UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

CONSTRUCCIÓN Y VALIDACIÓN DE UN MODELO DE PROPENSIÓN DE COMPRAS POR SEGMENTO A PARTIR DEL USO DE MODELOS DE MACHINE LEARNING Y EXPERIMENTOS DE CAMPO

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERA CIVIL INDUSTRIAL

MARCELA PAZ MALDONADO GONZÁLEZ

PROFESOR GUÍA:
JUAN PABLO ROMERO GODOY

MIEMBROS DE LA COMISIÓN:
FELIPE DE LA FUENTE DÍAZ
BLAS DUARTE ALLEUY

SANTIAGO DE CHILE
2022

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL
POR: **MARCELA PAZ MALDONADO GONZÁLEZ**
FECHA: 2022
PROF. GUÍA: JUAN PABLO ROMERO G.

CONSTRUCCIÓN Y VALIDACIÓN DE UN MODELO DE PROPENSIÓN DE COMPRAS POR SEGMENTO A PARTIR DEL USO DE MODELOS DE MACHINE LEARNING Y EXPERIMENTOS DE CAMPO

La era digital está cambiando los hábitos de las personas, las mantiene informadas y las hace cada vez más exigentes. Esto genera que las empresas tengan el reto de reinventarse para entregar buenas experiencias, responder a las demandas y así, aumentar sus utilidades. La nueva unidad de negocio del Grupo Falabella, Falabella.com nace con el objetivo de que el usuario encuentre en un solo sitio toda la oferta perteneciente a los negocios del mismo. Es en esta empresa donde se realizará el trabajo de título, específicamente, en el área de Customer Behaviour, Segmentation & CRM.

La oportunidad identificada busca analizar los clientes de Sodimac, quienes no se encuentran segmentados, para comprender su comportamiento respecto a las subcategorías de producto similares en el sitio, buscando aumentar la tasa de conversión

El objetivo es generar y validar experimentalmente un modelo que permita aumentar la tasa de conversión de ventas de usuarios segmentados en las subcategorías de producto dentro del sitio web utilizando modelos de Machine Learning y experimentos de campo.

Para cumplir lo anterior, se utiliza la metodología CRISP-DM extrayendo el conocimiento de los datos, pero manteniendo un enfoque aplicado al negocio. Se comienza comprendiendo el negocio, luego se segmenta a los clientes, se corre el modelo de propensión y finalmente, se realiza el experimento.

La segmentación genera 5 clusters donde las variables más relevantes para éstos fueron el género y el tipo de pago relativo al programa de Lealtad que existe en el Grupo Falabella. El modelo de propensión, entrenado con el 70 % de la base y testeado con el 30 % restante, obtiene un rendimiento del 73,1 % y un Recall de 84,7 %. La propensión de compra para cada segmento tiende a ser mayor para las subcategorías que representan en conjunto más de el 50 % de la distribución y se cumple para todos los casos que las subcategorías con mayor propensión son aquellas que pertenecen al top 3 más compradas por cada segmento.

La experimentación consiste en el envío de 10 campañas vía mail donde el 80 % de cada segmento recibe el correo y el 20 % restante no lo recibe mostrando mejores resultados de cara a la tasa de conversión para los clientes que reciben el mail, siendo significativa estadísiticamente en el general de las campañas y para aquellas relacionadas a las subcategorías con mayor propensión de compra por segmento. Lo anterior significa un aumento de 0,7 millones de dólares la venta trimestral de la empresa.

A todas las niñas, adolescentes y mujeres de Chile y el mundo,
para que nunca más nos digan que no podemos.
Somos capaces de todo, no dejemos de soñar y luchar.

Agradecimientos

A mi mamá y papá, por su amor incondicional, por ser los motores de mi vida y por cuidarme y entregarme todo lo necesario para poder llegar hasta donde estoy hoy. Gracias por nunca dudar de mis capacidades, incluso en los momentos más difíciles. A mis hermanas, Victoria y Amelia, por ser las mayores alegrías de mi vida y por enseñarme el arte de ser hermana mayor. Espero que el resultado de este trabajo las motive a pelear por sus sueños; si yo pude, ustedes aún más. A mi Dobby, por mejorar mi vida con su existencia, porque su colita ilumina el mundo entero. Los amo infinitamente.

A los Maldonado y los González, por ser parte de mi vida, por no preguntarme cuánto me faltaba para salir de la u, por extrañarme en todas las juntas familiares que no estuve porque tenía que estudiar y por quererme mucho. Mención especial a la Polyn, por ser la mejor madrina del universo.

A Javier, por ser mi contención, mi apoyo y mi refugio, por siempre creer en mí y por amarme tanto. Gracias por hacer mi vida más bonita y más simple, por calmarme, escucharme y celebrar mis logros como si fueran tuyos. Gracias por darme ánimo y tranquilidad. Todo es más hermoso contigo en mi vida, eres magia. Te amo siempre.

A mis amigas del colegio, Pame y Norma, por enseñarme que las amistades bacanes existen. Gracias por tantas risas y buenos momentos. A mis amigas de la U, Isi y Fer, por hacer más fáciles y divertidos estos años universitarios, por haber sido parte de los momentos más épicos que viví siendo estudiante y por su amitad que vale millones. A Edgar y Javi Fajardo por ser tremendos amigos, gracias por todo el cariño que me entregan. Y gracias, Rodrigo, por tantos aprendizajes.

Gracias a la Feria Empresarial 2016-2017 y el CEIN 2019 por sacarme de la monotonía universitaria y por toda la felicidad que se sentía al sacar los proyectos adelante. Al equipo de Fcom, por permitirme realizar este trabajo de título, por la buena onda y la preocupación constante. A Juan Pablo Romero y Felipe de la Fuente, por la paciencia y el apoyo que fue muy necesario para terminar esta memoria. Agradecimientos también, a Pablito Lastra por bancarme en cada cosa que le pedí con respecto a la memoria y a Ignacio Letelier por impulsar mi acercamiento a los datos y los análisis.

Y por último (pero no menos importante), gracias a mí, por creer en mí. Porque no me la ganó la universidad, ni Mecánica, ni Termo, ni el cansancio. Me agradezco por entender y respetar mis tiempos y necesidades, por ponerme como prioridad y por luchar todos los días para convertirme en una gran profesional y una mejor mujer.

Tabla de Contenido

Índice de Tablas

Índice de Ilustraciones

Capítulo 1

Introducción

1.1. Antecedente generales

1.1.1. Características de la Empresa

El presente trabajo de título es desarrollado en la empresa Falabella.com, la nueva unidad de negocio del Grupo Falabella. Esta empresa pertenece a la industria del Retail en su formato e-commerce (comercio en línea).

Falabella está formada a través de una Sociedad Anónima Abierta (la informaión bursatil de la organización se encuentra en la sección A.1 del anexo) y cuenta con presencia de diversos negocios en más de 4 países. El principal objetivo es: "Satisfacer las necesidades de sus clientes, desde lo más básico presente en la vida cotidiana" [1]. Para lo anterior, cuenta con 6 principales negocios, mostrados en la figura 1.1. El negocio en el cual se desarrolla este trabajo de título nace desde Linio.

Figura 1.1: Negocios de Falabella S.A. Fuente: [2]

Sus valores son cuatro pilares fundamentales que tienen como objetivo llevar a cabo de la mejor manera todo el proceso de venta y relación con los clientes, éstos se exponen en la figura A.2 de la sección A.1 del anexo.

Su organigrama, dimensionamiento de la actividad ralizada y ventaja competitiva en el mercado, se encuentran en las secciones A.2, A.3 y A.4 del anexo, respectivamente.

Con respecto a su estrategia, se desea construir un ecosistema físico-digital, que replique y expanda las relaciones que se tienen con los clientes en el mundo digital, con una propuesta diferenciada, que se apalanca en los activos únicos que han desarrollado en el mundo físico.

En la actualidad, se destaca el esfuerzo que tienen los retailers por conocer las preferencias y necesidades de los clientes a través del tiempo, ya que estos individuos modifican su comportamiento de manera constante. Es así como el consumidor tiende a ser cada vez más exigente con respecto a los productos y al nivel de servicio entregados por las organizaciones, busca una mayor cantidad de variedades y tiende a complementar sus compras realizadas en espacios físicos con las compras realizadas por una plataforma online.

Ante un acelerado crecimiento del canal digital, que continúa creciendo con fuerza en un 123 % respecto al año 2019 [3] y alcanzando un 24 % de participación en el total de ventas, se anunció en septiembre 2020 la transformación de Falabella.com en la única plataforma tecnológica de e-commerce del Grupo, haciendo converger en ella toda su oferta de retail, Marketplace y los productos de Sellers, junto con el desarrollo de nuevas y mejores funcionalidades en las apps de retail y financieras.

Actualmente la empresa se encuentra en una etapa de crecimiento que tiene como objetivo potenciar la experiencia de compra online de sus clientes ampliando sus funcionalidades a través de los canales digitales. Así mismo, continúa trabajando y escalando las capacidades logísticas para ofrecer entregas más rápidas, junto con más y mejores servicios a los Sellers. En paralelo, se continúa fortaleciendo el programa de lealtad como pilar fundamental que proporciona transversalmente importantes beneficios para los clientes, permite conocerlos y generar soluciones personalizadas de acuerdo con sus necesidades.

Falabella.com entonces, trae consigo una gran gama de productos, que concentrará toda la oferta de Falabella Retail, Sodimac, Tottus y Linio, junto a los productos de Sellers. De esta forma, se incluirán productos de tiendas por departamento, mejoramiento del hogar y Marketplace, contando con más de 7 millones de SKUs [3].

Además de esto, se hace cargo de la entrega de los productos comprados dentro de la plataforma la que puede ser de dos maneras, entrega en tienda o despacho a domicilio. Para ello, se están acelerando las inversiones y desarrollando capacidades para apalancar la escala y volumen, buscando tener la red logística más eficiente de sus mercados, ofreciendo servicios de almacenamiento y distribución a retailers y Sellers; con una infraestructura tecnológica escalable y flexible.

Con respecto a sus usuarios y clientes existe una distinción entre ellos. Mientras los usuarios corresponden a cualquier persona que ingrese al sitio web y navegue en éste, los clientes corresponden a usuarios que generaron una orden de compra en el sitio, es decir, que pagaron por uno o más productos. Actualmente el Holding de Falabella posee más de 30 millones de clientes en los más de 4 países donde tiene actividades [3].

Además de lo aterior, se identifican los actores, su relación con la empresa y las regulaciones relevantes en el Anexo B de este documento.

Capítulo 2

Descripción del Proyecto y Justificación

2.1. Información del área de la Empresa

El trabajo de título se realizará el Área de Customer Behaviour, Segmentation & CRM (desde ahora, CB, S & CRM), que pertenece a la gerencia de Loyalty & Personalization. Esta gerencia, además del área antes mencionada, incorpora dos áreas más, On Site Experience y Marketing Directo como se puede ver en la figura C.1 de la sección C del anexo.

El área de CB, S & CRM está orientada a analizar el comportamiento de los clientes de Falabella.com, generar segmentaciones de clientes utilizando la base de datos disponible, realizar reportería, analizar resultados de experimentos y también, analizar o generar campañas respectivas al programa de lealtad de clientes. De esta manera se desea encontrar insights para tomar acciones comerciales en base a la información generada por la explotación de los datos que maneja la organización.

El trabajo de título impactará y beneficiará directamente al área de CB, S & CRM ya que se estará trabajando en un proyecto que actualmente no se está llevando a cabo. Esto principalmente ya que al ser un área relativamente nueva el personal se está acomodando a sus tareas y muchas veces se encuentra sobre exigido con éstas. Además, es parte fundamental de las labores del área y de la empresa identificar segmentos de clientes para poder enviar y analizar campañas y eventos. También, impactará y beneficiará al área de Marketing Directo a quienes se les entregará los segmentos para poder realizar de forma más directa, rápida y óptima el envío de campañas y su posterior análisis.

Por otra parte, el área de One Site Experience, se podrá ver beneficiada ya que con los segmentos y sus propensiones de compra se puede personalizar de mejor forma el sitio web, logrando así que los usuarios puedan tener una mejor experiencia de navegación y compra. Finalmente, Falabella.com será beneficiada indirectamente ya que se le entregará una mejor experiencia de compra a sus usuarios y/o clientes, lo que se alinea con el propósito que tiene la empresa de tener al cliente siempre al centro de toda su gestión, además de aumentar la probabilidad de compra lo que significaría un aumento de sus utilidades.

2.2. Oportunidad identificada

Falabella.com declara que el cliente está en el centro de las decisiones, buscando simplificar sus vidas, transformando sus experiencias de compra, a través de soluciones simples y personalizadas. Además, se enfocan los esfuerzos en conocer y servir a los clientes en todas sus necesidades de consumo, lo que lleva a reinventar continuamente e innovar para darles soluciones flexibles y fáciles para desarrollar relaciones permanentes en el tiempo.

Los clientes son diferentes, tienen necesidades y gustos distintos por lo que reconocer esas diferencias es una oportunidad para el negocio, sobre todo porque se encuentra en una etapa de construcción y crecimiento, pudiendo optimizar sus recursos y aumentando la conversión de los usuarios, su interacción con el sitio y las ventas.

Se debe considerar también, que la atención del usuario es reducida, por lo que, al pasar a distintas páginas del sitio, la cantidad de usuarios que pasan de etapa a etapa se reduce cada vez más, creándose un "funnel" o embudo de conversión de ventas, como se muestra en la figura C.2 de la sección C.2 del anexo, que ejemplifica como los clientes van dejando el sitio mientras avanzan desde la página de inicio (fase "awareness") a páginas más profundas del sitio (fase "desire"), hasta finalmente llegar al carro de compra y el pago (fase "action"). Por lo tanto, será importante buscar estrategias en este medio que permitan aprovechar sus características únicas para mejorar la experiencia de los usuarios, y traducirlo en una mayor penetración, conversión y finalmente, venta del sitio.

Se desea que la unificación, además de contener toda la cartera de productos del holding, generando un gran e-commerce latinoamericano, logre aumentar la venta de productos de los negocios independientes. Es decir, que los clientes de Falabella Retail, Sodimac, Linio y Tottus compren productos de otras unidades de negocio (2P-3P), pero que también logre aumentar la venta por visita de cada negocio individualmente (1P). Es importante mencionar que a pesar de que los productos pertenecen a negocios diferentes, en el sitio web los productos similares estarán agrupados por líneas y subcategorías definidas por Falabella.com.

Actualmente la empresa está trabajando en responder ¿cómo se comportarán los clientes en este nuevo sitio web?, ¿cuál es la probabilidad de que un cliente compre una cierta línea o subcategoría de productos?, ¿qué se debe ofrecer en una campaña para aumentar la conversión de un cliente en el sitio?, ¿a quién le debo enviar cierta campaña para aumentar la compra?.

Es así como este proceso de unificación posee varios desafíos y oportunidades. Uno de ellos es poder segmentar a los clientes que han comprado en alguno de los negocios del holding. Esto permitiría apoyar el trabajo que se realiza en pro de una de las estrategias del negocio para aumentar la tasa de conversión de venta y las visitas a través del envío de campañas de email. Estas campañas se envían a clientes que poseen caracterísiticas similares, agrupándolos en segmentos para hacer más simple el diseño de la campaña y su posterior envío. Específicamente se desea conocer a aquellos clientes que tienen historial de compras en algunos de los negocios independientes del Grupo para poder evaluar cómo se comportan y/o interactúan con la nueva marca y las subcategorías de productos y así, lograr aumentar la tasa de conversión y la venta del sitio.

Por lo tanto, los principales esfuerzos de la empresa y en específico del área donde se llevará a cabo el trabajo de título, se concentran en entender el comportamiento de los clientes que vienen de los negocios independientes del grupo y saber cuál es la probabilidad de que compren una línea o subcategoría de producto en Falabella.com. Está información será entregada al área de Marketing Directo para habilitar las campañas de email marketing, que son enviadas periódicamente según el calendario y las necesidades del Área Comercial.

Se evidencia entonces la oportunidad de realizar un estudio de los clientes de la empresa, en específico de los clientes de Sodimac, identificando distintos tipos de segmentos y sus características para posteriormente, experimentar, en base a un modelo de propensión de compra, con el objetivo de realizar acciones comerciales focalizadas y potenciar la experiencia de compra de estos clientes, aumentando su conversión de venta en el sitio web y también, facilitar y apoyar el proceso de envío de campañas de mail.

En el caso de obtener resultados positivos, la empresa lograría entender el comportamiento de los clientes segmentados de Sodimac con respecto a:

1. **C**orreos enviados desde la nueva marca.
2. **S**ubcategorías de productos similares a las de Sodimac pero pertenecientes a Falabella.com, con una probabilidad de compra respecto a cada una.

Lo anterior podría impulsar el avance de los proyectos de la empresa y el área en pro de aumentar la tasa de conversión del sitio, mejorar la personalización de éste, entregar una mejor experiencia al usuario y comenzar a generar cercanía con esta nueva marca para que sea la primera opción de los clientes de todos los negocios del holding, sobre todo ya que en un futuro será la única plataforma web de éste.

2.3. Propuesta de valor o impacto de la solución planteada

Con la alternativa previamente propuesta para desarrollar el presente trabajo de título, se espera mejorar la pertinencia dentro del sitio web mediante campañas focalizadas por segmento que se cuantifique con el aumento de la tasa de conversión y la venta en el sitio web, siguiendo los pasos mencionados en el apartado anterior.

