

Mejora del Proceso de Casos de Uso Mediante la Segmentación Estratégica

Monserrat Baeza Muñoz

Ingeniería Civil Industrial

14 de noviembre de 2023

Resumen Ejecutivo

En la actualidad, el ecommerce se hace cada vez más presente en la industria minorista. Es por esto que la empresa Walmart pone especial atención a los indicadores claves en este tipo de comercio. El presente informe de este proyecto se realiza con el fin de explicar cómo se llevó a cabo el proceso para potenciar estos indicadores, centrándose en mejorar el proceso de casos de uso a través de la implementación de análisis avanzados de datos enfocado en la generación de insights al equipo de envíos de correos electrónicos. El proyecto se llevó a cabo en el área de Marketing Operacional, encargada de optimizar procesos internos de gestión de Marketing e inversión en campañas, agilizando la entrega de valor al cliente omnicanal, basando la toma de decisiones en datos.

Para el desarrollo del proyecto se siguió la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), comenzando por la comprensión del negocio donde se logró identificar la oportunidad de aumentar el CTOR (Click Through Open Rate) para mejorar la tasa de conversión de ecommerce a través de la mejora continua en la identificación y segmentación de la audiencia. Esta oportunidad se logra visualizar a través de los síntomas detectados que se atribuyen a que la persuasión deseada para realizar la acción en la estrategia de marketing no se logra debido a esta falta de eficacia en la comunicación. En base al análisis realizado y al problema detectado, se revelan las causas: falta de analítica avanzada, falta de análisis en métricas resultantes, elección de canal de contacto inadecuado y la falta de la elección adecuada del contenido enviado.

Una vez comprendido el negocio, se prosiguió con los demás pasos de la metodología, llevando a implementar la solución escogida. Las modificaciones de las estrategias de marketing basadas en los insights generados se realizaron en el equipo de Salesforce, que es responsable de la comunicación vía email. Es gracias a la acogida de estas recomendaciones que se logró una segmentación más efectiva de la audiencia y una mayor personalización del contenido gracias al enfoque en variables claves del comportamiento y preferencias del cliente. Durante la fase de implementación, los insights clave generados que se aplicaron fueron algunos como la búsqueda de clientes en diferentes segmentos de valor y la toma de decisiones basada en el comportamiento de compra en relación con el gasto mensual y el monto de boleta promedio en distintos horarios. Las métricas de conversión superaron las expectativas, obteniendo una tasa mensual atribuida del 0,09%. La migración de clientes físicos a omnicanales fue exitosa, generando una atribución mensual de \$5.912.315 y una diferencia mensual de gasto notable. El cierre del proyecto refleja un ROI del 13%, consolidando la rentabilidad tras la implementación de este. Gracias al proyecto, se considera aún más la mejora continua del proceso y las estrategias de crecimiento.

Abstract

Currently, e-commerce is becoming increasingly prevalent in the retail industry. That's why Walmart pays special attention to key indicators in this type of commerce. This report on this project is carried out to explain how the process was carried out to enhance these indicators, focusing on improving the use case process through the implementation of advanced data analysis focused on providing insights to the email delivery team. The project was carried out in the Operational Marketing area, responsible for optimizing internal marketing management processes and campaign investment, streamlining the delivery of value to the omnichannel customer, basing decision-making on data.

For the development of the project, the CRISP-DM methodology (Cross-Industry Standard Process for Data Mining) was followed, starting with understanding the business where the opportunity to increase CTOR (Click Through Open Rate) was identified to improve the e-commerce conversion rate through continuous improvement in audience identification and segmentation. This opportunity is visualized through the symptoms detected, attributed to the desired persuasion for action in the marketing strategy not being achieved due to a lack of effectiveness in communication.

Based on the analysis and the identified problem, the causes are revealed: lack of advanced analytics, lack of analysis on resulting metrics, inappropriate choice of contact channel, and the lack of appropriate content selection.

Once the business was understood, the methodology's other steps were followed, leading to the implementation of the chosen solution. Modifications to marketing strategies based on the generated insights were made in the Salesforce team, which is responsible for email communication. It is thanks to the adoption of these recommendations that more effective audience segmentation and greater content personalization were achieved through a focus on key customer behavior and preferences. During the implementation phase, key insights applied included searching for customers in different value segments and making decisions based on purchasing behavior in relation to monthly spending and average ticket amount at different times. Conversion metrics exceeded expectations, achieving an attributed monthly rate of 0.09%. The successful migration of physical to omnichannel customers generated a monthly attribution of \$5,912,315 and a notable monthly spending difference. The project's closure reflects a 13% ROI, consolidating profitability after its implementation. Thanks to the project, continuous improvement of the process and growth strategies is even more considered.

1. Índice

Resumen Ejecutivo.....	2
Abstract	3
2. Introducción	5
a) La empresa:	5
b) Formatos de la empresa:.....	5
c) Áreas de interés:	5
d) Competidores:	7
e) Casos de Uso:.....	7
f) Dolor y/o brecha:.....	8
g) Causas:	12
h) Oportunidad:	13
3. Objetivos	13
a) Objetivo SMART:	13
b) Objetivos específicos:	14
4. Metodología	15
5. Medidas de desempeño	16
6. Planificación.....	16
7. Estado del arte	17
8. Solución.....	18
c) Matriz de riesgos y mitigaciones.....	22
9. Evaluación económica.....	24
10. Desarrollo del proyecto	27
a) Fase 1: Preparación	27
b) Fase 2: Análisis y modelado de datos	29
c) Fase 3: Implementación y seguimiento	43
d) Fase 4: Evaluación y cierre:	43
11. Referencias.....	48
12. Anexos.....	49

2. Introducción

a) La empresa:

Walmart se destaca por ser una de las empresas con cadena minorista más grande del mundo, siendo una multinacional que se enfoca en la venta al por menor y que tiene tanto presencia física como presencia en el comercio electrónico. Desde sus inicios Walmart ha apostado por la estrategia de ofrecer a sus clientes precios bajos, impactando de manera directa en la vida de estos.

Walmart fue fundada en 1962 por Sam Walton en Arkansas, Estados Unidos. La primera tienda que abrió se llamó “Walton’s Five and Dime”, la cual se reconoció por otorgar precios bajos. En el año 1969, Walmart pasa a ser una empresa pública, cotizando en la Bolsa de Valores de Nueva York. En este año también abrió su primera tienda fuera de Arkansas, en Missouri. En la década de 1970, la empresa comienza a tener un gran crecimiento, expandiéndose por todo Estados Unidos. Luego, en la década de 1980, Sam Walton inserta la primera tienda de membresía al por mayor, llamándola “Sam’s Club”. En 1990, Walmart comienza a expandirse internacionalmente y logra ser el minorista más grande de EE. UU. En los años del 2000 se adquiere Walmart.com, por ende, la empresa comienza a operar en el comercio electrónico.

b) Formatos de la empresa:

Walmart posee diferentes formatos de sus tiendas, adaptándose a las necesidades de sus clientes, entre estos están: Líder, Express, SuperBodega aCuenta, Central Mayorista y Líder.cl.

c) Áreas de interés:

Walmart, como la mayoría de las empresas, cuenta con diversas áreas de trabajo, sin embargo, en el presente informe se dará especial atención al área de Marketing Operacional dado que es el área de relevancia para el proyecto propuesto. Se presenta el organigrama de Marketing y Clientes a modo de obtener una representación gráfica y general del área:

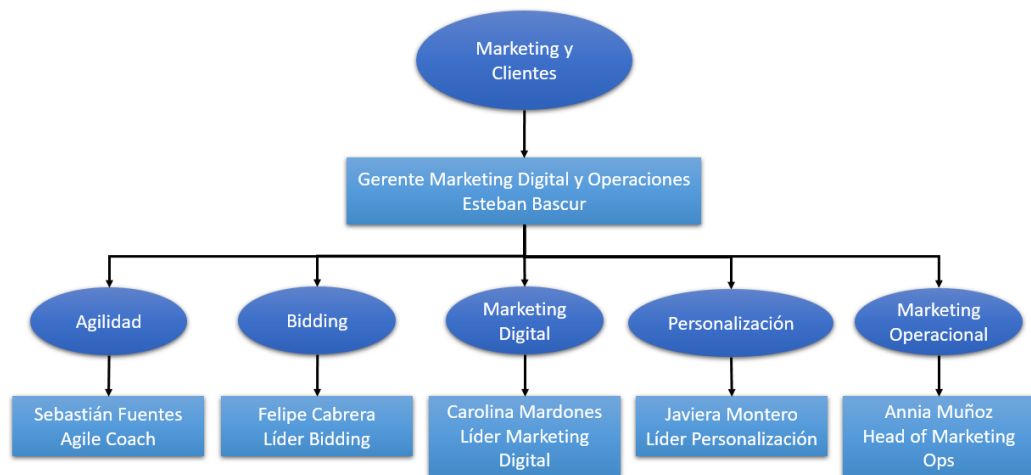


Imagen 1: Organigrama Marketing y Clientes.

El trabajo de Marketing Operacional se enfoca en la ejecución de las estrategias y la gestión de campañas de Marketing. Posee dos subáreas, por un lado, está Martech, que busca habilitar ecosistemas digitales optimizando procesos internos de gestión de Marketing e inversión en campañas, con el fin de agilizar la entrega de valor al cliente omnicanal, basando la toma de decisiones en datos. Por otro lado, se encuentra Control y Gestión que simplifica la gestión financiera de los distintos equipos de marketing a través de la optimización de procesos, lo cual a su vez otorga herramientas para la toma de decisiones.

A continuación, se muestra la estructura organizativa del área de Marketing Operacional en un organigrama, esto con el objetivo de clarificar los roles y relaciones dentro del equipo, de forma que se facilite la comprensión del proyecto a lo largo del informe:

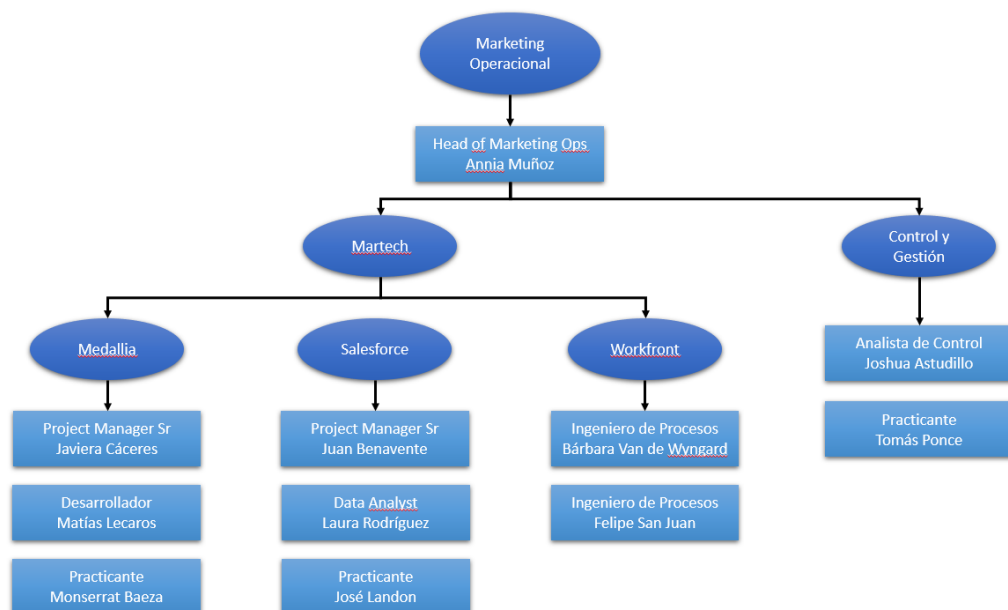


Imagen 2: Organigrama Marketing Operacional.

d) Competidores:

En Chile, las cadenas de supermercados más importantes del país como Walmart, Tottus, SMU y Unimarc, concentraron el 53% de las ventas minoristas del país el año 2022, según el Servicio Agrícola Exterior. En particular, Walmart tiene una cuota del 20%, convirtiéndolo en la cadena minorista más grande de Chile en cuanto a ingresos. En segundo lugar, le sigue Cencosud acaparando un 17,3%, luego SMU con un 11,1% y Tottus con un 4,1%. Comparando estos datos con el año 2021, se puede rescatar que este mercado presenta un crecimiento del 14,4% dado que estas cadenas de supermercados concentraban un 33,8% de las ventas minoristas del país en ese año.

e) Casos de Uso:

Para el proyecto del presente informe se seleccionó el proceso de casos de uso. El proceso de casos de uso es una estrategia utilizada en Marketing y Clientes que busca llegar al cliente con una comunicación personalizada, con el objetivo de que el cliente realice la acción planteada en la estrategia. Para dar mayor comprensión, se mostrará el flujo representado gráficamente y luego, se explicará cada etapa del proceso.

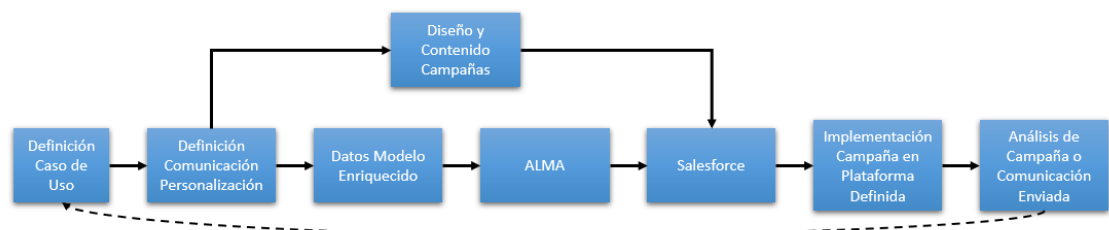


Imagen 3: Flujo Proceso Casos de Uso.

1. Definición de Casos de Uso:

En esta etapa participan los equipos de Kimun, Marketing Digital, Insights, Personalización, Comunicaciones y Experiencia de Cliente (contactabilidad). Estos equipos forman una mesa de trabajo en la cual se busca gestionar el caso de uso; tomando decisiones en base a datos (historia, comportamiento, kpis, entre otros) y definiendo el indicador que se analizará una vez implementada la campaña (¿Dónde quiero generar impacto?).

2. Definición comunicación-personalización:

Se define el diseño y el contenido para dar dirección al equipo de Diseño y Contenido. También se aterrizan los canales de comunicación luego de la evaluación de métricas del canal (Responde a las preguntas: ¿Qué quiero comunicar? ¿Cómo lo quiero comunicar? ¿Por dónde lo quiero comunicar?).

3. Se habilitan dos procesos:

a) Diseño y Contenido Campañas:

Se ejecutan las definiciones tomadas en el paso 2), comenzando a armar las gráficas del contenido por medio de la vía de comunicación elegida.

b) Datos/Modelos Enriquecidos:

Se realizan cruces de datos que segmentan a los clientes, logrando de esta forma la audiencia final.

