "left", pad=30)
plt.show()
#distribución de edades
plt.figure(figsize=(10, 6))
sns.histplot(df_clientes['edad'], bins=20, color='#9525ad', edgecolor=None)
plt.title('Distribución de Edades', fontsize=16, loc="left", pad=30)
plt.xlabel('Edad', fontsize=12)
plt.ylabel('Frecuencia', fontsize=12)
plt.grid(False)
plt.ylabel("")
for spine in plt.gca().spines.values():
    spine.set_visible(False)
plt.yticks([])
plt.show()
#Distribución por sexo
orden_sexo = df_clientes['sexo'].value_counts().sort_values(ascending=False).index
plt.figure(figsize=(8, 6))
ax = sns.countplot(
    x='sexo',
    data=df_clientes,
    order=orden_sexo, # Aplicar el orden de mayor a menor
    palette={'F': '#f03a8f', 'M': '#5b5bd7'}
)
plt.title('Distribución de Sexo', fontsize=16, loc="left", pad=30)
plt.xlabel('Sexo', fontsize=12)
plt.ylabel('Cantidad', fontsize=12)

```

```

plt.grid(False)

for spine in plt.gca().spines.values():
    spine.set_visible(False)
plt.yticks([])
#etiquets de datos
for p in ax.patches:
    height = p.get_height()
    ax.text(
        p.get_x() + p.get_width() / 2,
        height + 5,
        f'{int(height)}',
        ha='center', va='bottom', fontsize=11, color='black'
    )
plt.tight_layout()
plt.show()
#Ventas por sucursal
ventas_por_sucursal = df_trans.groupby("sucursal", as_index=False)["venta_total"].sum()
ventas_por_sucursal.sort_values("venta_total", ascending=False, inplace=True)

plt.figure(figsize=(10, 6))
ax = sns.barplot(
    data=ventas_por_sucursal,
    x="sucursal",
    y="venta_total",
    color="#9525ad"
)
plt.title("Ventas Totales por Sucursal", fontsize=16, loc="left", pad=30)
plt.xlabel("Sucursal")
plt.ylabel("")
plt.yticks([])
plt.xticks(rotation=45)
plt.grid(False)
for spine in plt.gca().spines.values():
    spine.set_visible(False)
for p in ax.patches:
    height = p.get_height()
    ax.text(
        p.get_x() + p.get_width() / 2,
        height + 500,
        f"Q{int(height):,}",
        ha='center',
        va='bottom',
        fontsize=10,
        color='black'
    )

plt.tight_layout()
plt.show()
df_trans['fecha'] = pd.to_datetime(df_trans['fecha'])
#Ventas por mes
df_trans['año_mes'] = df_trans['fecha'].dt.to_period('M')
ventas_por_mes = df_trans.groupby('año_mes')['venta_total'].sum().reset_index()
ventas_por_mes['año_mes'] = ventas_por_mes['año_mes'].astype(str)
plt.figure(figsize=(10, 6))

```

```

plt.plot(
    ventas_por_mes['año_mes'],
    ventas_por_mes['venta_total'],
    marker='o',
    color='#9525ad',
    linewidth=2
)
plt.title('Ventas Totales por Mes', fontsize=16, loc="left", pad=30)
plt.xlabel('Mes', fontsize=12)
plt.ylabel('Ventas Totales', fontsize=12)
plt.xticks(rotation=45)
plt.grid(False) # Quitamos la cuadrícula