Una forma de medir el impacto de la propuesta es a través de métricas clave del negocio. Esto se puede hacer a través de Lift Analysis que es una forma de medir como un cambio afecta a una métrica clave. El Lift del modelo es calculado como el porcentaje de crecimiento o decrecimiento de una métrica, para usuarios que reciben este cambio versus un grupo de control para este caso.

Se ha determinado que, durante el primer trimestre del año pasado, las tasas de conversión totales del sitio web han presentado un Lift cercano al 2,5 %[1]. Esto puede significar un aumento de un 2,5 % en las ventas del sitio, si la cantidad de visitas y el gasto promedio por orden se mantiene constante.

Tomando en consideración que las ventas de Falabella.com durante el primer trimestre año 2021 ascienden a cerca de 175 millones de dólares[2] un incremento de un 1 % en la tasa de conversión bajo estas condiciones y tomando como supuesto que en las experimentaciones realizadas el 80 % de los clientes corresponde al grupo de tratamiento y el 20 % al grupo de control, esto significarían aumentar en aproximadamente 1,4 millones de dólares la venta trimestral. En el caso de que en las experimentaciones realizadas el 50 % de los clientes correspondieran al grupo de tratamiento y el 50 % al grupo de control, esto significaría aumentar en aproximadamente 0,88 millones de dólares la venta trimestral.

[1] Datos entregados por el Squad Analytics, 2021
[2] Datos entregados por el Squad Analytics, 2021

Capítulo 3

Objetivos y Alcances

3.1. Objetivo General

El objetivo general es: **"Generar y validar experimentalmente un modelo que permita aumentar la tasa de conversión de usuarios segmentados en las subcategorías de producto de Falabella.com, usando modelos de Machine Learning y análisis experimentales"**

3.2. Objetivos Específicos

Por otra parte, los objetivos específicos se definen a continuación:

1. Conocer el negocio, para identificar los objetivos del área y cómo se trabaja actualmente.
2. Identificar variables y subcategorías relevantes al trabajo mediante el análisis de la data de productos y clientes, lo que permitirá realizar los modelos de mejor manera.
3. Identificar y caracterizar los segmentos de clientes de Sodimac en base a la información transaccional y sociodemográfica de éstos.
4. Determinar la propensión de compra por subcategoría de los segmentos previamente definidos para encontrar la probabilidad de compra de cada segmento por subcategoría.
5. Diseñar y realizar la experimentación de email marketing definiendo el grupo de tratamiento (recibirá el correo) y control (no recibirá el correo).
6. Evaluar el impacto del experimento comparando métricas clave para analizar si se logró un aumento en la tasa de conversión.

3.3. Alcances y resultados esperados

Se define el alcance del trabajo de título:

- Identificación de segmentos con la data sociodemográfica, transaccional y de frecuencia de compra que posee la empresa. Dado que Falabella.com aún no se encuentra puesta en marcha al 100 %, se utilizarán datos de navegación solo para evaluar el experimento pero no se utilizará para generar segmentos.
- Se utilizarán solo una muestra de los datos de Sodimac pues en la actualidad es aquel que tiene una oportunidad de análisis mayor a los otros negocios.
- Se utilizarán datos correspondientes a una ventana de tiempo de no más de 2 años.
- Se utilizarán de Falabella.com solo las subcategorías de producto similares a las de Sodimac, es decir, construcción, herramientas, maquinarias, terrazas, jardín y decoración principalmente. Esto, porque aún no se definen bien los lineamientos con respecto a las líneas blandas y hogar de Falabella Retail y podría requerir un esfuerzo adicional que se escapa de los plazos de la memoria.
- Por temas de tiempo, sólo se generará un modelo de propensión en base a segmentos, puesto que la elaboración de un modelo de propensión para cada cliente requerirá un esfuerzo adicional que se escapa de los plazos de la memoria.
- Los planes o personalizaciones que se deseen hacer con los segmentos y resultados del modelo no formarán parte del trabajo de título y serán las áreas One Site Experience y Marketing Directo quienes velarán por el uso eficiente de esta información.
- Los KPIs para evaluar los experimentos serán la tasa de conversión y la venta por visita, dado que son las principales métricas de evaluación declaradas por el equipo de Analytics, pero se calculará la significancia solo sobre la primera.

Se espera como resultado que el trabajo de título logre generar un modelo validado experimentalmente que permita aumentar la tasa de conversión de los usuarios en las subcategorías de productos del sitio web de Falabella.com, a través de modelos de Machine Learning y análisis experimentales.

Por otra parte, los entregables al área de CB, Segmentation & CRM se resumen en 2 grandes elementos: los segmentos encontrados y la probabilidad de compra de estos segmentos para las líneas de producto de Falabella.com.

En específico, acorde a cada objetivo, los entregables se detallan a continuación:

1. **Identificar variables y subcategorías relevantes al trabajo :** El entregable será una lista de las variables a utilizar para la segmentación.
2. **Identificar y caracterizar los segmentos de clientes:** El entregable será una lista de los segmentos encontrados con su respectiva caracterización y descripción.
3. **Determinar la propensión de compra por línea de los segmentos previamente definidos:** El entregable será la probabilidad de compra de los segmentos para las diferentes subcategorías de productos.

4. **Diseñar y realizar la experimentación de email marketing para probar experimentalmente el efecto de la propensión anterior:** El entregable será el análisis de los resultados del experimento realizado.

5. Además de lo anterior, se entregarán los códigos SQL, Python, la documentación respectiva al trabajo realizado y una propuesta de trabajo futuro con sus recomendaciones respectivas.

Capítulo 4

Marco Conceptual

Se abordará el problema desde las áreas de Marketing Cuantitativo, Minería de Datos o Machine Learning y la evaluación experimental, incluyendo también, nociones de Marketing Digital. Las temáticas principales serán la segmentación y clasificación de clientes, modelos de predicción, experimentación y e-commerce.

A continuación, se realiza un estudio del estado del arte en las temáticas antes mencionadas, con énfasis en las metodologías que existen para generar conglomerados de datos, las métricas usadas en los procesos y los últimos descubrimientos de la academia.

4.1. Segmentación y clasificación de clientes

La segmentación de clientes consiste en clasificar los clientes en grupos homogéneos de comportamiento, de manera que permita a las empresas diferenciar actuaciones a lo largo del ciclo de vida del cliente en función de los segmentos de clientes.

Es el proceso de dividir un grupo de clientes en grupos uniformes más pequeños que tengan características y necesidades semejantes. Los segmentos son grupos homogéneos y que tienen como particularidad que, dentro de cada grupo, es probable que respondan de modo similar a determinadas estrategias de marketing [4].

Los grupos encontrados deben ser lo más heterogéneos posible respecto a los elementos pertenecientes a los otros grupos y lo más homogéneos posible respecto a los elementos que pertenecen al mismo grupo. Además, los segmentos encontrados deben significativos, es decir, tener un tamaño considerable que haga justificable que se traten de manera distinta que los demás segmentos. También deben ser identificables, esto se traduce que contar con segmentos que puedan distinguirse unos de otros.

Finalmente deben ser alcanzables, lo que significa que los segmentos deben ser accesibles para la empresa y pueda implementar estrategias de trabajo distintas para cada uno de ellos [5].

Tradicionalmente se han distinguido dos formas básicas de segmentar un mercado [6]:

- **Segmentación a priori:** Tanto el número de segmentos como su descripción se establece antes de que el estudio se lleve a cabo. El investigador elige primero una base a partir de la cual segmentar el mercado y luego clasifica a los compradores en segmentos de acuerdo con esa designación y estudia en qué medida se relacionan estos criterios con otras variables. La experiencia del responsable de marketing o el conocimiento del mercado son factores que ayudan a la hora de conocer de antemano cuales son los criterios base y los segmentos relevantes. Especialmente relevante para este es la existencia de criterios estándares y otros estudios realizados.

- **Segmentación a posteriori:** Cuando se desconocen las características de mercado resulta más eficaz realizar una segmentación a posteriori. En este modelo, el número de segmentos, su tamaño, y su descripción forman parte del análisis. Primero se realiza una exploración cualitativa para conocer en profundidad el mercado y a continuación se aplica una técnica estadística conocida como análisis de clústeres, que agrupa a los sujetos de acuerdo con la similaridad de sus perfiles respecto a algunas variables preestablecidas.

Existen distintos tipos de segmentación, las que variarán de acuerdo con las variables seleccionadas en cada caso. Existen variables que son muy simples y se manifiestan fácilmente, como el sexo, mientras que hay otras más complejas que están ocultas en los clientes como la opinión de éstos en algún tema. Algunos ejemplos de tipos de Segmentación pueden ser Demográficas como género, edad, estado civil; Sociodemográficas como país, región, comuna; Frecuencia, como el tiempo promedio entre compras o la cantidad de compras en un periodo de tiempo; entre otros.

Los métodos de clustering se dividen en 2 grandes grupos [7]:

- **Algoritmos jerárquicos:** Método que entrega una jerarquía de las divisiones del conjunto de elementos en clústeres.

  Este grupo de métodos tiene su origen en el estudio de las taxonomías, basándose principalmente en la noción de aglomeración y división de conjuntos. Por ejemplo, un método jerárquico de división se inicia con un sólo grupo, que incluye todas las observaciones presentes en la data, para luego en cada paso buscar el mejor corte que separa al grupo en dos subgrupos, y luego este proceso se repite para los subgrupos creados. De manera opuesta, los métodos aglomerativos parten con un grupo en cada una de las observaciones de la data, y luego en cada paso se van juntando pares de grupos que se consideren cercanos. Estos métodos son apropiados cuando se intuye o requiere una estructura jerárquica en la agrupación de los datos, ya que el resultado entrega la información completa de la jerarquía resultante, la cual típicamente se visualiza mediante un dendograma. Algunas aplicaciones típicas de este método son la categorización de especies animales, proteínas y ADN, como también la generación de índices temáticos de sitios web por categorías, e.g., DMOZ y Yahoo! [8].

  Los métodos jerárquicos son algoritmos iterativos en donde de una iteración a otra, se modifica el valor de pertenencia a grupos de un único objeto, En este método no se

requiere fijar un número de segmentos a priori, y su principal ventaja es que la segmentación entregada es fácil de comprender, ya que posee criterios de agrupación bien definidos.

- **Algoritmos de Partición:** Método de dividir el conjunto de observaciones en k clústeres, en donde k lo define inicialmente el usuario.

  Este grupo de métodos se caracteriza por ser en general el más preciso (pero no necesariamente el más válido) para generar un criterio de clustering. En particular, los algoritmos de partición requieren de un parámetro entregado por el usuario, que indique una cantidad fija de clústeres a encontrar dentro del data set en estudio. Luego, a través de una función objetivo, típicamente WGSS (Within-Group Sum of Square errors) se evalúan las particiones candidatas a través de la varianza de los subgrupos resultantes. De esta manera, al encontrar un mínimo local de la función objetivo, se considera haber llegado a un óptimo de clustering para el criterio planteado. Clásicamente estos métodos han estado asociados únicamente a poder encontrar clústeres con formas esféricas y de tamaños similares, no obstante, con el cambio de la métrica de distancia utilizada, es posible encontrar estructuras muy diversas. Algunos métodos clásicos en esta área son el algoritmo K-Means y Fuzzy C-Means [8].

  Son algoritmos iterativos, pero en cada iteración ubican a los objetos en el grupo más cercano a él, en este caso si se necesita un número predeterminado de segmentos donde serás asignados todos los datos. Son asignaciones menos sencillas que las jerárquicas, pero son muy eficientes cuando se trabaja con un gran número de casos. Dentro de esta segmentación se distinguen dos métodos de asignación, los discretos y los difusos. El primero define que cada elemento pertenece a un segmento y el segundo, que cada elemento tiene una probabilidad de pertenecer a un segmento.

  A continuación, se detalla el método de partición más usado [5]:

  - **KMEANS:** El algoritmo de K-Means es una conocida aproximación para particionar un conjunto de datos en K distintos grupos no solapados entre sí.

    La idea básica detrás de k-Means consiste en definir clústeres para que la variación total dentro del clúster (conocida como variación total dentro del clúster) se minimice. Hay varios algoritmos k-Means disponibles. El algoritmo estándar es el algoritmo de Hartigan-Wong (1979), que define la variación total dentro del clúster como la suma de las distancias al cuadrado Distancias euclidianas entre los elementos y el centroide correspondiente:

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \tag{4.1}$$

Donde x_i describe un punto de datos perteneciente al clúster C_k y μ_k es el valor medio de los puntos asignados al clúster C_k

Cada observación (x_i) se asigna a un clúster dado de tal manera que la suma de los cuadrados (SS) de distancia de la observación a sus centros de clúster asignados μ_k un mínimo. Se define la variación total dentro del cúmulo de la siguiente manera:

$$tot.withinss = \sum_{k=1}^{k} W(C_k) = \sum_{k=1}^{k} \sum_{x_i \in C_k} (x_i - \mu_k)^2 \tag{4.2}$$

La suma total del cuadrado mide la bondad del clustering y se desea que sea lo más pequeño posible.

Un requisito para el uso del algoritmo consiste en especificar a priori el número k de clúster que se desea encontrar, asignando a cada observación uno y solo uno de los clústeres encontrados. El algoritmo funciona en base a una iteración de distancias, minimizando las distancias intra-clúster y maximizando las distancias inter-clúster.

Su algoritmo general es el siguiente [5]:

1. Se inicia la K-partición aleatoriamente o basado en algún conocimiento previo. Calcular la matriz $M = m_1, ..., m_k$
2. Asignar cada observación del set de datos al clúster más cercano C_l, i.e.

$$x_j \in C_l, si \parallel x_j - m_l \parallel < \parallel x_j - m_i \parallel$$
$$i \neq j, i = 1, ..., K$$

3. Recalcular M basado en la actual partición
4. Repetir (2) y (3) hasta que no haya cambios en cada clúster

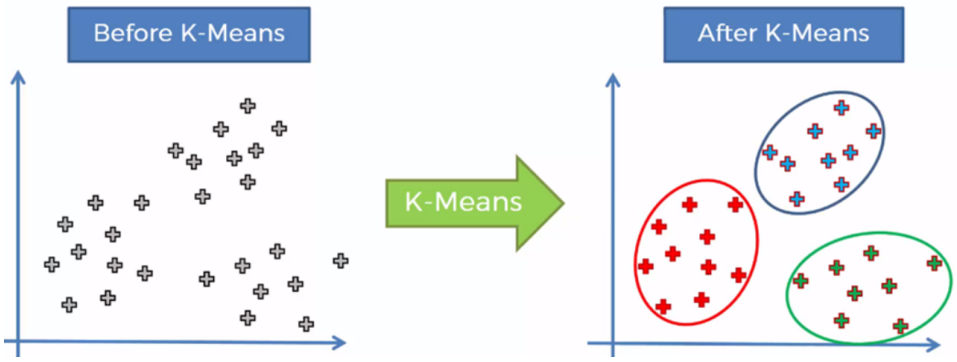


Figura 4.1: Antes y después del algoritmo K-Means. Fuente : [**?**]

En la actualidad existen diferentes aplicaciones para esto tipos de agrupamiento, que van desde medicina (reconocimiento de patrones en imágenes para agrupar células cancerígenas), hasta ciencias sociales agrupando perfiles psicológicos. Sin embargo, también existen desventajas en estas formas de agrupamientos, ya que el análisis de clúster es una metodología descriptiva, ateórica y no inferencial, es decir, no tiene bases estadísticas sobre las que deducir inferencias estadísticas para una población a partir de una muestra. Es un método basado en criterios geométricos y se utiliza fundamentalmente como una técnica exploratoria, descriptiva pero no explicativa [9].

Para evaluar el rendimiento de un algoritmo no supervisado, en específico de un clúster, la dificultad radica en poder determinar si lo que considera "cerca" o "lejos" el algoritmo desarrollado, es realmente "cerca" o "lejos" en la realidad. En específico, se debe buscar una forma de medir cuantitativamente si los elementos ubicados dentro de un clúster pertenecen o pertenecen a dicho conjunto, o más general aún, si dicho clúster tiene sentido. La literatura habla de dos formas diferentes para medir lo anterior, la primera es utilizando datos etiquetados que no se han usado para el clúster, también llamada evaluación externa, o en caso de no poseer estos datos, usar el mismo modelo para evaluar (evaluación interna) [**?**]. El coeficiente de la silueta es una métrica de medida interna que tiene como objetivo encontrar el número óptimo de clúster. La agrupación externa compara el algoritmo generado por el modelo de clustering, con otro que es considerado como "el mejor agrupamiento". Para poder realizar una validación externa, es necesario contar con estos datos de comparación, además de presentar una buena calidad de estos.

4.2. Árbol de Decisiones

Un árbol de decisiones es una representación esquemática que facilita la toma de decisiones al representar visualmente las diferentes posibilidades que existen ante un escenario, además de las posibles consecuencias que cada escenario podría traer. Su nombre se da debido al parecido que tiene el esquema con las ramas de un árbol. Además, puede utilizarse en cualquier aspecto de la vida cotidiana, desde decisiones difíciles en la familia, hasta aplicaciones complejas en los negocios y en la inteligencia artificial.

Su objetivo principal es el aprendizaje inductivo a partir de observaciones y construcciones lógicas. Son muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva para la solución de un problema. Constituyen probablemente el modelo de clasificación más utilizado y popular [10].