4. ALMA:

Permite la disponibilidad de los datos de Walmart anonimizados, los cuales son extraídos desde el paso 3.b), a las plataformas disponibles para el envío de comunicación (ETL: Extracción, transformación y carga).

5. Salesforce:

Se encarga de la comunicación relacional con el cliente. Por un lado, envía la comunicación de mailing y notificaciones push y, por otro lado, disponibiliza la audiencia en medios pagados.

6. Implementación Campaña Plataforma Definida:

Ocurre en Salesforce o bien, en los medios pagados.

7. Análisis de Campaña o Comunicación Enviada:

Una vez implementada la comunicación, se evalúan las métricas definidas en el punto 1) y se analizan métricas transversales como sesiones, tasa de conversión, ROAS y la tasa de rebote. Por último, se retroalimenta el flujo según los resultados obtenidos de la vía de comunicación.

f) Dolor y/o brecha:

El análisis posterior a cada implementación de caso de uso nos brinda la oportunidad de identificar la brecha generada en el proceso, centrándonos especialmente en la tasa de conversión, que refleja el porcentaje de usuarios que llevan a cabo la acción deseada en el caso de uso implementado. Esta métrica, de constante medición para la empresa, es esencial para evaluar la efectividad de las estrategias y optimizar el uso de recursos. Los datos recopilados de enero a agosto de 2023 revelan que la tasa de conversión promedio en el comercio electrónico de Walmart es del 1,58%, mientras que la media en todas las industrias es del 2,1% (Asociación Nacional Avisadores de Chile, 2022). Esta brecha

sugiere ingresos potencialmente perdidos, una cuota de mercado reducida y, específicamente en este proyecto, una eficacia inferior en la implementación de casos de uso. Esto señala una oportunidad de crecimiento que no solo mejoraría la rentabilidad y eficiencia de los recursos, sino también la eficacia actual del trabajo de la empresa.

Adicionalmente a la comparación de tasas de conversión, se llevó a cabo un análisis de la situación actual de la empresa, utilizando el canal de comunicación por correo electrónico debido a su desarrollo avanzado en personalización y gobernanza de datos. Este análisis se basó en métricas clave como Emails enviados, Open Rate (OR), Click Through Rate (CTR) y Click Through Open Rate (CTOR).

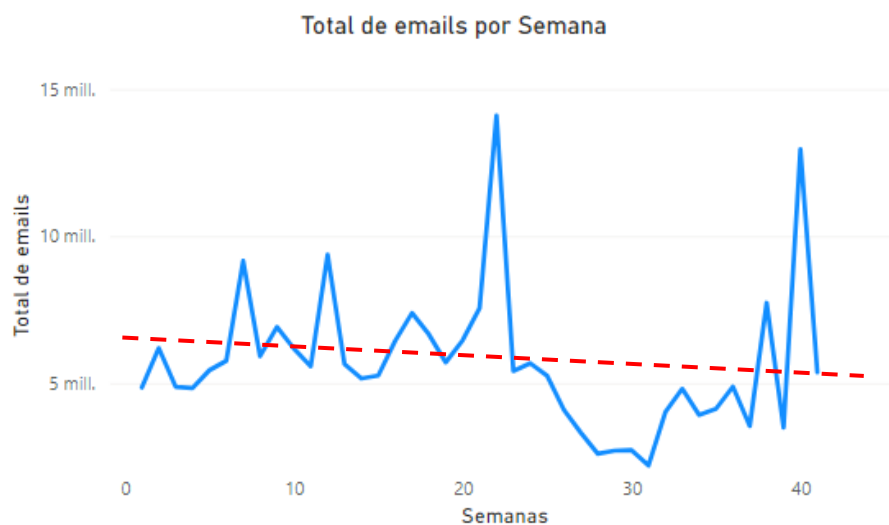


Gráfico 1: Mails enviados por semana 2023 (enero – octubre).

Del gráfico 1 que representa el total de emails enviados por semana se extrae que en las semanas 22 y 40 se enviaron más comunicaciones mientras que entre las semanas 27 y 37 hubo una baja en los envíos. La línea de tendencia indica que se están enviando menos correos, lo cual se debe al cambio en la estrategia de marketing que ahora busca mayor personalización y reducir el envío de mails no efectivos al cliente.

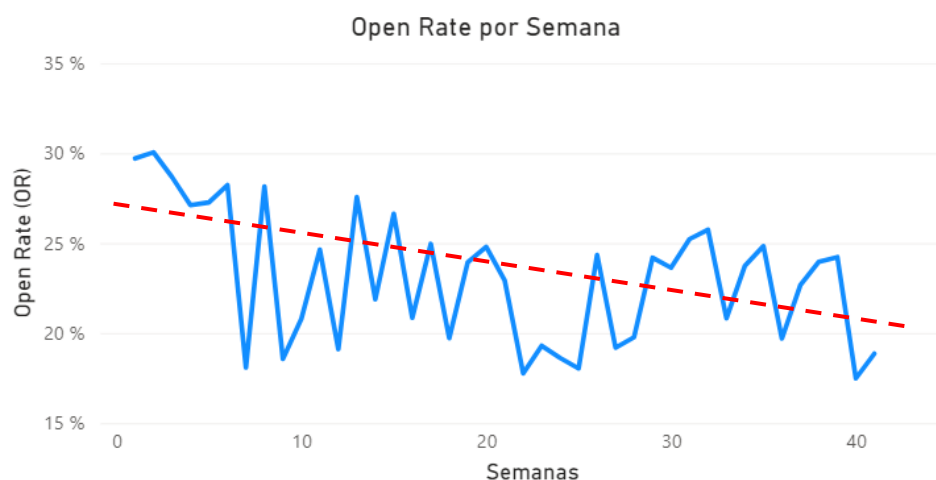


Gráfico 2: Open Rate por semana 2023 (enero – octubre).

Comparando el gráfico 1 con el gráfico 2, es decir, los emails enviados versus el Open Rate (OR) por semana, se puede notar que no hay indicios de que exista alguna relación entre ellas. La línea de tendencia tiene pendiente negativa lo cual puede significar que el usuario no está tan interesado en abrir los correos como antes.

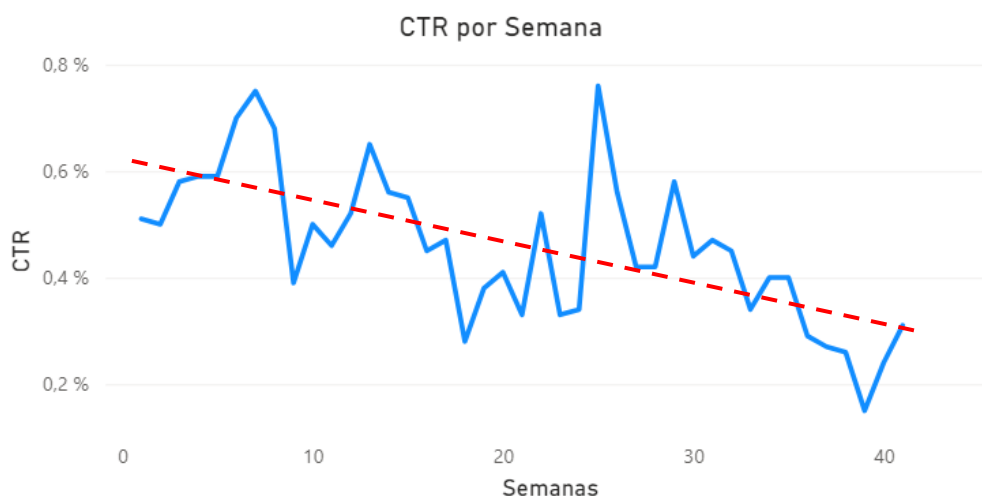


Gráfico 3: Click Through Rate (CTR) por semana 2023 (enero – octubre).

Del gráfico 3 correspondiente al CTR se puede decir que no presenta una relación clara con el gráfico 2 de Open Rate, por lo que parecen no estar relacionados. La tendencia es a la baja.

CTOR por Semana

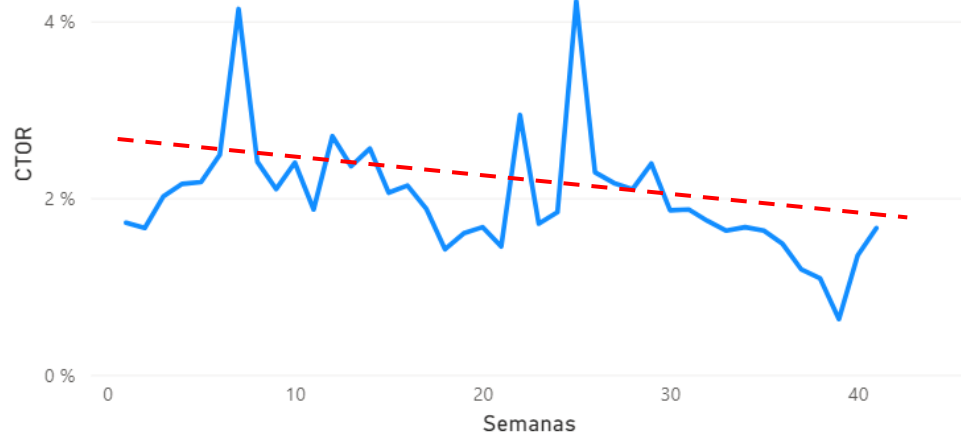


Gráfico 4: Click Through Open Rate (CTOR) por semana 2023 (enero – octubre).

Del gráfico 4 y el gráfico 3 se puede ver una similitud en sus comportamientos, pudiendo existir una relación entre ellos. La línea de tendencia es negativa.

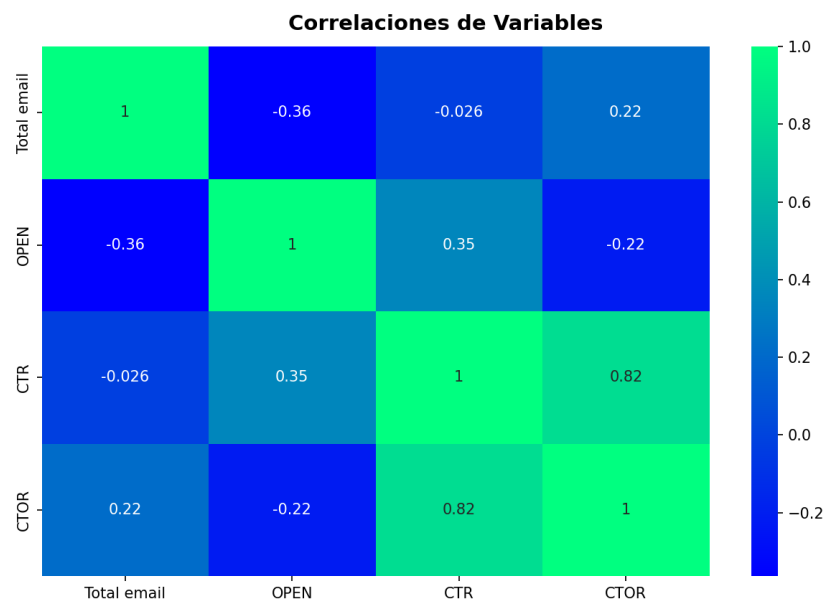


Gráfico 5: Correlaciones entre variables Total de emails, OR, CTR y CTOR.

El Gráfico 5 revela las correlaciones esperadas a partir de los análisis anteriores. Se destaca la alta correlación (0,82) entre CTR y CTOR, indicando una fuerte asociación entre estas métricas. Sin embargo, se subraya que la correlación no implica causalidad, y aunque estas métricas están estrechamente vinculadas, no se puede concluir que cambios en una causen cambios en la otra. Además, otras correlaciones entre variables no resultan significativas. Se observa una posible relación inversa entre la cantidad de correos

enviados y la apertura de estos, sugiriendo que el envío masivo puede no ser efectivo, abriendo la puerta a la efectividad de estrategias de envío más personalizadas.

El análisis de los gráficos proporciona insights sobre la respuesta del cliente a los correos enviados. Aunque un alto Open Rate indica interés inicial, tanto el CTR como el CTOR son considerablemente más bajos y menos afectados por variaciones. Esto sugiere que, aunque el usuario muestra interés al abrir el correo, no demuestra un fuerte deseo de interactuar con el contenido.

Se identifica el problema central: una brecha significativa en la tasa de conversión de la empresa en comparación con la industria, atribuible a una comunicación no efectiva con el cliente. La persuasión deseada para realizar la acción en la estrategia de marketing no se logra debido a esta falta de eficacia en la comunicación. Es fundamental recordar que, aunque se encuentren correlaciones, se debe evitar asumir automáticamente una relación causal directa.

g) Causas:

En base al análisis realizado y al problema detectado, se revelan las causas del problema:

1. Segmentación de audiencia:

La falta de implementación de analítica avanzada (análisis de comentarios del cliente, segmentación en grupos específicos, estudio del comportamiento a través de diversos factores, entre otros) ha llevado a la falta de identificación de clientes potenciales de manera eficaz.

2. Seguimiento y análisis:

La falta del análisis de métricas resultantes luego de la implementación del caso de uso limita la capacidad de obtener información valiosa para la mejora continua de la estrategia de marketing.

3. Canal de comunicación:

La elección del canal inadecuado sin considerar según su comportamiento y preferencias ha resultado en una falta de llegada efectiva a la audiencia deseada.

4. Personalización:

La falta de la elección adecuada del contenido según los intereses y necesidades del cliente ha afectado la capacidad de llegar de manera efectiva a dicho cliente.

h) Oportunidad:

La oportunidad detectada es el aumento del CTOR (Click Through Open Rate) para mejorar la tasa de conversión de ecommerce a través de la mejora continua en la identificación y segmentación de la audiencia.

En cuanto al impacto en el equipo se espera que el proyecto permita generar constantes recomendaciones respecto a la audiencia considerada en los distintos casos de uso para habilitar futuras iteraciones que permitan cumplir el objetivo planteado, logrando que el proceso siempre esté en mejora continua.

3. Objetivos

a) Objetivo SMART:

Para la definición del objetivo SMART se consideraron los valores obtenidos del CTOR (Click Through Open Rate) durante el año 2023 (enero-octubre):

Mes	CTOR
Enero	1,89%
Febrero	2,66%
Marzo	2,33%
Abril	2,01%
Mayo	1,92%
Junio	2,52%
Julio	2,08%
Agosto	1,67%
Septiembre	1,10%
Octubre	1,51%

Tabla 1: CTOR, enero a octubre 2023.

Luego, se puede notar que el CTOR ha tenido valores en el rango de 1,10% a un 2,66% con un promedio de 1,97% y un promedio de 1,42% en los últimos tres meses (agosto, septiembre y octubre). Por esto, buscando una meta realista en base a los datos históricos, se propone aumentar CTOR a un 1,85%.

