

Mejora de campañas y tendencias a través de modelos
predictivos.



Alumno: Jens Krarup

Universidad Adolfo Ibáñez

Carrera: Ingeniería Civil Industrial

Empresa/Área: Justo/Business

Profesor: Fernando Vásquez Acuña

Fecha: 2do semestre 2023

Resumen ejecutivo

Justo, una startup chilena fundada en 2018, brinda soluciones integrales para comercios y restaurantes, abarcando desde la toma de pedidos hasta el servicio de delivery. La empresa experimentó un crecimiento exponencial de más del 100% en 2020 lo que la llevó a expandirse a cinco países más en Latinoamérica.

Sin embargo, Justo enfrenta desafíos significativos en la predicción de tendencias y comportamientos de compra, con una baja en pedidos generados mediante alianzas de un 38%, una efectividad decreciente en sus campañas de marketing por correo electrónico y otros factores, se ha generado una disminución en el revenue de la empresa en por lo menos 5% respecto al año pasado.

Para abordar este problema se establecieron como objetivos generales aumentar la tasa de conversión en un 3% y mejorar el ROAS de las campañas de marketing en por lo menos un 20% para cada marca en un período de 5 meses.

Para lograr lo anterior se propusieron dos soluciones en base a los objetivos y recursos disponibles. Posteriormente y en base a una serie de parámetros evaluados en forma conjunta con el equipo de trabajo, se determinó utilizar técnicas de machine learning (ML) para afrontar el problema.

Para tener un trabajo eficaz y ordenado, se adoptó la metodología DMAIC la cual permite un enfoque estructurado y sistemático de mejora continua que abarca las fases de definir, medir, analizar, mejorar y controlar para optimizar procesos y reducir la variabilidad en la gestión de calidad, especialmente en el marco de Six Sigma. También, para el desarrollo de los modelos de ML se siguió una metodología CRISP-DM.

Además se estableció un plan de implementación en el cual se asignaron las etapas y el proceso de implementación del proyecto, para así asegurar una correcta puesta en marcha. Sumado a esto, se realizó un análisis de riesgo para ver las potenciales causas que podrían afectar negativamente al proyecto.

Finalmente y luego de haber sido implementado el proyecto, se pueden destacar los resultados obtenidos en este. Pese a no cumplir con el objetivo de mejorar la tasa de conversión en un 3% se pudo cumplir con una mejora promedio del 30% para el ROAS y una mejora promedio del 39% en la tasa de conversión del mailing.

En base a los resultados obtenidos se concluye que el proyecto cumple satisfactoriamente pero aún quedan desafíos en la implementación completa de este, por lo que a modo de recomendación a la empresa se le sugiere trabajar en esto cuando el contexto y situación de la empresa lo amerite.

Abstract

Justo, a Chilean startup founded in 2018, provides comprehensive solutions for shops and restaurants, ranging from order taking to delivery services. Experiencing exponential growth of over 100% in 2020, the company expanded to five more countries in Latin America.

However, Justo faces significant challenges in predicting trends and buying behaviors, with a 38% decrease in orders generated through alliances, declining effectiveness in email marketing campaigns, and other factors contributing to a revenue decrease of at least 5% compared to the previous year.

To address this issue, the overarching objectives were set to increase the conversion rate by 3% and improve the ROAS (Return on Advertising Spend) of marketing campaigns by at least 20% for each brand within a 5-month period.

Two solutions were proposed based on the objectives and available resources, and after evaluating parameters collaboratively with the team, machine learning (ML) techniques were determined as the approach to tackle the problem.

To ensure an effective and organized workflow, the DMAIC methodology was adopted, offering a structured and systematic approach to continuous improvement, particularly within the Six Sigma framework. Additionally, the development of ML models followed the CRISP-DM methodology.

A detailed implementation plan was established, outlining the project stages and the implementation process to ensure a successful rollout. Furthermore, a risk analysis was conducted to identify potential factors that could negatively impact the project.

After the project's implementation, noteworthy results were achieved. Despite not meeting the goal of improving the conversion rate by 3%, there was an average improvement of 30% in ROAS and a 39% average improvement in the mailing conversion rate.

Based on the obtained results, it is concluded that the project has been successful, but challenges remain in its full implementation. As a recommendation to the company, it is suggested to address these challenges when the context and situation of the company allow for it.

Índice

Introducción.....	5
Contexto.....	5
Descripción del problema.....	6
Objetivos.....	11
Objetivo general.....	11
Objetivos específicos.....	11
Estado del arte.....	12
Solución.....	15
Alternativas de solución.....	15
Elección de solución.....	16
Desarrollo de la solución.....	18
1. Limpieza y transformación de los datos.....	18
2. Análisis RFM.....	20
3. Clusterización.....	21
4. Modelos de predicción de respuesta a campañas de marketing.....	24
5. Pronósticos de demanda para alianzas.....	25
Metodología.....	27
Métricas de desempeño.....	29
Plan de implementación.....	31
Análisis de riesgo.....	33
Evaluación económica.....	37
Resultados.....	42
Conclusiones y recomendaciones.....	46
Referencias.....	48
Anexos.....	50
1. Detalle de los servicios ofrecidos por Justo.....	50
2. Cantidad de órdenes de alianzas por mes en el año 2022.....	51
3. Cantidad de órdenes de alianzas por mes en el año 2023.....	51
4. Códigos desarrollo de solución.....	52
5. Métricas resultados de modelos.....	52
7. Marco metodológico y representación gráfica metodología CRISP-DM.....	54
8. Estimaciones flujo de caja.....	56
9. Funnel de compra y ejemplo costos productos vs delivery Buffalo Waffles.....	57
10. Funnel de compra clientes.....	58

Introducción

Contexto

Justo es una startup chilena fundada en 2018 que se encarga de todo el proceso de venta en comercios y restaurantes de manera integral, esto es desde la toma de pedidos hasta el servicio de delivery. La misión de la empresa es ser un partner que ayude a comercios y restaurantes a tomar el control de su propio negocio teniendo una solución end to end que le permita conectar con los clientes. Para lograr esto, Justo ofrece distintos productos que permiten a sus distintas marcas crecer y escalar su negocio en toda su cadena de valor, estos son:

- **JustoDelivery:** Logística de repartos.
- **JustoEcommerce:** Plataforma para vender los productos del restaurante. (sitio web, app propia y app Justo)
- **JustoMarketing:** Servicio de mailing que contacta a los consumidores de forma rápida y sencilla con mailing personalizado.
- **JustoReserve:** Optimiza la rotación de mesas sin la intervención de un garzón.
- **JustoCommander:** Plataforma para centralizar pedidos.
- **JustoCrisp :** Herramienta de gestión efectiva para operaciones, servicio y control de ventas. (POS)
- **JustoLoyalty:** Sistema que fomenta la fidelización de los clientes.

Más información acerca de los productos de Justo se pueden encontrar en el anexo 1.

En sus primeros años de operación, la empresa tuvo un crecimiento exponencial el cual fue triplicado mes a mes durante 2020. Asimismo el tamaño de empleados pasó de 10 en 2019 a 150 en 2020. Este crecimiento fue debido en gran parte a la pandemia que potenció el delivery y la necesidad de los restaurantes de poder hacer entregas de manera eficiente. Dado este crecimiento exponencial, Justo pudo expandirse a cinco países más de latinoamérica los cuales son México, Perú, Ecuador, Colombia y Costa Rica. Actualmente Justo tiene 123 empleados en total para los seis países y a nivel local ya cuenta con más de 2000 marcas adheridas.

El área en la cual se desarrollará el proyecto corresponde a la de Business Chile en la que participan un total de 13 personas. El objetivo del área es aumentar el Net Revenue de Justo a través del manejo de cuentas (account management), marketing y alianzas.

Para esto mantienen, mejoran y buscan nuevas relaciones comerciales con las marcas ya sea ofreciendo un mix de productos y herramientas de venta y comunicación a marcas que ya son clientes, o también buscando a nuevos partners comerciales que quieran ser parte del ecosistema Justo.

Lo mencionado anteriormente también aplica para las alianzas estratégicas que realiza la empresa en la cual se busca integrar nuevos partner comerciales como entidades bancarias con el fin de obtener y fidelizar nuevos clientes.

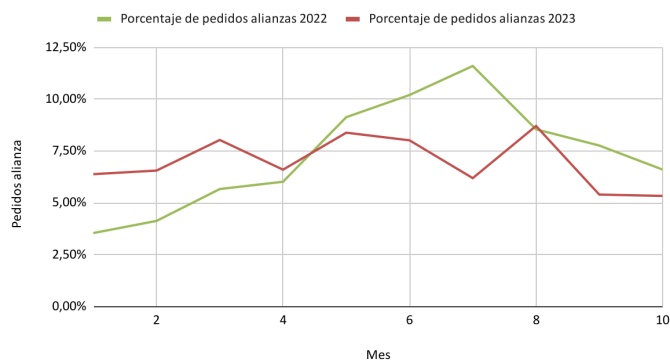
Descripción del problema

Hoy en día, la empresa presenta una falta de capacidad para prever tendencias y comportamientos de compra en las distintas categorías de Justo. La efectividad de las campañas a través del mailing ha ido a la baja durante el año lo cual disminuye la tasa de conversión de la empresa. Esto último, sumado a una baja en órdenes generadas por alianzas ha provocado una disminución de al menos un 5% del revenue respecto al año pasado.

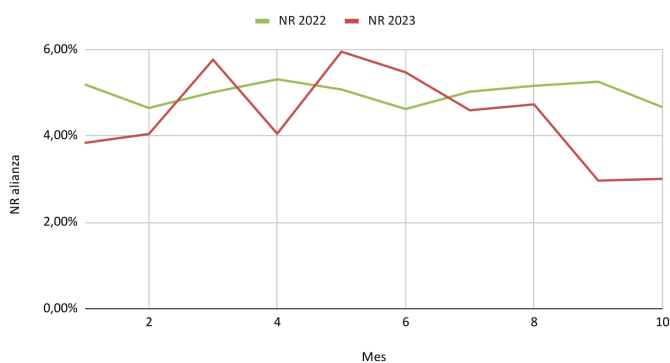
Como se mencionó anteriormente, Justo tiene relaciones comerciales con distintas entidades bancarias que desean participar en campañas de alianza que realiza Justo. Con estas campañas se benefician ambas partes, las entidades bancarias se benefician generando atracción de nuevos clientes gracias a los beneficios de sus tarjetas participantes en las alianzas con Justo, asimismo, Justo se beneficia con mayores ingresos por ventas gracias a los descuentos que las tarjetas tienen en las distintas marcas lo que incitan a los consumidores a realizar compras en las distintas plataformas (sitio web de la marca, app de la marca y app Justo). Por esto último, es clave determinar cuál será la respuesta de los clientes a una campaña ya que si esta tiene malos resultados probablemente el partner bancario dejara de ver como una opción llamativa el tener una alianza con Justo.

Por temas de confidencialidad de la empresa, todas las cifras monetarias presentadas en el documento estarán en formato porcentaje (%) del total de la variable que se esté analizando. Dicho esto, es importante recalcar que en lo que va del año las órdenes generadas a partir de alguna alianza representan un 7,04% del total de órdenes y un 4,42% del revenue total de la empresa para el año 2023. La evolución de la cantidad de órdenes y del net revenue de las alianzas se pueden ver en los gráficos a continuación en los cuales se compara el periodo comprendido entre el 1 de enero y el 31 de octubre para el año 2022 y 2023.

Pedidos a partir de alianzas frente a mes


Gráfico 1: Porcentaje del total de pedidos generados a partir de alianzas.¹

Net Revenue alianzas frente a Mes


Gráfico 2: Porcentaje de net revenue a partir de alianzas.²

Como se puede observar en ambos gráficos la comparativa en porcentajes entre el año 2022 y 2023 nos entrega un valor parecido tanto en porcentaje de órdenes como en porcentaje de net revenue, pero al ver la evolución por número de órdenes se tiene una reducción de 72.488 órdenes respecto al año pasado lo que afecta de manera importante el net revenue de la empresa. Esta información se puede verificar en las tablas presentadas en el anexo 2 y 3.

¹ Krarup, 2023

² Krarup, 2023

Actualmente el sistema de alianzas funciona mediante un acuerdo entre las partes en el que se establece un presupuesto y un número de canjes disponibles por consumidor. Por ejemplo, se establece una campaña con un presupuesto de CLP \$10.000.000. La idea de la alianza es agotar todo el presupuesto para alcanzar la maximización de ingresos y que no quede dinero sin gastar que se traducen en potenciales ingresos para la empresa. Si bien Justo realiza y estudia cuando son momentos oportunos para establecer las alianzas muchas veces este presupuesto no es usado al 100% y no se alcanzan los resultados esperados. Esto se debe a diversos factores como la época del año, preferencias de los consumidores y mala asignación de presupuestos y canjes. A continuación se muestra una tabla resumen de la utilización de presupuesto de las principales alianzas que maneja Justo.