El conocimiento obtenido durante el proceso de aprendizaje inductivo se representa mediante un árbol. Un árbol gráficamente se representa por un conjunto de nodos, hojas y ramas. El nodo principal o raíz es el atributo a partir del cual se inicia el proceso de clasificación; los nodos internos corresponden a cada una de las preguntas acerca del atributo en particular del problema.

Cada posible respuesta a los cuestionamientos se representa mediante un nodo hijo. Las ramas que salen de cada uno de estos nodos se encuentran etiquetadas con los posibles valores

del atributo. Los nodos finales o nodos hoja corresponden a una decisión, la cual coincide con una de las variables clase del problema a resolver.

Las ventajas de los árboles de decisión es que son buenos clasificadores y predictores listos para usar. También son útiles para la selección de variables, con los predictores más importantes que generalmente aparecen en la parte superior del árbol. Además, requieren relativamente poco esfuerzo de los usuarios en los siguientes sentidos: En primer lugar, no hay necesidad de transformación de variables (cualquier transformación monótona de las variables dará los mismos árboles). En segundo lugar, la selección de subconjuntos variables es automática, ya que forma parte de la selección dividida.

Los árboles también son intrínsecamente robustos a los valores atípicos, ya que la elección de una división depende del orden de los valores de observación y no de las magnitudes absolutas de estos valores. Sin embargo, son sensibles a los cambios en los datos, e incluso un ligero cambio puede causar divisiones muy diferentes[10].

A diferencia de los modelos que asumen una relación particular entre la respuesta y los predictores (por ejemplo, una relación lineal como en la regresión lineal y el análisis discriminante lineal), la clasificación y los árboles de regresión son no lineales y no paramétricos [10]. Esto permite una amplia gama de relaciones entre los predictores y la respuesta. Sin embargo, esto también puede ser una debilidad: Dado que las divisiones se realizan en predictores individuales en lugar de en combinaciones de predictores, es probable que el árbol pierda las relaciones entre los predictores, en particular las estructuras lineales como las de los modelos de regresión lineal o logística [10].

Existen dos tipos de árboles de decisión, los árboles de clasificación en los cuales la variable respuesta Y es cualitativa y los árboles de regresión, en los cuales la variable respuesta Y es cuantitativa. Ambos son explicados a continuación [10]:

4.2.1. Árboles de Clasificación

En la clasificación, la variable de resultado será una variable categórica. La partición recursiva divide el espacio p-dimensional de las variables x en rectángulos multidimensionales no superpuestos. Las variables X aquí se consideran continuas, binarias u ordinales. Esta división se realiza de forma recursiva (es decir, operando sobre los resultados de divisiones anteriores). Para evaluar la precisión de un árbol en la clasificación de nuevos casos, se comienza dividiendo los datos en conjuntos de entrenamiento y validación. El conjunto de entrenamiento se usa para hacer crecer el árbol y el conjunto de validación se usa para evaluar su rendimiento.

4.2.2. Árboles de Regresión

Funcionan de la misma manera que los árboles de clasificación. La variable de salida, Y, es una variable numérica en este caso, pero tanto el principio como el procedimiento son los mismos: se intentan muchas divisiones, y para cada una, medimos la “impureza” en cada rama del árbol resultante. Hay tres detalles que son diferentes en los árboles de regresión que

en los árboles de clasificación: predicción, medidas de impurezas y evaluación del rendimiento.

Se describe a continuación:

1. **Predicción:** Predecir el valor de la respuesta Y para una observación se realiza de una manera similar al caso de clasificación: la información del predictor se utiliza para "soltar" el árbol hasta llegar a un nodo hoja. En los árboles de regresión, el valor del nodo hoja viene determinado por el promedio de los datos de entrenamiento de esa hoja.

2. **Medición de Impurezas:** Se describen dos tipos de medidas de impurezas para los nodos en los árboles de clasificación: el índice de Gini y la medida basada en la entropía. En ambos casos el índice en una función de la relación entre las categorías de las observaciones en ese nodo. En los árboles de regresión una medida de impureza típica es la suma de las desviaciones cuadráticas de la media de la hoja. Esto es equivalente a los errores al cuadrado porque la media de la hoja es exactamente la predicción. La impureza más baja posible es cero, cuando todos los valores del nodo son iguales.

3. **Evaluación del Rendimiento:** Como se indicó anteriormente, las predicciones se obtienen promediando los valores de las respuestas en los nodos. Por lo tanto, tenemos la definición habitual de predicciones y errores. El rendimiento predictivo de los árboles de regresión se puede medir de la misma manera que se evalúan otros métodos predictivos, utilizando medidas de resumen como el RSS (Residual sum of Squares) que es una medida de la discrepancia entre los datos reales y los predichos por el modelo. Un RSS bajo indica un buen ajuste del modelo a los datos, es decir, se busca minimizar el RSS. Se define el RSS como XXXXXX donde y es el valor real de la variable a predecir y y es el valor predicho:

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{4.3}$$

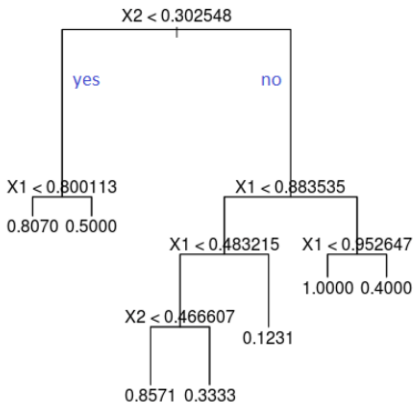


Figura 4.2: Ejemplo Árbol de Decisión.

Existen diferentes criterios para detener el crecimiento del árbol antes de que comience a sobre ajustar los datos, lo que podría traer inconvenientes al momento de obtener resultados del mismo. Algunos ejemplos son la profundidad del árbol (es decir, el número de divisiones), el número mínimo de observaciones en un nodo y la reducción mínima de la impureza [10].

Los métodos anteriores desarrollados se basaban en la idea de la partición recursiva, utilizando reglas para evitar que el árbol crezca excesivamente y sobreajuste los datos de entrenamiento. Un método popular llamado CHAID (detección automática de interacción chi-cuadrada) es un método de particionamiento recursivo que precede a los procedimientos de árbol de clasificación y regresión (CART) por varios años y es ampliamente utilizado en aplicaciones de marketing de bases de datos hasta el día de hoy[10]. Utiliza una prueba estadística de chi-cuadrado para la independencia, logrando evaluar si la división de un nodo mejora la pureza en una cantidad estadísticamente significativa.

La fuerza de asociación se mide por el valor p de una prueba de chi-cuadrado de independencia. Si para el mejor predictor la prueba no muestra una mejora significativa, la división no se lleva a cabo, y el árbol se termina. Este método es más adecuado para predictores categóricos, pero se puede adaptar a predictores continuos agrupando los valores continuos en grupos categóricos.

4.3. Experimentación con AB test

Los AB test en línea comenzaron a utilizarse a finales de la década de 1990 con el crecimiento de Internet. Hoy en día, muchos sitios grandes, incluidos Amazon, Bing, Facebook, Google, LinkedIn y Yahoo! ejecutan de miles a decenas de miles de experimentos cada año probando cambios en la interfaz de usuario (UI), mejoras en los algoritmos (búsqueda, anuncios, personalización, recomendación, etc.), cambios en las aplicaciones, el sistema de administración de contenido, etc. [11].

En un A/B Test, los usuarios se dividen aleatoriamente entre las variantes de manera persistente (un usuario recibe la misma experiencia en varias visitas). Sus interacciones con el sitio se instrumentan y se calculan métricas clave. La figura 11 muestra la estructura de un experimento A/B. En la práctica, se puede asignar cualquier porcentaje al tratamiento y control, pero el 50 % proporciona al experimento la máxima potencia estadística [11].

En un sentido general, el análisis probará si la distribución estadística del tratamiento es diferente de la del control. En la práctica, la prueba más común es si las dos medias son iguales o no. Para este caso, el efecto de la versión B (o efecto de tratamiento) se define como [11]:

$$E(B) = \bar{X_B} - \bar{X_A} \tag{4.4}$$

Donde X es una métrica de interés y x es la media de la variante B. Sin embargo, para la interpretabilidad, el cambio porcentual normalmente se informa con un adecuado (por ejemplo, 95 %) intervalo de confianza. La variante que muestra una mejora estadísticamente

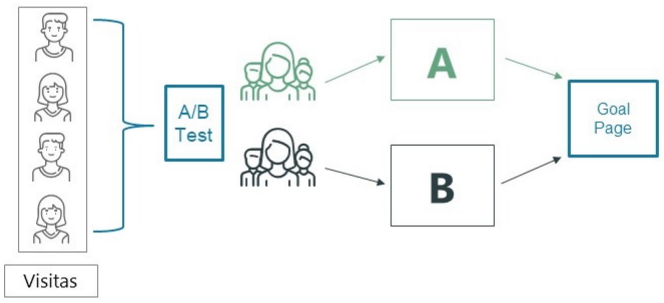


Figura 4.3: Proceso A/B Test. Fuente: Elaboración propia.

significativa se conserva, mientras que la otra se descarta.

La prueba de Hipótesis está dada por:

$$\begin{aligned} H_0 &: \bar{x_A} = \bar{x_B} \\ H_1 &: \bar{x_A} \neq \bar{x_B} \end{aligned} \tag{4.5}$$

Cuando se realiza una prueba de si el tratamiento tuvo un efecto o no, a menudo se produce un valor de p que es la probabilidad obtener un efecto igual o más extremo que el observado dado que la hipótesis nula es verdadera (Biau, Jolles y Porcher 2010).

4.4. E-commerce y Marketing Digital

Existen diversas definiciones de e-commerce, pero todas ellas se engloban en la misma filosofía. El comercio electrónico es definido por Del Águila (2000) como "el desarrollo de actividades económicas a través de las redes de telecomunicaciones". Por tanto, puede ser considerado como "todo tipo de negocio, transacción administrativa o intercambio de información que utilice cualquier tecnología de la información y las comunicaciones" (Comisión Europea, 1999) o simplemente "hacer negocios electrónicamente" (Comisión Europea, 1997).

El e-commerce se caracteriza por su alcance global, ya que a diferencia de los mercados tradicionales o tiendas físicas no existe una la barrera relacionada al lugar, permitiendo ofrecer y comprar productos sin importar donde se esté físicamente; solo se necesitará que la empresa llegue al lugar solicitado y que cuente con un dispositivo conectado a internet el cual es indispensable ya que corresponde al medio para poder realizar la venta.

El e-commerce es una realidad actual para las empresas que desean mantenerse en el mercado, debido a que permite poder llegar a más personas de manera rápida. Adicionalmente, permite poder obtener una data más precisa del perfil del consumidor logrando así ofrecer un producto más adecuado para éste.

Cuando los consumidores compran en línea, pasan por una experiencia de compra y navegación por etapas en la que tanto el producto como el sitio web afectan los resultados de

navegación y la duración de la visita, que luego influye en la decisión de compra (Mallapragada, Chandukala, Liu, 2016). La forma de comprar del consumidor se ve influenciada por relaciones psicológicas, sociales y económicas, sumado a esto, se tiene en cuenta que el consumidor posee tres tipos de motivaciones: racional, emocional y de lealtad o apoyo a la marca (Koumbis, 2015).

Es por lo anterior que el marketing digital ha sido de gran relevancia en el último tiempo, debido al auge de las redes sociales y el aumento del tráfico de visitas en internet. De manera preliminar, el marketing digital se puede definir como la introducción de la tecnología a la realización de acciones publicitarias en los distintos medios digitales. Por otro lado, se define como "estrategias de mercadeo que se realizan en la web, para que un usuario en un determinado sitio concrete su visita tomando una acción planteada con anterioridad" (Habyb Selman, 2017) [12]. Junto con lo anterior, se señala que la conversión es otro concepto relevante para el marketing digital y lo define como el "proceso por el que se logra que un usuario que visita un sitio web realice la acción comercial que uno desea, por ejemplo, comprar un producto, suscribirse a un boletín, otorgar información de contacto, entre otras" (Habyb Selman, 2017) [12]. Según lo planteado, el marketing digital cuenta con 2 aspectos fundamentales:

- **Personalización:** El marketing digital es personalizable debido a que los sistemas digitales que manejan estas estrategias permiten la creación de perfiles de usuarios más allá de las características demográficas, pudiendo recopilar cualquier tipo de preferencia, gustos e intereses de los visitantes de una determinada página web.
- **Masividad:** El marketing digital puede alcanzar a una mayor audiencia objetivo con un menor presupuesto que en el caso del marketing tradicional, es decir, para el mismo gasto, el marketing digital tiende a ser más masivo que el tradicional. El marketing digital cuenta con las 4 F's, las que ayudan a definir el flujo de relación que debe existir entre una empresa y los prospectos que visiten sus páginas web y redes, las cuales son:
  - **Flujo:** Se definen aspectos estratégicos relacionados a las iniciativas de marketing, qué se ofrece, qué grado de interactividad se buscará con los prospectos, qué estrategia online se buscará, diferenciación con respecto a la competencia, entre otros.
  - **Funcionalidad:** Se define la estructuración de los canales digitales a implementar, ya sean anuncios, páginas web, videos, etc. Además, se establecen los llamados a la acción, la propuesta para captar la atención de los prospectos y la acción deseada por parte de ellos.
  - **Feedback:** Se define cómo se atiende la necesidad de los usuarios y prospectos, estableciendo canales de atención, entablando cómo se llevará a cabo el diálogo con los clientes y además establecer qué se hará con la información entregada por los prospectos.
  - **Fidelización:** Se debe establecer cómo se realizará la fidelización del cliente para evitar su abandono, definiendo herramientas y acciones necesarias para mantener la relación empresa-cliente, qué beneficios pueden obtener los clientes, y cómo se entregan los mensajes publicitarios a los usuarios que ya son clientes fidelizados.

Capítulo 5

Metodología

Para llevar a cabo el trabajo de título y cumplir los objetivos planteados, se utilizará la metodología Cross Industry Standard Process for Data Mining (CRISP-DM) con herramientas aplicadas en Marketing Cuantitativo, Minería de Datos y Experimentación. Esta metodología surge a partir del término KDD, utilizado para extraer conocimiento de los datos, pero manteniendo un enfoque aplicado en el negocio. La metodología contempla como eje central el análisis de datos y presenta una estructura moldeable acorde a las necesidades del proyecto.

CRISP DM cuenta 6 etapas que se presentan a continuación, las cuales se seguirán para la realizar el trabajo de título [13]:

1. **Comprensión del negocio:** Etapa que se enfoca en el conocimiento y entendimiento de los propósitos y requerimientos del negocio y del proyecto, con el objetivo de transformarlos en metas de tipo técnico y desarrollar un plan de proyecto de minería de datos.
2. **Comprensión de los datos:** Etapa que busca familiarizarse con los datos, a través de su recopilación inicial, descripción y exploración, buscando verificar calidad y usabilidad de ellos.
3. **Preparación de los datos:** Etapa que tiene como objetivo construir el conjunto de datos que ser utilizado en la etapa siguiente. Para ello se debe seleccionar los datos relevantes, limpiarlos, integrarlos y formatearlos según corresponda.
4. **Modelado:** En esta etapa se selecciona y aplica las técnicas de modelado que sea más apropiada para el problema o proyecto, abarcando el diseño, construcción y evaluación de éste.
5. **Evaluación:** Se busca evaluar los resultados obtenidos en la etapa anterior, considerando si son útiles para las necesidades del negocio, la fiabilidad de los resultados y estableciendo los siguientes pasos o acciones a seguir.
6. **Implementación:** A esta etapa se llega una vez que el modelo ha sido validado. Consiste en la transformación del conocimiento obtenido en acciones dentro del proceso de negocio, a través de dos mecanismos, recomendaciones o la aplicación del modelo en otros conjuntos de datos.

Como se puede ver en la figura 5.1, esta metodología no sigue un único flujo y permite un movimiento más libre entre etapas, permitiendo así generar modelos que aporten valor de acuerdo con las necesidades del negocio utilizando de la mejor manera la información disponible. Es por esto que se utilizará esta metodología en la construcción del trabajo de título, ya que además de poder ser aplicada a proyectos de minería de datos, posee una etapa que incorpora al negocio y, además, permite la iteración o repetición de etapas de ser necesario.