Aumentar la eficacia considerando el aumento del CTOR (Click Through Open Rate) a un 1,85% en el canal de Mailing en los siguientes 3 meses.

Cabe destacar que se elige este canal dado que se encuentra más desarrollado en temas de personalización dentro de la empresa, resultando ser más manipulable y cercano para el proyecto.

b) Objetivos específicos:

1. Recopilación y limpieza de datos.
 - Identificar las fuentes de datos relevantes.
 - Creación de plan de recopilación de datos.
 - Establecer pasos en la limpieza de datos, además identificar y manejar valores atípicos, datos faltantes y errores.
2. Análisis exploratorio.
 - Hacer un análisis descriptivo de las variables claves.
 - Visualizar distribución de datos.
 - Identificar patrones preliminares.
3. Técnicas de análisis estadístico y/o minería de datos.
 - Aplicar técnicas estadísticas adecuadas que respondan preguntas.
 - Incluir algoritmos de minería de datos.
 - Interpretación de resultados y generación de conclusiones.
4. Propuesta de audiencia.
 - Análisis de resultados para identificar segmentos de audiencia.
 - Desarrollo de estrategias específicas para abordar cada segmento.
 - Creación de perfiles de la audiencia objetivo.
5. Monitoreo de métricas.
 - Identificar métricas claves relacionadas con los objetivos.
 - Seguimiento de la implementación.
 - Establecer orden en el monitoreo (periodos de monitoreo).
6. Comunicación de resultados.
 - Desarrollo de informes claros y atractivos.
 - Adaptar de presentación de resultados para gente de la empresa.
 - Preparar presentaciones y sesiones de retroalimentación.

4. Metodología

Dado que se quiere enfocar el proyecto en lo analítico, se eligió la metodología CRISP-DM (Cross-Industry Standar Process for Data Mining) ya que se quiere realizar iteraciones que refinan el análisis de manera estructurada y centrada en los objetivos.

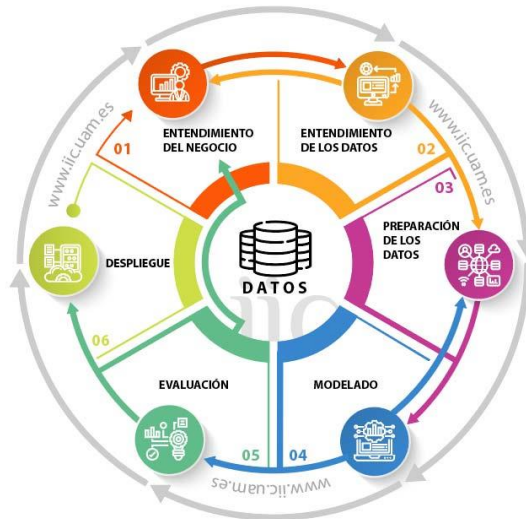


Imagen 4: Metodología CRISP-DM.

La imagen 4 se inserta como un apoyo visual para la explicación de la metodología. Se detallan las 6 fases que posee:

1. Comprensión del negocio: Se estudia el problema y se definen los objetivos.
2. Comprensión de los datos: Recopilación y exploración de datos. Se identifican las fuentes relevantes, se evalúa la calidad de los datos y se realizan análisis exploratorios.
3. Preparación de los datos: Limpieza, transformación y preparación de los datos para el análisis.
4. Modelado: Se entrenan modelos de minería de datos y se evalúan.
5. Evaluación: Se evalúa la capacidad de los modelos para lograr los objetivos propuestos y se mide su calidad.
6. Despliegue: Implementación de solución y posterior monitoreo. Además, se documentan los resultados y se comunican al equipo.

5. Medidas de desempeño

Considerando que las medidas de desempeño serán evaluadas a partir del caso de uso que se implementará en el momento, se definen:

- 1) CTOR (Click Through Open Rate): Proporción de usuarios que deciden hacer clic en el elemento mostrado versus usuarios que abrieron el contenido.

$$CTOR = \frac{Clics}{Comunicaciones\ abiertas} \times 100$$

- 2) BR (Bounce Rate): Porcentaje de usuarios que llegan al sitio web y abandonan la página sin interactuar versus usuarios que llegan al sitio web.

$$BR = \frac{Número\ de\ usuarios\ que\ rebotan}{Numero\ total\ de\ usuarios\ que\ visitan\ la\ página} \times 100$$

- 3) OR (Open Rate): Porcentaje de comunicaciones abiertas versus comunicaciones enviadas.

$$OR = \frac{Comunicaciones\ abiertas}{Comunicaciones\ enviadas} \times 100$$

6. Planificación

El proyecto se estructura en cuatro fases: Preparación, Análisis y Modelado de Datos, Implementación y Seguimiento, y Evaluación y Cierre, con un hito de Finalización. La Fase 1, Preparación, abarca la definición de objetivos, recopilación e identificación de fuentes de datos, y la planificación estratégica. En la Fase 2, Análisis y Modelado de Datos, se realiza la limpieza de datos, análisis exploratorio, aplicación de técnicas de minería de datos, propuesta de audiencia y validación de resultados. La Fase 3, Implementación y Seguimiento, incluye el apoyo en la personalización e implementación de casos de uso, monitoreo de medidas de desempeño y ajuste de estrategia. La Fase 4, Evaluación y Cierre, evalúa el proyecto en comparación con los objetivos, documenta los resultados y prepara un informe final con conclusiones y recomendaciones.

7. Estado del arte

La mejora de procesos en las empresas ha sido un tema relevante para el alcance de la excelencia operativa, más aún, en los últimos años es fundamental que las empresas estén en constante adaptación y mejora para asegurar su supervivencia y éxito. En particular, para la mejora del proceso de casos de uso que busca mejorar el CTOR y con ello, llegar de forma más efectiva al cliente, es necesario contextualizar sobre las estrategias que se ocupan hoy en día en la conversión y el marketing. Dentro de ejemplos de empresas que sortearon problemas similares a este y aplicaron soluciones, se encuentran:

1. Spotify:

Spotify buscaba que sus usuarios gratuitos se convirtieran en suscriptores. Para lograrlo, se enfocaron en tres pilares: Multicanalidad, Marketing Automation y personalización. La multicanalidad se encargaba de contactar a los usuarios desde distintos puntos, llegando a estos en el momento preciso. La implementación del Marketing Automation segmentaba la audiencia, permitiendo darles a los usuarios gratuitos ofertas personalizadas. Por último, la personalización ofrecía beneficios especiales por ser suscriptor, también se ofrecieron pruebas gratuitas. El conjunto de estas estrategias logró el aumento en la tasa de conversión (Santamaría-Rodríguez, 2017).

2. Amazon:

Amazon notó que muchos de sus usuarios abandonaban el carrito de compras, por esto, buscaban alcanzar la compra. Se utilizaron estrategias respecto a la Mejora de Experiencia del Usuario (UX), Marketing Automation y personalización. Para UX se rediseñó la página y se dispuso un proceso de pago más simple. Marketing Automation enviaba correos recordando el abandono del carrito de compras. La personalización permitía dar recomendaciones de productos relevantes para cada cliente. El resultado que se obtuvo fue una baja en el abandono de carritos y un aumento en la tasa de conversión (Medina-Velandia, 2018).

3. Netflix:

Netflix buscaba retener a sus suscriptores a largo plazo. Dado este objetivo, implementó estrategias como: Personalización y recomendaciones, y contenido exclusivo. En cuanto a la personalización y recomendaciones, se analizó el comportamiento de lo que veían los usuarios, logrando ofrecer recomendaciones específicas para cada cliente. El contenido exclusivo también fue una parte fundamental para alcanzar el objetivo ya que permitía que los usuarios pudieran ver

ese contenido solo en Netflix, esto además de retener a los suscriptores, atrajo nuevos. También se ofrecieron pruebas gratuitas. El alineamiento de estas estrategias aumento la retención de usuarios a largo plazo (Fernández-Manzano, Neira y Clares-Gavilán, 2016).

En estos ejemplos se puede que empresas líderes tuvieron que afrontar problemáticas de conversión similares al problema detectado en el proyecto, permitiendo visualizar tendencias que se emplean hoy en día para abordar el desafío propuesto.

8. Solución

Para determinar qué solución sería la más adecuada en el contexto de este proyecto, se analizaron distintas propuestas de solución en base a la literatura y las investigaciones encontradas:

1. Inbound Marketing: Estrategia que atrae al cliente a través de contenido relevante y de valor para el cliente, lo hace de una forma no intrusiva.

➤ Ventajas:

- Atracción con alta segmentación de público.
- Mayor lealtad y confianza del cliente.
- Leads de calidad.
- Medición de ROI y resultados.

➤ Desventajas:

- Alto nivel de esfuerzo en la creación de contenido.
- Visualización de resultados sólidos conlleva bastante tiempo.
- Gran competencia en generación de contenido.

2. Marketing Automation: Es la automatización de tareas en marketing, tiene por objetivo la segmentación y personalización de la comunicación enviada.

➤ Ventajas:

- Disminución de recursos gracias a la automatización de tareas.
- Aumento de la eficiencia en estrategias de marketing.
- Personalización de comunicaciones.
- Enriquecimiento de leads.

➤ Desventajas:

- Necesaria inversión de software (\$3.000.000 aproximadamente) y capacitación del personal (1.000.000 aproximadamente).
- Depende de la calidad de los datos empleados.

3. Social CRM (Customer Relationship Management) y Big Data: Análisis del cliente mediante el estudio de su interacción en redes sociales y los datos que posee la empresa de este.

➤ Ventajas:

- Mayor precisión en cuanto a estrategias y ofertas.
- Seguimiento preciso del comportamiento del cliente.
- Toma de decisiones informada.

➤ Desventajas:

- Necesaria inversión en software CRM y herramientas de análisis (más de \$100.000.000).
- Preocupante ética respecto a la privacidad del cliente.
- Gestión de datos de gran magnitud.

4. DMP (Data Management Platform): Recopilación y segmentación de datos a gran escala.

➤ Ventajas:

- Permite la segmentación precisa.
- Decisiones basadas en datos.
- Posible mejora en efectividad.

➤ Desventajas:

- Necesaria inversión en herramienta DMP (más de \$100.000.000).
- Regulaciones de privacidad de datos del cliente.
- Requiere precisión y calidad de datos.

5. Análisis de datos avanzados: Implica un enfoque más minucioso y profundo en el análisis de datos para la obtención de conocimientos más detallados y precisos. Contempla el uso de técnicas y métodos más complejos que van superan los análisis descriptivos básicos.

➤ Ventajas:

- Profundidad en la comprensión de patrones y datos.
- Precisión en la predicción del comportamiento.
- Decisiones basadas en datos.

➤ Desventajas:

- Habilidades técnicas del análisis de datos.
- Requiere calidad de datos.
- Implementación de modelos es costosa (\$1.500.000 aproximadamente) y conlleva tiempo (2 a 3 meses).

Luego del estudio de cada una de las cinco propuestas de solución donde se consideró tanto la necesidad de llegar de manera más efectiva al cliente como las limitaciones presentes, se decidió dirigir el proyecto hacia el análisis de datos avanzados (5). Respecto a esta propuesta, se enfocará más bien en la generación de insights al equipo encargado del envío de correos (Salesforce) para la toma de acción, en vez de la creación de un modelo predictivo.

Tras la evaluación exhaustiva de las alternativas, a continuación, se da a conocer en detalle la justificación del descarte de las demás propuestas de solución:

1. Inbound Marketing: Su enfoque radica en la atracción de clientes potenciales mediante la creación de contenido y estrategias de marketing, mientras que el proyecto busca la mejora del proceso de casos de uso mediante la segmentación estratégica. A su vez, requiere la creación de contenido, estando esto fuera del alcance del equipo debido a las restricciones de recursos y habilidades específicas necesarias para generar contenido de calidad. Además, su implementación conlleva tiempo, pudiendo ser un obstáculo para la culminación del proyecto.
2. Marketing Automation: Si bien busca la personalización de la comunicación enviada, tiene un enfoque externo ya que se basa en la gestión de campañas de marketing, preocupándose por la automatización de tareas. Este enfoque no aborda directamente la mejora del proceso interno. Los costos asociados a esta propuesta son elevados dado que se necesita invertir en software, recursos y capacitación, pudiendo resultar como un obstáculo financiero significativo.

3. Social CRM (Customer Relationship Management) y Big Data: La implementación es compleja y costosa. Se necesitan integrar sistemas, además de la obtención y el procesamiento de grandes volúmenes de datos, pudiendo no ser alcanzable la implementación exitosa dado las limitaciones de tiempo. Además, resultan necesarias habilidades respecto a la configuración del sistema CRM y gestión de BigData las cuales pueden requerir una curva de aprendizaje extensa.
4. DMP (Data Management Platform): Implementación compleja y costosa, pudiendo impedir las limitaciones de tiempo en el término del proyecto. También se necesitan habilidades especiales en cuanto a la gestión y configuración de la DMP. Se deben considerar las regulaciones de privacidad que pueden requerir tiempo adicional y recursos legales.

La elección de la solución, es decir, el análisis avanzado de datos enfocado en la generación de insights al equipo encargado del envío de correos electrónicos, se basó en las siguientes razones:

- **Alcance de equipo:**

La implementación y creación de un modelo predictivo como tal está fuera del alcance de marketing operacional dado que es el equipo de data el encargado de estas tareas. No obstante, se puede emplear el análisis de datos avanzados para la generación de insights al equipo de Salesforce, logrando que se tomen acciones en el envío de correos en base a las recomendaciones generadas.

- **Enfoque:**

Se puede enfocar el análisis de datos para la mejora del proceso de casos de uso.

- **Tiempo:**

Dado el tiempo disponible para la realización del proyecto, se busca una solución que sea logable y medible en este periodo.

- **Costos:**

Gracias a las habilidades y recursos del equipo, se escatima en costos y se mantienen bajo control.

- **Resultados:**

Obtención de resultados de forma rápida y visualización de beneficios a corto plazo.

Entre los beneficios que se perciben de la solución escogida se encuentran:

- ✓ Retroalimentación continua y constante.
- ✓ Información clave.
- ✓ Control de costos.
- ✓ Rápida obtención de resultados.
- ✓ Flexibilidad en la alineación con objetivos.
- ✓ Adaptabilidad de habilidades y recursos existentes.
- ✓ Comunicación clara.
- ✓ Agilidad en la toma de decisiones informada.