Alianzas principales	Promedio de presupuesto utilizado
Onepay	87,25%
Fpay	77,40%
Cencosud Scotiabank	85,67%
Cencosud Black	74,40%

Tabla 1: Promedio de utilización de presupuesto para cada alianza³

Las alianzas mostradas en la tabla son las cuatro más importantes y las que más revenue aportan a la empresa, como se puede observar no se llega al 100% de los presupuestos asignados por lo que hay un porcentaje de ingresos no captado.

Sumado a las alianzas, las campañas de marketing mediante mailing son claves para tener una mejor tasa de conversión y respuesta de los clientes. Como se observa en el siguiente gráfico la conversión del mailing presenta una tendencia a la baja a lo largo de este año por lo que se generan menos ventas a partir de campañas en comparación a principio de año. Esto, a su vez, le genera un ROAS más bajo a la empresa ya que si bien se puede estar realizando la misma inversión, las ventas son menores.

³ Krarup, 2023

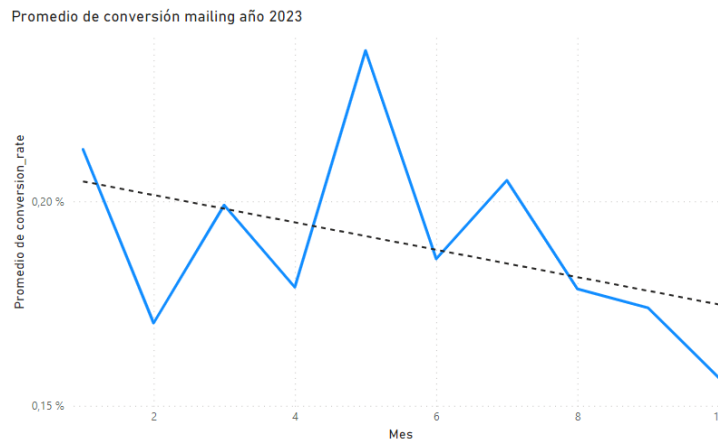


Gráfico 3: Conversión mailing durante año 2023⁴

Es importante subir los niveles de conversión ya que las marcas clientes de Justo pagan por el servicio de mailing. Este servicio tiene un costo variable para las marcas que depende de los contratos establecidos entre la marca y Justo. Estos valores van desde CLP \$1 + IVA hasta los CLP \$5 + IVA por cada email enviado a un cliente. A modo de ejemplo, si un email de \$1 es enviado a una marca que tiene 100.000 clientes en su base de datos a la marca le costará \$100.000 + IVA el envío del email. Si la marca ve que el mailing no está teniendo el efecto deseado puede dejar de usar este servicio lo que se traduce en menos revenue para Justo y una desconexión entre los restaurantes y sus clientes.

La oportunidad que se presenta para la empresa es poder obtener una mejora significativa en el proceso y enfoque que se le da a las campañas de marketing para así poder obtener mayores beneficios de estas y mejorar la relación con las marcas. Hoy en día no se está teniendo en cuenta el comportamiento del cliente y cuales son las necesidades específicas de este. Teniendo en cuenta que con el crecimiento y nuevas posibilidades de la era digital es clave tener buenas campañas de marketing que sean personalizables, visibles para el cliente y que estén alineadas con sus preferencias para poder mostrar un contenido más direccionado. Lo anterior trae un valor agregado para la empresa ya que al tener una mejora en la gestión de campañas de marketing se puede a su

⁴ Krarup, 2023

vez mejorar la gestión de recursos destinados a las campañas. Sumado a lo anterior, el mostrar un contenido más personalizado mejora notablemente la experiencia del cliente.

Objetivos

Los objetivos para este proyecto se definieron en conjunto con el líder del área para garantizar que los resultados esperados sean números realmente alcanzables y realistas conforme a la duración del proyecto.

Objetivo general

Aumentar la tasa de conversión de las ventas en 3 puntos porcentuales para cada marca y mejorar el ROAS de las campañas en al menos un 20% para cada marca en un período de 5 meses.

Objetivos específicos

1. Realizar análisis de datos y segmentación de audiencia. Esto se llevará a cabo para identificar distintos segmentos de audiencia, permitiendo dirigir estrategias de marketing específicas para satisfacer las necesidades de cada grupo.
2. Realizar pruebas A/B en las campañas de marketing para determinar qué estrategias generan las mejores tasas de conversión y ROAS.
3. Incrementar la tasa de conversión del mailing un 30% para cada marca en los próximos 5 meses.
4. Evaluar y ajustar el presupuesto de marketing para asignar recursos de manera eficiente y mejorar el ROAS.
5. Comparar el rendimiento de las alianzas antes y después de la implementación de la solución. Medir un aumento mínimo del 20% en el ROAS en las diversas campañas y un aumento de 3 puntos porcentuales en la tasa de conversión para cada marca.

Estado del arte

Durante los últimos años y más aún producto de la pandemia, las ventas por delivery han aumentado exponencialmente en Chile y el mundo, según se constata en el sitio web de la Cámara De Comercio De Santiago “las ventas de delivery de restaurantes y locales de comidas pasaron de 295,6 millones de dólares en 2014 a 483,1 millones de dólares en 2019, es decir, un incremento de 63,4% en cinco años. Crecimiento importante, aunque ampliamente rebasado por las cifras de 2020: 820 millones de dólares en venta, lo que se traduce en un alza de 69,7% respecto a 2019.”⁵ Si bien estas cifras dejaron de crecer en forma exponencial debido al término de las restricciones sanitarias, las diversas aplicaciones de delivery expandieron su base de clientes enormemente y pudieron captar una gran cantidad de clientes que ahora deben retener.

Como se mencionó anteriormente, el problema encontrado consiste en la baja capacidad de Justo para predecir las tendencias y comportamientos de compra de sus consumidores lo que se traduce en campañas de alianzas y marketing que no siempre resultan ser las mejores. Para tener una idea general de qué hacen los competidores y la industria alrededor del mundo y cómo enfrentan el tema de las campañas de marketing direccionadas, se han investigado en distintas fuentes de información como papers y sitios web, algunas técnicas y herramientas que actualmente usan grandes empresas para poder lograr captar de mejor manera a sus consumidores y lograr campañas eficientes.

Como primer acercamiento se tiene el caso de UberEATS India quienes utilizan datos basados en la ubicación del cliente para un marketing eficiente. “UberEATS utiliza sus propios datos de viajes y clientes de Uber en India para orientar su marketing por ubicación, destino y hora del día para aumentar los pedidos de manera más eficiente.”⁶ India es un país con una cultura alimentaria muy localizada por lo que tener una sola estrategia de marketing para toda la región no sirve y con este tipo de solución pudieron aumentar la conversión 12 veces respecto a como era antes que se implementara este sistema. Este tipo de mecanismo es una ventaja que presenta UberEATS ya que utiliza la misma infraestructura de la aplicación de transportes que al estar conectado a los viajes en tiempo real del consumidor en la aplicación, puede generar distintas estrategias de acuerdo a la zona del país en la que se encuentra el cliente haciendo que el marketing sea focalizado en este.

⁵ *Servicio De Delivery: Las Empresas Se Adaptaron En Pandemia*, n.d.

⁶ *UberEATS Uses Location Data for Efficient Marketing*, 2018

Otro caso es el de Rappi, empresa con alta presencia en Chile y Latinoamérica. Rappi ideó su forma de llegar al consumidor de una forma más eficiente mediante la creación de una plataforma de marketing digital B2B que ayuda a la marca a conocer mejor a sus clientes, anticipando sus necesidades y así pudiendo crear campañas de marketing mejor segmentadas que permitan tener un mejor rendimiento. “La plataforma de Brands By Rappi funciona a través del análisis de datos y la aplicación de estos en la toma de decisiones y creación de campañas de tu empresa o marca. Para el análisis de datos, Brands by Rappi utiliza los datos inferidos en base a los patrones de consumo de FMCG de sus usuarios y también datos declarados como sexo y edad que el usuario confirma al crear su perfil.”⁷ Actualmente Brands by Rappi ya cuenta con más de 468 empresas que utilizan la plataforma en su día a día. Esta solución entrega una serie de datos e insights que ayudan a las marcas clientes de Rappi para la creación de campañas correctamente segmentadas y así las marcas puedan tener más información respecto al comportamiento de los consumidores y poder conectar mejor con estos ofreciéndoles lo que ellos buscan.

Por otra parte, Netflix, uno de los servicios de streaming más grande del mundo que ofrece una gran variedad de películas, series y documentales, incorpora e invierte fuertemente en técnicas de machine learning en distintas áreas de la compañía para lograr una mejor experiencia del usuario y optimizar el servicio de Netflix end-to-end. “El aprendizaje automático impacta muchas áreas interesantes en toda nuestra empresa. Históricamente, la personalización ha sido el área más conocida, donde el aprendizaje automático impulsa nuestros algoritmos de recomendación. También utilizamos el aprendizaje automático para ayudar a dar forma a nuestro catálogo de películas y programas de televisión al aprender las características que hacen que el contenido sea exitoso... También impulsa nuestra inversión en publicidad, combinación de canales y creatividad publicitaria para que podamos encontrar nuevos miembros que disfruten de Netflix.”⁸

Vistos estos tres casos se puede notar que las tres empresas utilizan el big data aplicado de distintas maneras dado los recursos que se tienen para poder cumplir sus objetivos. Teniendo en cuenta la situación actual de Justo y las posibilidades que el big data y los distintos mecanismos y tecnologías

⁷ Brands by Rappi, *¡La Herramienta Ideal De Marketing B2B!*, 2021

⁸ Netflix Research, 2018



que hay disponibles se puede evaluar el utilizar técnicas similares para resolver el problema anteriormente planteado.

Solución

Alternativas de solución

Después de lo investigado en el inciso anterior y de conversar con las distintas áreas que se ven involucradas en el proyecto se establecieron dos posibles soluciones relacionadas con lo propuesto en estado del arte.

1. Plataforma de marketing digital B2B:

Esta plataforma cumpliría la misma función que la comentada anteriormente llamada “Brands by Rappi” pero con una implementación completamente diferente. La creación de una plataforma de este estilo permitiría que la marca pueda acceder fácil y cómodamente a la información y comportamiento de compra que tienen sus clientes para así conocer mejor las preferencias de clientes y entregar contenido y ofertas personalizadas de acuerdo a las necesidades de cada uno. El acceso a este tipo de información es vital para que así las marcas puedan establecer una conexión más eficiente con sus clientes. Esta solución fue pensada como una integración al “admin” de Justo en el que están todas las marcas. Aquí cada marca tiene un resumen de sus ventas, clientes, reservas, etc. y además desde esta herramienta se controlan todas las configuraciones de la marca como su página web, cupones de descuento, menús, entre otros. A continuación se muestra una imagen que muestra un “admin” tipo demo en este caso para una marca ficticia.

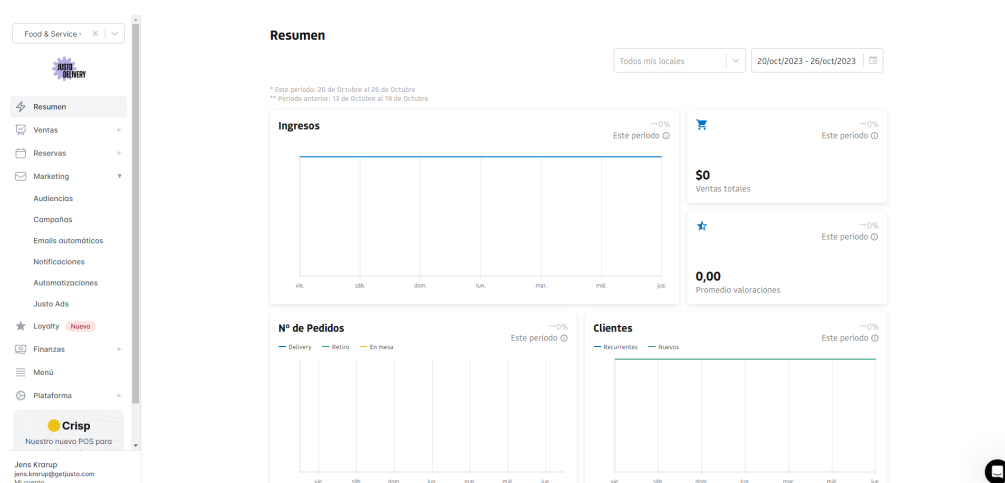


Imagen 1: Admin demo⁹

⁹ Justo, 2022

Se buscará implementar la plataforma dentro del inciso de marketing que se puede observar en Imagen 1, esto con el fin de que las marcas no tengan que manejar más de una plataforma para la visualización de sus datos, de esta forma se mantiene la simplicidad para las marcas a la hora de manejar las herramientas que Justo les proporciona.