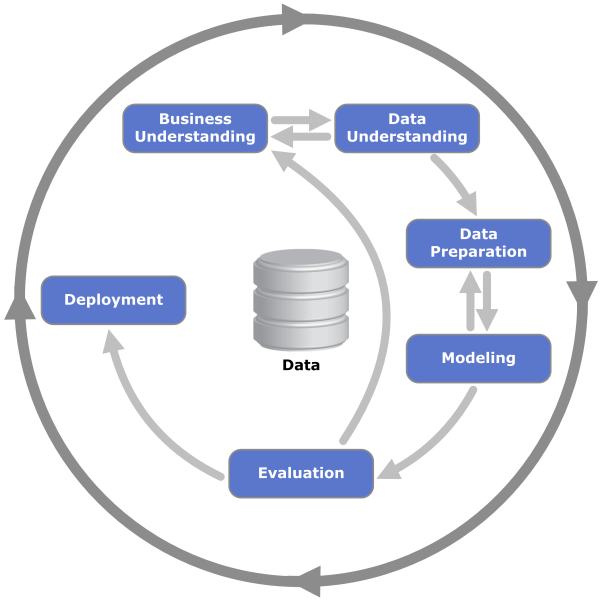


Figura 5.1: Flujo de proceso de la metodología CRISP-DM. Fuente: [**?**]

En base a las etapas de la metodología ya descrita, se propone a continuación el desarrollo de cada uno de los pasos aplicados al contexto del presente trabajo de título:

1. **Comprensión del negocio:** Para lograr esta etapa será fundamental la comunicación con la gerencia de Loyalty & Personalization, en específico con el área de Customer Behaviour, Segmentation & CRM, quienes aportarán con el conocimiento del negocio necesario para lograr la comprensión completa del proyecto a realizar. Las principales tareas por realizar son:

    - Entender cuáles son las labores de los equipos de la gerencia de Loyalty & Personalization y la relación que existe entre ellos y con los clientes.
    - Comprender los insights significativos con respecto a la segmentación de clientes.
    - Entender la forma en que se aplican reglas o modelos para cumplir con los objetivos de los equipos y la definición de los resultados esperados.
    - Navegar en el sitio web de Falabella, Sodimac y Linio para comprender cuáles son sus fortalezas y debilidades y, cómo se relacionan con el usuario.
    - Se realizarán reuniones día por medio de avance, para mostrar resultados y recibir feedback oportuno, además de una comunicación contante mediante las plataformas de comunicación que posee la empresa.

2. **Compresión de los datos:**

   - Se solicitará acceso a los servidores de Falabella.com, Google Cloud Plataform para poder tratar los datos. Con este acceso se podrá acceder a toda la información transaccional que posee Falabella.com, los datos de compras realizadas por los clientes, los datos sociodemográficos construidos y manejados por la empresa.
   - Se realizará una comprensión de la arquitectura en la cual se encuentran alojados los datos, para poder facilitar su posterior utilización.
   - Revisar la data disponible desglosando las variables que poseen, su antigüedad, la cantidad de datos que se tienen y las relaciones entre tablas deser necesario, todo esto a través de querys en lenguaje de programación SQL.

3. **Preparación de los datos:**

   - Realizar un análisis exploratorio de los datos.
   - Se realizará una normalización de los datos, para evitar problemas de unidades.
   - Se definirá entre que ventana de tiempo se utilizarán los datos.
   - Se generarán tablas con las variables más relevantes de información de clientes, transacciones realizadas y productos comprados.
   - Definir los segmentos a utilizar para correr el modelo de propensión del punto siguiente.

4. **Modelado:**

   - Se comenzará utilizando Árboles de decisión y según los resultados obtenidos, se realizará otra iteración con un algoritmo diferente, de ser necesario.
   - Según los resultados obtenidos, se iterará modificando o manteniendo la ventana de tiempo definida en el punto anterior o los hiperparámetros del mismo.
   - Se procederá a correr el modelo.

5. **Evaluación:**

   - Analizar la pertinencia de los resultados en base a las características del negocio e iterar nuevamente la etapa de modelado hasta encontrar una solución aparentemente óptima.
   - Seleccionar KPIs relevantes para el negocio, con los que se medirán los resultados de la experimentación.
   - Se diseñará un experimento de campo para medir los resultados en la variación de la conversión obtenida para las nuevas segmentaciones, comparado con un grupo de control del mismo segmento.

6. **Implementación:**

   - Implementar experimentación definida anteriormente y analizar resultados obtenidos y de ser favorables se podrían utilizar los segmentos de forma permanente para enviar campañas de mail y personalizar el sitio web de Falabella.com.

Capítulo 6

Desarrollo metodológico

De acuerdo a la metodología descrita en el capítulo anterior, se presenta a continuación el desarrollo realizado para cada uno de los pasos asociado al CRISP-DM, aplicado al caso de estudio.

6.1. Comprensión del Negocio

Esta etapa fue fundamental para definir los alcances del trabajo de título y los pasos a seguir. Se comenzó comprendiendo las labores de los equipos de la gerencia de Loyalty & Personalization y la relación que existe entre ellos para posteriormente, comprender los insights significativos con respecto a la segmentación de clientes y envío de campañas. Además, se pudo entender la forma en que se aplican reglas o modelos para cumplir con los objetivos de los equipos y la definición de los resultados esperados.

Paralelo a esto, se comprendió cuál es la relación de la empresa con usuarios y clientes, cuáles son los momentos en que ellos interactúan con el sitio, la comunicación y los diferentes canales que existen para poder brindar un servicio durante todo el proceso de compra, desde su primera visita a la post venta. Finalmente, se conocieron todos los proyectos que se están diseñando o desarrollando en pro de mejorar la experiencia en la página web y así, lograr que los usuarios lleguen a convertirse en clientes.

Todo lo anterior sirvió para entender las reales necesidades del área y definir cómo el trabajo de título aportaría en su estrategia y en las tareas a realizar. Paralelo a esto, se navegó en los sitios web de Falabella, Sodimac y Linio para comprender cuáles son sus fortalezas y debilidades y, cómo se relacionan con el usuario.

Después de toda esta exploración, se decidió en conjunto con el Squad de Analytics que se trabajaría con los clientes de Sodimac, pues es el negocio que está más atrasado con segmentación y el envío de sus campañas versus los otros negocios del grupo, según la planificación realizada para el segundo semestre del 2021. Además, es importante poder saber las preferencias de estos segmentos de clientes ya que en su mayoría compran de forma offline, lo que genera una oportunidad de atraerlos también al canal online ofreciéndoles productos que sean de su interés como lo podrían ser las subcategorías de productos similares a la cartera de productos de Sodimac.

6.2. Obtención, comprensión y preparación de los datos

Para la realización de este trabajo se utilizaron varias bases de datos obtenidas desde el Google Cloud Platform mediante lenguaje SQL y trabajadas con Python. Específicamente se trabajó con bases de datos transaccionales y sociodemográficas de los clientes de Sodimac.

Las variables seleccionadas para trabajar son algunas de las usadas por la empresa como variables de interés al momento de definir segmentos y trabajar con clientes. Se seleccionaron aquellas que tenían más del 90 % de valores no nulos dado que este tipo de datos podía ensuciar mucho la muestra. En específico, se descartaron variables educacionales y de empleo, de localidad y las de navegación (lo que fue explicado en los alacances de este trabajo). Se describen en la tabla 6.1.

Tabla 6.1: Variables que se utilizarán. Fuente: Elaboración Propia.

Nombre	Descripción
Órdenes	*Cantidad de compras mensuales*
Venta	*Cantidad monetaria gastada mensualmente*
Fl_hijos	*Variable dummy, 1 si tiene hijos, 0 si no*
Cust_age	*Edad del cliente*
Loyalty_points	*Cantidad de puntos del programa de lealtad*
Channel_Online	*1 si posee compra online, 0 si no*
Channel_Offline	*1 si posee compra offline, 0 si no*
Prod_sub_categorías_desc	*Subcategoría comprada*
Género	*Género del cliente*
Loyalty_type	*Tipo de pertenencia del cliente al programa de lealtad de Falabella*

Las variables *órdenes* y *venta* son las más importantes al momento de analizar cualquier evento pues de estas dos nacen los KPIs más relevantes para el trabajo del área, como lo son la tasa de conversión y la venta por visita.

Luego de elegir las variables, se realizó su normalización respectiva. Primero, todas aquellas variables tipo "string" (subcategorías, channel, género, loyalty type) se convirtieron en variables dummy. En aquellas filas con 'NA', las columnas respectivas se completaron con 0.

Para las variables de edad y loyalty points se realizaron regresiones lineales con el objetivo de distribuir los valores conocidos en aquellos 'NA'.

La correlación ($> 0,15$) entre las variables se muestra en la figura 6.1. Las variables con mayor correlación (0,73) son "venta" con "órdenes". A pesar de ser mayor que el umbral decidido para la discriminación que fue de 0,5 se dejaron ambas variables por ser variables de completo interés para la empresa. Finalmente, el dataset final a utilizar posee 1.233.490 clientes únicos.

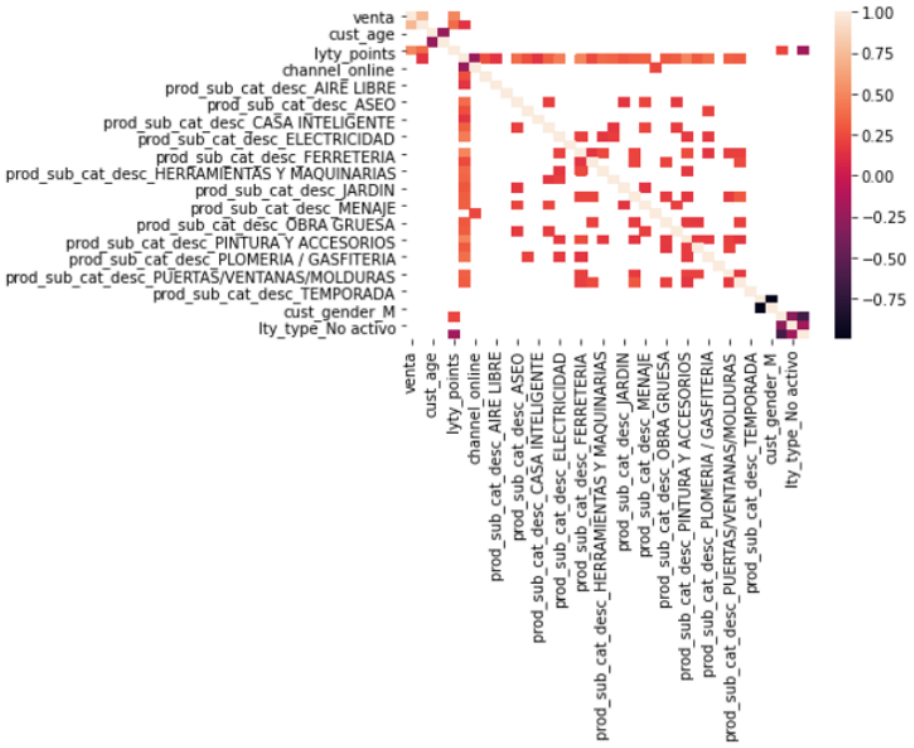


Figura 6.1: Correlación entre Variables $> 0,15$. Fuente: Elaboración propia.

6.2.1. Análisis Exploratorio

Posterior a la selección de variables a utilizar se genero el respectivo análisis exploratorio de ellas, donde los principales hallazgos de los datos se detallan a continuación.

Con respecto al canal de venta, como se puede ver en la figura 6.2 el 88,7 % de los clientes compra en tienda (canal offline) y 11,3 % lo hace en el sitio web (canal online). Además, se puede ver la distribución por sexo, donde el 59,8 % de los clientes son hombres y el 40,2 % corresponden a mujeres.

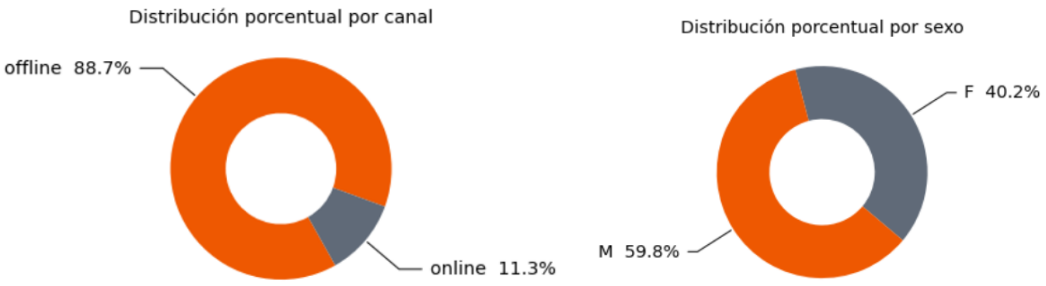


Figura 6.2: Distribución porcentual por canal y sexo. Fuente: Elaboración propia.

La distribución por edad de los clientes se comporta parecido a un gráfico de distribución normal como se aprecia en la figura 6.3, cuyo promedio se encuentra en los 44 años.

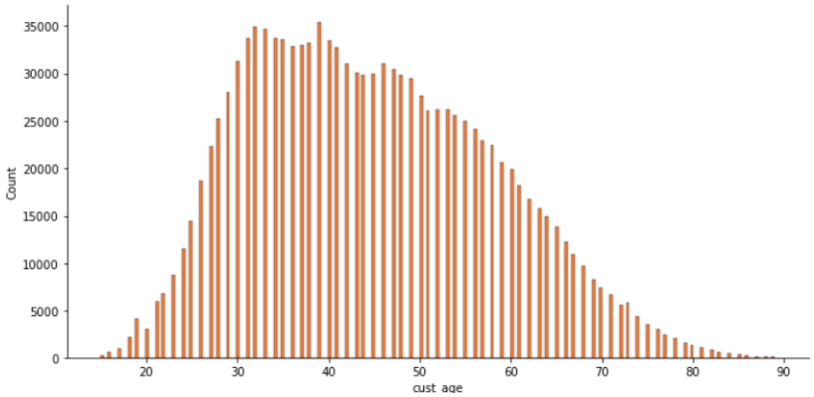


Figura 6.3: Distribución por edad. Fuente: Elaboración propia.

Las subcategorías de productos de Sodimac son 26 y la distribución de compra se muestra en la figura D.1 de la sección D.1 en los anexos, destacando las subcategorías de Ferretería (10 %), Pintura y Accesorios (9 %), Electricidad (6 %), Herramientas y Maquinaria (6 %), Decoración (6 %) y Jardín (6 %) que significan en conjunto más de el 40 % de las ventas del negocio.

Finalmente, la distribución por tipo de loyalty muestra que el 55 % de los clientes pertenecen al programa de fidelización del tipo CMR, el 37 % pertenecen al tipo Todo Medio de Pago (TMP) y el 9 % no está inscrito en el progama, lo que se define como "no activo", lo que se puede ver en la figura 6.4.

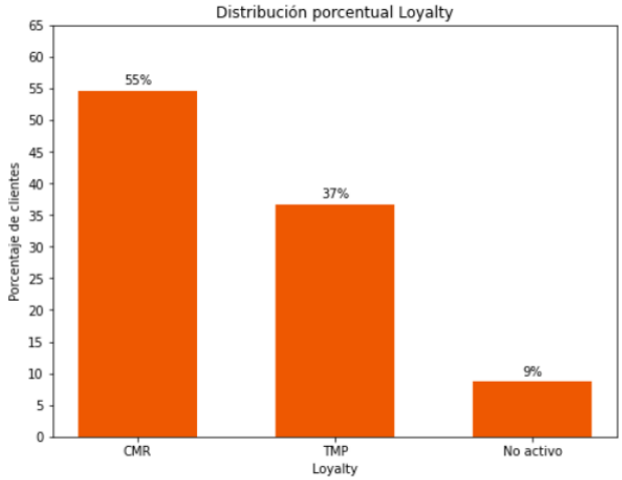


Figura 6.4: Distribución porcentual por tipo de Loyalty. Fuente: Elaboración propia.

6.2.2. Segmentación

Para realizar los segmentos se utilizó el modelo K-means. Primero fue necesario decidir cuántos segmentos se utilizarían para lo que se realizó un diagrama del codo como se aprecia

en la figura 6.5. Se elegirá la cantidad de segmentos según el primer punto que tenga la menor distancia con su sucesor.

Para este caso, el diagrama del codo se muestra en la figura 6.5.

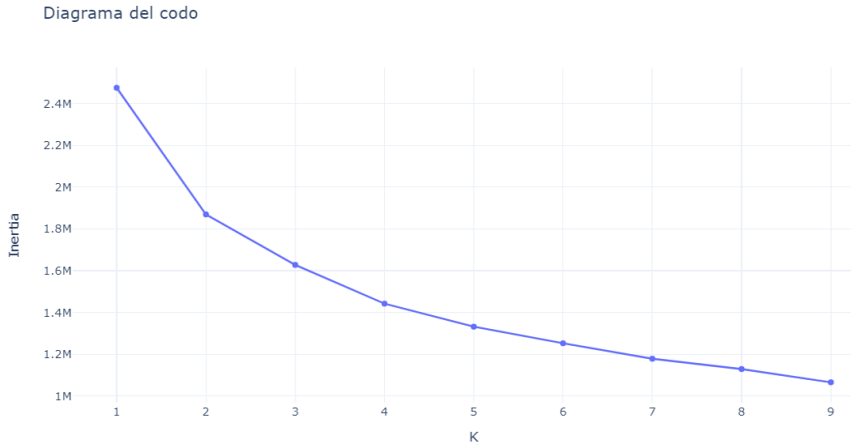


Figura 6.5: Diagrama del codo. Fuente: Elaboración propia

La diferencia entre los puntos se puede ver en la tabla D.1 de la sección anexos, donde la primera menor distancia está entre el punto 5 y 6, por lo que se decide utilizar solo 5 segmentos para el K-means.

Al correr el modelo de clusterización, cada segmento se diferencia del anterior por sus centroides y el peso de diferentes variables con respecto a él mismo. En la figura 6.6 se puede notar el peso de las variables para cada segmento, el diccionario de las variables está adjunto en el apartado D.2 de la sección anexos.

Las variables más relevantes para estos clústeres son las que se muestran en la tabla 6.2, 2 de ellas relacionadas al género del cliente y las otras 2, al tipo de Loyalty.

Tabla 6.2: Variables con mayor relevancia para los clústeres. Fuente: Elaboración Propia.