En síntesis, la solución escogida se considera la más apropiada dado el contexto del proyecto y se espera que esta logre abordar el problema detectado.

c) Matriz de riesgos y mitigaciones

Se llevó a cabo una evaluación de riesgos de la solución elegida con el fin de prevenir posibles obstáculos que puedan surgir en la implementación del proyecto. A continuación, se presenta una matriz de riesgos asociada a la gravedad y probabilidad de que ocurran:

		Gravedad →				
		1 Insignificante	2 Menor	3 Moderada	4 Importante	5 Catastrófica
↑ Probabilidad	5 Muy probable	5	10	15	20	25
	4 Probable	4	8	12	16	20
	3 Posible	3	6	9	12	15
	2 No es probable	2	4	6	8	10
	1 Muy improbable	1	2	3	4	5
				Falta de recursos y presupuesto	Falta de alineación con los objetivos	

Imagen 5: Matriz de riesgos.

Se detallan los riesgos y sus mitigaciones:

- **Resistencia al cambio:** Impacto moderado (12)

La falta de apoyo por parte del equipo puede retrasar la implementación del proyecto y terminar por la no adopción de la solución. En consecuencia, dificulta la obtención tanto de insights como de resultados.

✓ Mitigaciones:

- Comunicación de beneficios de la solución a implementar.
- Involucramiento de partes afectadas en la toma de decisiones.

- **Falta de recursos y presupuesto:** Impacto bajo (3)

Limitaría la implementación eficaz del proyecto.

✓ Mitigaciones:

- Estudiar y establecer presupuesto necesario para el proyecto.
- Asignar recursos suficientes.
- Priorización de tiempo y esfuerzo.

- **Falta de habilidades analíticas:** Impacto moderado (12)

Puede impedir la generación de insights efectivos y, en consecuencia, la toma de decisiones.

✓ Mitigaciones:

- Capacitación en análisis de datos avanzados.
- Colaboración y/o apoyo con expertos en analítica.

- **Falta de acceso a datos de calidad:** Impacto moderado (8)

El trabajo con datos deficientes en calidad no permitiría la generación de insights precisos, pudiendo provocar decisiones erróneas.

✓ Mitigaciones:

- Garantizar la calidad y coherencia de los datos.
- Limpiar y normalizar los datos.
- Garantizar la integridad de los datos.

- **Falta de alineación con los objetivos:** Impacto bajo (4)

Si no se está alineado con los objetivos, los resultados esperados pueden estar desviados del proyecto y se corre el riesgo de no cumplir lo que se desea.

✓ Mitigaciones:

- Seguimiento continuo de la implementación y los resultados, de ser necesario, se debe ajustar la estrategia.

- **Problemas de privacidad de datos:** Impacto alto (10)

Puede tener repercusiones legales y reputacionales.

✓ Mitigaciones:

- Proteger la identidad de los clientes.
- Establecer procedimientos para el trato de datos sensibles.
- Respetar regulaciones de privacidad.

9. Evaluación económica

Para realizar la evaluación económica del proyecto se hizo un análisis de sensibilidad con escenario pesimista, normal y optimista. Las variables de sensibilidad seleccionadas fueron escogidas dada la opinión del negocio y también el impacto que tienen estas en el proceso.

Meses	Julio-Agosto
Monto boleto promedio	\$27.201
Cantidad de clientes	146.271
Tasa de conversión mensual atribuida	0,07%
Ganancia cliente físico - cliente omnicanal	\$40.000
Costo por adquisición	\$30.246

Tabla 2: Datos actuales del caso de uso implementado.

En la tabla 2 se presentan los datos actuales del caso de uso a analizar, en base a este caso se generarán insights y se relanzará el caso incluyendo las recomendaciones entregadas para así evaluar su rendimiento.

Este caso de uso fue evaluado durante los meses de julio y agosto, considerando una audiencia de 146.271, la cual tiene un monto de boleto promedio de \$27.201, en base a esto y al gasto realizado por la misma audiencia en ecommerce la ganancia obtenida por convertir un cliente físico a uno omnicanal es de \$40.000. Se obtuvo una tasa de conversión mensual atribuida de 0,07%. Además, el costo por adquisición de un cliente fue de \$30.246. Se presenta la situación actual:

Meses	Cantidad de clientes	Situación actual
Septiembre-Octubre	205	\$8.191.176
Noviembre-Diciembre	205	\$8.202.644
Enero-Febrero	205	\$8.202.660
Marzo-Abril	205	\$8.202.660
Costo analista		-\$9.600.000
Costo por adquisición		-\$24.801.069
Ganancia total		-\$1.601.930

Tabla 3: Proyección de situación actual.

De la tabla 3 se obtiene que no hubo ganancias en el caso de uso actual. Es más, se supera el costo en \$1.601.930.

El cálculo de los ingresos atribuidos al caso de uso fue por la fórmula:

$$\text{Ingreso} = \text{Cantidad de clientes captados} * \text{Ganancia de conversión}$$

El costo del analista se calculó como el sueldo del analista por la cantidad de meses que será necesario, en este caso son 8 meses.

El costo por adquisición multiplicó la suma de los clientes convertidos desde el canal físico a ecommerce en los meses proyectados.

Por último, la ganancia total fue la suma de los ingresos atribuidos al caso de uso desde septiembre hasta abril menos los costos asociados.

A continuación, se presentan las variables sensibles en los escenarios pesimista, normal y optimista:

Escenarios posibles

Escenario	Pesimista	Normal	Optimista
Tasa de conversión mensual atribuida	0,08%	0,09%	0,10%
Monto boleto promedio	\$25.000	\$27.201	\$28.000
Ganancia cliente físico - cliente omnicanal	\$41.000	\$45.000	\$50.000
Costo por adquisición	\$30.300	\$29.000	\$28.000

Tabla 4: Escenarios pesimista, normal y optimista con variables sensibles.

Como se puede ver en la tabla 4, en el escenario pesimista se considera una tasa de conversión diaria atribuida al caso de uso del 0,08%, un monto de boleto promedio de \$25.000 y una ganancia de cliente convertido de \$41.000 y un costo por adquisición de cliente de \$30.300. En el caso normal se toma el monto de boleto promedio como \$27.201 y ganancia de convertir a un cliente de físico a omnicanal de \$45.000, además, el costo por adquisición desciende \$29.000. La tasa de conversión diaria atribuida se fija en un 0,09%. Respecto al caso optimista, la tasa de conversión diaria atribuida es de un 0,1%, el monto de boleto promedio es de \$28.000, la ganancia de conversión es de \$50.000 y el costo por adquisición es de \$28.000. Cabe destacar que la variable monto boleto promedio no se considera una variable sensible, solo se presenta a modo de entender la ganancia por cliente convertido.

Costos asociados

Se considera como un costo significativo del proyecto la contratación de un analista de datos y el costo por adquisición de un cliente desde el canal físico al ecommerce, evaluando el primer costo en \$1.200.000 mensuales y el segundo se consideró una variable sensible que va desde el rango de los \$30.300 hasta los \$28.000. Se descartan los gastos de implementación de herramientas y migración de datos ya que el negocio ya disponía de estos recursos. No se consideran los costos de consumo de datos ya que son despreciables para el área de marketing operacional.

Escenario Pesimista:

Meses	Sept-oct	Nov-dic	En-Feb	Mar-abr
Monto boleto promedio	\$25.000	\$25.000	\$25.000	\$25.000
Cantidad de clientes captados	234	234	234	234
Ganancia cliente físico - cliente omnicanal	\$41.000	\$41.000	\$41.000	\$41.000
Costo por adquisición	\$30.300	\$30.300	\$30.300	\$30.300

Tabla 5: Clientes capturados en el escenario pesimista.

En la tabla 5 se presenta la cantidad de clientes captados de la audiencia escogida considerando una tasa de conversión mensual atribuida del 0,08%.

Escenario Normal:

Meses	Sept-oct	Nov-dic	En-Feb	Mar-abr
Monto boleto promedio	\$27.201	\$27.201	\$27.201	\$27.201
Cantidad de clientes	263	264	264	264
Ganancias cliente físico - cliente omnicanal	\$45.000	\$45.000	\$45.000	\$45.000
Costo por adquisición	\$29.000	\$29.000	\$29.000	\$29.000

Tabla 6: Clientes capturados en el escenario normal.

Se calcula la cantidad de clientes captados por el escenario normal, considerando la tasa de conversión mensual atribuida fijada en 0,09%.

Escenario Optimista:

Meses	Sept-oct	Nov-dic	En-Feb	Mar-abr
Monto boleto promedio	\$28.000	\$28.000	\$28.000	\$28.000
Cantidad de clientes	293	293	293	293
Ganancia cliente físico - cliente omnicanal	\$50.000	\$50.000	\$50.000	\$50.000
Costo por adquisición	\$28.000	\$28.000	\$28.000	\$28.000

Tabla 7: Clientes capturados en el escenario optimista.

En el escenario optimista se propone una tasa de conversión de 0,1%. En la tabla se muestra la audiencia capturada.

Luego, se calculan las ganancias potenciales del proyecto en los tres escenarios, considerando tanto ingresos como costos:

Meses	Pesimista	Normal	Optimista
Septiembre-Octubre	\$9.595.378	\$11.847.951	\$14.627.100
Noviembre-Diciembre	\$9.610.730	\$11.869.277	\$14.656.354
Enero-Febrero	\$9.610.755	\$11.869.316	\$14.656.413
Marzo-Abril	\$9.610.755	\$11.869.316	\$14.656.413
Costo analista	-\$9.600.000	-\$9.600.000	-\$9.600.000
Costo por adquisición	-\$28.398.947	-\$30.582.665	-\$32.813.917
Ganancia potencial	\$428.671	\$7.273.195	\$16.182.363

Tabla 8: Proyecciones en los escenarios pesimista, normal y optimista.

En la tabla 8 se puede apreciar que en los tres escenarios proyectados el proyecto obtiene ganancias potenciales, siendo estas de \$428.671, \$7.273.195 y \$16.182.8363 en los escenarios pesimista, normal y optimista, respectivamente. Se puede decir que el proyecto es rentable incluso en el escenario pesimista, superando la ganancia potencial al costo en cada periodo. La evaluación de los 3 escenarios propuestos respalda la viabilidad y rentabilidad del proyecto, abriendo la posibilidad de obtener mayores oportunidades de ganancia y mejorando la eficacia en la conversión de clientes físicos a omnicanal.

10.Desarrollo del proyecto

El proyecto seguirá la metodología CRISP-DM, abordando pasos como la comprensión del negocio, comprensión de los datos, preparación de datos, modelado, evaluación y despliegue. Además, se cumplirá con la planificación establecida, que incluye etapas de preparación, análisis y modelado de datos, implementación y seguimiento, y evaluación y cierre. El avance de la implementación se detalla a continuación.

a) Fase 1: Preparación

1. Comprensión del negocio:

Se realizó un análisis profundo de la situación actual de la empresa comprendiendo con ello los dolores de esta, pero, sobre todo, los dolores del área de marketing. Se indaga en las métricas relevantes del marketing operacional, siendo la que recibe mayor interés la tasa de conversión. En consecuencia, se estudia el comportamiento de la métrica y se compara con la industria, descubriendo que existe una brecha significativa.

En base a esto, se analizan en profundidad métricas relacionadas a la tasa de conversión y al marketing operacional resultando ser: Total de envío de emails, Open Rate (OR), Click Through Rate (CTR) y Click Through Open Rate (CTOR). Gracias a la calidad de los datos que provee, se decide estudiar específicamente el canal de mail.

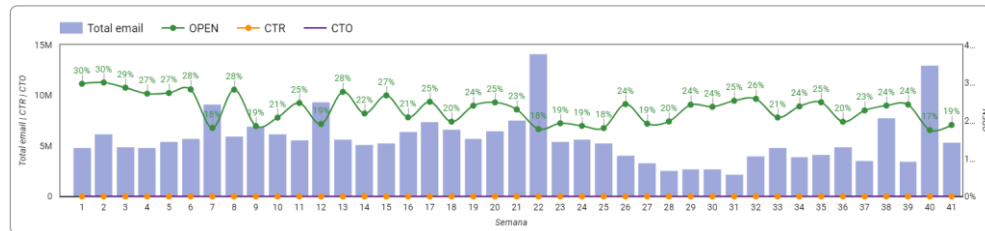


Imagen 6: Dashboard de Walmart, métricas de canal mail 2023 (Walmart).

Luego del análisis detallado de las métricas mencionadas, se infiere qué está causando esta brecha y se define el objetivo SMART. Posterior a esto, se establece la planificación del proyecto y se comienza a estudiar el estado del arte para posteriormente, evaluar las posibles soluciones. Una vez seleccionada la solución del proyecto, se da mayor enfoque a los datos disponibles.

2. Comprensión de los datos:

Una vez definido el objetivo, se comienzan a identificar fuentes relevantes de datos para hacer conexiones entre distintas tablas y poder obtener más información del cliente. Se crean variables de interés para el análisis a partir de las variables ya existentes. A continuación, se presentan las relaciones entre las tablas de las cuales se extrajo información:

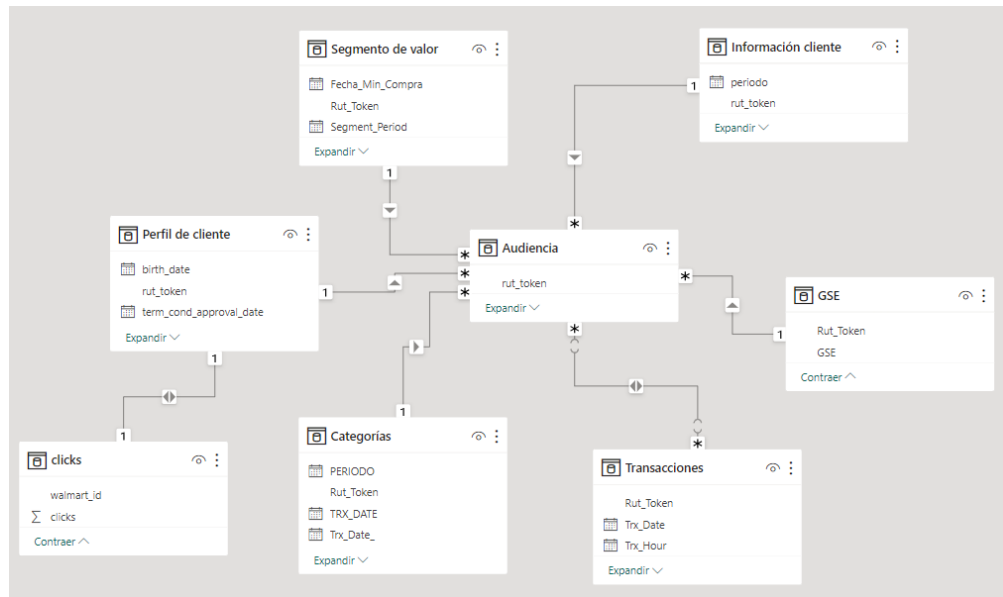


Imagen 7: Tabla relacional.

Las tablas utilizadas se detallan en el anexo 4 y el glosario de las variables que serán mencionadas más adelante en el anexo 5.