2. Modelos de segmentación de clientes y predicción utilizando machine learning:

Esta solución se basa a nivel muy generalizado a lo que hace Netflix pero adaptado a las necesidades y recursos disponibles en Justo. Para abordar el desafío de mejorar la conversión es clave entender de mejor manera al cliente por lo que en primera instancia se hará una segmentación de clientes mediante técnicas de machine learning y así poder determinar el comportamiento de compra de los clientes en las distintas marcas y catalogarlos de acuerdo a sus patrones de compra. Posterior a la segmentación se harán modelos de predicción con el fin de predecir la respuesta de los clientes frente a campañas publicitarias y así poder lograr mejorar la conversión y el ROAS de las campañas ya que se les dará un enfoque mucho más direccionado dado que la campaña será dirigida a clientes que tienen un buen historial de compra en base a campañas. Esta proyección irá acompañada de pronósticos de demanda para que la marca pueda tener una idea de la cantidad de pedidos que podrían haber durante alguna campaña. Esto permitirá también a Justo tener a disposición más repartidores disponibles cercanos al local de la marca para poder afrontar un posible aumento de la demanda sin tener mayores complicaciones.

Elección de solución

Para la elección de la solución se discutieron los puntos detallados en la tabla presentada a continuación para los cuales se estableció una escala de Likert que va desde el 1 al 5 siendo 1 muy malo y 5 muy bueno. Cabe destacar que para los criterios costo de implementación y tiempo de implementación se utilizó una escala inversa, es decir, 1 muy bueno y 5 muy malo. Estos puntajes se establecieron entre el alumno, el supervisor, equipo de business intelligence (BI) y equipo de TI. Sumado a los puntajes que dió cada equipo se conversó posteriormente respecto a la escalabilidad y posibilidad de implementación en el corto plazo debido a las prioridades de la empresa.

Criterio	Ponderación
Factibilidad de implementación	10%
Impacto del proyecto	25%
Alcance del proyecto	25%
Costo de implementación	20%
Tiempo de implementación	20%

Tabla 2: Criterios elección solución¹⁰

Una vez discutidos, analizados y calificados los criterios anteriores se obtuvieron los siguientes resultados:

Solución / Criterio	Impacto del proyecto (25%)	Alcance del proyecto (25%)	Factibilidad de implementación (10%)	Costo de implementación (20%)	Tiempo de implementación (20%)	Ponderación final
Plataforma digital B2B	5	4	2	3	2	3,6
Modelos machine learning	4	3	4	4	3	3,75

Tabla 3: Asignación de puntaje y ponderación para cada solución.¹¹

Como se observa en Tabla 3, la solución que obtuvo una mayor ponderación fue la de incorporar modelos de machine learning.

¹⁰ Krarup, 2023

¹¹ Krarup, 2023

Posterior a los resultados obtenidos en la matriz anterior se reunió al equipo para ver los principales insights de porqué se eligió esta como la solución a realizar. Si bien todos concordaron que ambas soluciones brindan un valor agregado a lo que hoy se tiene en la empresa y cómo se manejan las campañas de marketing, la solución de la plataforma digital B2B demanda más tiempo y personas que trabajen en su desarrollo e implementación y hoy en día la empresa tiene destinados esos recursos a proyectos que fueron puestos en marcha antes que el proyecto que se describe en este documento. Esto no quiere decir que en un futuro no lejano no se de la posibilidad de incorporar una plataforma como esta dentro de Justo ya que brinda mucha información de valor que ayudaría a cumplir los objetivos previamente planteados.

Dicho esto, con la solución escogida se espera cumplir con la mejora en la tasa de conversión en 3 puntos porcentuales y un aumento del ROAS de las campañas en un 20% mediante la utilización de técnicas de machine learning que se usarán para la segmentación de clientes y una posterior predicción de respuestas a campañas de marketing para cada grupo de clientes.

Desarrollo de la solución

Para el desarrollo del proyecto se utilizó Jupyter Notebook como la plataforma para realizar los códigos y obtener los resultados. En este inciso del documento se mostrarán imágenes generales del desarrollo de los códigos y sus resultados. El código completo se puede encontrar en el anexo 4. Cabe destacar que el código fue anexado en formato pdf ya que los outputs contienen información de los clientes que la empresa no permite revelar. El proyecto fue realizado en un plan piloto para 5 marcas de comida rápida las cuales son Chicken Love You, Mozzarella, Shawarma King, Buffalo Waffles y Pizzería Capri. La elección de las marcas y categoría se eligió en conjunto con la empresa en base a su rendimiento del último año, la efectividad de su mailing, su ticket promedio y el revenue que estas marcas aportan a la empresa. El proceso realizado para las 5 marcas fue el mismo por lo que a modo de demostración se utilizara el desarrollo de la marca Chicken Love You.

1. Limpieza y transformación de los datos

La limpieza de datos es esencial para garantizar la calidad y la integridad de estos, lo que afecta directamente la precisión de los análisis posteriores. Los datos limpios son fundamentales para tomar decisiones informadas por lo que el primer paso es asegurarse de que los datos estén limpios

y listos para su análisis. Esto implica eliminar valores atípicos, valores faltantes y corregir inconsistencias en los datos.

El data frame original consta de 26 variables detalladas en la imagen a continuación

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110659 entries, 0 to 110658
Data columns (total 27 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   _id                                         110659 non-null object
1   userId                                     110659 non-null object
2   websiteId                                  110659 non-null object
3   email                                       110659 non-null object
4   firstLogin                                 110659 non-null datetime64[ns]
5   fullName                                   110659 non-null object
6   importedAt                                65 non-null     datetime64[ns]
7   lastLogin                                  107297 non-null datetime64[ns]
8   ordersCount                               110659 non-null int64
9   amount                                     110659 non-null int64
10  phone                                       106527 non-null float64
11  search                                     110659 non-null object
12  firstOrder                                 95287 non-null  datetime64[ns]
13  lastOrder                                  95287 non-null  datetime64[ns]
14  blockMarketingEmail                       12584 non-null  float64
15  cityId                                     73182 non-null  object
16  lastVisitedStoreId                        87936 non-null  object
17  categoriesIds                             0 non-null     float64
18  productsIds                               0 non-null     float64
19  channelTypeUsed                           0 non-null     float64
20  paymentMethodsUsed                        0 non-null     float64
21  latestPersonalDataSignedTerms             0 non-null     float64
22  firstLoyaltyRedemption                    2733 non-null  datetime64[ns]
23  blockMarketingNotifications               170 non-null   float64
24  blockMarketingSMS                         257 non-null   float64
25  isNotFavorite                             75 non-null    float64
26  subscribedFromEMenuAt                     0 non-null     float64
dtypes: datetime64[ns](6), float64(11), int64(2), object(8)
memory usage: 22.8+ MB
```

Imagen 2: Variables y valores nulos originales¹²

Al analizar los datos se pudo constatar información de las variables relevantes que no deben ser eliminadas sino transformadas. Estas corresponden a “blockMarketingEmail” y “lastOrder”. Partiendo por la primera esta representa si un cliente se dio de baja del mailing de la marca. Los valores no nulos están representados por un 1 por lo que se procedió a llenar los valores con 0 con el fin de obtener una variable binaria que representa si el cliente está habilitado o no para recibir mails. Continuando con la variable “lastOrder” ésta representa la fecha en la que el cliente hizo un pedido por última vez. En la base de datos hay clientes registrados que no han hecho compras por lo que para efectos del proceso que se realizará después se eliminaron a todos los clientes registrados pero sin órdenes con el fin de enfocar el proyecto en clientes activos. Posteriormente se eliminaron las variables consideradas como irrelevantes para el proceso por lo que se trabajó con las variables presentadas en la siguiente tabla.

¹² Krarup, 2023

Número de variable	Nombre variable	Descripción
1	userId	Id del cliente
2	email	Email del cliente
3	firstLogin	Fecha de registro del cliente
4	fullName	Nombre completo del cliente
5	lastLogin	Última fecha de ingreso del cliente
6	ordersCount	Cantidad de pedidos del cliente
7	amount	Cantidad de dinero gastado del cliente
8	phone	Teléfono celular del cliente
9	firstOrder	Fecha del primer pedido del cliente
10	lastOrder	Fecha del último pedido del cliente
11	blockMarketing Email	Cliente habilitado para recibir mails (binaria)
12	cityId	Ciudad del cliente

Tabla 4: Variables de data frame actualizado¹³

Una vez ya con los datos listos para ser trabajados se pasa a la siguiente etapa del desarrollo de la solución la cual corresponde a un análisis mediante la técnica RFM.

2. Análisis RFM

Para la segunda parte de la solución correspondiente a la segmentación de clientes, se utilizara una técnica de segmentación llamada RFM (Recencia, Frecuencia y valor monetario). Esta segmentación utiliza la información de cuándo fue la última compra, cuántas veces ha comprado y cuánto dinero ha gastado un cliente en una determinada marca.

En base a esto último y a la información disponible en los “admin” de cada marca es que se optó por esta técnica de segmentación que servirá para determinar a los distintos segmentos de clientes que existen en cada marca.

Para poder realizar la segmentación se agregaron columnas al data frame que calculan los días que pasaron hasta el 11 de noviembre del 2023 según corresponda para cada fila. Estas columnas fueron llamadas “Dias_firstLogin”, “Dias_firstOrder” y “Dias_lastOrder”. Una vez obtenidos estos valores,

¹³ Krarup, 2023

fueron escalados del 1 al 5 con el fin de poder realizar un trabajo más pulcro y rápido en las etapas posteriores.

```
In [14]: scaler = MinMaxScaler(feature_range=(1, 5))
df_rfm["Recencia_Score"] = scaler.fit_transform(df_rfm[["Dias_lastOrder"]])
df_rfm["Frecuencia_Score"] = scaler.fit_transform(df_rfm[["ordersCount"]])
df_rfm["ValorMonetario_Score"] = scaler.fit_transform(df_rfm[["amount"]])
df_rfm["Recencia_Score"] = 6 - df_rfm["Recencia_Score"]

df_rfm["RFM_Score"] = df_rfm["Recencia_Score"] + df_rfm["Frecuencia_Score"] + df_rfm["ValorMonetario_Score"]