# Quitar solo los bordes del gráfico (no el eje Y)
for spine in plt.gca().spines.values():
    spine.set_visible(False)
plt.tight_layout()
plt.show()
df_trans["dia_del_mes"] = df_trans["fecha"].dt.day
df_trans["mes"] = df_trans["fecha"].dt.month
#ventas totales por mes y dia
df_agg = df_trans.groupby(["mes", "dia_del_mes"], as_index=False)["venta_total"].sum()
#grafico de dispersión
plt.figure(figsize=(10, 6))
sns.scatterplot(
    data=df_agg,
    x="dia_del_mes",
    y="venta_total",
    color="#9525ad",
    alpha=0.6,
    s=80
)
plt.title("Día del Mes vs. Ventas (por Mes)", fontsize=16, loc="left", pad=30)
plt.xlabel("Día del Mes")
plt.ylabel("Venta Total Agregada del Día")
plt.tight_layout()
plt.show()
ticket_data = df_trans['venta_total'].dropna()
#ticket promedio
plt.figure()
sns.histplot(ticket_data, kde=False, color='#9525ad', bins=30, edgecolor=None) # sin borde en barras
plt.title('Distribución de monto por factura', fontsize=16, loc="left", pad=30)
plt.xlabel("Monto del ticket (Q)")
plt.ylabel("")
for spine in plt.gca().spines.values():
    spine.set_visible(False)
plt.grid(False)
plt.yticks([])
plt.show()
# Resumen estadístico
print("Resumen estadístico del monto por factura")
summary_ticket = df_trans['venta_total'].describe(percentiles=[0.2, 0.4, 0.6, 0.8])
print(summary_ticket)
df_freq = df_trans.groupby("telefono_cliente")["numero_orden"].count().reset_index(name="frecuencia_compra")
#frecuencia de compra

```

```

plt.figure(figsize=(8, 5))
sns.histplot(df_freq["frecuencia_compra"], bins=22, color='#9525ad', edgecolor=None) # rosado oscuro, sin bordes
plt.title("Distribución de la Frecuencia de Compra (Órdenes por Cliente)", fontsize=16, loc="left", pad=30)
plt.xlabel("Número de Órdenes (Frecuencia de Compra)")
plt.ylabel("")
for spine in plt.gca().spines.values():
    spine.set_visible(False)
plt.yticks([])