Es importante destacar que luego de los cruces de las tablas se conservaron los datos de comportamiento de 110.189 clientes en la tabla principal, la cual contiene 47 variables.

b) Fase 2: Análisis y modelado de datos

1. Preparación de los datos:

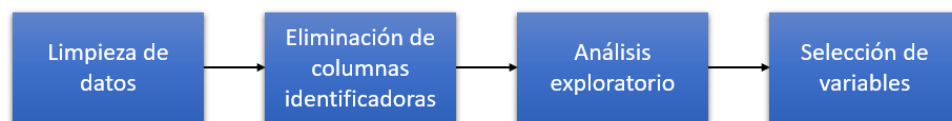


Imagen 8: Pasos del proceso de preparación de datos.

En la imagen 8 se muestran los pasos que se siguieron para la preparación de los datos, se explican a continuación:

1. Limpieza de datos:

Se comienza por entender la forma de la data creada, es decir, se ve el tipo de variables que trae y la cantidad de filas y columnas. Luego, se miran los valores nulos en cada variable y se decide eliminar las filas con valores nulos en las variables de edad y genero ya que pueden afectar el posterior análisis. Además, se establece un

rango para la columna edad, siendo este desde 18 años hasta 100 años y se aplican condiciones para que las variables relacionadas con montos pagados por el cliente sean mayores a ceros. Por último, se rellenan los valores nulos en la data con ceros.

2. Eliminación de columnas identificadoras:

Dado el motivo de protección de los datos del cliente y baja relevancia en el análisis, se eliminan variables identificadoras de este, como lo son el rut_token asignado y el walmart_id. Entonces, quedan 45 variables y 95.359 filas de datos.

3. Análisis exploratorio:

Se realiza la separación de variables categóricas y numéricas y se procede con el análisis.

Variables numéricas

Se realiza un análisis descriptivo de la mayoría de las variables (Anexo 2), donde lo más relevante que se obtiene es:

- ✓ La media de la edad de la audiencia seleccionada es de 40 años. El 25% de los clientes tiene 35 años y el 75% tiene 44 años.
- ✓ Es común que los clientes compren todos los meses del año en SOD (supermercado online líder).
- ✓ Todos los clientes de la audiencia tienen un índice de valor mayor a 0,4.
- ✓ La media del gasto mensual de los clientes es de \$220.660, sin embargo, la desviación estándar existente es de \$187.416. También se puede decir que los clientes suelen gastar más en el horario 06:00 – 12:00 horas (Anexo 3).
- ✓ La media de los ítems por boleta es de 26 ítems con una desviación estándar de 25 ítems, mientras que el monto de boleta promedio es de \$50.882 con una desviación estándar de \$47.721.
- ✓ La media de la cantidad de clicks es de 0,02, lo cual rectifica la baja tasa de conversión en el CTOR del caso de uso implementado. La máxima de clics realizados por un cliente es de 51. Aún así, la desviación estándar es de 0,31.

Luego, se grafican las variables numéricas en histogramas para obtener una comprensión más visual de la distribución de los datos. A continuación, se muestran los histogramas que generaron más información:

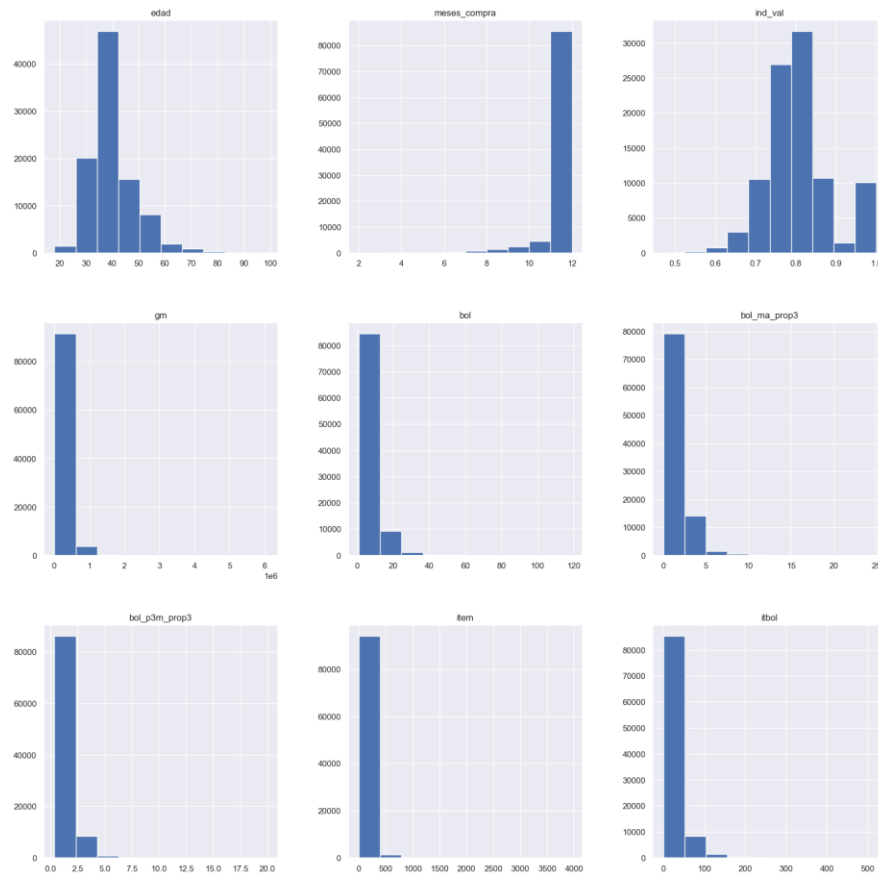


Gráfico 6: Histograma de variables numéricas.

Del gráfico 6 se puede decir que tanto el gasto mensual, el monto de boleta promedio, la cantidad de ítems y los ítems por boleta tienen datos concentrados hacia la izquierda, por tanto, sus datos son poco dispersos. Mientras, la edad y el índice de valor tienen datos más dispersos. Cabe destacar que la cantidad de boletas suele ser mayor en los horarios 06:00 – 12:00 horas y 10:00 – 18:00 horas.

En el análisis también se incluyeron representaciones de las variables numéricas en gráficos de cajas y bigotes, para esto fue necesario normalizar las variables.

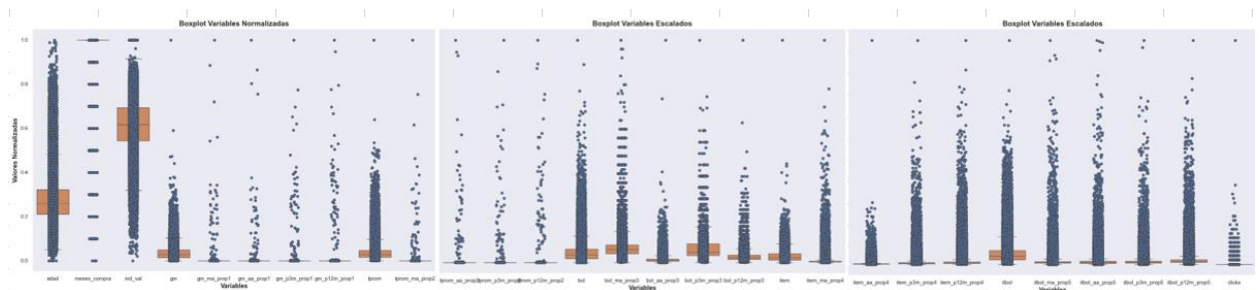


Gráfico 7: Boxplots de variables numéricas.

Del gráfico 7 se obtiene que hay gran presencia de datos atípicos. Sin embargo, se decide conservarlos dado que en el contexto del problema podrían arrojar patrones de comportamiento de los clientes.

Se analizó la correlación entre variables:

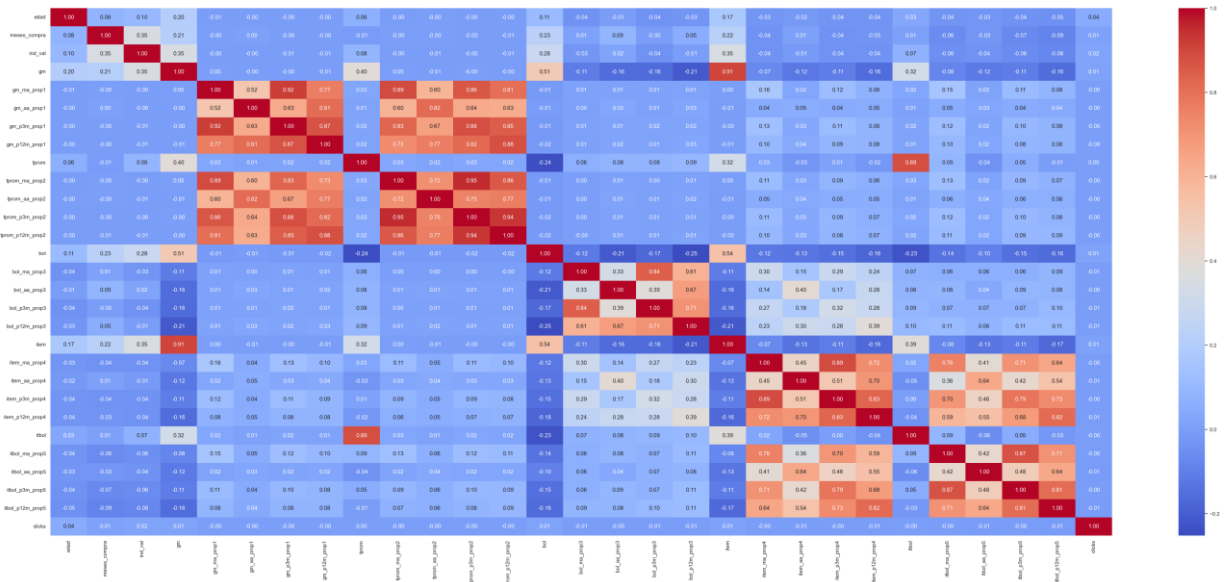


Gráfico 8: Matriz de correlación entre variables numéricas.

Se obtiene de la matriz de correlaciones (gráfico 8) que existe una gran cantidad de variables altamente correlacionadas por lo cual se deciden eliminar variables con un índice de correlación mayor a 0,7, procurando preservar variables que arrojan información útil en base a los análisis anteriores. Las variables numéricas que se mantuvieron fueron: edad, meses_compra, ind_val, bol, bol_aa_prop3, bol_p12m_prop3, item_aa_prop4 e itbol_aa_prop5. Se obtiene el nuevo gráfico de correlación:

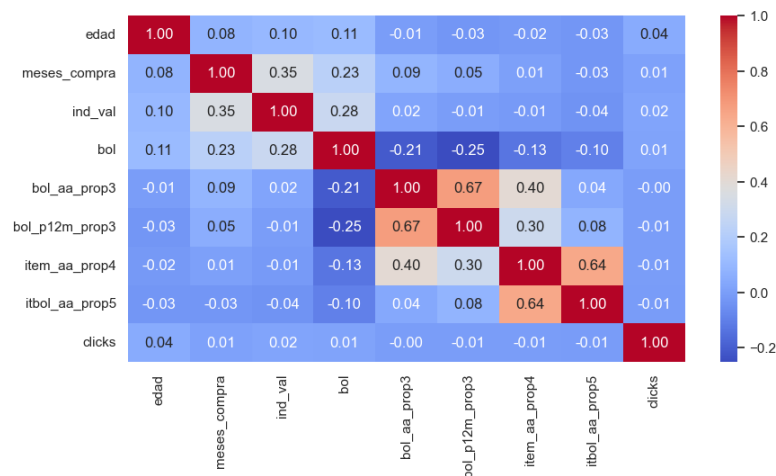


Gráfico 9: Matriz de correlación entre variables numéricas seleccionadas.

El gráfico 9 muestra las variables mantenidas con sus respectivas correlaciones, todas con valores menores a 0,7.

Variables categóricas

Al igual que para las variables numéricas, se comienza con un análisis descriptivo de las variables. En primera instancia, se dividen en variables categóricas nominales y ordinales.

Variables categóricas nominales:

Se crea una tabla de frecuencias y porcentajes (anexo 6) de la que se obtiene que el mayor porcentaje de perfil de usuario es de un 16,88% los cuales poseen las características de ser del género femenino, tener como categoría de lealtad el vestuario, ser clientes nuevos y un tener formato de compra usual en líder. En segundo lugar, con un 16,73% se encuentran clientes que difieren con los recientemente mencionados en la característica de que no son clientes nuevos.

Resumiendo, en el top 6 de los mayores porcentajes se puede rescatar que se encuentra una proporción entre mujeres y hombres del 76% y 23% respectivamente, que todos los usuarios tienen categoría de lealtad de vestuario, y se dividen entre el formato de compra más usual entre líder y líder express.

Se añade una visualización de la distribución de las categorías categóricas nominales para mayor comprensión:

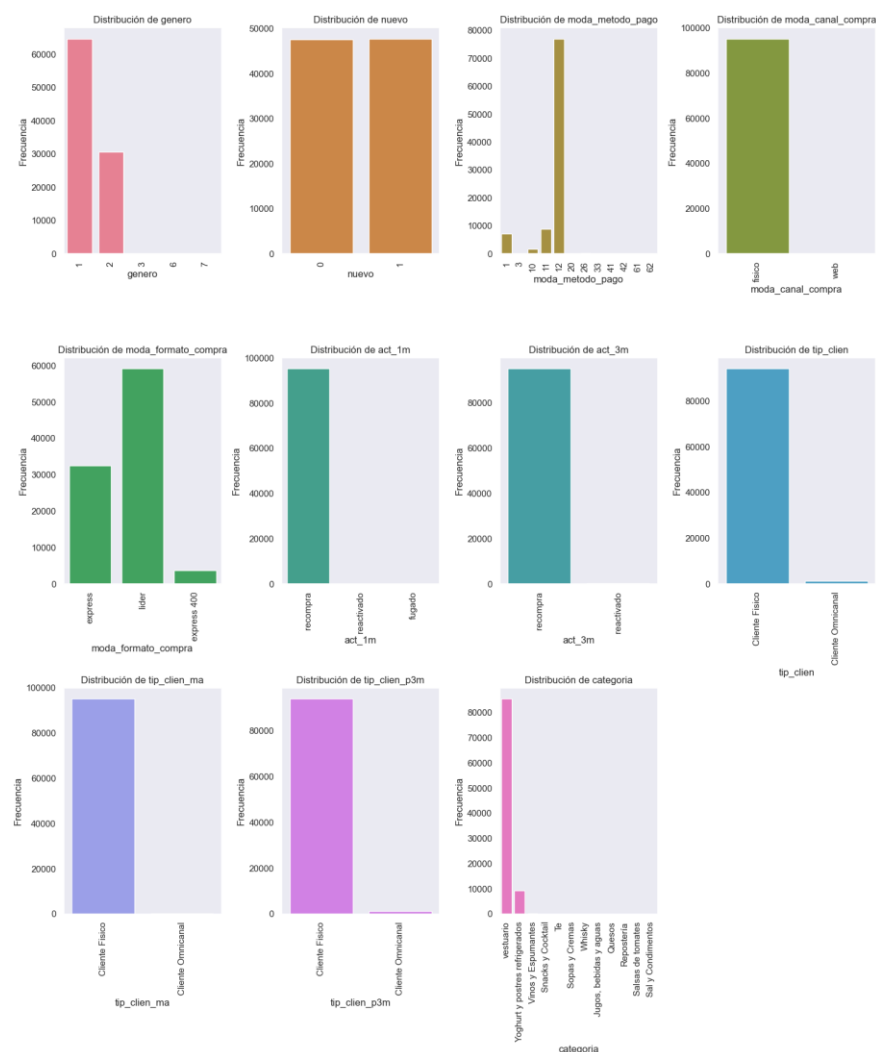


Gráfico 10: Frecuencias de variables categóricas nominales.