df_rfm["Segmento"] = "Otros"
df_rfm.loc[df_rfm["RFM_Score"] >= 9, "Segmento"] = "Cliente VIP"
df_rfm.loc[(df_rfm["RFM_Score"] >= 7.5) & (df_rfm["RFM_Score"] < 9), "Segmento"] = "Cliente fidelizado"
df_rfm.loc[(df_rfm["RFM_Score"] >= 5) & (df_rfm["RFM_Score"] < 7.5), "Segmento"] = "Cliente ocasional"
df_rfm.loc[(df_rfm["RFM_Score"] >= 2.5) & (df_rfm["RFM_Score"] < 5), "Segmento"] = "Cliente en riesgo"
df_rfm.loc[(df_rfm["RFM_Score"] >= 0) & (df_rfm["RFM_Score"] < 2.5), "Segmento"] = "Cliente perdido"
```

Imagen 3: Creación variables para segmentación RFM y segmento de cliente correspondiente¹⁴

Una vez se obtuvieron los resultados de “RFM_Score” para cada cliente, se segmentó en 5 grupos que son los que se observan en la imagen 3. Luego se elaboraron representaciones gráficas para visualizar de mejor manera los resultados de cada segmento. Estas están presentes en el código del anexo 4.

Una vez obtenidos la cantidad de clientes por segmento y los resultados de recencia, frecuencia y valor monetario por cliente se desarrolló la siguiente etapa del proyecto.

3. Clusterización

La clusterización permite agrupar segmentos de clientes de manera automática, lo que es valioso para personalizar estrategias de marketing y adaptar las campañas a las necesidades y preferencias de grupos específicos. El algoritmo escogido para esta parte del proyecto es k-means ya que este es el que mejor se adecua al formato de los datos que ya están escalados y fueron previamente obtenidos del análisis RFM. K-means sirve para poder dividir un conjunto de datos en grupos o también llamados clusters. Estos grupos contienen en este caso a clientes con características similares en cuanto al score de recencia, frecuencia y valor monetario. Los principales componentes del algoritmo y los procesos de un modelo k-means están dados por:

- **Centroides:** son puntos ficticios que representan el centro de un cluster. En el contexto del k-means, se selecciona un número predefinido de centroides (k) al azar al inicio del algoritmo.

¹⁴ Krarup, 2023

- **Asignación de puntos a clusters:** Cada punto de datos en el conjunto se asigna al cluster cuyo centroide está más cercano a él en términos de distancia euclidiana. Esto crea una partición inicial de los datos.
- **Repetición del proceso:** Los pasos de asignación y recálculo de centroides se repiten iterativamente hasta que se alcanza la convergencia. La convergencia ocurre cuando los centroides ya no cambian significativamente entre iteraciones.

Como primera parte de esta etapa del proyecto se elaboró la curva elbow para obtener el número de centroides óptimo de manera automática obteniendo el siguiente resultado.

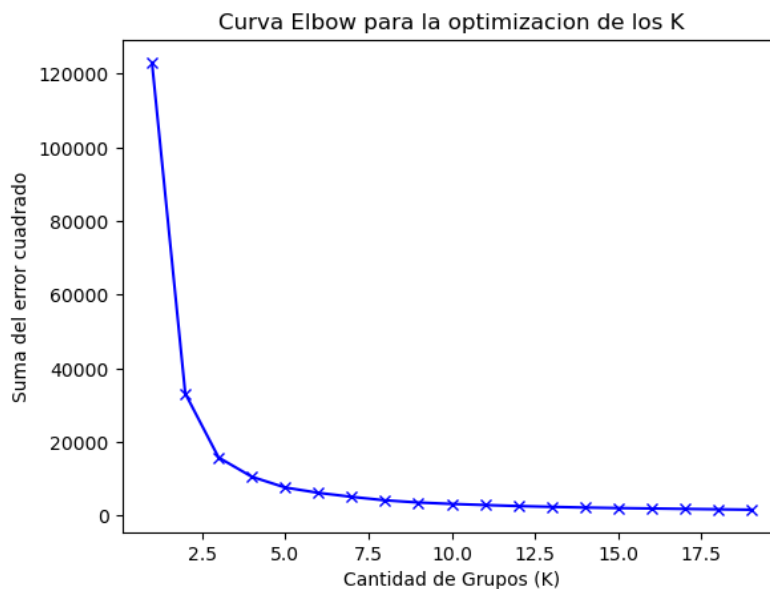


Imagen 4: Curva elbow modelo k-means¹⁵

En base a la imagen anterior se observa que a partir de $K=6$ la suma del error cuadrado deja de cambiar considerablemente por lo que para esta marca el número de clusters que habrá será de 6. Una vez ya ejecutado el algoritmo para esta marca en específico, se elaboraron representaciones gráficas para las variables de recencia, frecuencia y valor monetario según el cluster al que pertenece un cliente.

¹⁵ Krarup, 2023

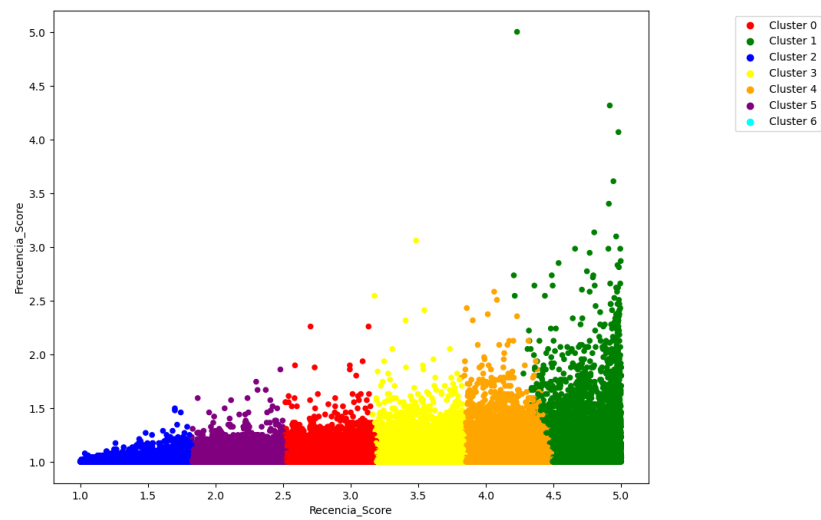


Imagen 5: Clusterización Recencia_Score vs Frecuencia_Score¹⁶

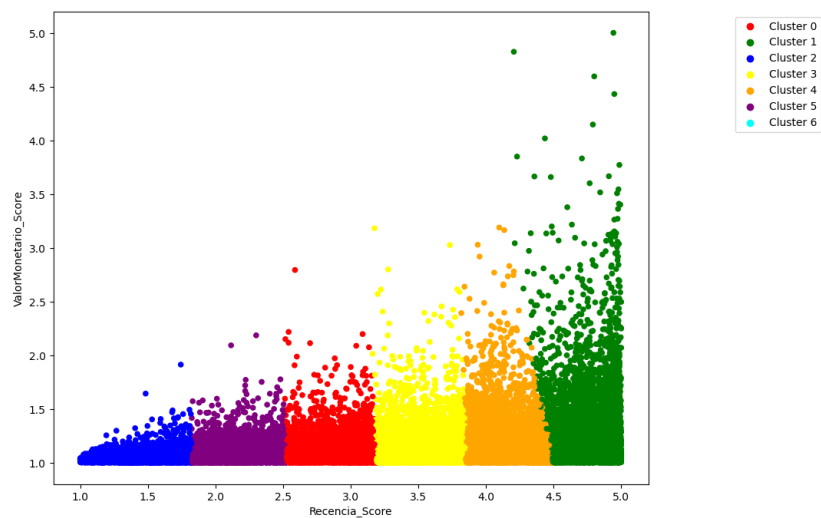


Imagen 6: Clusterización Recencia_Score vs ValorMonetario_Score¹⁷

Las imágenes anteriores ratifican que el modelo segmenta adecuadamente a los clientes en base a los datos entregados ya que como se puede observar prácticamente no hay superposición de datos y los conjuntos pueden ser diferenciados claramente.

¹⁶ Krarup, 2023

¹⁷ Krarup, 2023

Con esta segmentación ya realizada se puede tener una noción más general acerca de qué campañas se deben orientar a qué tipo de clientes.

4. Modelos de predicción de respuesta a campañas de marketing

Los modelos de predicción de respuesta ayudan a identificar cuáles clientes tienen más probabilidades de responder positivamente a las distintas campañas. Esto ayudará a mejorar los gastos, la inversión en marketing y aumentar la tasa de conversión. Es por esto que a modo de complementar los modelos de segmentación realizados previamente, también se desarrollaron modelos predictivos que utilicen la información que se ha recopilado en las etapas anteriores para predecir la respuesta de los clientes a campañas de marketing.

Para el desarrollo del proyecto se consideraron 3 tipos de modelos que fueron puestos a prueba con el fin de decidir en base a los resultados y los recursos computacionales que requieren cual es el más apto para este proyecto. Se probó un modelo de gradient boosting, k-nearest neighbors y support vector machine. A modo de resumen, los resultados de las pruebas se presentan en la siguiente tabla

Modelo	Accuracy	F1 score	Curva ROC	Tiempo de ejecución (segundos)
Gradient Boosting (XgBoost)	0,95	0,96	0,98	15,94
K- Nearest Neighbors (KNN)	0,95	0,95	0,98	38,72
Support vector machine (SVM)	0,94	0,94	0,97	184,81

Tabla 5: Resultados prueba de modelos¹⁸

La explicación detallada de qué representa cada métrica se puede encontrar en el anexo 5. Sumado a las métricas anteriores, se graficaron las curvas de aprendizaje que se encuentran en el anexo 6 y se graficaron las matrices de confusión de los modelos para las cuales sus resultados se encuentran resumidas en la siguiente tabla

¹⁸ Krarup, 2023

Modelo	Número de true positives	Número de true negatives	Número de false positives	Número de false negatives
Gradient boosting (XgBoost)	9.602	8.575	15	866
K- Nearest Neighbors (KNN)	9.363	8.561	254	880
Support vector machine (SVM)	9.428	8.450	189	991

Tabla 6: Resultados matrices de confusión para cada modelo¹⁹

Tomando en cuenta todo lo descrito para esta etapa es que se escogió el modelo KNN como mejor opción. Si bien al mirar la tabla 6 se ve que XGBoost pareciera tener mejores resultados, es importante destacar que como se puede observar en el anexo 6, el modelo no se comporta adecuadamente con los datos de prueba y su precisión está bajo lo deseado. En cambio KNN aumenta su capacidad predictora a la hora de introducir más datos al conjunto de prueba y entrenamiento. Por otra parte, SVM quedó descartado básicamente por el tiempo de ejecución el cual es muy alto en comparación a los otros dos modelos. Con el modelo a utilizar ya seleccionado, se procede a la última etapa del desarrollo de la solución.

5. Pronósticos de demanda para alianzas

Como se mencionó anteriormente, los pronósticos de demanda servirán para que la marca tenga noción de cuántos pedidos esperar durante una alianza con algún banco. Esto generará una mejora operacional en el proceso de entregas de pedidos ya que se podrá aproximar de mejor manera la cantidad de repartidores que deben haber en una zona específica. Para realizar estos pronósticos se utilizó la técnica de regresión lineal debido a la correlación alta entre la cantidad de pedidos totales con los pedidos de alianzas lo cual se puede ver evidenciado en la siguiente tabla.

¹⁹ Krarup, 2023

	<i>Pedidos totales</i>	<i>Pedidos alianza</i>	<i>Pedidos normales</i>
Pedidos totales	1	0,70426867	0,936937093
Pedidos alianza	0,70426867	1	0,411735053
Pedidos normales	0,936937093	0,411735053	1

Tabla 7: Correlación entre variables²⁰

Como se observa en la tabla 7 la correlación entre pedidos totales y pedidos de alianza es bastante alta por lo que un método de regresión lineal sería adecuado para predecir el comportamiento de esos pedidos. A continuación se presenta un gráfico que representa los resultados del pronóstico, los círculos en color naranja corresponden al pronóstico para noviembre y diciembre.

Pronóstico pedidos totales y pedidos alianza

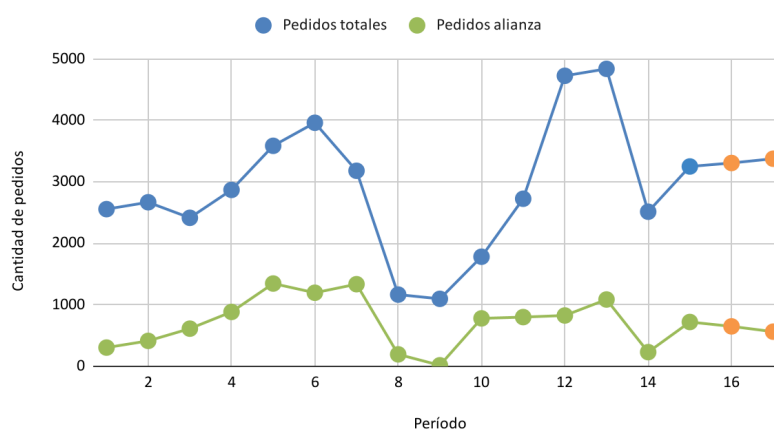


Gráfico 4: Valores pronosticados para noviembre y diciembre 2023²¹

²⁰ Krarup, 2023

²¹ Krarup, 2023

Metodología

Con el fin de cumplir con los objetivos del proyecto es que se utilizará la metodología de trabajo DMAIC (Definir, Medir, Analizar, Mejorar y Controlar) que forma parte de la metodología Six Sigma. Esta brinda un enfoque estructurado utilizado para mejorar procesos y es especialmente efectivo en proyectos de mejora continua. Sumado a esto, se utilizó CRISP-DM. (Cross-Industry Standard Process for Data Mining) únicamente para el desarrollo de la solución. El marco metodológico y las etapas de esta metodología se encuentran en el anexo 7. Dicho esto, la metodología DMAIC aplicada a cada objetivo es la siguiente

1. Definir:

- a. Objetivo específico 1: Realizar análisis de datos y segmentación de audiencia
 - i. Definir el alcance: Establecer claramente los límites de la segmentación de audiencia y su relación con las estrategias de marketing.

2. Medir:

- a. Objetivo específico 2: Realizar pruebas A/B en campañas de marketing
 - i. Medir el rendimiento actual: Recopilar datos históricos sobre las tasas de conversión y ROAS de las campañas existentes.
 - ii. Establecer KPI: Definir indicadores clave de rendimiento para evaluar el éxito de las pruebas A/B.

3. Analizar:

- a. Objetivo específico 3: Incrementar la tasa de conversión del mailing en un 30%
 - i. Análisis de causa raíz: Identificar las causas subyacentes de la tasa de conversión actual y las oportunidades para mejorarla.
 - ii. Evaluación de datos: Realizar un análisis detallado de los datos de mailing para comprender patrones y tendencias.

4. Mejorar:

- a. Objetivo específico 4: Evaluar y ajustar el presupuesto de marketing
 - i. Diseñar estrategias de mejora: Desarrollar estrategias para asignar recursos de manera más eficiente y mejorar el ROAS