plt.grid(False)
plt.show()
# Resumen estadístico
freq_summary = df_freq["frecuencia_compra"].describe(percentiles=[0.2, 0.4, 0.6, 0.8])
print("\n=== Resumen de la Frecuencia de Compra ===")
print(freq_summary)
#recencia, días desde la ultima compra
df_trans["fecha"] = pd.to_datetime(df_trans["fecha"])
max_date = df_trans["fecha"].max()
df_last_purchase = df_trans.groupby("telefono_cliente")["fecha"].max().reset_index(name="ultima_compra")
df_last_purchase["recencia_dias"] = (max_date - df_last_purchase["ultima_compra"]).dt.days
plt.figure(figsize=(8, 5))
sns.histplot(df_last_purchase["recencia_dias"], bins=30, color='#9525ad', edgecolor=None)
plt.title("Distribución de la Recencia (Días desde la Última Compra)", fontsize=16, loc="left", pad=30)
plt.xlabel("Días desde la última compra")
plt.ylabel("")
for spine in plt.gca().spines.values():
    spine.set_visible(False)
plt.yticks([])
plt.grid(False)
plt.show()
# Resumen estadístico
recencia_summary = df_last_purchase["recencia_dias"].describe(percentiles=[0.2, 0.4, 0.6, 0.8])
print("\n=== Resumen de la Recencia (días) ===")
print(recencia_summary)
#sucursales con mas redenciones
df_prod_stats = df_trans.groupby("producto", as_index=False).agg({
    "cantidad_vendida": "sum",
    "venta_total": "sum"
})
df_prod_stats.sort_values("venta_total", ascending=False, inplace=True)
# Gráfico
plt.figure(figsize=(8, 5))
ax = sns.barplot(
    data=df_prod_stats.head(10),
    x="producto",
    y="venta_total",
    color="#9525ad"
)
plt.title("Marketing 10 Productos por Ingreso (Venta Total)", fontsize=16, loc="left", pad=30)
plt.xlabel("Producto")
plt.ylabel("") # Quitar etiqueta del eje Y
plt.yticks([]) # Quitar los ticks del eje Y
plt.xticks(rotation=45)
plt.grid(False)
for spine in plt.gca().spines.values():
    spine.set_visible(False)

```

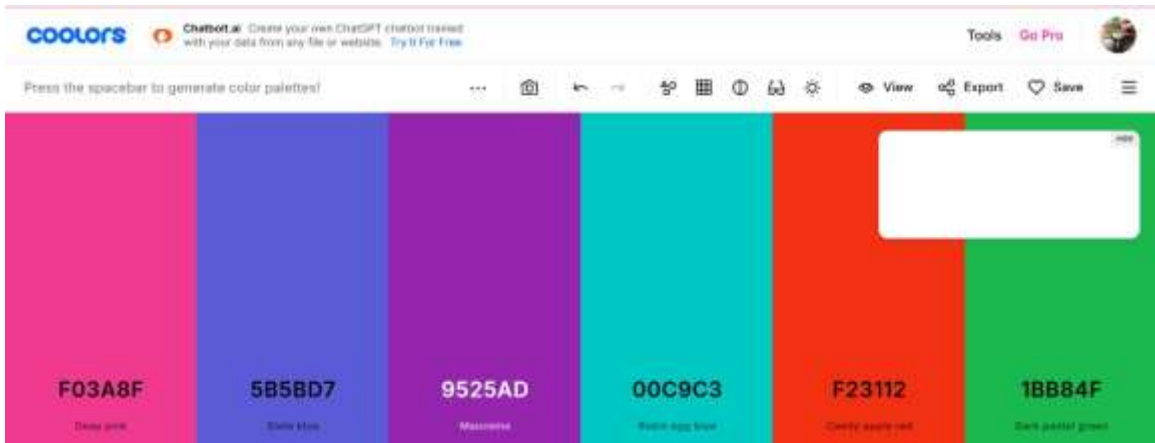
```