Variable	**Descripción**
v34	*Género Femenino*
v35	*Género Masculino*
v36	*Loyalty CMR*
v38	*Loyalty Todo Medio de Pago*

A continuación, se presenta la caracterización de los segmentos según sus variables de mayor peso:

1. **Segmento 1 - "Hombre Hogar":** Este segmento está compuesto por un 99 % de hombres, donde el 37 % no están activos en el programa Loyalty. En promedio compran

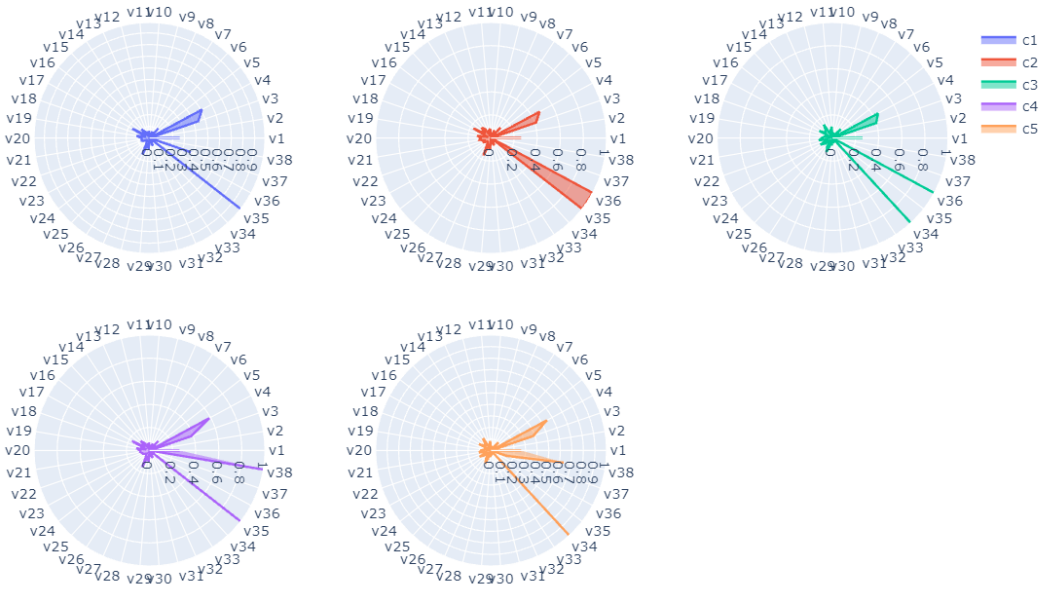


Figura 6.6: Distribución de los centroides por segmento. Fuente: Elaboración propia.

al mes 2,9 veces en el canal offline, el 50 % tiene hijos y su edad promedio es de 47 años. Su venta, órdenes y puntos promedio son $130.411, 3,8 y 5.198 respectivamente. Finalmente, las 3 subcategorías más compradas por este segmento son Electrohogar (11,3 %), Pintura y Accesorios (10,1 %) y Fierro/Hierro/Acero (7,8 %).

2. **Segmento 2 - "Hombre Constructor CMR":** Este segmento está compuesto en su totalidad por hombres y el 100 % pertenecen al programa Loyalty del tipo CMR. En promedio compran 3,2 veces en el canal offline, el 48 % tiene hijos y su edad promedio es de 45,5 años. Su venta, órdenes y puntos promedio son $169.919, 4,8 y 15.321 respectivamente. Finalmente, las 4 subcategorías más compradas por este segmento son Ferretería (10,7 %), Pintura y Accesorios (9,1 %), Herramientas y Maquinaria (7,2 %) y Electricidad (7,2 %).

3. **Segmento 3 - "Mujer Deco CMR":** Este segmento está compuesto en su totalidad por mujeres y el 100 % pertenecen al programa Loyalty del tipo CMR. En promedio compran al mes 2,7 veces en el canal offline, el 45 % tiene hijos y su edad promedio es de 44,5 años. Su venta, órdenes y puntos promedio son $131.922, 4,1 y 12.097 respectivamente. Finalmente, las 4 subcategorías más compradas por este segmento son Decoración (8,4 %), Pintura y Accesorios (7,8 %), Ferretería (7,3 %) y Menaje (7,3 %).

4. **Segmento 4 - "Hombre Constructor TMP":** Este segmento está compuesto en su totalidad por hombres donde el 100 % pertenece programa Loyalty de tipo Todo Medio de Pago (TMP). En promedio compran al mes 2,9 veces en el canal offline, el 59 % tiene hijos y su edad promedio es de 43 años. Su venta, órdenes y puntos promedio son $146.132, 4,4 y 2.287 respectivamente. Finalmente, las 3 subcategorías más compradas por este segmento son Ferretería (10,5 %), Pintura y Accesorios (9,5 %), Herramientas y Maquinaria (7,1 %).

5. **Segmento 5 - "Mujer Deco TMP":** Este segmento está compuesto en su totalidad

por mujeres donde el 63,3 % pertenecen al programa Loyalty del tipo Todo Medio de Pago (TMP). En promedio compran al mes 2,5 veces en el canal offline, el 55 % tiene hijos y su edad promedio es de 43,5 años. Su venta, órdenes y puntos promedio son $104.728, 3,6 y 2.217 respectivamente. Finalmente, las 4 subcategorías más compradas por este segmento son Decoración (8,3 %), Ferretería (8,3 %), Pintura y Accesorios (8,0 %), Menaje (7,8 %) y Jardín (7,8 %).

La distribución porcentual por segmento se muestra en la figura D.3 donde el 27 % (328.297 clientes) pertenecen al segmento 2, el 22 % (266.383 clientes) al segmento 3, el 20 % (249.550 clientes) al segmento 5, el 18 % (220.047 clientes) al segmento 4 y el 14 % (169.213 clientes) al segmento 1.

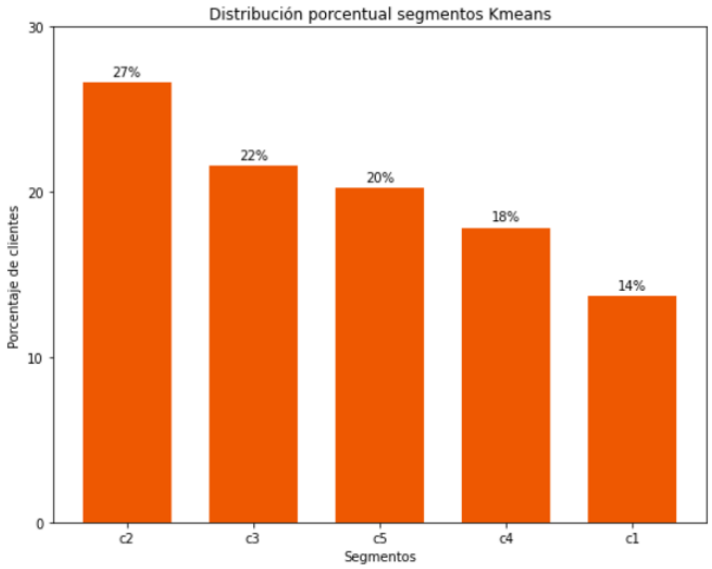


Figura 6.7: Distribución porcentual por segmento. Fuente: Elaboración propia.

6.3. Modelado

La base de datos necesaria para poder aplicar el modelo de propensión se generó agregando la columna segmento a la base de datos de clientes obtenida luego de la limpieza anteriormente realizada, contando finalmente con 5 segmentos y 23 subcategorías.

Antes de ejecutar el modelos de clasificación, la base de datos se divide en dos grupos de similares características para poder validar de manera adecuada el desempeño de éste. Para realizar este proceso el 70 % de la base se utilizará como entrenamiento y el otro 30 % para testear. Esta división se realiza ya que en el entrenamiento se pueden sobre ajustar los resultados haciendo necesario observar errores en datos diferentes a los del entrenamiento.

Con el fin de que el proceso cumpla con su propósito se debe verificar que la división de los datos cumpla con la misma distribución de observaciones etiquetadas por segmento,

para evitar un desbalanceo en el entrenamiento que conllevaría a errores en el testeo. La distribución por segmento para ambos grupos se puede ver en la tabla 6.3.

Tabla 6.3: Distribución por segmento del Entrenamiento y Testeo. Fuente: Elaboración Propia

Segmento	**Entrenamiento**	**Testeo**
1	*26,5 %*	*26,9 %*
2	*22,6 %*	*22,3 %*
3	*19,7 %*	*19,8 %*
4	*17,9 %*	*17,6 %*
5	*13,3 %*	*13,4 %*

Luego de varias pruebas para afinar el modelo, se definen los hiperparámetros a utilizar. Como medida de pureza de clases se utilizará el índice de Gini, la profundidad máxima se fijó en 25 y el nivel de confianza en 0,25. Este parámetro indica que la clasificación errónea no puede ser mayor al 25 % en cada nodo.

Las métricas obtenidas después de correr el modelo se muestran en la tabla 6.4 identificando que modelo tiene un rendimiento general de 73.1 % y el Recall obtuvo un valor igual a 84,7 %. Este porcentaje representa que tan bien se ajusta la probabilidad de compra por cada segmento para cada subcategoría.

Tabla 6.4: Métricas del modelo. Fuente: Elaboración Propia

Accuracy	**Recall**
0,731	*0,847*

6.4. Evaluación

Esta etapa de la metodología tiene como objetivo poder diseñar la experimentación a realizar en base a los resultados del modelo.

Los resultados obtenidos en el modelo se presentan el la figura 6.8, donde se puede ver la probabilidad de compra por cada línea de producto de Falabella.com por segmento. Las mayores probabilidades de compra tienden a concentrarse en aquellas líneas que poseen mayor distribución, a excepción del segmento 1 con las líneas de Electrohogar y Fierro.

Se destacan a continuación, aquellas probabilidades de compra mayores a 10 % por segmento:

- El segmento 1 tiene un 25 % de probabilidad de comprar Electrohogar, 24 % de probabilidad de comprar Pintura y Accesorios y 18 % de comprar productos de la subcategoría de fierro. Esto quiere decir que de 100 compras que haga este segmento, 67 podrían ser en

		Probabilidad de compra				
Subcategoría	**Distribución**	**S1**	**S2**	**S3**	**S4**	**S5**
FERRETERÍA	10%	4%	15%	11%	15%	12%
PINTURA Y ACCESORIOS	9%	24%	12%	12%	14%	11%
ELECTRICIDAD	6%	3%	10%	3%	9%	4%
HERRAMIENTAS Y MAQUINARIA	6%	3%	10%	4%	8%	4%
DECORACIÓN	6%	1%	4%	12%	5%	12%
JARDIN	6%	1%	7%	10%	6%	10%
MENAJE	5%	3%	7%	10%	6%	10%
PLOMERIA/GASFITERÍA	5%	1%	5%	1%	4%	1%
ILUMINACION Y VENTILADORES	5%	2%	5%	6%	4%	5%
ASEO	5%	0%	1%	4%	2%	3%
ORGANIZACIÓN	4%	0%	1%	5%	2%	5%
BAÑOS Y COCINA	4%	4%	5%	5%	5%	5%
TABIQUERIA/TECHUMBRE/AISLACION	4%	1%	2%	1%	4%	1%
ACCESORIOS AUTOMOVILES	4%	2%	2%	1%	3%	0%
ELECTROHOGAR	4%	25%	2%	3%	2%	3%
MUEBLES	3%	1%	1%	2%	2%	2%
MADERA Y TABLEROS	3%	0%	1%	1%	2%	2%
PISOS	3%	1%	2%	1%	1%	1%
PUERTAS/VENTANAS/MOLDURAS	2%	2%	2%	1%	1%	1%
OBRA GRUESA	2%	1%	1%	1%	1%	1%
AIRE LIBRE	2%	1%	2%	3%	1%	2%
FIERRO/HIERRO/ACERO	1%	18%	2%	1%	2%	2%
CASA INTELIGENTE	1%	2%	1%	2%	1%	3%

Figura 6.8: Resultados Modelo de Propensión. Fuente: Elaboración Propia

alguna de estas subcategorías.

- En el caso del segmento 2, se diversifican un poco las probabilidades. Existe un 15 % de probabilidad de compra Ferretería, 12 % de probabilidad de comprar Pintura y Accesorios y 10 % de comprar Electricidad o Herramientas y Maquinaria.

- Para el segmento 3, existe un 12 % de probabilidad de comprar Pintura o Decoración, 11 % de probabilidad de comprar Ferretería y 10 % de comprar Jardín o Menaje.

- El segmento 4, tiene un 15 % de probabilidad de comprar Ferretería y un 14 % de probabilidad de comprar Pintura y Accesorio.

- Finalmente, el segmento 5, tiene un 12 % de probabilidad de comprar Ferretería o Decoración, 11 % de probabilidad de comprar Pintura y 10 % de comprar Jardín o Menaje.

Con estos resultados ya es posible obtener un primer insight sobre el comportamiento de los segmentos por subcategoría de productos, lo que se desea probar experimentalmente a través de envíos de campañas creadas por el Área Comercial.

6.4.1. Diseño Experimentación

Se realizará un experimento de campo del tipo AB test para medir los resultados de la conversión obtenida para las nuevas segmentaciones, comparado con un grupo de control del mismo segmento. Para este caso, el grupo A será el grupo de tratamiento, quienes serán expuestos a una acción específica y el grupo B será el grupo de control o grupo de tratamiento base, quienes no serán expuestos a lo anterior. El experimento será el envío de correos con campañas diseñadas por el área comercial.

El proceso de envío de campañas vía mail funciona de la siguiente forma:

1. El área comercial genera el calendario semanal de campañas, lo que se desea comunicar y el contenido para cada una de ellas.
2. El área comercial le solicita al área de Marketing Directo que envíen una tabla con los clientes que interacturán con la campaña, es decir, el grupo de tratamiento y control y cuál será la proporción de cada uno. En la mayoría de los casos utilizan 90 % para el tratamiento y 10 % para el control[3]
3. El área de Marketing Directo solicita al área de CB, S & CRM tablas con clientes segmentados o no, que puedan entregar valor a cada campaña en específico.
4. Se entregan las tablas y se envían al Área Comercial quienes asignan un código a la campaña para poder analizarla posteriormente.
5. Se envía la campaña.
6. Se analizan los resultados con una ventana de tiempo de hasta 7 días después de envíado el correo.

El experimento a realizar va a trabajar los puntos 2, 3 y 4 anteriormente mencionados.

Previo a entregar la tabla con los clientes que participarán en el experimento, se definieron, en el área de CB, S & CRM, ciertos criterios comerciales para este envío. Es decir, del universo de clientes segmentados con los que se trabajó anteriormente, se seleccionarán solo aquellos que tengan contactabilidad (correo electrónico no nulo en la tabla de contactabilidad) y su tasa de apertura de correos sea mayor al 10 %. Estas limitaciones se utilizan para poder optimizar los costos del envío de campañas, sobre todo al momento de hacer experimentos como el que se desea realizar en este trabajo de título. También se decidieron las campañas que se ocuparían para experimentar, seleccionando 10 de ellas que se encontraban calendarizadas en un intervalo de 16 días como se ve en la tabla 6.5.

Además de esto, se definió que el 80 % de los clientes por segmento recibirán el correo (GT) y el otro 20 % no recibirá el correo (GC), seleccionados aleatoriamente. En un mundo ideal, esta composición porcentual estaría dada por dos grupos idénticos, o sea, 50 % para GT y 50 % para GC, para poder hacer comparaciones con la misma cantidad de clientes por grupo

[3] Datos entregados por el área de Marketing Directo y revisión de campañas enviadas anteriormente, 2021.

Tabla 6.5: Fecha envío de campañas. Fuente: Elaboración Propia

Campaña	Fecha Envío
Herramientas Eléctricas	*5 de Noviembre de 2021*
Terraza	*6 de Noviembre de 2021*
Construcción	*9 de Noviembre de 2021*
Menaje Cocina	*12 de Noviembre de 2021*
Duchas y Cabinas	*13 de Noviembre de 2021*
Iluminación	*14 de Noviembre de 2021*
Especial Bauker	*17 de Noviembre de 2021*
Jardín y Terraza	*19 de Noviembre de 2021*
Pinturas	*20 de Noviembre de 2021*
Escaleras	*21 de Noviembre de 2021*

y maximizar el poder estadístico. Pero en la realidad, y como se mencionó anteriormente, el grupo de tratamiento corresponde a un 90 % por lo que usar un 50 % para realizar estos experimentos significaría un gran riesgo para la empresa. Es más, se podrían llegar a perder 0,52 millones de dólares de venta trimestral (utilizando los datos descritos en el punto 2.3 del Capítulo 2 del presente documento) haciendo 50-50 versus 80-20.

Posteriormente, se generó el cruce de la tabla de contactabilidad con la de los segmentos utilizando como restricción que el OR fuera mayor a 10 % y los resultados finales se pueden apreciar en la tabla 6.6, donde también se ve la cantidad de personas por segmento que recibirán el correo y aquellas que pertenecerán al grupo de control.