Del gráfico 10 se extrae que la audiencia seleccionada por el caso de uso se divide en las categorías de vestuario y de yogurt y postres refrigerados. Además, la mayoría de la audiencia corresponde a mujeres y en menor proporción a hombres. En cuanto al formato de compra usual predomina líder, seguido por líder express y líder 400. El método de pago más común es el 12.

Variables categóricas ordinales:

Se calcularon los cuartiles y se obtiene la tabla:

Métrica	gse	seg_agrup_rfm	seg_rfm	seg_valor
unique	7	3	4	6
top	C3	Leal	Champions	platino
freq	44.485	94.629	64.313	43.890
count	95.359	95.359	95.359	95.359
porc	47%	99%	67%	46%

Tabla 9: Cuartiles de variables categóricas ordinales.

Luego, se puede notar de la tabla 9 que la moda de gse es la clase C3 lo que, correspondiendo este nivel a la clase baja, por tanto, se puede decir que los clientes no tienen grandes ingresos. Los clientes leales representan a un 99% de la audiencia. En el caso de seg_rfm el 67% de los clientes están catalogados como Champions. En seg_valor el 46% de los clientes son platino.

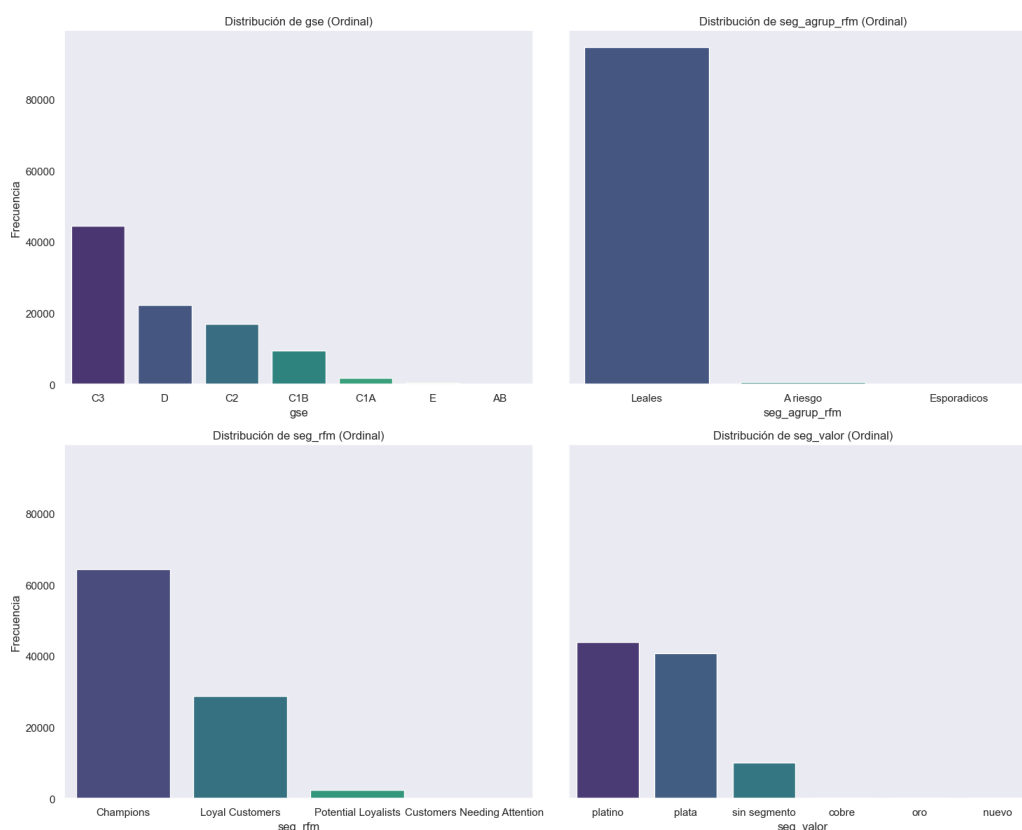


Gráfico 11: Histograma de variables categóricas ordinales.

En el gráfico 11 se puede apreciar el histograma de las variables categóricas ordinales pudiendo obtener su distribución. Se ve que la audiencia suele tener un alto nivel de lealtad según la clasificación de Walmart. En cuanto al gse se puede decir que los clientes seleccionados se concentran en clases sociales bajas.

Relaciones entre variables categóricas nominales y ordinales

De la relación entre los distintos tipos de variables categóricas, se hayan 2 relaciones que entregan información útil sobre el comportamiento de los clientes, se muestran los gráficos:

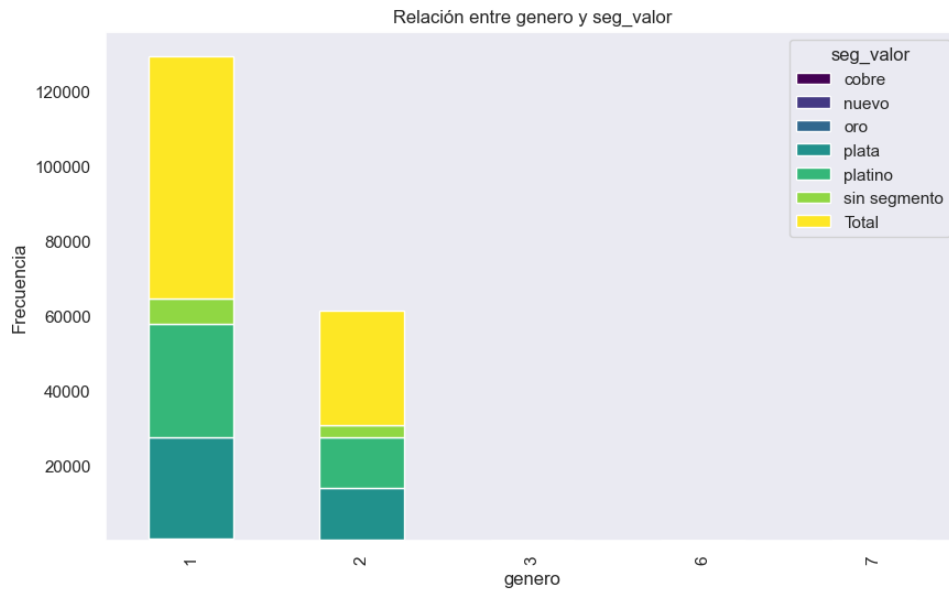


Gráfico 12: Relación entre variable categórica género y segmento de valor.

Del gráfico 12 de las variables de género y seg_valor se obtiene que las mujeres tienden a ser clasificadas mayoritariamente como clientes platinos y en segundo lugar como clientes plata, mientras que los hombres son repartidos equitativamente entre estos dos segmentos.

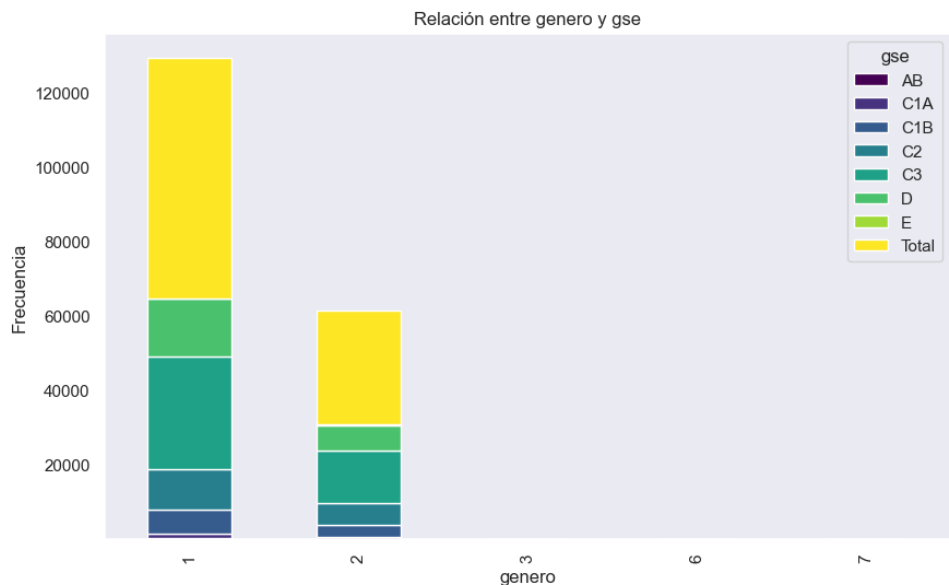


Gráfico 13: Relación entre variable categórica género y gse.

En el gráfico se puede apreciar la relación entre género y gse, de la cual se puede decir que las mujeres pertenecer principalmente a la clase social C3 y a la E. En cambio, los hombres pertenecer principalmente a la C3.

3. Selección de variables:

En la selección de variables se utilizaron distintos métodos para evaluar la importancia tanto de las variables numéricas como categóricas y su contribución a la predicción de la variable clics.

Variables numéricas

Se utilizaron dos algoritmos: Random Forest Regressor y RFECV. Por un lado, la prueba del primer algoritmo se estableció un umbral de 0,07. Este modelo seleccionó las variables: edad, ind_val, bol_aa_prop3, item_aa_prop4, itbol_aa_prop5. Por otro lado, RFECV dio como resultado de selección las variables: edad, ind_val, bol, bol_aa_prop3, bol_p12m_prop3, item_aa_prop4 e itbol_aa_prop5. Dado que estos algoritmos solo difirieron en la incorporación de las variables bol y bol_p12m_prop3, se deciden descartar estas. En consecuencia, las variables numéricas seleccionadas fueron: edad, ind_val, bol_aa_prop3, item_aa_prop4 e itbol_aa_prop5.

Variables categóricas

Para la selección de variables categóricas en primer lugar se codificaron las variables. En el caso de las variables categóricas nominales se aplicó el One Hot Encoding y en el caso de las categóricas ordinales se ocupó la codificación ordinal. Una vez codificadas las variables, se realizó la prueba ANOVA, en base a esta prueba se consideraron significativas las variables que arrojaron un resultado de p-value menor a 0,05. En consecuencia, las variables seleccionadas para el proceso fueron: seg_valor, genero_2, nuevo_1, moda_metodo_pago_10, moda_metodo_pago_12, moda_metodo_pago_41, moda_formato_compra_lider, act_1m_reactivado, tip_clien_Cliente Omnicanal y tip_clien_p3m_Cliente Omnicanal.

2. Modelado y evaluación:

La etapa de modelado se dividió en tres partes; entrenamiento y prueba, clusterización y análisis bivariable:

1. Entrenamiento y prueba:

Una vez obtenida la selección de variables, siendo esta: seg_valor, genero_2, nuevo_1, moda_metodo_pago_10, moda_metodo_pago_12, moda_metodo_pago_41, moda_formato_compra_lider, act_1m_reactivado, tip_clien_Cliente Omnicanal, tip_clien_p3m_Cliente Omnicanal, edad, ind_val,

bol_aa_prop3, item_aa_prop4 e itbol_aa_prop5, se entrenan dos modelos en torno a la variable a predecir (clics). El primer modelo en entrenar es regresión lineal, este modelo arrojó un error cuadrático medio de $3,56e^{-5}$ y una raíz de error cuadrático medio de 0,005. Luego, se entrenó el modelo de árbol de decisión, obteniendo un error cuadrático medio de $6,71e^{-5}$ y una raíz de error cuadrático medio de 0,008. Dado que ambos modelos arrojan errores cercanos a cero, se considera que la selección de variables predice bien la variable objetivo.

2. Clusterización:

En la clusterización de los datos se aplicaron los modelos de DBSCAN, Mean Shift, Aglomerativo, Affinity Propagation, K Means y OPTICS. Para la selección del algoritmo de clusterización se realizó la prueba de silueta a cada uno de ellos, resultando ser el algoritmo de OPTICS el que obtuvo mayor puntuación (0,57).

El gráfico de clusterización realizado por el algoritmo OPTICS (Anexo 7), de esta se obtiene un solo cluster representado por la etiqueta -1 y datos atípicos o ruido marcados en amarillo por la etiqueta 0. Aunque se observa solo un cluster, esto hace sentido ya que la audiencia escogida proviene de una segmentación previa realizada por el equipo de marketing, por tanto, es esperable que los datos pertenezcan a solo un grupo de clientes. Sin embargo, se logró detectar algunos puntos que no pertenecen al cluster de clientes.

3. Análisis bivariable (variables categóricas vs numéricas):

El análisis bivariable se enfocó en la interpretación del comportamiento de las variables numéricas frente a las categóricas.

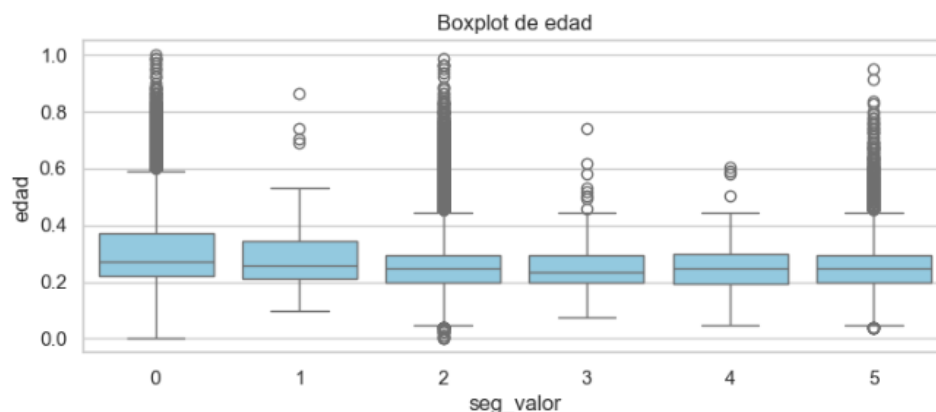


Gráfico 14: Boxplot variable edad versus segmento de valor.