for p in ax.patches:
    height = p.get_height()
    ax.text(
        p.get_x() + p.get_width() / 2,
        height + 500,
        f"Q{int(height):,}",
        ha='center',
        va='bottom',
        fontsize=10,
        color='black'
    )

plt.tight_layout()
plt.show()
#productos más redimidos
df_prod_canje = df_red.groupby("producto_canjeado",
as_index=False).size().rename(columns={"size":"conteo_canjes"})
df_prod_canje.sort_values("conteo_canjes", ascending=False, inplace=True)
plt.figure(figsize=(8, 5))
ax = sns.barplot(
    data=df_prod_canje,
    x="producto_canjeado",
    y="conteo_canjes",
    color="#9525ad"
)
plt.title("Productos Más Canjeados", fontsize=16, loc="left", pad=30)
plt.xlabel("Producto Canjeado")
plt.ylabel("")
plt.yticks([]) # Ocultar valores del eje Y
plt.xticks(rotation=45)
plt.grid(False)
for spine in plt.gca().spines.values():
    spine.set_visible(False)
for p in ax.patches:
    height = p.get_height()
    ax.text(
        p.get_x() + p.get_width() / 2,
        height + 1,
        f"{int(height)}",
        ha='center',
        va='bottom',
        fontsize=10,
        color='black'
    )
plt.tight_layout()
plt.show(

```

Anexo VIII: Paleta de colores para el tablero



Anexo IX: Creación del tema para los tableros



## Anexo X: Relaciones entre los cuadros en Power BI



## Anexo XI: Prompt para generación de cuadro de meta de ventas

“Se recomienda que el siguiente paso, antes de lanzar oficialmente el programa de lealtad sea realizar una proyección financiera, tomando en cuenta tanto los ingresos esperados como los gastos e inversiones que se deberán hacer para llevar a cabo el proyecto. Esta proyección permitirá analizar la rentabilidad estimada del programa y ajustar la estrategia si fuera necesario para garantizar su viabilidad económica.

Es fundamental que las acciones de *marketing* no se realicen de manera generalizada, por eso se recomienda generar estrategias dirigidas a segmentos específicos identificados mediante el modelo RFM y la construcción de *buyer personas*. Por ejemplo; se recomienda diseñar una campaña de reactivación para el segmento “No podemos perder”, ofreciendo beneficios exclusivos dentro de los canales propios, una campaña que genere un sentido de urgencia con mensajes como: “¡Solo por hoy!”. Además de comunicar un mensaje como: “Te extrañamos”, para reactivar el vínculo emocional que el cliente ya había establecido con la marca

Se recomienda Invertir los esfuerzos de *marketing* en segmentos de alto impacto, priorizando la inversión de recursos y campañas en los segmentos que representen un peso significativo tanto en las ventas como en el volumen de transacciones, basándose en los análisis realizados a través del *dashboard*. De esta manera, se podrá maximizar el retorno de la inversión enfocándose en los perfiles de clientes que generan mayor valor para el negocio.

Debido a que el comportamiento de los clientes la mayoría de veces puede variar con el tiempo, se sugiere actualizar los análisis de segmentación y perfilación de forma periódica, utilizando la nueva información recolectada a través del programa de fidelización una vez esté en marcha. Esto permitirá mantener estrategias relevantes y adaptadas a la evolución del mercado y las preferencias de los consumidores.”

Anexo XII: Prompt para generar la simulación de gastos de *marketing*

“estas son las ventas mensuales reales de mi data:

Año	Mes	Suma de venta_total	sucursal
2025	abril	Q65436.15	Sucursal 1
2025	abril	Q69658.05	Sucursal 2
2025	abril	Q65574.80	Sucursal 3
2025	abril	Q70178.70	Sucursal 4
2025	abril	Q46725.80	Sucursal 5
2025	abril	Q44829.75	Sucursal 6
2025	abril	Q41491.55	Sucursal 7
2025	abril	Q45842.85	Sucursal 8