Tabla 6.6: Contactabilidad de clientes. Fuente: Elaboración Propia

Segmentos	Clientes	Cruces CC y > 10 % OR	% Cruces	80 % CC	20 % CC
1	*169.213*	*16.284*	*9,62 %*	*13.027*	*3.257*
2	*328.297*	*150.091*	*45,72 %*	*120.073*	*30.018*
3	*266.383*	*129.808*	*48,73 %*	*103.846*	*25.962*
4	*220.047*	*62.506*	*28,41 %*	*50.005*	*12.501*
5	*249.550*	*65.971*	*26,44 %*	*52.777*	*13.194*
Total	*1.233.490*	*424.660*	*34,43 %*	*339.728*	*84.93*

Finalmente, en base a los analisis de campañas anteriores y conversaciones con el área de CB, S & CRM se definen los KPIs a utilizar para analizar el resultado del experimento, que serán la tasa de conversion (órdenes sobre visitas) y venta por visita (venta total sobre visitas). Se analizarán los resultados por campaña para el general de los clientes y para cada segmento en particular, en una ventana de tiempo de hasta 7 días después de enviada la campaña.

Para el caso de la conversión se medirá la significancia estadística de su resultado utilizando un nivel de confianza del 95 % a través de un test de dos colas, donde la prueba de Hipótesis estará dada por (siendo A el GT y B el GC):

$$H_0 : \bar{x_A} = \bar{x_B}$$
$$H_1 : \bar{x_A} \neq \bar{x_B} \tag{6.1}$$

6.5. Implementación

Tal como se declaró en la metodología, es en este paso donde se implementará el experimento diseñado en el paso anterior y se analizarán los resultados obtenidos en base a los KPIs principales: la Tasa de Conversión (TC) y la Venta por Visita (VxV). En específico, se analizará si la Tasa de Conversión es significante estadísticamente y cómo se comporta la Venta por Visita para algún segmento en el general de los envíos y en alguna campaña particularmente.

La cantidad de correos que fueron efectivamente enviados por campaña se muestran en la tabla 6.7, y varían con respecto a la tabla del apartado anterior dado que algunos correos rebotan o algunos clientes se desuscriben de la tabla de contactabilidad en alguna campaña. Además, dependiendo de lo específico de la campaña, el equipo comercial puede definir un máximo de correos por enviar, lo que puede explicar la baja en los correos enviados en el caso de Especial Bauker.[4]

Tabla 6.7: Cantidad de correos enviados y grupo de control por campaña.
Fuente: Elaboración Propia

Campaña	Enviados	Grupo de Control
Herramientas Eléctricas	*329.492*	*82.373*
Terraza	*255.889*	*56.472*
Construcción	*339.322*	*84.831*
Menaje Cocina	*254.536*	*63.634*
Duchas y Cabinas	*194.641*	*48.660*
Iluminación	*333.879*	*83.470*
Especial Bauker	*84.747*	*21.187*
Jardín y Terraza	*191.023*	*47.756*
Pinturas	*188.112*	*47.028*
Escaleras	*222.136*	*55.534*

6.5.1. Resultados Obtenidos

Para el caso general, es decir, para todos los segmentos pero diferenciando el Grupo de Tratamiento y Control, los resultados obtenidos se pueden ver en la figura 6.9. En promedio, las visitas al sitio web por campaña fueron 8.607 para el grupo de tratamiento y 2.068 para el grupo de control. Con respecto a los KPIs principales, la venta por visita, en promedio, fue de $2.205 para el grupo de tratamiento y $1.870 para el grupo de control y, la tasa de conversión del total, fue 2,1 % y 1,6 % para tratamiento y control, respectivamente, obteniendo significancia estadística, lo que quiere decir que en general, enviar un campaña via mail a un grupo

[4] Bauker es una marca de herramientas eléctricas. En este caso es una campaña muy específica pues se quiere comunicar sobre una marca dentro de una subcategoría.

de clientes mejora la tasa de conversión dentro del sitio versus quienes no recibieron el correo.

Si se revisan los mismos resultados por campaña, lo que también se puede ver en la figura 6.9, las campañas de Terraza, Construcción, Iluminación, Jardín y Terraza y, Pinturas obtuvieron significancia estadística para la tasa de conversión. Las demás campañas, a pesar de que la tasa de conversión es mayor para el grupo de tratamiento (a excepción de Especial Bauker), no se logra la significancia por lo que no se puede asegurar estadísticamente que el grupo de tratamiento sea mejor que el grupo de control en conversión.

Haciendo un desglose de los resultados por segmento, las tasas de conversión estadísticamente significativas las obtiene el segmento 2 en las campañas de Construcción y Pinturas, el segmento 3 en la campaña de Jardín y Terraza, el segmento 4 en las campañas de construcción, iluminación y pinturas y, el segmento 5 en la campaña de jardín y terraza y en la de pinturas. El segmento 1 no mostró significancia en ninguna de las campañas enviadas por lo que no se puede asegurar estadísticamente que para este segmento, el grupo de tratamiento sea mejor que el grupo de control al menos por conversión.

El desglose de la venta por visita de cada campaña por segmento se presenta en el apartado D.3 del anexo. Sin embargo, la figura 6.11 muestra la diferencia que existe entre la venta por visita del grupo de tratamiento y control por segmento y por campaña. Las campañas significativas estadísticamente muestran una diferencia positiva de cara al grupo de tratamiento, pero en el caso de la campaña de Iluminación esto ocurre al revés, siendo la venta por visita mayor para el grupo de control.

			Venta por Visita		Tasa de Conversión	
Campañas	**Visitas GT**	**Visitas GC**	**GT**	**GC**	**GT**	**GC**
Herramientas Electricas	11.494	2.883	$ 1.976	$ 1.576	1,8%	1,5%
Terraza	7.129	1.977	$ 3.251	$ 2.345	2,2%	1,5%
Construcción	4.528	2.969	$ 1.187	$ 362	0,9%	0,3%
Menaje Cocina	14.325	2.227	$ 2.026	$ 2.606	2,4%	2,2%
Duchas y Cabinas	11.277	1.703	$ 1.863	$ 2.467	1,3%	1,7%
Ilumación	22.936	2.921	$ 2.920	$ 4.584	3,0%	2,4%
Especial Bauker	3.910	742	$ 1.987	$ 2.096	2,4%	2,6%
Jardín y Terraza	4.271	1.671	$ 2.103	$ 1.075	2,7%	1,4%
Pinturas	2.112	1.646	$ 2.484	$ 638	2,4%	0,6%
Escaleras	4.088	1.944	$ 2.250	$ 946	2,1%	1,4%
Promedio	**8.607**	**2.068**	**$ 2.205**	**$ 1.870**	**2,1%**	**1,6%**

Figura 6.9: Resultados Generales experimentación. Fuente: Elaboración Propia

	SEGMENTO 1		SEGMENTO 2		SEGMENTO 3	
CAMPAÑA	GT	GC	GT	GC	GT	GC
Herramientas Electricas	0,97%	0,77%	2,76%	2,20%	0,87%	0,69%
Terraza	1,25%	0,90%	1,16%	0,84%	3,63%	2,62%
Construcción	0,33%	0,10%	1,25%	0,38%	0,40%	0,12%
Menaje Cocina	2,02%	2,00%	1,16%	1,19%	1,71%	1,84%
Duchas y Cabinas	3,15%	4,18%	1,51%	1,99%	0,81%	1,07%
Ilumación	3,29%	2,28%	3,21%	2,58%	2,52%	2,96%
Especial Bauker	1,14%	1,20%	3,18%	3,36%	0,93%	0,98%
Jardín y Terraza	1,21%	0,62%	0,88%	0,45%	5,32%	2,72%
Pinturas	1,75%	0,45%	2,50%	0,64%	1,94%	0,50%
Escaleras	1,27%	1,14%	4,89%	3,11%	0,33%	0,23%

	SEGMENTO 4		SEGMENTO 5	
CAMPAÑA	GT	GC	GT	GC
Herramientas Electricas	3,53%	2,81%	0,70%	0,55%
Terraza	1,25%	0,90%	3,30%	2,02%
Construcción	1,72%	0,52%	0,55%	0,17%
Menaje Cocina	3,76%	2,99%	3,77%	3,37%
Duchas y Cabinas	0,94%	1,24%	0,75%	1,00%
Ilumación	3,44%	1,90%	2,79%	2,05%
Especial Bauker	5,26%	5,54%	1,41%	1,48%
Jardín y Terraza	1,56%	0,80%	4,33%	2,21%
Pinturas	3,16%	0,81%	2,37%	0,61%
Escaleras	2,31%	1,71%	0,49%	0,21%

Figura 6.10: Resultados con respecto a la Tasa de Conversión por Segmento y Grupo de Tratamiento y control. Fuente: Elaboración Propia

CAMPAÑA	S1	S2	S3	S4	S5
Herramientas Electricas	$ 173	$ 621	$ 177	$ 814	$ 136
Terraza	$ 286	$ 343	$ 1.778	$ 331	$ 1.629
Construcción	$ 216	$ 1.221	$ 309	$ 1.703	$ 490
Menaje Cocina	$ -342	$ -245	$ -909	$ -1.099	$ -357
Duchas y Cabinas	$ -1.290	$ -692	$ -432	$ -511	$ -298
Ilumación	$ -1.401	$ -2.107	$ -2.028	$ -1.353	$ -1.138
Especial Bauker	$ -21	$ -174	$ -33	$ -232	$ -53
Jardín y Terraza	$ 305	$ 297	$ 2.238	$ 513	$ 1.619
Pinturas	$ 1.106	$ 1.907	$ 1.585	$ 2.557	$ 1.918
Escaleras	$ 525	$ 3.450	$ 29	$ 1.523	$ 189

Figura 6.11: Diferencia de la Venta por visita entre el Grupo de Tratamiento y Control por Segmento. Fuente: Elaboración Propia

Capítulo 7

Conclusiones

Con todo lo expuesto a lo largo de este trabajo y luego de obtener los resultados de la experimentación, se declaran las siguientes conclusiones:

• La metodología CRISP-DM fue óptima para realizar este trabajo. Los seis pasos descritos en el capítulo 5 ordenaron el trabajo de buena manera además de proporcionar tareas a corto plazo para avanzar e iterar de ser necesario en pro del objetivo general. El paso de comprensión del negocio fue fundamental, ya que definió los lineamientos para todo lo realizado. Esto principalemente porque el desarrollar este trabajo en una empresa real limita el espacio de acción y lo encausa al objetivo del negocio y obviamente, a sus intereses ya sean económicos o estratégicos, lo que hay que tener presente en cada paso de la metodología.

• La base de datos para realizar la segmentación y el modelo de propensión fue suficiente para una primera iteración de este proyecto, pero, es imperativo incorporarle prontamente variables de navegación y comportamiento dentro del sitio web. Esto podría cambiar totalmente los resultados, sobre todo los de la segmentación realizada ya que probablemente, las variables más relevantes para clusterizar sean aquellas relacionadas a la cantidad de visitas por usuario en el sitio o las compras dentro del mismo.

• La segmentación realizada mediante el Algoritmo K-Means sirvió para identificar de buena manera los segmentos de clientes de Sodimac, a pesar de tener dos variables con correlación mayor a la decidida para discriminar, como lo fue el caso de venta y órdenes. Las variables más relevantes para estos clústeres (género y tipo de Loyalty) dividieron a los clientes en grupos diferentes en el macro pero con algunas similitudes en otras variables como fue el caso de la variable relativa a los hijos para todos los segmentos y las subcategorías más compradas para los segmentos de hombres y mujeres respectivamente.

• El segmento 1 parece ser prescindible ya que no aportó realmente a definir un grupo de clientes, sino que cumplió más una labor de recibir y agrupar a todos aquellos que no lograron entrar a ninguno de los segmentos restantes. A pesar de esto, fue necesario tenerlo ya que de lo contrario, los otros cuatro segmentos no quedaban tan bien definidos.

• Usar un árbol de clasificación arrojó buenos resultados para identificar la propensión de compra por subcategoría de producto obteniendo un 73,1 % de rendimiento y 84,7 % de Recall. Esto pudo ser mejor utilizando algún método para generar mayor precisión como

Bootstrapping.

- La propensión de compra para cada segmento tiende a ser mayor para las subcategorías que representan en conjunto más de el 50 % de la distribución. Y se cumple para todos los casos que las subcategorías con mayor propensión, son aquellas que pertenecen al top 3 de compras visto en la caracterización de los segmentos. A pesar de esto, la probabilidad de compra para los segmentos 2, 3, 4 y 5 no se concentra en esas 3 o 4 subcategorías a diferencia del segmento 1 donde 3 de ellas se llevan más de el 65 % de la propensión. Esto quiere decir que el segmento 1 posee poco interés en comprar productos de subcategorías diferentes a las que compra recurrentemente, ya que existe una diferencia de 12 % entre las propensiones en el tercer y cuarto lugar, a diferencia del resto de los segmentos donde existe una probabilidad de interesarse por una categoría diferente a las usuales cuya diferencia entre los top 10 de subcategorías no supera el 5 %.

- El hecho que los resultados de la experimentación hayan arrojado resultados significativos estadísticamente denotan que ésta fue útil para aumentar la tasa de conversión de venta dentro del sitio. Para el caso general (solo analizando GT vs GC), el grupo de tratamiento, es decir aquellos clientes que recibieron el mail con alguna campaña se comportó de mejor manera en relación a la conversión en comparación a quienes no lo recibieron. Haciendo un zoom a este resultado, las campañas de Terraza, Construcción, Iluminación, Jardín y Pinturas obtuvieron significancia estadística, lo que quiere decir que la acción de recibir un correo genera una diferencia positiva de cara a la conversión en el sitio.

- En el caso de los resultados por segmento, se concluye que para aquellas subcategorías de productos con mayor propensión de compra para cada segmento, el recibir una campaña afín aumenta la tasa de conversión de venta. Esto se ve reflejado en las campañas con significancia estadística para cada segmento, por ejemplo para el caso del segmento 2, las campañas de Construcción (Ferretería) y Pinturas tuvieron significancia y ambas subcategorías se encuentran en el top 3 de propensión de compra para este segmento. La únicas campañas en las que no se condice totalmente el modelo de propensión son la campaña de Pinturas para el Segmento 1 (24 % de propensión),donde no se logra la significancia e Iluminación para el Segmento 4 (4 % de propensión) donde sí se logra. Lo anterior muestra que el modelo puede no acertar totalmente para todos los segmentos, pero que en su mayoría sí logra predecir de buena manera el comportamiento del cliente, sobre todo frente a la exposición de productos de interés para los segmentos.

Tomando en consideración todos los puntos mencionados se concluye que el trabajo realizado mediante la metodología CRISP-DM logró cumplir con el objetivo general propuesto en el capítulo 3.

Capítulo 8

Trabajo Futuro

Se proponen 4 ideas para mejorar y/o utilizar el trabajo realizado con la intención de aportar en los proyectos actuales y futuros que realice el área, cuyo objetivo es lograr un aumento en los resultados de las métricas clave del negocio:

1. Utilizar datos de navegación de los clientes como variable para segmentar. De esta manera existirá una variable no transaccional que impulsaría el análisis del comportamiento de los clientes en base a sus preferencias en la página web, logrando también, incorporar todas las líneas de producto de Falabella.com.

2. Utilizar los segmentos y los resultados del modelo para personalizar el sitio web. De esta forma se podrían optimizar los espacios en el home, PLPs (Product Listing Pages), PDPs (Product Detail Pages) o carros de compra dentro del sitio, crear carruseles con productos específicos o bundles con las preferencias de las personas por segmento.

3. Analizar con mayor profundidad las variables utilizadas para la segmentación y el modelo de propensión. Por ejemplo, se podría estudiar si el tipo de Loyalty se diferencia por nivel de engagement a una tarjeta (CMR específicamente), es decir, si solo la tiene y paga con ella en algún negocio del grupo en específico o es su tarjeta principal. Esto principalmente para obtener insights significativos para el negocio y sus opciones de ofertas o campañas (como CMR Days u otras).

4. Optimizar el envío de campañas y la comunicación de los clientes por segmento, es decir, si un segmento mostró resultado significativos estadísticamente para cierta línea en la experimentación (como es el caso del segmento 3 con Jardín y Terraza), enviar campañas más específicas a estos segmentos sobre cierta línea y/o ofrecer ofertas o promociones con respecto a ésta.

5. Realizar el modelo de propensión con mayor granularidad, en específico, lograr hacerlo por cliente único o visita. De esta manera se puede lograr crear experiencias más personalizadas que por segmento, identificando las preferencias de cada cliente. Además se podría utilizar otro tipo de modelo de aprendizaje de máquinas o una regresión logística para clasificar.

Bibliografía

[1] *FALABELLA S.A.* Memoria Anual 2019 [en línea] <https://www.cmfchile.cl/> [Consulta 12/11/2021].

[2] *FALABELLA. S.A.* Quiénes somos, Nuestros Negocios [en línea] <https://investors.falabella.com/Spanish/quienes-somos/default.aspx> [Consulta 12/11/2021].

[3] *FALABELLA S.A.* Memoria Anual 2020 [en línea] <https://s22.q4cdn.com/Falabella-Memoria-Anual-2020.pdf> [Consulta 12/11/2021].

[4] *NUÑEZ, S. 2010.* Segmentación de clientes de una cadena de supermercados en base a estilos de vida. Memoria Ingeniería Civil Industrial. Santiago. Universidad de Chile, 66-78.

[5] *BOSCH, M. y GOIC, M. 2010.* Modelos de Segmentación [Diapositivas]. Santiago.

[6] *TORRES, S. 2011.* Segmentación de clientes de menor escala para una empresa distribuidora de maquinaria. Memoria Ingeniería Civil Industrial. Santiago. Universidad de Chile, 22-69.

[7] *DE LA FUENTE, S. 2011.* Análisis de Conglomerados. Madrid. UAM. 52p.