 - ii. Implementar cambios: Aplicar ajustes en la asignación de presupuesto basados en las estrategias más efectivas.

5. Controlar:

- a. Objetivo específico 5: Comparar el rendimiento de las alianzas
 - i. Establecer controles continuos: Implementar un sistema de monitoreo continuo para evaluar el ROAS y la tasa de conversión de las alianzas.
 - ii. Procedimientos de ajuste: Definir procedimientos para realizar ajustes inmediatos ante desviaciones no deseadas.

Métricas de desempeño

Para poder determinar si la solución planteada cumple el objetivo planteado anteriormente, es vital poder establecer métricas que permitan medir cuantitativamente el desempeño de la solución.

Estas métricas fueron establecidas en base a una clara relación y congruencia entre el posible impacto y alcance del proyecto. A continuación se detallan las cuatro principales métricas que se utilizarán para medir el impacto del proyecto, qué representa cada una y como se calcula.

- 1) **Tasa de conversión en plataformas (TCP):** Proporción de visitantes que realizan una compra en la plataforma (sitio web, app propia de la marca y app Justo)

$$TCP = \frac{\text{Cantidad de compras realizadas}}{\text{Cantidad de visitas en la plataforma}} \cdot 100$$

Imagen 7: Fórmula tasa de conversión en plataformas²²

Actualmente la conversión que se tiene es de un 6,5% a nivel general en Justo. Como se detalló en el objetivo general, se buscará aumentar la conversión en 3 puntos porcentuales para cada marca.

- 2) **ROAS de campañas:** Mide el retorno de las campañas.

$$ROAS = \frac{\text{Ingresos totales de la campaña}}{\text{Coste total de la campaña}} \cdot 100$$

Imagen 8: Fórmula cálculo de ROAS de campañas²³

También como se mencionó en el objetivo general, se buscará mejorar para cada marca en por lo menos un 20% el ROAS de las campañas. Actualmente este valor está en un 8% a nivel general en Justo. Además servirá para comparar las pruebas A/B que se mencionan en los objetivos específicos.

²² Krarup, 2023

²³ Krarup, 2023

- 3) **Tasa de conversión mailing (TCM)** : Proporción de mails enviados que derivan en una compra.

$$TCM = \frac{\text{Cantidad de compras realizadas durante la campaña}}{\text{Cantidad de mails enviados durante la campaña}} \cdot 100$$

Imagen 9: Fórmula tasa de conversión mailing²⁴

Esta métrica actualmente está en un 0,19% a nivel general en Justo. Esta métrica busca cumplir con el objetivo específico de mejorar la tasa de conversión del mailing en un 30% para cada marca.

- 4) **Precision, Recall y F1 score:** Mide la precisión en las predicciones de los modelos.

$$Precision = \frac{TP}{TP + FP}$$

TP = True positive

TN = True negative

$$Recall = \frac{TP}{TP + FN}$$

FP = False positive

FN = False negative

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

Imagen 10: Fórmulas precision, recall y f1 score²⁵

Estas 3 métricas explicadas en anexo 5 permiten definir si un modelo está teniendo un buen rendimiento o no. Para efectos de este proyecto como kpi esperado se tiene que para el desempeño de los modelos se espera un 0,93 de acierto (93%) como meta para F1 score.

²⁴ Krarup, 2023

²⁵ Krarup, 2023

Plan de implementación

Para la implementación del proyecto, se deberán considerar las siguientes etapas:

1. Etapa de integración piloto

En esta etapa se pondrá en marcha la implementación de la solución de forma gradual para las cinco marcas pertenecientes al plan piloto con el fin de poder adaptarlos si es que estos presentan deficiencias o errores. Se comenzará en noviembre con algunas de las cinco marcas para corroborar que efectivamente los modelos funcionen adecuadamente.

2. Etapa de capacitación

Al ser una incorporación de una nueva herramienta, se debe capacitar al personal para que pueda hacer uso correcto de esta. Por esto mismo es importante enseñar y realizar capacitaciones durante las primeras semanas de implementación del proyecto.

3. Etapa de integración completa

Una vez teniendo los modelos en las condiciones adecuadas, estos se implementarán para el resto de las marcas.

4. Etapa de monitoreo

Como bien se mencionó anteriormente, se busca que los modelos cumplan un cierto estándar, por lo que se deben estar monitoreando constantemente para que estos no queden desactualizados y el desempeño sea el correcto.

A continuación se muestra la planificación más detallada por etapas para todo el proyecto, incluyendo las fases previas hasta la implementación del proyecto.

N° Etapa	Meses / Semanas	Agosto					Septiembre					Octubre					Noviembre					Diciembre					Post Dicie m bre
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	
	Definición, planificación inicial y recolección de datos																										
1.1	Identificar los objetivos específicos y las expectativas del proyecto.																										
1.2	Establecer el presupuesto y los recursos disponibles.																										
1.3	Definir los plazos y cronogramas.																										
1.4	Recopilar y analizar los datos existentes, incluyendo datos de clientes y campañas anteriores.																										
1.5	Identificar y corregir problemas en los datos, como valores atípicos y datos faltantes.																										
1.6	Unificar diversas fuentes de datos en una única base de datos.																										
1.7	Realizar la ingeniería de características para crear variables relevantes para el modelo de machine learning.																										
2	Desarrollo e implementación de la solución																										
2.1	Evaluar posibilidades y seleccionar los algoritmos de machine learning adecuados para la segmentación de clientes y la predicción de respuestas.																										
2.2	Utilizar datos históricos para entrenar el modelo.																										
2.3	Validar el modelo utilizando técnicas como validación cruzada y ajustar hiper parámetros.																										
2.4	Realizar pruebas exhaustivas para asegurarse de que el sistema funcione correctamente.																										
2.5	Evaluar el rendimiento del sistema en términos de tasa de conversión y ROAS utilizando datos de prueba y comparándolos con los resultados deseados.																										
2.6	Evaluar el impacto del proyecto en términos de mejora en la tasa de conversión y ROAS.																										
2.7	Realizar ajustes finales en el modelo de machine learning según sea necesario.																										
2.8	Establecer un sistema de monitoreo continuo para supervisar el rendimiento del modelo y la precisión de las predicciones.																										
2.9	Capacitar al personal																										
2.9.1	Integrar completamente los modelos																										
2.9.2	Realizar ajustes periódicos en el modelo y la estrategia de marketing basados en los resultados y el feedback.																										

Imagen 11: Planificación de proyecto e implementación de la solución²⁶

²⁶ Krarup, 2023

Análisis de riesgo

Una vez definido el plan de implementación es importante realizar un análisis de riesgo para cerciorarse que los riesgos que el proyecto pueda tener sean mitigables y tengan el menor impacto posible en el desarrollo de la solución. Con el fin de lograr lo anterior, se expone a continuación el procedimiento para catalogar las tareas del plan de implementación de acuerdo a una escala de valores de 1 a 3 tanto para la probabilidad de ocurrencia como para el impacto que tendrá en el proyecto una determinada tarea, siendo 1 una probabilidad baja y 3 una probabilidad alta, así como 1 para un impacto bajo y 3 un impacto alto. La clasificación de riesgo de cada tarea se guiará de acuerdo a la matriz de riesgo presentada a continuación.

Probabilidad de ocurrencia	Impacto		
	Bajo (1)	Medio (2)	Alto (3)
Alta (3)	3	6	9
Media (2)	2	4	6
Baja (1)	1	2	3

Tabla 8: Matriz de probabilidad de ocurrencia - impacto²⁷

Para el análisis, se definió en conjunto con la empresa cuales son los eventos de riesgo y las potenciales consecuencias de estos, las cuales se detallan en la siguiente tabla

²⁷ Krarup, 2023

Riesgo	Potencial consecuencia
1. Objetivos cambiantes o mal definidos	Desviación del proyecto de su objetivo principal
2. Riesgos no identificados o mal evaluados	Errores inesperados durante la implementación
3. Desalineación de partes involucradas	Dificultad en toma de decisiones y cambios no deseados
4. Problemas en la obtención de datos	Retraso en implementación y baja calidad de modelos
5. Problemas en el rendimiento de los modelos	Ineficiencia de la solución en campañas de marketing
6. Problemas de escalabilidad	Dificultad para implementar el proyecto a más marcas
7. Problemas inesperados durante la implementación	Interrupción del proyecto o pérdida de datos
8. Falta de seguimiento post-implementación	Dificultad para identificar y corregir problemas de rendimiento

Tabla 9: Eventos de riesgo y potenciales consecuencias.²⁸

Ya teniendo claros cuales son los eventos de riesgo, estos fueron asignados en una matriz de riesgo que se exhibe a continuación

²⁸ Krarup, 2023

Evento	Probabilidad de ocurrencia	Impacto	Nivel Riesgo	Nivel Riesgo
1. Objetivos cambiantes o mal definidos	1	3	3	Bajo
2. Riesgos no identificados o mal evaluados	1	3	3	Bajo
3. Desalineación de partes involucradas	1	2	2	Bajo
4. Problemas en la obtención de datos	2	3	6	Moderado
5. Problemas en el rendimiento de los modelos	2	3	6	Moderado
6. Problemas de escalabilidad	2	3	6	Moderado
7. Problemas inesperados durante la implementación	2	3	6	Moderado
8. Falta de seguimiento post-implementación	2	2	4	Moderado

Tabla 10: Matriz de riesgo para cada evento²⁹