2025	abril	Q45512.95	Sucursal 9
2025	mayo	Q84463.61	Sucursal 1
2025	mayo	Q73419.77	Sucursal 2
2025	mayo	Q78562.64	Sucursal 3
2025	mayo	Q76177.56	Sucursal 4
2025	mayo	Q49220.65	Sucursal 5
2025	mayo	Q52873.14	Sucursal 6
2025	mayo	Q53748.24	Sucursal 7
2025	mayo	Q51225.58	Sucursal 8
2025	mayo	Q54759.86	Sucursal 9
2025	junio	Q71114.77	Sucursal 1
2025	junio	Q80210.00	Sucursal 2
2025	junio	Q73897.01	Sucursal 3
2025	junio	Q73427.71	Sucursal 4
2025	junio	Q51182.52	Sucursal 5
2025	junio	Q42479.75	Sucursal 6
2025	junio	Q47841.72	Sucursal 7
2025	junio	Q60242.84	Sucursal 8
2025	junio	Q46493.10	Sucursal 9
2025	julio	Q79955.49	Sucursal 1
2025	julio	Q78572.00	Sucursal 2
2025	julio	Q75350.12	Sucursal 3
2025	julio	Q78072.12	Sucursal 4
2025	julio	Q49871.15	Sucursal 5
2025	julio	Q52331.69	Sucursal 6
2025	julio	Q44721.40	Sucursal 7
2025	julio	Q50013.59	Sucursal 8
2025	julio	Q52029.08	Sucursal 9
2025	agosto	Q78647.18	Sucursal 1
2025	agosto	Q84136.94	Sucursal 2
2025	agosto	Q84286.75	Sucursal 3
2025	agosto	Q86447.13	Sucursal 4
2025	agosto	Q51725.28	Sucursal 5

2025	agosto	Q64852.22	Sucursal 6
2025	agosto	Q57194.68	Sucursal 7
2025	agosto	Q56362.80	Sucursal 8
2025	agosto	Q48982.74	Sucursal 9
2025	septiembre	Q75586.16	Sucursal 1
2025	septiembre	Q72086.19	Sucursal 2
2025	septiembre	Q73605.45	Sucursal 3
2025	septiembre	Q81387.85	Sucursal 4
2025	septiembre	Q55747.27	Sucursal 5
2025	septiembre	Q53343.88	Sucursal 6
2025	septiembre	Q46130.84	Sucursal 7
2025	septiembre	Q47595.19	Sucursal 8
2025	septiembre	Q50413.42	Sucursal 9
2025	octubre	Q88069.19	Sucursal 1
2025	octubre	Q90139.66	Sucursal 2
2025	octubre	Q90755.54	Sucursal 3
2025	octubre	Q97297.48	Sucursal 4
2025	octubre	Q49979.01	Sucursal 5
2025	octubre	Q57044.17	Sucursal 6
2025	octubre	Q61834.38	Sucursal 7
2025	octubre	Q72190.63	Sucursal 8
2025	octubre	Q59301.77	Sucursal 9
2025	noviembre	Q79801.60	Sucursal 1
2025	noviembre	Q87325.54	Sucursal 2
2025	noviembre	Q83739.44	Sucursal 3
2025	noviembre	Q91791.64	Sucursal 4
2025	noviembre	Q52583.93	Sucursal 5
2025	noviembre	Q60356.60	Sucursal 6
2025	noviembre	Q51938.72	Sucursal 7
2025	noviembre	Q58453.34	Sucursal 8
2025	noviembre	Q62439.70	Sucursal 9
2025	diciembre	Q96635.07	Sucursal 1
2025	diciembre	Q95603.71	Sucursal 2
2025	diciembre	Q91520.06	Sucursal 3
2025	diciembre	Q99707.64	Sucursal 4
2025	diciembre	Q58780.67	Sucursal 5
2025	diciembre	Q61892.13	Sucursal 6
2025	diciembre	Q51839.61	Sucursal 7
2025	diciembre	Q62053.01	Sucursal 8
2025	diciembre	Q51689.72	Sucursal 9
2026	enero	Q104655.22	Sucursal 1
2026	enero	Q97192.56	Sucursal 2
2026	enero	Q90115.45	Sucursal 3
2026	enero	Q90671.07	Sucursal 4
2026	enero	Q71250.28	Sucursal 5

2026	enero	Q68758.05	Sucursal 6
2026	enero	Q71212.81	Sucursal 7
2026	enero	Q58615.51	Sucursal 8
2026	enero	Q56921.09	Sucursal 9
2026	febrero	Q95337.12	Sucursal 1
2026	febrero	Q97276.21	Sucursal 2
2026	febrero	Q94469.61	Sucursal 3
2026	febrero	Q92825.49	Sucursal 4
2026	febrero	Q60232.91	Sucursal 5
2026	febrero	Q53568.31	Sucursal 6
2026	febrero	Q55117.29	Sucursal 7
2026	febrero	Q55685.30	Sucursal 8
2026	febrero	Q60088.94	Sucursal 9
2026	marzo	Q103485.76	Sucursal 1
2026	marzo	Q106440.09	Sucursal 2
2026	marzo	Q103714.90	Sucursal 3
2026	marzo	Q90520.59	Sucursal 4
2026	marzo	Q55233.27	Sucursal 5
2026	marzo	Q62107.06	Sucursal 6
2026	marzo	Q50506.41	Sucursal 7
2026	marzo	Q76153.83	Sucursal 8
2026	marzo	Q63940.05	Sucursal 9

ahora necesito simular una base de datos con las metas de ventas mensuales por sucursal en el periodo de tiempo que se tiene ue es del 1 de abril al 31 de marzo, las metas de cada sucursal adecúalas según sus ventas mensuales, no tienen que ir siempre por arriba, puede que esten parecidas, más bajas o así haz esa variación.