[8] *BECA, S. 2007.* Clustering Difuso con Selección de Atributos. Memoria Ingeniería Civil Industrial. Santiago. Universidad de Chile, 15-25.

[9] *KASSAMBARA, A. 2017.* Practical Guide to Cluster Analisys in R. STHDA. 187p.

[10] *Shmueli, G., Patel, N. R., & Bruce, P. C. 2010.* Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner. (2nd ed.). John Wiley & Sons.

[11] *TAMBURRELLI, G y MARGARA, A. 2014.* Towards Automated A/B Testing. Faculty of Informatics. University of Lugano, Switzerland.

[12] *SELMAN, H. 2017.* Marketing Digital. California, EE.UU. Editorial Ibukku.

[13] *Chapman, Pete (NCR); Clinton, Julian (SPSS); Kerber, Randy (NCR); Khabaza, Thomas (SPSS); Reinartz, Thomas (DaimlerChrysler); Shearer, Colin (SPSS); Wirth, Rüdiger (DaimlerChrysler). 2000.* Step-by-step data mining guide. 78p.

[14] *Cámara de comercio de Santiago.* E-commerce B2C en Chile [En línea] <https://www.ecommerceccs.cl/eCommerce-B2C-en-Chile-2020-FEB.pdf> [Consulta 12/11/2021].

[15] *SANCHEZ, D.* Ventas del retail cerraron el 2020 en terreno negativo, pero mejor a lo esperado: el canal online compensó la caída [En línea]. La Tercera en línea. 29 de

enero, 2021. <https://www.latercera.com/pulso/noticia/ventas-del-retail-cerraron-el-2020-en-terreno-negativo-pero-mejor-a-lo-esperado-el-canal-online-compenso-la-caida/> [Consulta 12/11/2021].

[16] *Ármate Abogados.* Aspectos legales del E-commerce [En línea] <https://armate.cl/web/2020/01/30/aspectos-legales-del-e-commerce/> [Consulta 12/11/2021].

[17] *Brújula: Investigación y Estrategia.* IX Encuesta de Accesos y usos de Internet, Subsecretaría de Telecomunicaciones de Chile [En línea] <https://www.subtel.gob.cl/Informe-Final-IX-Encuesta-Acceso-y-Usos-Internet-2017.pdf>.

[18] *CENTRO ECONOMÍA DIGITAL CCS.* Tendencias del comercio electrónico [el línea] <https://www.ccs.cl/html/estudios/LEVER-ECD2018-CCS.pdf> [Consulta 13/11/2021].

[19] *Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A. Y., Bouras, A. 2014.* A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis. IEEE Transactions on Emerging Topics in Computing, 2(3), 267-279.

[20] *Baarsch, J., Celebi, M. E. 2012.* Investigation of internal validity measures for K-means clustering. Proceedings of the international multiconference of engineers and computer scientists, 1, s.n.

[21] *Berger, P. D., Maurer, R. E., Celli, G. B. 2017.* Experimental Design. Springer Publishing.

[22] *Kohavi, R., Longbotham, R., Sommerfield, D., Henne, R. M. 2008.* Controlled experiments on the web: survey and practical guide. Data Mining and Knowledge Discovery, 18(1), 140-181.

[23] *Dmitriev, P., Gupta, S., Kim, D. W., Vaz, G. 2017.* A Dirty Dozen: Twelve Common Metric Interpretation Pitfalls in Online Controlled Experiments . Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1427-1436.

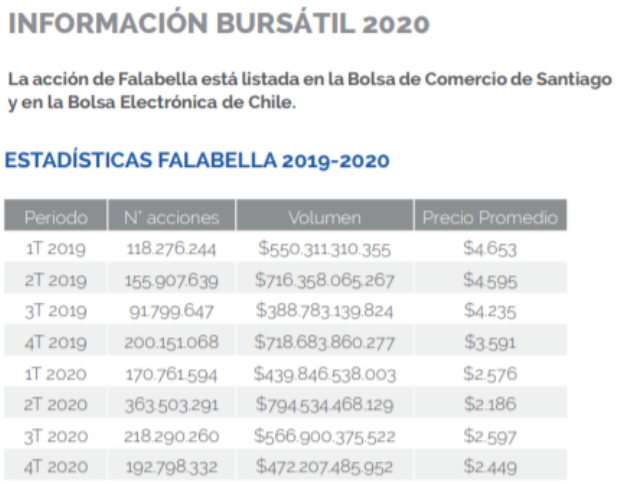
Anexos

Anexo A

Antecedentes generales

A.1. Información Bursátil y Valores

En el cuarto trimestre de 2020 Falabella emitió un total de 192 millones de acciones, comercializables en la Bolsa de Comercio de Santiago y en la Bolsa Electrónica de Chile. Cada una de estas unidades se transó a un precio promedio de $2.449 CLP [3] en dicho periodo, mostrando una disminución de su valor en los últimos dos años (ver A.1 con evolución bursátil de los últimos dos años).



INFORMACIÓN BURSÁTIL 2020

La acción de Falabella está listada en la Bolsa de Comercio de Santiago y en la Bolsa Electrónica de Chile.

ESTADÍSTICAS FALABELLA 2019-2020

Periodo	N° acciones	Volumen	Precio Promedio
1T 2019	118.276.244	$550.311.310.355	$4.653
2T 2019	155.907.639	$716.358.065.267	$4.595
3T 2019	91.799.647	$388.783.139.824	$4.235
4T 2019	200.151.068	$718.683.860.277	$3.591
1T 2020	170.761.594	$439.846.538.003	$2.576
2T 2020	363.503.291	$794.534.468.129	$2.186
3T 2020	218.290.260	$566.900.375.522	$2.597
4T 2020	192.798.332	$472.207.485.952	$2.449

Figura A.1: Registro de información bursátil para el periodo 2019-2020. Fuente: [3]

A.2. Organigrama

El organigrama simplificado de Falabella.com se muestra en la figura A.3. Éste funciona a través de una estructura jerarquizada, al mando de un Gerente General, quien maneja la empresa en su conjunto. Bajo este cargo se encuentran diferentes gerencias que poseen objetivos independientes unas de otras y que, a su vez, tienen equipos por países (Chile, Colombia Perú y México) y un equipo corporativo que realiza tareas para los cuatro países en paralelo.

Figura A.2: Valores Falabella. Fuente: [3]

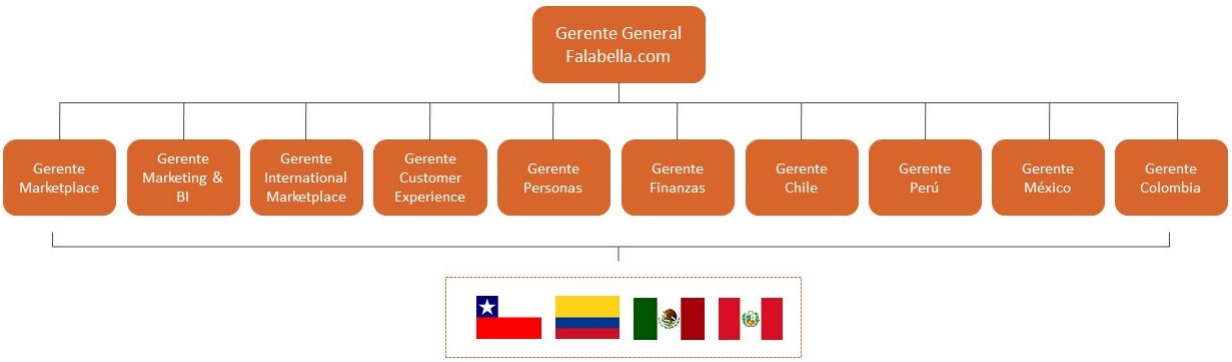


Figura A.3: Organigrama simplificado de Falabella.com. Fuente: Elaboración propia.

A.3. Dimensionamiento de la actividad ralizada por la empresa

Durante el año 2020 el e-commerce de Falabella obtuvo un Gross Merchandise Value de US$ 3.335MM, lo que significa un 123 % más con respecto al 2019, trabajando con más de 10.000 Sellers y 30 millones de órdenes despachadas [3].

Por otra parte, el Gross Merchandise Value (GMV) para 1P (clientes pertenecientes a solo una unidad de negocio de Falabella) fue de US$2.572 MM lo que significó un aumento de 114 % con respecto al año anterior y el GVM para 3P (clientes pertenecientes a tres unidades de negocio de Falabella), es decir, clientes Cross Formato fue de US$763 MM y tuvieron un aumento de 157 % con respecto al 2019 como se puede ver en la figura A.4.

Si bien aún no existe consenso respecto a cómo medir la economía digital ni cuáles son exactamente sus indicadores, expertos indican que es una economía en la cual la información, experiencia y habilidades de las personas son un factor fundamental al momento de generar valor.

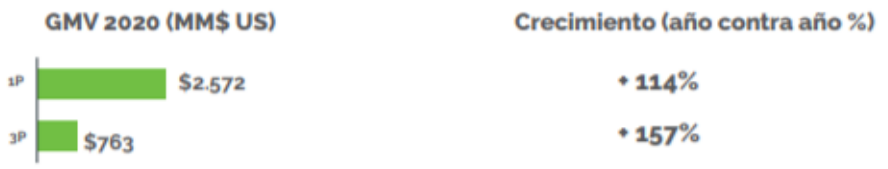


Figura A.4: GVM 2020, formatos 1 y 3P. Fuente: [6]

En el caso de Chile, según datos entregados por Accenture y Oxford Economics en el estudio denominado "El avance de la Economía Digital en Chile" (2018), ésta representa el 22,2 % del PIB, lo que se traduce en casi US$55 mil millones del producto nacional. Bajo esta misma línea, se prevé que para 2021 crezca tres puntos porcentuales, alcanzando entre 25,3 % o 26,3 % del PIB [14].

De acuerdo con la Cámara de Comercio de Santiago, las ventas totales del comercio electrónico B2C (business-to-consumer) alcanzaron los US$ 9.400 millones [15] durante el año 2020. Falabella.com por otra parte, durante el 2020 tuvo ingresos de aproximadamente US$1.300 millones [5], representando un 13,8 % del comercio electrónico nacional.

A.4. Ventaja competitiva en el mercado

El comercio en línea se encuentra en un crecimiento exponencial desde aproximadamente el año 2006 (ver figura B.3). Las razones principales de este aumento son: el acceso a internet, que ya llegó al 87,4 % de los hogares a nivel total de la población en el año 2017 [3] y acelerados por esta pandemia, se están haciendo cada vez más inversiones para mejorar aún más la necesaria conectividad a costos accesibles; y la pandemia mundial por Covid-19, lo que ha cambiado la conducta del consumidor aumentando su confianza en la compra online y la necesidad de compra por la cuarentenas obligatorias existentes en Chile.

Se desea ofrecer a los consumidores una propuesta comercial única, donde puedan encontrar una oferta variada de productos y una valiosa experiencia de compra. Contar con surtido exclusivo y profundo es diferenciador.

Por lo que el valor diferenciador de la propuesta es la combinación única de productos, canales y servicios que se ofrece a los consumidores y Sellers desde distintas propuestas que se complementan y potencian: centros comerciales, tiendas, productos, e-commerce, Marketplace, servicios logísticos, pagos y servicios financieros. Además, las variadas interacciones con el cliente permiten conocerlo mejor y personalizar propuestas que sean más relevantes y le generen más valor, buscando ser el E-commerce líder en Latinoamérica.

Por otra parte, continúan enfocados en el desarrollo de marcas propias y exclusivas que permitan fortalecer el posicionamiento de la empresa, lo que permite [3]:

[5] Información interna de Falabella.com, 2021

- **Diferenciarse de otros retailers y comercios online:** Una oferta exclusiva con un posicionamiento robusto. Se desarrolla una propuesta que se ajusta a grupos específicos de clientes, que se conoce y entiende; lo que ha permitido posicionar muchas marcas propias como líderes en su segmento y con un reconocimiento independiente del negocio.
- **Ofrecer una propuesta conveniente:** Con una alta relación calidad/precio, se permite competir, por ejemplo, con marcas de fast-fashion en moda; marcas líderes en supermercados y mejoramiento del hogar.
- **Aumentar la creación de valor del negocio:** Con costos más competitivos que los obtenidos a través de marcas de terceros. En los distintos formatos se aprovechan las oportunidades que abre el mundo digital para ampliar sustancialmente el surtido de productos y personalizar las ofertas de acuerdo con las preferencias de los consumidores, incrementando también las áreas de retiro en tienda de compras online.

El valor diferenciador de esta nueva plataforma para los Sellers radica en entregar la mejor propuesta de valor en términos de [3]:

- Acceder a una mayor base de tráfico aumentando el potencial de ventas.
- Plataforma simple y única, que les permitirá segmentar diferentes canales para captar clientes.
- Acceder a soluciones de pagos y financiamiento.
- Soluciones logísticas, incluyendo servicios de entrega e infraestructura para gestionar cambios y devoluciones.

Y para los consumidores, se aspira a ofrecer la mejor propuesta en términos de [3]:

- Mayor profundidad de surtido.
- Experiencia de compra única y simplificada, pero manteniendo segmentaciones de propuestas.
- Soluciones de pago y financiamiento.
- Alternativas de despacho ampliadas.

Anexo B

Mercado y/o Marco Institucional

B.1. Actores y relación con la empresa

Los principales actores y su relación con la empresa se describen a continuación [3]:

- **Clientes:** Se aspira a simplificar la vida de los clientes buscando transformar las experiencias de compra, procurando asegurar la calidad y seguridad de los productos y servicios que se ofrecen y de las instalaciones que operan. Los canales de diálogo principales son las redes sociales, tiendas físicas y online, equipos de Atención al Cliente, evaluaciones de satisfacción, entre otros.

- **Proveedores:** Se busca mantener relaciones constructivas y de largo plazo con los proveedores. Se cree que la colaboración y confianza permite el crecimiento de ellos como de la empresa. Los canales de diálogo son la plataforma digital B2B, reuniones entre compradores y proveedores, capacitaciones a proveedores, programas de apoyo para mejorar estándares de servicio, programas de apoyo a proveedores de menor tamaño, entre otros.

- **Colaboradores:** Se reconoce la importancia de los colaboradores en el crecimiento de la empresa, por lo que se está comprometido con su desarrollo profesional y su bienestar integral buscando activamente atraer y retener en cada uno de los equipos al mejor talento y equipo humano, valorando la diversidad e inclusión. Los canales de diálogos son el canal de integridad, encuestas de clima internas, canal de comunicaciones corporativas, programas de voluntariado, programas de movilidad interna, Falanet (intranet), reuniones con sindicatos, equipos de calidad de vida, entre otros.

- **Accionistas e Iversionistas:** Se está comprometido en generar valor sostenible, manteniendo canales de comunicación transparentes con sus accionistas y potenciales inversionistas. Los canales de diálogo son el equipo de relación con inversionistas, eventos y reuniones con inversionistas, Junta Ordinaria de Accionistas, reportes financieros y sitio web de inversionistas.

- **Comunidad, Sociedad Civil:** Se promueve la generación de vínculos con la comunidad y el aporte a su desarrollo y calidad de vida buscando colaborar en el desarrollo social económico y cultural de las comunidades en América Latina, ante lo cual se pone

en práctica programas sociales cuyo impacto sea medible y significativo para la comunidad. Los canales de diálogo son el programa de acción social, voluntariado corporativo, miembros de Green Building Council, Cámaras de Comercio, entes gubernamentales, entre otros.

- **Competencia:** Se reconoce la existencia de otras organizaciones dedicadas a satisfacer las mismas necesidades del cliente, llevando a cabo acciones similares a la empresa y manteniendo una cartera de proveedores similar, lo que podría significar un riesgo en la empresa si son captados por este actor. Es por esto que se debe cuidar las relaciones con clientes, manteniendo en secreto su información privada y entregando un servicio de calidad y también, cuidar la relación económica, legal y de comunicación con proveedores.

B.2. Regulaciones relevantes

En Chile, no existe una ley que regule directamente al e-commerce, con excepción de la ley del consumidor, esto significa, que cuando se tiene un e-commerce con domicilio en Chile, se debe tener en consideración las mismas leyes que se aplicarían a un negocio tradicional con existencia en una tienda física, la cual tiene que al menos cumplir con [16]:

- Informar a los clientes el precio, características relevantes del producto o servicio, modalidades y plazos de entrega, costos asociados al despacho, formas de pago y cualquier otra información relevante.
- Informar a los clientes los datos de contacto de la empresa, e idealmente los datos de quién es el representante legal de la misma.
- Respetar la garantía 3×3, en el caso de que falle el producto vendido dentro de los 3 primeros meses, se debe reparar, cambiar o devolver lo pagado a elección del consumidor.
- Respetar lo ofrecido en la publicidad.
- Para las compras realizadas por Internet, los clientes tienen el derecho a retracto dentro de los 10 días siguientes desde que recibió el producto o contrató el servicio, salvo que se informe lo contrario.
- Se tiene la obligación de enviar una confirmación por escrito, de lo contrario clientes tendrán el derecho a retracto por 90 días.

Dentro de la empresa existen, además, políticas y programas relacionados con lo económico, ambiental, ético, legal y de diversidad como lo son [3]:

- **Política y programa de libre competencia:** Establecen, comunican y guían las conductas esperadas de los colaboradores para prevenir la ocurrencia de cualquier conducta contraria a la normativa de libre competencia.

- **Programa de ética y código de integridad:** Establece, comunica y guía las conductas esperadas de los colaboradores para conectar con un estándar ético único para todo Falabella, mantener relaciones de confianza con los clientes, proveedores y colaboradores, proteger el valor y la reputación de la compañía y, promover una actuar basado en principios.
- **Política y modelo de prevención de delitos:** Implementan una forma de organización y los procesos destinados a evitar la comisión de delitos que generen responsabilidad penal al interior de la compañía.
- **Política de gestión de conflictos de intereses:** Establece principios y criterios para la efectiva declaración de intereses y la gestión de los conflictos de interés por parte de los colaboradores
- **Política y programa de Medioambiente y cambio climático:** Establecen, comunican y guían las conductas esperadas de los colaboradores y proveedores en relación con el medio ambiente.
- **Política de diversidad e inclusión:** Establece los principios básicos por los que debe regirse Falabella y todos sus colaboradores, con el objetivo de promover una cultura organizacional y un ambiente laboral diverso e inclusivo.

La integración de las plataformas de e-commerce está enfocada en potenciar el efecto flywheel, donde mayor tráfico impulsará más ventas, atrayendo cada vez a más Sellers a la plataforma, incrementando así la oferta de productos enriqueciendo la propuesta, haciéndola más atractiva a los clientes, como se puede apreciar en la figura B.1.

Figura B.1: Flywheel de crecimiento de Falabella.com. Fuente: [3]

Con respecto al Gross Merchandise Value (GVM) se puede notar que existe un gran crecimiento con respecto al año anterior en todos los países como se aprecia en la figura B.2.

La tecnología continúa siendo el foco prioritario de sus inversiones, apuntando al desarrollo de una plataforma tecnológica, flexible y robusta que potencie la integración de nuestras capacidades de e-commerce como motor de crecimiento de nuestro ecosistema físico-digital. Para el año 2021, el foco de las inversiones en tecnología se concentra en [3]:

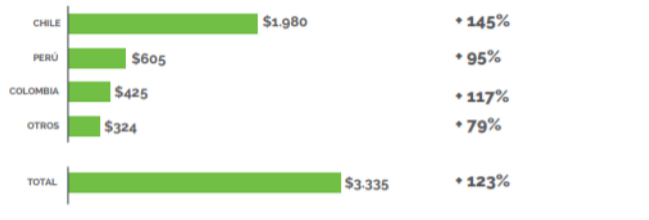


Figura B.2: GVM y crecimiento % a/a de Falabella.com. Fuente: [3]

- Lanzamiento de "Integrated seller center", plataforma única para que los Sellers administren las publicaciones de sus productos y gestionen sus pagos y facturaciones.
- Motor de e-commerce único para Falabella.com.
- Continuar escalando las capacidades e integraciones de nuestra plataforma única de procesamiento de pagos para el e-commerce unificado.
- Potenciar el motor de gestión de entregas express en Falabella.com.
- Desarrollos tecnológicos para potenciar integraciones de sistemas logísticos y en potenciar la visibilidad de la última milla. Tecnología, como un habilitador transversal a la organización, es clave para potenciar las siguientes dimensiones de la organización: Cultura, Herramientas y desarrollos y Capacidades TI.

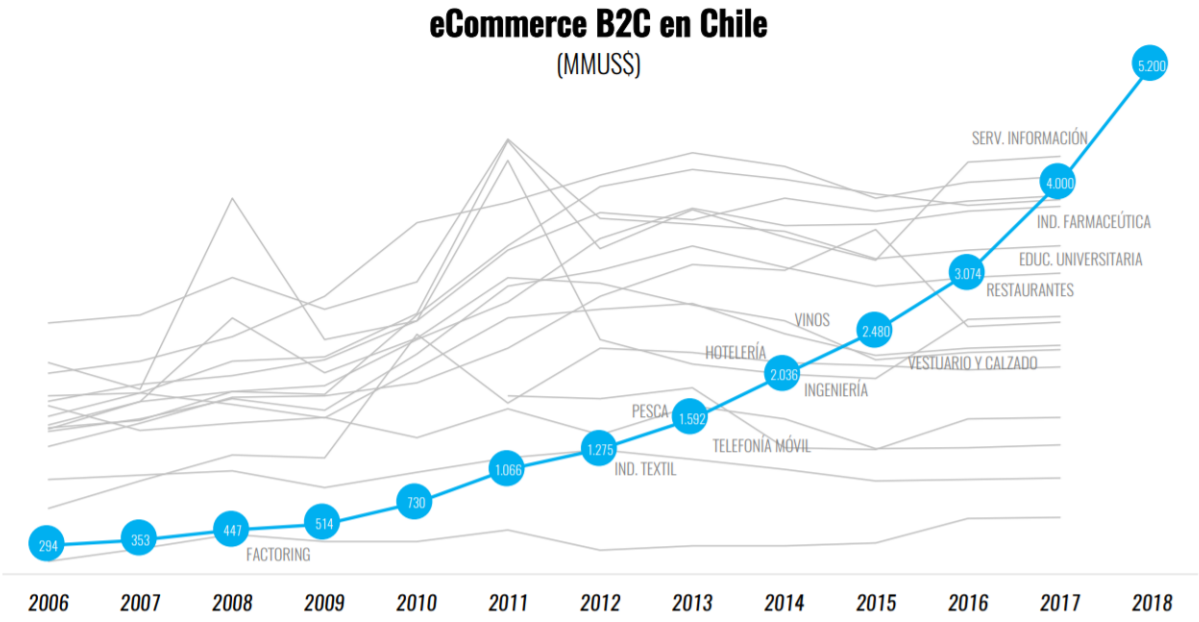


Figura B.3: Tendencia E-commerce B2C en Chile (MMUS$) Fuente: [18]

Anexo C

Descripción del Proyecto y Justificación

C.1. Área de Trabajo

El área de CB, S & CRM posee dos equipos, el Squad Analytics y el Squad Loyalty. El primero, y donde se desarrollará el trabajo de título, se preocupa de generar procesos a dos niveles. En primer lugar, procedimientos operativos que ya están planificados y coordinados con otras áreas o equipos de la empresa (periódicos), de forma que se generan reportes (lo más automático posible) para un cierto objetivo de negocio. Ya sea para los comités de Gerencias/Directorio, para áreas en específico o para encargadas/os más enfocados a líneas de productos o marcas de todos los países del corporativo. Y, en segundo lugar, se trabaja en función de pedidos de distintos clientes internos, generando análisis o reportes de una manera similar a lo descrito previamente, pero para analizar algún fenómeno, campaña, experimento o evento en particular que se solicite.

En la mayoría de los casos, los clientes del área corresponden a las áreas de Marketing Directo y One Site Experience y también, la gerencia de Loyalty & Personalization. Esto principalmente porque los análisis y reportes que se generan en Analytics, se pueden utilizar como input para generar campañas a distintos niveles, principalmente a través del emailing y también, generar recomendaciones y/o personalizaciones dentro del sitio web para mejorar la experiencia de los usuarios dentro éste. A pesar de esto, cualquier área o gerencia podría solicitarle datos al área sin problema.

C.2. Funnel de Conversión de Ventas

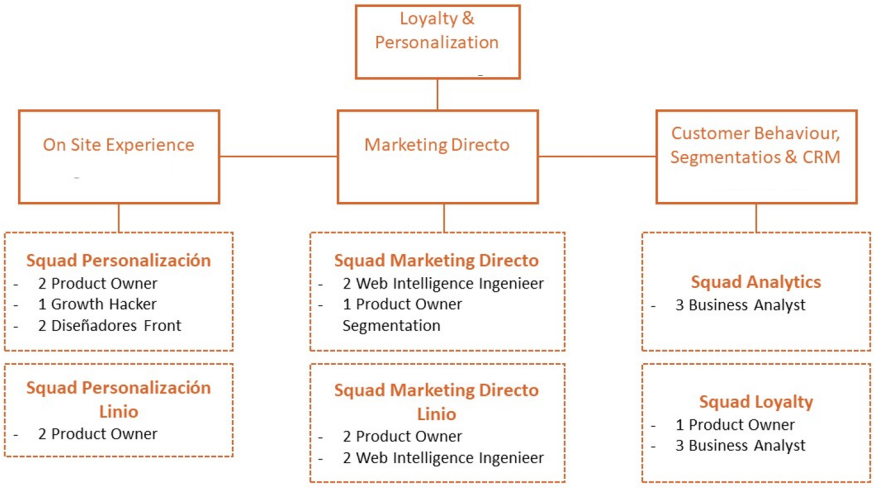


Figura C.1: Organigrama Gerencia Loyalty & Personalization. Fuente: Elaboración propia.

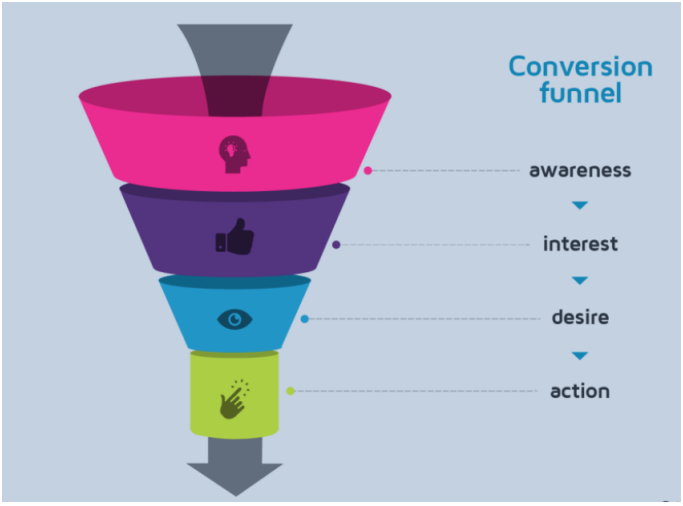


Figura C.2: Funnel de Conversión de ventas E-commerce. Fuente: [**?**]

Anexo D

Desarrollo Metodológico

D.1. Análisis Exploratorio

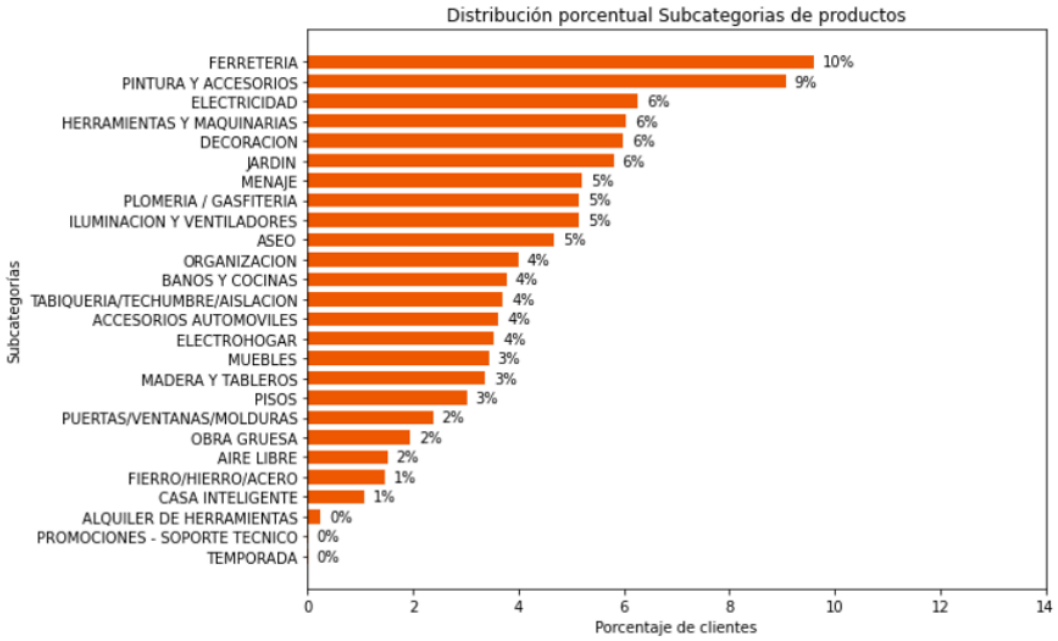


Figura D.1: Distribución porcentual por subcategoría de productos. Fuente: Elaboración propia.

Tabla D.1: Diferencias entre nodos. Fuente: Elaboración Propia.

Nodo	Diferencia
1 a 2	*605.750*
2 a 3	*241.580*
3 a 4	*185.463*
4 a 5	*110.171*
5 a 6	*79.344*
6 a 7	*73.983*
7 a 8	*49.349*
8 a 9	*63.322*

D.2. Segmentación

```
{'venta': 'v1',
 'ordenes': 'v2',
 'cust_age': 'v3',
 'fl_hijos': 'v4',
 'lyty_points': 'v5',
 'channel_offline': 'v6',
 'channel_online': 'v7',
 'prod_sub_cat_desc_ACCESORIOS AUTOMOVILES': 'v8',
 'prod_sub_cat_desc_AIRE LIBRE': 'v9',
 'prod_sub_cat_desc_ALQUILER DE HERRAMIENTAS': 'v10',
 'prod_sub_cat_desc_ASEO': 'v11',
 'prod_sub_cat_desc_BANOS Y COCINAS': 'v12',
 'prod_sub_cat_desc_CASA INTELIGENTE': 'v13',
 'prod_sub_cat_desc_DECORACION': 'v14',
 'prod_sub_cat_desc_ELECTRICIDAD': 'v15',
 'prod_sub_cat_desc_ELECTROHOGAR': 'v16',
 'prod_sub_cat_desc_FERRETERIA': 'v17',
 'prod_sub_cat_desc_FIERRO/HIERRO/ACERO': 'v18',
 'prod_sub_cat_desc_HERRAMIENTAS Y MAQUINARIAS': 'v19',
 'prod_sub_cat_desc_ILUMINACION Y VENTILADORES': 'v20',
 'prod_sub_cat_desc_JARDIN': 'v21',
 'prod_sub_cat_desc_MADERA Y TABLEROS': 'v22',
 'prod_sub_cat_desc_MENAJE': 'v23',
 'prod_sub_cat_desc_MUEBLES': 'v24',
 'prod_sub_cat_desc_OBRA GRUESA': 'v25',
 'prod_sub_cat_desc_ORGANIZACION': 'v26',
 'prod_sub_cat_desc_PINTURA Y ACCESORIOS': 'v27',
 'prod_sub_cat_desc_PISOS': 'v28',
 'prod_sub_cat_desc_PLOMERIA / GASFITERIA': 'v29',
 'prod_sub_cat_desc_PROMOCIONES - SOPORTE TECNICO': 'v30',
 'prod_sub_cat_desc_PUERTAS/VENTANAS/MOLDURAS': 'v31',
 'prod_sub_cat_desc_TABIQUERIA/TECHUMBRE/AISLACION': 'v32',
 'prod_sub_cat_desc_TEMPORADA': 'v33',
 'cust_gender_F': 'v34',
 'cust_gender_M': 'v35',
 'lty_type_CMR': 'v36',
 'lty_type_N': 'v37',
 'lty_type_TMP': 'v38'}
```

Figura D.2: Diccionario Variables. Fuente: Elaboración propia.

D.3. Implementación

	SEGMENTO 1		SEGMENTO 2		SEGMENTO 3	
CAMPAÑA	GT	GC	GT	GC	GT	GC
Herramientas Electricas	$ 852	$ 679	$ 3.064	$ 2.443	$ 873	$ 696
Terraza	$ 1.026	$ 740	$ 1.229	$ 887	$ 6.381	$ 4.603
Construcción	$ 310	$ 95	$ 1.757	$ 536	$ 445	$ 136
Menaje Cocina	$ 1.194	$ 1.535	$ 855	$ 1.099	$ 3.173	$ 4.082
Duchas y Cabinas	$ 3.978	$ 5.267	$ 2.134	$ 2.826	$ 1.331	$ 1.763
Ilumación	$ 2.457	$ 3.858	$ 3.695	$ 5.801	$ 3.556	$ 5.584
Especial Bauker	$ 380	$ 401	$ 3.185	$ 3.359	$ 603	$ 636
Jardín y Terraza	$ 623	$ 318	$ 607	$ 310	$ 4.577	$ 2.339
Pinturas	$ 1.487	$ 382	$ 2.565	$ 658	$ 2.132	$ 547
Escaleras	$ 906	$ 381	$ 5.955	$ 2.505	$ 50	$ 21

	SEGMENTO 4		SEGMENTO 5	
CAMPAÑA	GT	GC	GT	GC
Herramientas Electricas	$ 4.016	$ 3.202	$ 670	$ 534
Terraza	$ 1.190	$ 858	$ 5.846	$ 4.217
Construcción	$ 2.451	$ 748	$ 704	$ 215
Menaje Cocina	$ 3.838	$ 4.937	$ 1.248	$ 1.605
Duchas y Cabinas	$ 1.577	$ 2.088	$ 918	$ 1.216
Ilumación	$ 2.374	$ 3.727	$ 1.996	$ 3.134
Especial Bauker	$ 4.258	$ 4.490	$ 965	$ 1.018
Jardín y Terraza	$ 1.050	$ 536	$ 3.311	$ 1.692
Pinturas	$ 3.439	$ 883	$ 2.579	$ 662
Escaleras	$ 2.629	$ 1.106	$ 327	$ 138

Figura D.3: Resultados con respecto a la Venta por Visita por Segmento y Grupo de Tratamiento y control. Fuente: Elaboración Propia