Del gráfico 14 se obtiene el comportamiento de la variable edad frente al segmento de valor. Se puede decir que los 6 segmentos de valor contienen

cantidades de clientes parecidas y con edades similares. Aún así, se aprecia que en el segmento de valor 1 y 2 se concentran más clientes.

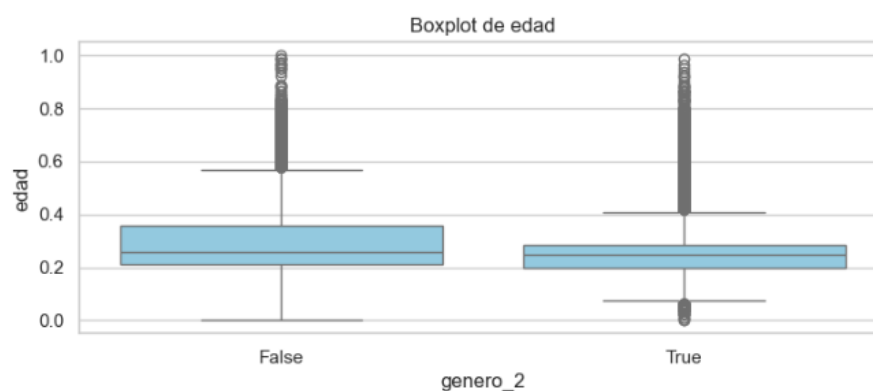


Gráfico 15: Boxplot variable edad versus género 2 (masculino).

En el boxplot de las variables edad versus género masculino (gráfico 15) se puede deducir que, considerando que solo hay hombres y mujeres, los hombres suelen tener edades más cercanas en comparación con las mujeres que presentan un rango más amplio de edad.

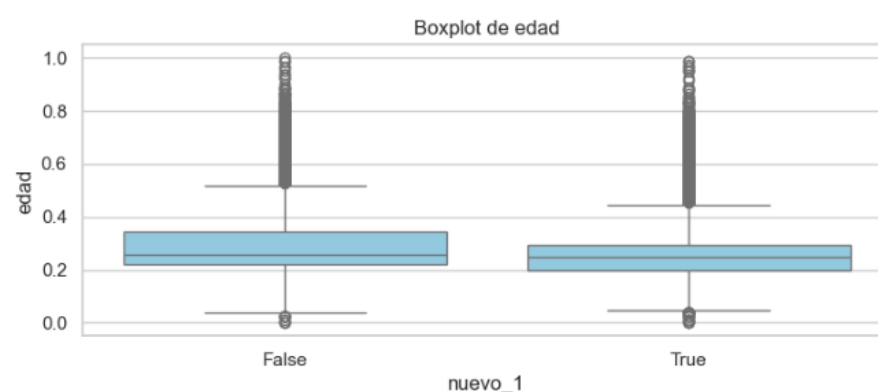


Gráfico 16: Boxplot variable edad versus cliente nuevo.

Del gráfico 16 se obtiene que los clientes nuevos suelen tener menos edad que los clientes que no lo son.

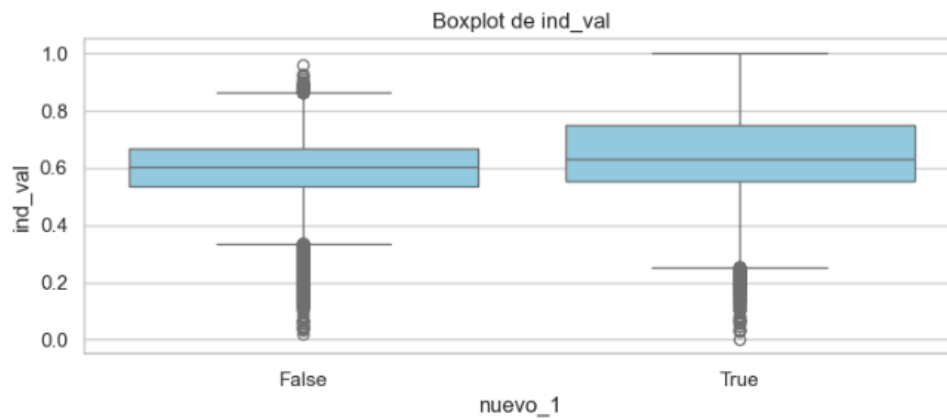


Gráfico 17: Boxplot variable índice de valor versus cliente nuevo.

Se puede inferir del gráfico 17 que los clientes nuevos suelen tener mayor índice de valor comparado con los clientes antiguos.

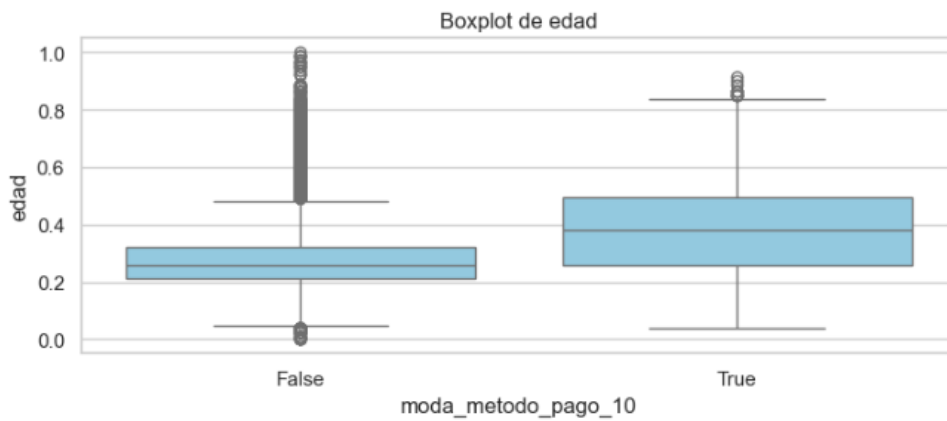


Gráfico 18: Boxplot variable edad versus método de pago 10.

Del gráfico 18 se rescata que gente de mayor edad suele pagar con el método de pago 10 (efectivo).

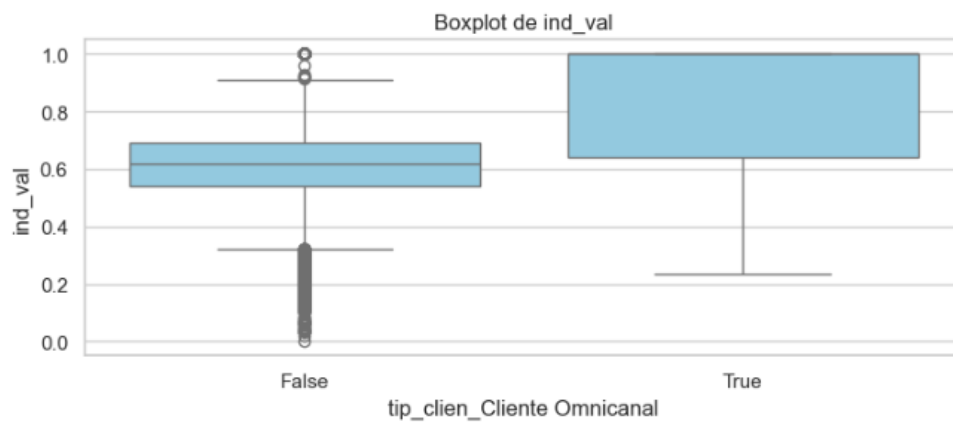


Gráfico 19: Boxplot variable índice de valor versus cliente omnicanal.

Se obtiene del gráfico 19 que los clientes omnicanales son mejor valorados que los clientes que no lo son.

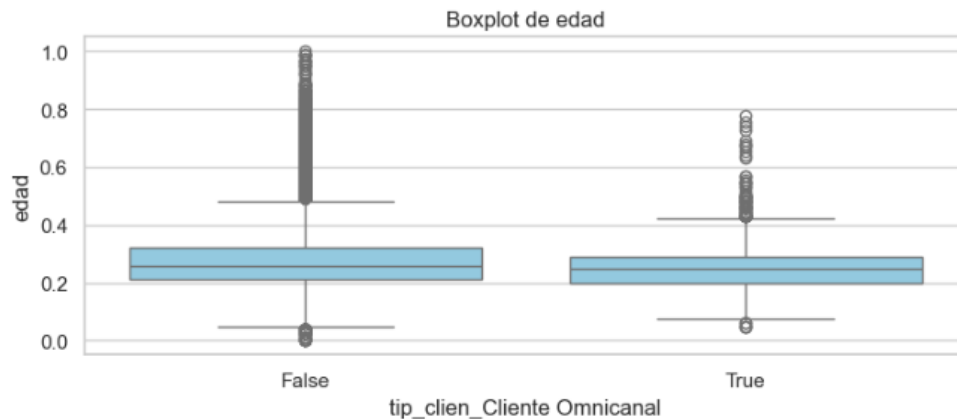


Gráfico 20: Boxplot variable edad versus cliente omnicanal.

Se extrae del gráfico 20 que los clientes omnicanales son más jóvenes que los clientes físicos.

Además, se analizaron los boxplots de las variables edad e índice de valor versus cliente omnicanal horario p3m arrojando estos gráficos comportamientos similares a los gráficos 19 y 20.

Insights

1. Segmentación de Audiencia:

Hay posibilidades de mejora en la segmentación actual mediante análisis avanzados utilizando herramientas de Salesforce. Se puede ver la posibilidad de incorporar variables como el índice de valor, la edad y el comportamiento de compra para obtener una segmentación más precisa.

Dado que la audiencia se concentra en los segmentos de valor 1 y 2 (platino y plata), podría ser beneficioso atraer a clientes en segmentos menos usuales.

2. Personalización de Contenido:

El contenido podría considerar factores como el perfil del cliente, género y comportamientos. Por ejemplo, las mujeres son mayoritariamente clasificadas como clientes platino, lo que puede influir en el tipo de productos que les podría parecer más atractivo.

La edad y el índice de valor son útiles para personalizar el contenido. Es decir, clientes más jóvenes pueden responder a estrategias de presentación de contenido diferentes en comparación con los que tienen mayor edad.

3. Comportamiento de Compra:

Por un lado, como es frecuente que los clientes compren durante todos los meses del año en SOD, se infiere que se necesitan estrategias de marketing continuas a lo largo del año.

Por otro lado, el gasto mensual y el monto de la boleta promedio del cliente son variables fundamentales que se deben considerar en la estrategia de marketing, especialmente durante los horarios de 06:00 a 12:00 horas, que es donde se observa un mayor gasto por parte del cliente.

4. Clusterización y Análisis de Datos:

El uso del algoritmo OPTICS rectifica que la audiencia escogida en el caso de uso pertenece a un único grupo, lo que es coherente con la segmentación previa realizada.

Sin embargo, se identificaron datos atípicos o ruido que podrían requerir una revisión adicional para entender su impacto.

5. Variables Predictivas:

Las variables seleccionadas para la predicción de la variable objetivo clics, como la edad, el índice de valor y la lealtad del cliente, afirman que tienen una capacidad predictiva sólida. Estas variables pueden ser clave para la personalización de las estrategias entorno al envío de correos electrónicos.

6. Cientes Omnicanal:

Los clientes omnicanal tienden a ser mejor valorados y más jóvenes en comparación con los clientes físicos. Este insight sugiere la importancia de fomentar la transición a un enfoque omnicanal.

7. Metodología de Pago:

Como se vio en el análisis, el método de pago más frecuente es el 12 (tarjeta de crédito o débito), seguido por el 10 (efectivo). Esta información puede ser útil en las estrategias de pago y promociones ofrecidas a los clientes a los que se les envían correos electrónicos.

Los insights obtenidos ofrecen una base sólida para la mejora de las estrategias de marketing en Salesforce, centrándose en los puntos mencionados anteriormente.

c) Fase 3: Implementación y seguimiento

1. Implementación:

Se comunicaron al equipo de Salesforce los insights generados, los cuales se convirtieron en aspectos fundamentales para la ejecución mejorada del caso de uso. Entre los insights que se tomaron en consideración por parte del equipo se encuentran la segmentación de clientes, la búsqueda de clientes pertenecientes a distintos segmentos de valor, la personalización de contenido en la estrategia de marketing en base a las variables mencionadas y la comprensión del comportamiento de compra entorno al gasto mensual y el monto boleto promedio del cliente en distintos horarios.

2. Seguimiento:

Se estableció un proceso de monitoreo para la evaluación de la efectividad de las estrategias implementadas. Se llevaron a cabo análisis periódicos (cada 2 días) de los indicadores clave de rendimiento entre ellos el Total de Mails enviados, el Open Rate (OR), el Click Through Open Rate (CTOR), el Click Through Rate (CTR) y el Bounce Rate (BR). Estos indicadores fueron claves para evaluar la implementación de los insights generados.

El proceso de seguimiento permitió realizar ajustes en tiempo real, mejorando las estrategias según las tendencias y patrones emergentes que se iban descubriendo durante el transcurso del caso de uso. Además, se llevaron a cabo sesiones con el equipo que tenían como tema central la retroalimentación con el equipo de Salesforce para concordar la comprensión compartida de los resultados y proporcionar agilidad ante cualquier cambio.

d) Fase 4: Evaluación y cierre:

1. Evaluación:

En la etapa de evaluación se analizaron minuciosamente los resultados obtenidos tras la implementación de las estrategias basadas en los insights entregados al área de Salesforce.

Grupo	tipo_comprador al 20-OCT 2023	tipo comprador al 20-NOV 2023	q_clientes
GC	FISICO	OMNICANAL	129
GC	FISICO	FISICO	24765
GO	FISICO	OMNICANAL	890
GO	FISICO	FISICO	145381

Tabla 10: Resultados de campaña implementada.

En la tabla 10 se compara el grupo de control con el grupo objetivo en el mismo periodo de tiempo (20 de octubre a 20 de noviembre). Los resultados que se entregan son que el grupo de control adquirió 129 clientes de un total de 24765 clientes físicos, mientras que el grupo objetivo adquirió 890 clientes de un total de 145.381.

TC	mensual
Tasa conversión GC	0,5209%
Tasa conversión GO	0,6122%

Tabla 11: Tasas de conversión mensual de GO y GC.

Los datos entregan una mejora del 17,5% en la tasa de conversión mensual para el grupo objetivo (GO) en comparación con el grupo de control (GC). Lo cual significa una tasa de conversión mensual atribuida de 0,0913%.

Luego, se calcula la audiencia atribuida por la fórmula:

$$\text{Audiencia atribuida} = \text{Audiencia total} * \text{Incremento atribuido mensual} * \text{Periodo}$$

Reemplazando los valores de audiencia total por 146.271 clientes, el incremento atribuido mensual de 0,09% y el periodo por un mes, da una audiencia atribuida de 133,5 clientes.