Con la matriz de riesgo ya establecida, se deben considerar las mitigaciones necesarias para que la implementación sea llevada a cabo de manera exitosa. Para esto, se establecieron las siguientes mitigaciones para cada evento de riesgo presentadas en la tabla 11.

²⁹ Krarup, 2023

Evento	Mitigación
1. Objetivos cambiantes o mal definidos	1. Realizar una reunión detallada para definir y documentar los objetivos del proyecto. 2. Establecer un proceso de cambio controlado para abordar cambios en los objetivos
2. Riesgos no identificados o mal evaluados	1. Realizar una evaluación exhaustiva de riesgos antes del inicio del proyecto. 2. Actualizar y revisar la evaluación de riesgos periódicamente
3. Desalineación de partes involucradas	1. Establecer un plan de comunicación claro y regular con todas las partes interesadas. 2. Asegurar que todas las partes comprendan y estén alineadas con los objetivos del proyecto.
4. Problemas en la obtención de datos	1. Realizar un plan de recolección de datos detallado y garantizar la calidad de los datos. 2. Tener un plan de contingencia para abordar problemas de calidad de datos.
5. Problemas en el rendimiento de los modelos	1. Realizar pruebas exhaustivas de los modelos antes de la implementación. 2. Definir criterios claros de rendimiento y realizar ajustes según sea necesario.
6. Problemas de escalabilidad	1. Planificar la escalabilidad de la solución desde el principio. 2. Identificar los puntos críticos y establecer planes para escalabilidad vertical u horizontal.
7. Problemas inesperados durante la implementación	1. Establecer un plan de contingencia para abordar problemas inesperados. 2. Realizar copias de seguridad regulares y pruebas de restauración de datos.
8. Falta de seguimiento post-implementación	1. Establecer un proceso de seguimiento post-implementación con revisiones regulares. 2. Definir métricas claras para evaluar el rendimiento y la satisfacción del usuario.

Tabla 11: Mitigaciones para cada evento de riesgo³⁰

³⁰ Krarup, 2023

Evaluación económica

Para evaluar correctamente el proyecto, se realizaron 2 flujos de cajas. Uno correspondiente al proyecto piloto que va desde agosto 2023 hasta diciembre 2023 y otro para el año 2024.

En primera instancia se obtuvieron los datos correspondientes para el cálculo de la tasa de descuento lo que posteriormente permite el cálculo del VAN y TIR de ambos flujos. Estos valores fueron obtenidos del Banco Central de Chile³¹ y del sitio del investigador financiero Aswath Damodaran³².

Tasas	
Tasa libre de riesgo	6,20%
Tasa de mercado	9,50%
Beta industria	1,27
Tasa descuento (r)	10,39%

Tabla 12: Tasas para el calculo de VAN y TIR³³

La elección de calcular el VAN y TIR para este proyecto se basa en la necesidad de tomar decisiones financieras fundamentadas y centradas en resultados. Estas métricas ofrecen una evaluación robusta de la rentabilidad y eficiencia económica del proyecto, al considerar el valor temporal del dinero y proporcionar insights claros sobre la atracción financiera a lo largo del tiempo.

A continuación se presentan ambos flujos, en anexo 8 se encuentra el proceso para realizar las estimaciones de ingresos, inversión y los costos asociados.

³¹ Damodaran, 2023

³² Banco Central, 2023

³³ Krarup, 2023

	Flujo proyecto piloto 2023					
	Mes					
	agosto	septiembre	octubre	noviembre	diciembre	
Ingresos				\$547.880	\$1.255.775	
Costos fijos						
UAI	\$0	\$0	\$0	\$547.880	\$1.255.775	
Impuesto a la renta	\$0	\$0	\$0	-\$147.927	-\$339.059	
UDI	\$0	\$0	\$0	\$399.952	\$916.715	
Flujo caja operacional	\$0	\$0	\$0	\$399.952	\$916.715	
Inversión fija	-\$1.149.425					
Flujo de caja de capitales	-\$1.149.425					
Flujo de caja privado	-\$1.149.425	\$0	\$0	\$0	\$399.952	\$916.715
VAN	-\$320.903					
TIR	2,94%					

Tabla 13: Flujo de caja proyecto piloto³⁴

Se observa una discrepancia entre el VAN y la TIR. Mientras que la TIR indica una tasa positiva, lo cual sugiere un retorno sobre la inversión, el VAN presenta un valor negativo.

Esta discrepancia podría ser atribuible a diversos factores, entre ellos, una tasa de descuento elevada o expectativas poco realistas en cuanto al rendimiento financiero. Aunque la TIR señala la posibilidad de un retorno positivo, el VAN negativo recalca que los flujos de efectivo para compensar la inversión inicial a la tasa de descuento utilizada son insuficientes.

Continuando con el flujo proyectado para 2024 se obtuvieron los siguientes resultados

³⁴ Krarup, 2023

Flujo proyecto año 2024													
Mes													
	0	enero	febrero	marzo	abril	mayo	junio	julio	agosto	septiembre	octubre	noviembre	diciembre
Ingresos		\$16.599.910	\$14.415.395	\$21.122.814	\$19.335.423	\$24.999.323	\$21.116.219	\$15.834.968	\$20.854.758	\$13.161.024	\$10.018.093	\$14.798.834	\$17.583.251
Costos fijos		-\$10.000.000	-\$10.000.000	-\$10.000.000	-\$10.000.000	-\$10.000.000	-\$10.000.000	-\$10.000.000	-\$10.000.000	-\$10.000.000	-\$10.000.000	-\$10.000.000	-\$10.000.000
UAI		\$6.599.910	\$4.415.395	\$11.122.814	\$9.335.423	\$14.999.323	\$11.116.219	\$5.834.968	\$10.854.758	\$3.161.024	\$18.093	\$4.798.834	\$7.583.251
Impuesto a la renta		-\$1.781.976	-\$1.192.157	-\$3.003.160	-\$2.520.564	-\$4.049.817	-\$3.001.379	-\$1.575.441	-\$2.930.785	-\$853.476	-\$4.885	-\$1.295.685	-\$2.047.478
UDI		\$4.817.935	\$3.223.238	\$8.119.654	\$6.814.859	\$10.949.506	\$8.114.840	\$4.259.527	\$7.923.973	\$2.307.547	\$13.208	\$3.503.149	\$5.535.773
Flujo caja operacional		\$4.817.935	\$3.223.238	\$8.119.654	\$6.814.859	\$10.949.506	\$8.114.840	\$4.259.527	\$7.923.973	\$2.307.547	\$13.208	\$3.503.149	\$5.535.773
Inversión fija	-\$8.302.600												
Flujo de caja de capitales	-\$8.302.600												
Flujo de caja privado	-\$8.302.600	\$4.817.935	\$3.223.238	\$8.119.654	\$6.814.859	\$10.949.506	\$8.114.840	\$4.259.527	\$7.923.973	\$2.307.547	\$13.208	\$3.503.149	\$5.535.773
VAN	\$30.044.325												
TIR	66,06%												

Tabla 14: Flujo de caja año 2024³⁵

Con un VAN positivo y una TIR que supera la tasa libre de riesgo se concluye que existe una rentabilidad favorable y respaldan la viabilidad financiera de la expansión a gran escala.

No obstante, es crucial mantener un monitoreo continuo de los resultados, considerar posibles contingencias y ajustar estrategias según las condiciones cambiantes del mercado. La combinación de un VAN positivo y una TIR superior a la tasa libre de riesgo indica un sólido potencial de retorno sobre la inversión en la etapa de implementación completa del proyecto.

Para evaluar cómo distintas variaciones tanto en los ingresos como costos fijos del proyecto afectan los resultados financieros de este, se realizó un análisis de sensibilidad que presenta los resultados del VAN de acuerdo a distintas variaciones en los ingresos y costos fijos del proyecto. Los resultados se presentan en la siguiente tabla

³⁵ Krarup, 2023

	VAN	Variaciones ingresos									
	\$30.044 .325	-25.0 0%	-20.0%	-15.0%	-10.0%	-5.0%	0.0%	5.0%	10.0%	15.0%	20.0%
Variaciones CF	20.0%	\$ -1,50 3,043	\$ 2,854,37 4	\$ 7,211,7 91	\$ 11,569, 208	\$ 15,926, 625	\$ 20,284, 042	\$ 24,641, 459	\$ 28,998, 876	\$ 33,356, 293	\$ 37,713, 710
	25.0%	\$ -3,94 3,113	\$ 414,304	\$ 4,771,7 21	\$ 9,129,1 38	\$ 13,486, 555	\$ 17,843, 972	\$ 22,201, 389	\$ 26,558, 805	\$ 30,916, 222	\$ 35,273, 639
	30.0%	\$ -6,38 3,184	\$ -2,025,7 67	\$ 2,331,6 50	\$ 6,689,0 67	\$ 11,046, 484	\$ 15,403, 901	\$ 19,761, 318	\$ 24,118, 735	\$ 28,476, 152	\$ 32,833, 569
	35.0%	\$ -8,82 3,255	\$ -4,465,8 38	\$ -108,42 1	\$ 4,248,9 96	\$ 8,606,4 13	\$ 12,963, 830	\$ 17,321, 247	\$ 21,678, 664	\$ 26,036, 081	\$ 30,393, 498
	40.0%	\$ -11,2 63,32 5	\$ -6,905,9 08	\$ -2,548,4 91	\$ 1,808,9 26	\$ 6,166,3 43	\$ 10,523, 759	\$ 14,881, 176	\$ 19,238, 593	\$ 23,596, 010	\$ 27,953, 427
	45.0%	\$ -13,7 03,39 6	\$ -9,345,9 79	\$ -4,988,5 62	\$ -631,14 5	\$ 3,726,2 72	\$ 8,083,6 89	\$ 12,441, 106	\$ 16,798, 523	\$ 21,155, 940	\$ 25,513, 357
	50.0%	\$ -16,1 43,46 7	\$ -11,786, 050	\$ -7,428,6 33	\$ -3,071, 216	\$ 1,286,2 01	\$ 5,643,6 18	\$ 10,001, 035	\$ 14,358, 452	\$ 18,715, 869	\$ 23,073, 286
	55.0%	\$ -18,5 83,53 7	\$ -14,226, 120	\$ -9,868,7 04	\$ -5,511, 287	\$ -1,153, 870	\$ 3,203,5 47	\$ 7,560,9 64	\$ 11,918, 381	\$ 16,275, 798	\$ 20,633, 215
	60.0%	\$ -21,0 23,60 8	\$ -16,666, 191	\$ -12,308, 774	\$ -7,951, 357	\$ -3,593, 940	\$ 763,477	\$ 5,120,8 94	\$ 9,478,3 11	\$ 13,835, 728	\$ 18,193, 144

Tabla 14: Resultados VAN del proyecto frente a variación de ingresos y costos fijos³⁶

³⁶ Krarup, 2023

El análisis de sensibilidad realizado ha demostrado que el VAN del proyecto es sensible tanto a las variaciones en los ingresos como en los costos fijos. Se observa que un incremento en los ingresos resulta en un aumento significativo del VAN, destacando la importancia de estrategias centradas en el crecimiento de ingresos para maximizar el valor del proyecto. Por otro lado, un aumento en los costos fijos tiene un impacto negativo en el VAN, aunque en una magnitud menor en comparación con los cambios en los ingresos.

Estos resultados indican que es crucial gestionar y controlar los costos, además de que las iniciativas que impulsen los ingresos podrían tener un efecto más pronunciado en el éxito financiero del proyecto. También, el análisis muestra que el proyecto mantiene un VAN positivo incluso bajo escenarios de disminución de ingresos hasta un cierto punto, lo que indica una cierta resiliencia frente a fluctuaciones negativas de estos.

Resultados

La solución fue implementada durante noviembre y diciembre en dos marcas, Pizzería Capri y Mozzarella. En primera instancia se analizaran los resultados obtenidos para las métricas del objetivo general correspondientes a una mejora en la tasa de conversión y mejora del ROAS de las campañas. El objetivo propuesto de mejorar la tasa de conversión de las plataformas en 3 puntos porcentuales no pudo cumplirse por una serie de factores externos al proyecto que influyen de gran manera en la tasa de conversión. Algunos de estos factores son la cantidad de clicks que debe hacer el cliente hasta comprar sus productos, la hora de entrega prometida al cliente, costo del despacho, etc. Los factores mencionados tienen una enorme importancia en la tasa de conversión de una marca ya que si por ejemplo se realiza una orden por CLP \$4950 y el costo de despacho es de CLP \$2980 es muy probable que un cliente no compre dado el alto valor del despacho en relación al valor de los productos. Para corroborar esto se puede observar en el anexo 9 cómo estos factores afectan la tasa de conversión final. Dicho esto, subir 3 puntos porcentuales por marca no se pudo lograr a pesar de cumplir con el resto de objetivos del proyecto.

Continuando con el objetivo de aumentar el ROAS de las campañas en por lo menos un 20% para cada marca, este fue cumplido de manera satisfactoria. A continuación se puede observar un gráfico para cada marca con la evolución del ROAS durante el año.

Evolucion ROAS Pizzería Capri

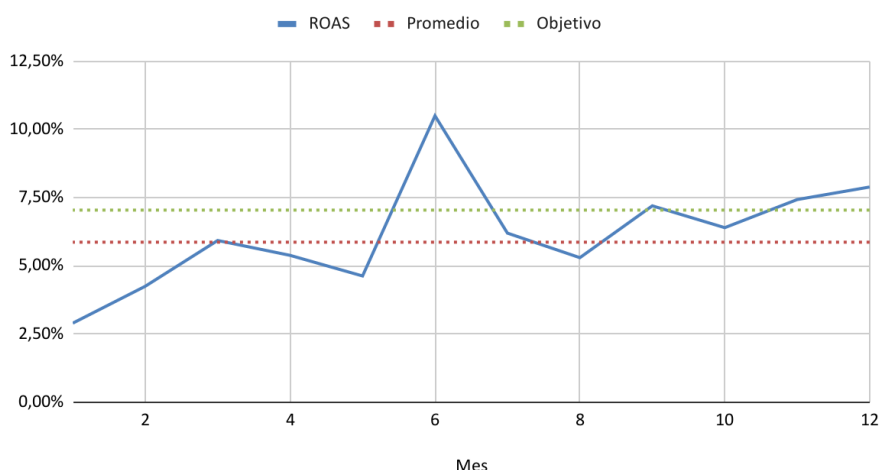


Gráfico 4: Evolución ROAS Pizzería Capri³⁷

³⁷ Krarup, 2023

Para esta marca se tenía un ROAS promedio de 5,87% por lo que se buscó aumentar a un 7,04% como mínimo lo que corresponde a una mejora del 20%. Durante el mes de noviembre el ROAS fue de 7,43% lo que significa un aumento del 26,5% y para diciembre el ROAS fue de 7,89% aumentando un 34,4% respecto al promedio y también hubo un aumento respecto al mes anterior.

Evolución ROAS Mozzarella

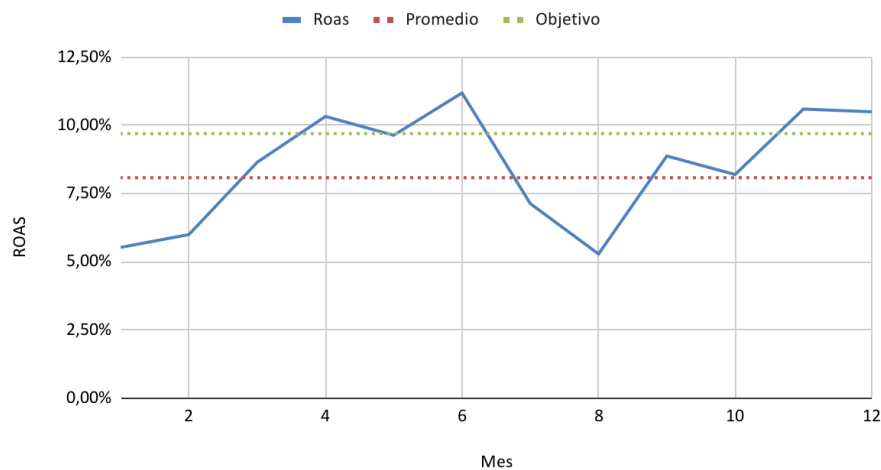


Gráfico 5: Evolución ROAS Mozzarella³⁸

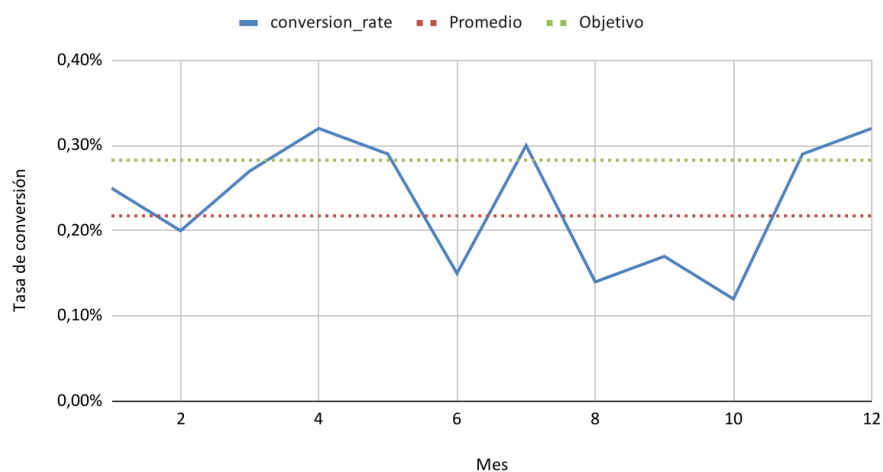
El promedio del ROAS para esta marca es de un 8,08% por lo que se buscó aumentarlo a un 9,7%. En este caso se logró una mejora del 31,1%, llegando a un ROAS de 10,6% en noviembre. Para diciembre el ROAS obtenido fue de un 10,5%.

Como queda evidenciado en los gráficos 4 y 5 es posible afirmar que se generó un aumento del ROAS en por lo menos un 20%, cumpliendo así el objetivo planteado inicialmente.

Sumado a lo anterior se evaluaron los resultados de las métricas asociadas a los objetivos específicos obteniendo los siguientes resultados

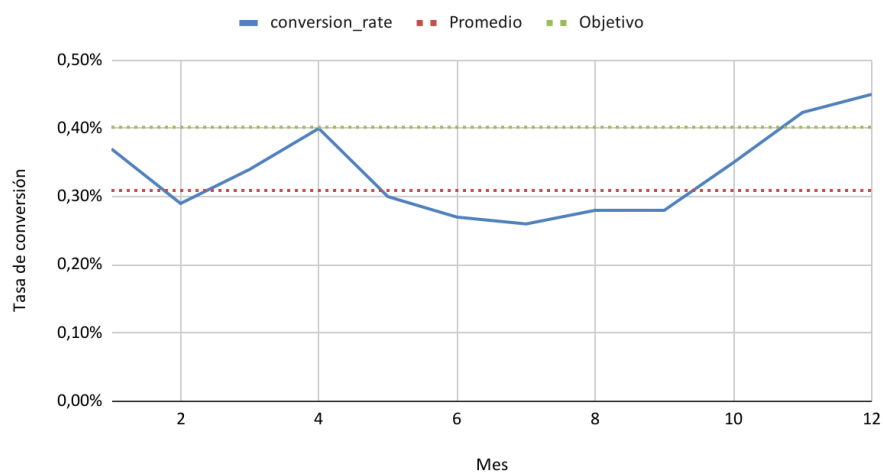
³⁸ Krarup, 2023

Tasa de conversión mailing Pizzería Capri


Gráfico 6: Tasa de conversión mailing Pizzeria Capri³⁹

Para la marca descrita en el gráfico se tenía un promedio anual de 0,22% en la tasa de conversión del mailing, este valor subió un 33,3% llegando a 0,29% en la tasa de conversión en noviembre, y en diciembre este valor alcanzó un aumento mayor obteniendo una tasa de conversión del 0,32% alcanzando así el objetivo propuesto.

Tasa de conversión mailing Mozzarella


Gráfico 7: Tasa de conversión mailing Mozzarella⁴⁰
³⁹ Krarup, 2023

⁴⁰ Krarup, 2023

Esta marca tiene un promedio anual del 0,31%. Con la mejora implementada, la tasa de conversión durante noviembre aumentó un 37% llegando a una tasa de conversión del 0,42%. Para diciembre la tasa de conversión alcanzó un valor de 0,45%.

Como se puede observar en ambos gráficos el objetivo de aumentar un 30% la tasa de conversión del mailing fue lograda en ambas campañas en las que se implementó la solución, pudiendo así notar que la solución cumple con las expectativas y estándares necesarios para cumplir el objetivo.

También, se evaluó el desempeño de los modelos para los cuales se obtuvieron los resultados resumidos en la siguiente tabla

Marca	F1 Score
Pizzería Capri	0,94
Mozzarella	0,96

Tabla 15: Resultados modelos ML para marcas en las que se ha implementado la solución⁴¹

Tal como se observa en la tabla anterior, se obtuvo un F1 Score igual o mayor a 0,93 en ambas marcas por lo que los modelos cumplen adecuadamente su función predictora la cual fue utilizada para hacer una mejora eficiente en las campañas y dirigirlas a aquellos que presentan más probabilidades de compra.

⁴¹ Krarup, 2023

Conclusiones y recomendaciones

Las startups forman parte de un mundo muy dinámico que está sometido a cambios constantes y desafíos imprevistos. El desarrollo del proyecto, desde su planteamiento hasta su implementación, no estuvo exento de complicaciones. A medida que se enfrentaron obstáculos, se evidenció la necesidad de adoptar un enfoque flexible y adaptativo. En particular, la empresa experimentó una significativa ola de despidos, lo que impactó directamente en las prioridades y disposiciones de la empresa, llevando a importantes modificaciones del proyecto en su alcance original.

Inicialmente la implementación sería en todas las marcas de Justo en Chile que tienen un KAM a su disposición pero las circunstancias imprevistas obligaron a una redefinición del proyecto, estableciendo un plan piloto para las cinco marcas mencionadas en este documento. La reducción notable en el equipo de trabajo participante llevó a cambios sustanciales en tareas y prioridades. Es por esto que la decisión de implementar la solución para el resto de las marcas quedó condicionada al rendimiento positivo de la solución durante el plan piloto y a la evaluación de la situación y contexto de la empresa en el futuro.

Si bien existieron imprevistos, el proyecto logró cumplir casi en su totalidad con los objetivos establecidos inicialmente, a pesar de la excepción de la mejora en la tasa de conversión en la plataforma. La falta de consideración de factores externos que influyen en esta métrica sirve como una lección clave para establecer correctamente los objetivos de un proyecto como este. Sin embargo, el proyecto demostró un impacto positivo en el resto de objetivos, proporcionando a la empresa una mejora en su enfoque de campañas, lo que, en última instancia, se tradujo en una mejora de ingresos.

La flexibilidad emergió como una clave del éxito en este entorno dinámico. La capacidad de adaptarse a circunstancias cambiantes permitió ajustes estratégicos y la mitigación efectiva de desafíos. Además, la comunicación constante y la colaboración fueron fundamentales para mantener la cohesión del equipo y alinear las expectativas con las nuevas realidades, más aún cuando ocurrieron los despidos.

En el transcurso de este proyecto, se destacan valiosos aprendizajes personales que han forjado un crecimiento significativo tanto en el ámbito personal como profesional. Al sumergirse en un entorno dinámico y en constante cambio, se ha experimentado la necesidad de tener una rápida capacidad de adaptación. La habilidad para ajustarse y responder eficazmente a las complejidades y desafíos inesperados ha emergido como una habilidad esencial.

Uno de los aprendizajes fundamentales ha sido la comprensión de que los proyectos, a menudo, divergen significativamente de las expectativas iniciales. La realidad ha demostrado que, incluso con una cuidadosa planificación y ejecución, los imprevistos son inevitables. Este reconocimiento ha llevado a una apreciación más profunda de la importancia de la flexibilidad y la adaptabilidad en la gestión de proyectos. Se ha interiorizado la noción de que la capacidad para ajustar estrategias y metas a medida que evolucionan las circunstancias es esencial para el éxito a largo plazo.

El proyecto ha sido importante para entender que la planificación inicial, aunque crucial, es solo el comienzo de un viaje empresarial. Aceptar la naturaleza cambiante de los proyectos ha sido clave para superar obstáculos y maximizar oportunidades emergentes.

Teniendo en cuenta los resultados expuestos anteriormente, se recomienda continuar con la implementación del proyecto realizando una evaluación continua de la solución a medida que se expande a otras marcas. Esto implica la recopilación constante de datos y realizar los ajustes necesarios para garantizar la eficacia continua y la adaptación a posibles cambios futuros.

Además, se sugiere una evaluación más detallada de los factores externos que puedan influir en los resultados y la inclusión de estrategias de mitigación. La sostenibilidad a largo plazo de la solución debe ser considerada, explorando posibles mejoras futuras y manteniendo una mentalidad de innovación constante.

Finalmente, se insta a la empresa a mantener un enfoque centrado en el usuario y en la marca. La retroalimentación de estos durante el plan piloto debe ser utilizada para informar mejoras continuas, asegurando que la solución siga siendo relevante y valiosa.

Estas recomendaciones buscan no solo consolidar los logros del proyecto, sino también preparar el terreno para futuras expansiones y mejoras, garantizando la continuidad del éxito en un entorno empresarial en constante evolución.

Referencias

Banco Central. (2023, octubre 31). *Tasas de Interés*. Banco Central. Retrieved November 30, 2023, from <https://www.bcentral.cl/web/banco-central/areas/estadisticas/tasas-de-interes>

Brands by Rappi, ¡la herramienta ideal de marketing B2B! (2021, May 13). Crehana. Retrieved September 12, 2023, from <https://www.crehana.com/blog/estilo-vida/conoce-brands-by-rappi-la-herramienta-ideal-de-marketing-b2b/>

CRISP-DM: La Brújula del mundo empresarial data driven. (n.d.). Medium. Retrieved October 31, 2023, from <https://medium.com/@datanauta/crisp-dm-la-br%C3%BAjula-del-mundo-empresarial-data-driven-1139b7623296>

Damodaran, A. (2023, January 5). *Total Beta by Industry Sector*. Damodaran. Retrieved November 30, 2023, from https://pages.stern.nyu.edu/~adamodar/New_Home_Page/datacurrent.html

Justo, G. (2022, July 2). *Dashboard admin demo*. Admin GetJusto. Retrieved October 30, 2023, from <https://admin.getjusto.com/pgFqoPatXfQbtt9Ze/admin/dashboard>

Krarup, J. (2023). *Desarrollo de proyecto*.

Netflix Research. (2018, agosto 8). *Machine Learning Learning - how to entertain the world*. Netflix Research. Retrieved October 26, 2023, from <https://research.netflix.com/research-area/machine-learning>

Servicio de delivery: las empresas se adaptaron en pandemia. (n.d.). e-certchile. Retrieved September 11, 2023, from <https://www.e-certchile.cl/noticias/servicio-de-delivery-las-empresas-se-adaptaron-en-pandemia>

UberEATS uses location data for efficient marketing. (2018, June 8). WARC. Retrieved September 11,

2023, from

<https://www.warc.com/newsandopinion/news/ubereats-uses-location-data-for-efficient-marketing/en-gb/40595>

Anexos

1. Detalle de los servicios ofrecidos por Justo



2. Cantidad de órdenes de alianzas por mes en el año 2022.

Mes	Órdenes alianzas 2022
1	8.857
2	9.886
3	15.213
4	16.379
5	26.160
6	28.082
7	31.341
8	21.077
9	16.770
10	15.093
Total	188.858

3. Cantidad de órdenes de alianzas por mes en el año 2023.

Mes	Órdenes alianzas 2023
1	9.911
2	9.593
3	13.802
4	11.767
5	16.085
6	14.303
7	10.223
8	14.599
9	9.120
10	6.967
Total	116.370

4. [Códigos desarrollo de solución](#)

5. [Métricas resultados de modelos](#)

Para entender mejor los resultados expuestos en tabla 7 es necesario comprender qué representan las métricas mostradas (Precision, Recall, F1-score y Curva ROC) las cuales son comúnmente utilizadas para evaluar el rendimiento de modelos de clasificación.

1. *Precision (Precisión):*

- La precisión se define como la proporción de instancias positivas correctamente clasificadas con respecto a todas las instancias clasificadas como positivas (tanto true positive como false positive)
- La precisión mide la calidad de las predicciones positivas del modelo.

2. *Recall (Sensibilidad):*

- El recall se define como la proporción de instancias positivas correctamente clasificadas con respecto a todas las instancias que son realmente positivas (tanto true positive como false negative).
- El recall mide la capacidad del modelo para "recuperar" todas las instancias positivas.

3. *F1 Score:*

- El F1-score es la media armónica de la precisión y el recall. Proporciona un equilibrio entre estas dos métricas.

4. *Curva ROC:*

- Es una representación gráfica de cómo varía el rendimiento de un modelo de clasificación binaria al ajustar el umbral de decisión. Muestra la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos, ayudando a evaluar la capacidad del modelo para distinguir entre clases. Un área bajo la curva ROC más cercana a 1 indica un mejor rendimiento, 0.5 sugiere una clasificación aleatoria y un valor cercano a 0 indica un peor rendimiento.

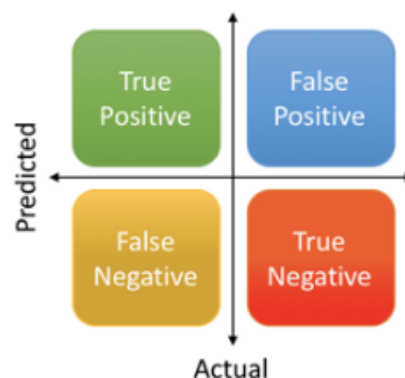
5. *Accuracy (Exactitud):*

- La exactitud se refiere a la proporción total de instancias clasificadas correctamente, independientemente de si son positivas o negativas. Mide cuántas de todas las predicciones son correctas.

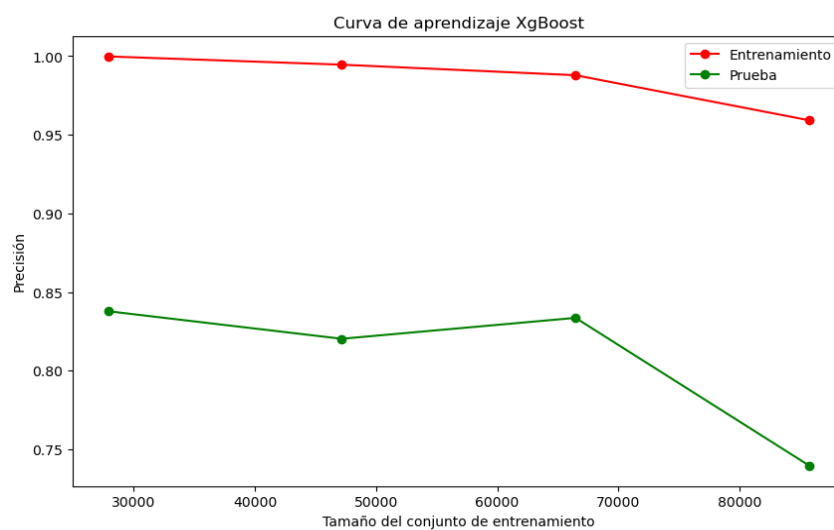
Además, se debe comprender que es un true positive, false positive, true negative, y false negative

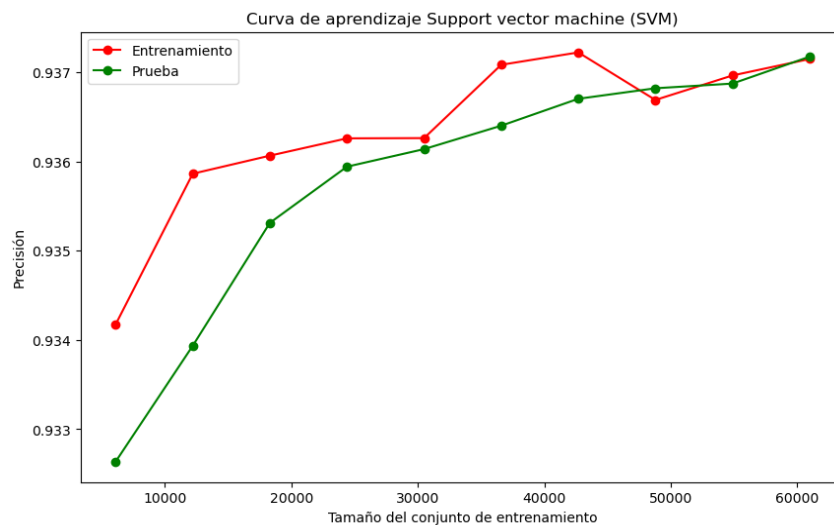
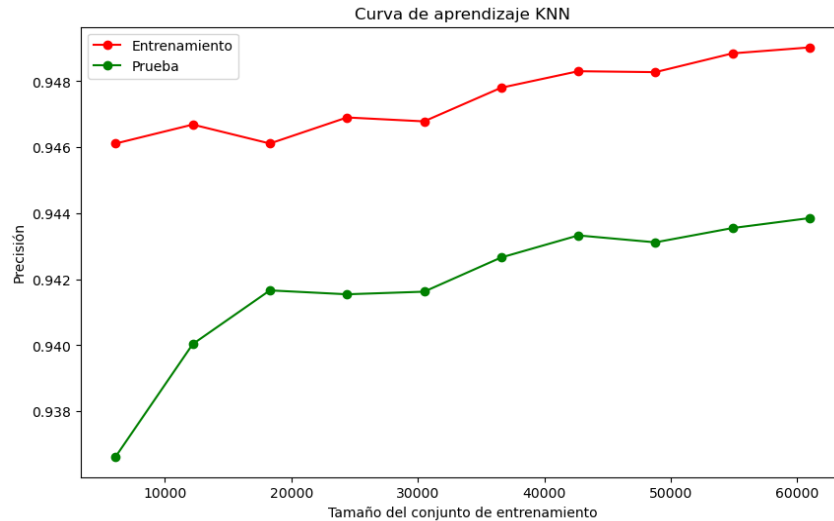
- True Positive (TP): Indica la cantidad de instancias positivas que el modelo clasifica correctamente como positivas.
- False Positive (FP): Representa la cantidad de instancias negativas que el modelo clasifica incorrectamente como positivas.
- True Negative (TN): Se refiere a la cantidad de instancias negativas que el modelo clasifica correctamente como negativas.
- False Negative (FN): Indica la cantidad de instancias positivas que el modelo clasifica incorrectamente como negativas.

Para entender mejor esto se muestra la siguiente representación gráfica.



6. Curvas de aprendizaje XGBoost, KNN y SVM





7. Marco metodológico y representación gráfica metodología CRISP-DM

Se eligió esta metodología porque es un marco de trabajo flexible que se adapta a proyectos de minería de datos y machine learning que ha sido utilizado en diversas industrias, además, facilita la organización y gestión de proyectos de análisis de datos, lo que ayuda a garantizar que se cumplan los objetivos y se obtengan resultados valiosos que van enfocados al entorno de negocio.

Algunas empresas que han usado esta metodología son Walmart y Netflix. “El gigante minorista Walmart utilizó este proceso para analizar y predecir los patrones de compra de sus clientes. Con esta información, Walmart pudo ajustar su gestión de inventario y maximizar las ventas. Otro caso

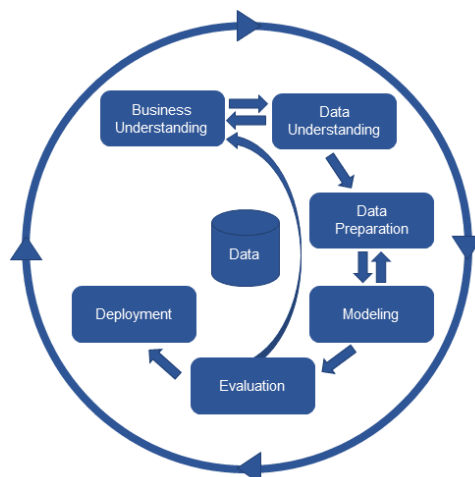
interesante es el de Netflix, que aplicó CRISP-DM para mejorar su algoritmo de recomendación. Mediante el análisis de los hábitos de visualización de sus usuarios, Netflix ha podido ofrecer

sugerencias personalizadas, lo que a su vez ha mejorado la experiencia de usuario y ha aumentado la retención de clientes.”⁴²

CRISP-DM cuenta con seis etapas clave que son descritas a continuación.

1. **Comprensión del Negocio (Business Understanding):** En esta etapa, se busca comprender a fondo los objetivos y necesidades del negocio. Se definen los problemas o las oportunidades a abordar con la minería de datos y se establecen los criterios de éxito. Esta fase es esencial para asegurar que el proyecto esté alineado con los objetivos del proyecto y empresa.
2. **Comprensión de los Datos (Data Understanding):** Aquí, se reúnen y exploran los datos disponibles. Se recopilan conjuntos de datos, se describen sus características y se identifican posibles problemas de calidad de datos. Esta fase proporciona información crítica sobre la calidad y la idoneidad de los datos para el análisis.
3. **Preparación de los Datos (Data Preparation):** En esta etapa, se realizan tareas de limpieza, transformación y manipulación de datos para asegurarse de que estén listos para el análisis.
4. **Modelado (Modeling):** En esta fase, se seleccionan algoritmos de machine learning apropiados, se entrenan modelos y se ajustan los hiper parámetros. El objetivo es construir modelos predictivos que sean capaces de abordar los objetivos del proyecto.
5. **Evaluación (Evaluation):** En esta etapa, se evalúa el rendimiento de los modelos contruidos en la fase de modelado. Se utilizan métricas de desempeño para medir la eficacia de los modelos y determinar si cumplen con los criterios de éxito definidos en la fase de comprensión del negocio.
6. **Despliegue (Deployment):** Una vez que los modelos han sido evaluados y se consideran satisfactorios, se implementan en el entorno de producción. Los modelos se utilizan para tomar decisiones o realizar predicciones en aplicaciones reales.

⁴² CRISP-DM: La Brújula Del Mundo Empresarial Data Driven, n.d.



8. Estimaciones flujo de caja

Todos los valores son en pesos chilenos.

Para proyecto piloto:

1. **Inversión fija:** Corresponde al valor monetario de las horas hombre dedicadas al desarrollo del proyecto.
2. **Ingresos:** Se realizó una estimación en base al porcentaje de mejora que brinda la solución a los procesos actuales. En promedio se mejora un 2,3% el revenue mensual de las órdenes generadas a partir de alguna campaña

Para proyecto completo:

1. **Inversión fija:**
 - a. **Capacitación y desarrollo del personal:** Se estimó un valor de \$412.600 correspondiente a las horas hombre dedicadas a capacitar al personal al uso de la solución.
 - b. **Adaptación infraestructura tecnológica:** Costo único que tiene la empresa para implementar la solución para cada marca dentro de la plataforma ya existente. Se estimó en un valor de \$7.890.000 en base a las horas hombre dedicadas a esto.
2. **Ingresos:** Para cada mes se estimaron los resultados esperados considerando los datos históricos para ese mes de los años anteriores con la solución implementada.
3. **Costos fijos:** Costos asociados al personal encargado del mantenimiento y actualización de los modelos.

9. Funnel de compra y ejemplo costos productos vs delivery Buffalo Waffles

		% de product added respecto a order page viewed	% de cart page viewed respecto a product added	% de continue to checkout respecto a cart page viewed	% de checkout page viewed respecto a continue to checkout	% de payment submitted respecto a checkout page viewed	% de payment completed respecto a payment submitted	
Pais	Marca							Conversion
CL	Buffalo Waffles	24,52%	81,40%	84,05%	99,70%	44,99%	61,53%	4,63%

La imagen anterior hace referencia al porcentaje de personas que avanza a la siguiente etapa del funnel respecto a la etapa anterior. Esto es, de todas las personas que visitan la página web de la marca, un 24,52% agrega un producto al carro. De ese 24,52% que agregó productos al carro, un 81,4% vio el carrito y así sucesivamente con el resto de etapas. Se puede observar que solo un 44,99% de las personas que llegaron hasta la etapa del pago, lo hicieron.

¿Tienes un cupón de descuento?

Tienes 496 Waffle Points
Canjear

Propina para el repartidor

5% \$248
10% \$495
15% \$743
Otro monto

Total Productos \$4.950
Despacho \$2.980
Propina para el repartidor \$0
Total a pagar \$7.930

Con esta compra acumularás 248 Waffle Points

Medio de pago

VISA VISA **** 6469
Otros

Pagar

Powered by JUSTO

Para esa misma marca, al pedir uno de sus productos se puede observar que el costo del despacho es más de la mitad del valor del producto por lo que se justifica la caída de clientes en esa etapa la cual como se ha mencionado anteriormente, influye enormemente en la tasa de conversión de la marca.

10. Funnel de compra clientes