Por último, se calculo la ganancia de la conversión del cliente:

canal AÑO ANTERIOR	canal_actual ESTE AÑO	gasto_x_cliente_ytd (oct-nov)	gasto promedio ytd
Excl. fisico	Excl. fisico	\$650.561	\$72.285
Excl. fisico	Omnicanal	\$1.049.148	\$116.572

Tabla 12: Diferencias de los consumidores comparado con el año 2022.

De la tabla 12 se obtiene que la ganancia de conversión del cliente fue de \$44.287.

Se procede a calcular la ganancia obtenida del ejercicio:

Período	Cantidad de clientes	Ingreso
Oct-Nov	133,5	\$5.912.315
Costo analista		-\$1.200.000
Costo por adquisición		-\$4.034.370
Ganancia		\$677.945

Tabla 13: Ganancias obtenidas del ejercicio.

De la tabla 13 se obtienen las ganancias netas del ejercicio considerando un costo por adquisición de cliente de \$30.220, resultando ser de \$677.945.

2. Cierre:

Tras la implementación y consideración de los insights generados, se destaca:

Logro del objetivo

Open Rate	CTR	CTOR	Bounce Rate
21,10%	0,40%	1,90%	25%

Tabla 14: OR, CTR, CTOR y Bounce Rate del ejercicio.

En la tabla 14 se pueden ver las métricas resultantes del caso de uso implementado.

Se informa que se alcanzó el objetivo del proyecto al alcanzar un CTOR (Click Through Rate) mayor que 1,85%.

Otros logros

Segmentación de Audiencia:

Se implementaron cambios que resultaron exitosos, permitieron identificar de manera más precisa los clientes potenciales a través del análisis avanzado.

Personalización del Contenido:

Se integraron estrategias efectivas de personalización del contenido, otorgando mejoras a la relevancia y el impacto de las comunicaciones de marketing.

Análisis de Comportamiento de Compra:

Se logró obtener una comprensión más profunda del comportamiento de compra del cliente, especialmente en relación con el gasto mensual y el monto de boleta promedio en diferentes horarios.

Cambio en el Canal de Compra:

La conversión exitosa de clientes físicos a omnicanal se traduce en una mejora significativa en las tasas de conversión y en la atribución de conversiones.

Resultados Financieros:

Ganancias totales del proyecto: \$677.945.

Retorno de Inversión (ROI):

$$ROI = \left(\frac{Ganancia\ neta}{Inversión\ total} \right) * 100$$

Considerando la ganancia neta de \$677.945 y una inversión total de \$5.234.379, el ROI da un valor de 13%. Entonces, por cada unidad de moneda invertida en el proyecto, se ha obtenido un rendimiento adicional del 13%. En consecuencia, el proyecto fue exitoso en términos financieros.

Aprendizajes

La importancia de la segmentación precisa para la efectividad y rentabilidad de las estrategias de marketing.

La necesidad de la adaptación del contenido según el comportamiento de compra y las preferencias del cliente.

La migración a canales omnicanal puede generar impactos positivos sustanciales.

Próximos Pasos

Gracias al proyecto se reconoce en mayor medida la importancia de mantener un proceso continuo de mejora en base a los casos de uso. Se espera que los aprendizajes obtenidos se integren en futuras estrategias de marketing operacional para seguir evolucionando y brindando un mayor valor tanto a los clientes como a la empresa.

11. Conclusiones

En cuanto al proyecto realizado durante la pasantía, se lograron alcanzar los objetivos propuestos, obteniendo un CTOR del 1,9%. El desarrollo del proyecto destaca la importancia de comprender el negocio antes de intentar mejorar un proceso, ya que esta comprensión resultó fundamental para que la solución implementada tuviera un impacto real en el área de trabajo y fuera útil para futuras iteraciones.

A pesar de que la selección de variables consideró una amplia gama de estas, se aprendió a través de análisis e insights generados que la inclusión de variables menos convencionales podría marcar una gran diferencia al estudiar el perfil del cliente, sus intereses y sus patrones de comportamiento de compra.

Los resultados obtenidos durante el estudio de las variables coincidían inicialmente con la segmentación de clientes previamente realizada por la empresa, donde la audiencia pertenecía solo a un clúster. Sin embargo, un análisis más detallado reveló factores cruciales para lograr una comunicación más efectiva con el cliente. Además, se observó la presencia de clientes considerados como ruido en la clusterización, lo que sugiere la existencia de características no identificadas que dificultan su asociación al grupo mayor de manera más precisa.

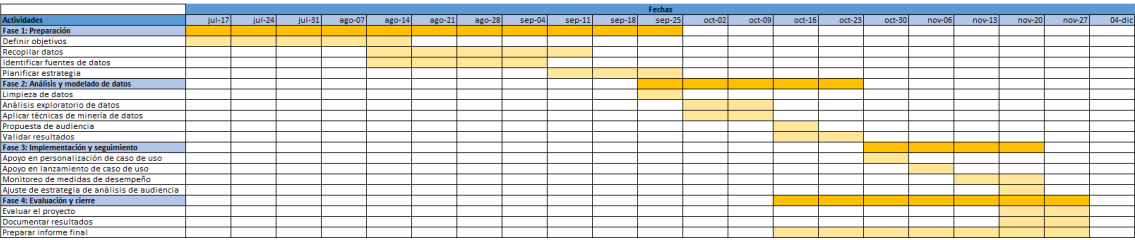
Finalmente, se abordó la dinámica de los equipos de trabajo en la realidad. A pesar de que el proyecto inicialmente se orientaba hacia la implementación de un modelo predictivo, factores políticos entre los equipos obstaculizaron su ejecución. Por lo tanto, fue necesario adaptarse a la dinámica de la empresa y redirigir el proyecto para asegurar que estuviera dentro del alcance y las habilidades del equipo. A pesar de las dificultades iniciales, el proyecto culminó con éxito, destacando la importancia de la adaptabilidad como una habilidad esencial para superar obstáculos en el entorno actual.

12. Referencias

1. Asociación Nacional Avisadores de Chile. (2022, junio 23). *Ecommerce: conoce la tasa de conversión como industria y cómo mejorarla*. <https://www.anda.cl/e-commerce-conoce-la-tasa-de-conversion-como-industria-y-como-mejorarla/#:~:text=De%20acuerdo%20a%20cifras%20compartidas,nivel%20de%20todas%20las%20industrias>
2. Diario Financiero/Chile. (2023, agosto 10). *Chile registró una mayor concentración del sector retail durante 2022*. <https://www.eleconomista.com.mx/empresas/Chile-registro-una-mayor-concentracion-del-sector-retail-durante-2022-20230810-0132.html>
3. Fernández-Manzano, E. P., Neira, E., & Clares-Gavilán, J. (2016). GESTIÓN DE DATOS EN EL NEGOCIO AUDIOVISUAL: NETFLIX COMO ESTUDIO DE CASO. *Profesional de La Información*, 25(4), 568–576. <https://doi.org/10.3145/epi.2016.jul.06>
4. Medina-Velandia, S. (2018). ESTUDIO DE LAS ESTRATEGIAS DE E-COMMERCE DESDE EL CASO DE LAS EMPRESAS MULTINACIONALES COMO EL GRUPO ÉXITO, FALABELLA, DAFITI, AMAZON, WALMART Y ZARA.
5. Santamaría-Rodríguez, D. (2017). EL STREAMING Y SUS IMPLICACIONES EN EL MARKETING. EL CASO DE SPOTIFY.
6. Andres, E., & Lescano, L. (2022). Estrategias de marketing para mejorar el embudo de conversión de ventas: una revisión sistemática Marketing strategies to improve the sales conversion funnel: a systematic review. 7(3), 2528–8083. <https://doi.org/10.5281/zenodo.8284898>
7. Comunica Web. (2022). Introducción ¿Por qué Inbound Marketing?
8. Hanssens, D. M., Pauwels, K. H., Srinivasan, S., Vanhuele, M., & Yildirim, G. (2014). Consumer Attitude Metrics for Guiding Marketing Mix Decisions. In *Source: Marketing Science* (Vol. 33, Issue 4). <https://www.jstor.org/stable/24544763>
9. Salcedo, R. A., & Arroyo López, M. (2019). Big Data: Applications of the data management in the different stages of the conversion funnel. In *Revista de Marketing y Publicidad* (Vol. 1).
10. Xu, L., Duan, J. A., & Whinston, A. (2014). Path to purchase: A mutually exciting point process model for online advertising and conversion. *Management Science*, 60(6), 1392–1412. <https://doi.org/10.1287/mnsc.2014.1952>

13.Anexos

1. Carta Gantt:



2. Análisis descriptivo de variables numéricas:

	edad	meses_compra	ind_val	gm	gm_ma_prop1	gm_aa_prop1	gm_p3m_prop1	gm_p12m_prop1
count	95359.0	95359.0	95359.000000	95359.0	95359.000000	95359.000000	95359.000000	95359.000000
mean	40.41197	11.57273	0.808057	220660.600017	80.293418	57.429364	52.588526	71.930484
std	8.49059	1.07278	0.084744	187416.253776	4596.713382	3608.140483	2621.194884	3228.303668
min	18.0	2.0	0.471100	5000.0	0.000000	0.000000	0.345377	0.150773
25%	35.0	12.0	0.758300	85385.5	1.097906	0.805063	1.050939	1.066341
50%	39.0	12.0	0.797600	172126.0	1.628263	1.396778	1.388150	1.559878
75%	44.0	12.0	0.837300	302747.5	2.900858	2.768076	2.224047	2.883783
max	99.0	12.0	1.000000	6079538.0	664453.000000	539318.000000	327524.000000	330530.000000

tprom	tprom_ma_prop2	...	item_ma_prop4	item_aa_prop4	item_p3m_prop4	item_p12m_prop4	itbol
95359.0	95359.000000	...	95359.000000	95359.000000	95359.000000	95359.000000	95359.000000
50882.94871	48.516212	...	3.951199	3.830815	2.907044	3.907620	26.182328
47721.734492	2967.308196	...	10.242230	11.518663	6.028011	8.497665	25.215322
1472.0	0.000000	...	0.000000	0.000000	0.340000	0.156667	0.333333
21331.0	1.062898	...	1.159693	0.826572	1.088889	1.106061	10.750000
36950.0	1.456292	...	1.727273	1.451429	1.460870	1.661290	18.625000
63290.5	2.231872	...	3.198958	2.941742	2.454545	3.225806	32.400000
1296436.0	478614.000000	...	568.000000	1395.000000	267.000000	279.000000	521.000000

itbol_ma_prop5	itbol_aa_prop5	itbol_p3m_prop5	itbol_p12m_prop5	clicks
95359.000000	95359.000000	95359.000000	95359.000000	95359.0
2.631680	2.667324	2.039723	2.487737	0.028702
4.812479	5.044890	2.838370	3.495811	0.313541
0.000000	0.000000	0.360000	0.163636	0.0
1.128205	0.955556	1.083871	1.144208	0.0
1.571429	1.522417	1.375000	1.566138	0.0
2.500000	2.634277	2.000000	2.461538	0.0
198.500000	203.000000	125.000000	118.000000	51.0

3. Siglas horarios Walmart:

Horario	Rango
AA (All Associates)	Todo el día
AM (Morning Associates)	06:00 - 12:00
P3M (Peak 3 Midday)	10:00 - 18:00
P12M (Peak 12 Midnight)	22:00 - 06:00

4. Tablas utilizadas:

Tabla	Descripción
Tabla Audiencia	Provee la audiencia a analizar, la cual corresponde a la que se utilizó en un caso de uso ya implementado. Tiene un total de 253.791 clientes.
Tabla Segmento de valor	Entrega información sobre el segmento de valor que le da la empresa al cliente. Estos pueden ser platino, oro, plata y cobre.
Tabla Perfil de cliente	Entrega información personal del cliente.
Tabla Clicks	Provee los clicks de los clientes durante el año.
Tabla Categorías	Permite obtener la categoría de lealtad del cliente, es decir, la más frecuente.
Tabla Información del cliente	Provee información sobre la actividad del cliente en la empresa.
Tabla GSE	Indica el nivel socioeconómico del cliente.
Tabla Transacciones	Detalla las transacciones del cliente.

5. Glosario de variables:

Concepto	Descripción
SOD	Supermercado en línea
aa	08:00-22:00 horas
ma	06:00-12:00 horas
p3m	10:00-18:00 horas
p12m	18:00-06:00 horas
rut_token	Rut codificado del cliente
walmart_id	Mail codificado del cliente
edad	Edad del cliente
meses_compra	Meses consecutivos que el cliente ha comprado
ind_val	Índice de valor otorgado por Walmart al cliente
gm	Gasto mensual del cliente
bol	Cantidad de boletas promedio del cliente
item	Cantidad de ítems promedio que lleva el cliente
itbol	Cantidad de ítems por boleta promedio del cliente
tprom	Valor de boleta promedio del cliente
genero	1: Femenino; 2: Masculino; 2
nuevo	1: Cliente nuevo; 0: Cliente antiguo
moda_metodo_pago	1: Cheque; 10: Efectivo; 12: Tarjeta crédito/débito
moda_canal_compra	Canal de compra más frecuente del cliente
moda_formato_compra	Formato de compra más frecuente del cliente
act_1m	Actividad del último mes del cliente
act_3m	Actividad de los últimos 3 meses del cliente
tip_clien	Tipo de cliente
categoria	Categoría de lealtad del cliente
gse	Clase social del cliente
seg_agrup_rfm	Segmento agrupado rfm
seg_rfm	Segmento rfm
seg_valor	Segmento de valor

6. Tabla de frecuencias y porcentajes de variables categóricas nominales:

genero	categori	nuevo	moda_metodo_pago	moda_canal_compr	moda_formato_compr	act_1m	act_3m	tip_cliente	tip_cliente_m	tip_cliente_p3m	Frecuenci	Porcentaj
1	vestuario	1	12	fisico	lider	recompra	recompra	Cliente Fisico	Cliente Fisico	Cliente Fisico	16100	16.88%
1	vestuario	0	12	fisico	lider	recompra	recompra	Cliente Fisico	Cliente Fisico	Cliente Fisico	15953	16.73%
1	vestuario	0	12	fisico	express	recompra	recompra	Cliente Fisico	Cliente Fisico	Cliente Fisico	7308	7.66%
2	vestuario	0	12	fisico	lider	recompra	recompra	Cliente Fisico	Cliente Fisico	Cliente Fisico	7001	7.34%
2	vestuario	1	12	fisico	lider	recompra	recompra	Cliente Fisico	Cliente Fisico	Cliente Fisico	6926	7.26%
1	vestuario	1	12	fisico	express	recompra	recompra	Cliente Fisico	Cliente Fisico	Cliente Fisico	6882	7.22%

7. Clusterización OPTICS:

