

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ingeniería en Electricidad y Computación

Precios dinámicos personalizados usando técnicas de aprendizaje
automático

PROYECTO DE TITULACIÓN

Previo la obtención del Título de:

Magister en Ciencias de Datos

Presentado por:

Ángel Luis Catagua González

José Alexi Mendoza Álava

GUAYAQUIL - ECUADOR

Año: 2021

AGRADECIMIENTO

En primer lugar, mi agradecimiento es a Dios. Luego, a la empresa actual donde laboro, Vitapro, especialmente a Kristel Méndez quien me pudo facilitar el problema de negocio abordado en esta tesis. Por último, pero no menos importante, a los directivos de la Sociedad Ecuatoriana de Estadística quienes siempre estuvieron prestos a cualquier consulta, así como también a mis amigos de Masapp, Christian, Néstor, Roberto y Fernanda.

Ángel Catagua

Le agradezco a Dios por la vida y los logros que siempre me ha permitido alcanzar. También, quiero agradecer a Ángel y a la empresa donde él labora, por haberme concedido ser parte del equipo de trabajo de este proyecto. Finalmente, quiero agradecer a Don Jerónimo y a Don Jaime por el apoyo constante que me han brindado en mis estudios.

Alexi Mendoza A.

DEDICATORIA

Este proyecto va dedicado a mi familia, en especial a mis hijos Amaia y Emiliano (Pichula, Pegajoso, Pepe y Rana), mis padres Ángel y Magdalena, mi esposa Vanessa y mis hermanos Karla y Luis con sus respectivas familias, además de una mención especial a Máxima. Además, nunca olvidaré dos grandes padres de profesión, Gaudencio Zurita y Francisco Cruz.

Ángel Catagua

Dedico este proyecto a mi familia, sobre todo a mi madre Galy, a mi querida esposa Joselyn, a mis bebés preciosos Emmita y Elliotcito, y a cada uno de mis hermanos: Yamile, Julio, Rosa, Mirian, Sabrina y Juana, y a sus correspondientes familias.

Alexi Mendoza A.

TRIBUNAL DE SUSTENTACIÓN



Dr. Christian Galarza Morales

PROFESOR TUTOR

Dr. José Córdova García

PROFESOR EVALUADOR

DECLARACIÓN EXPRESA

"Los derechos de titularidad y explotación, nos corresponde conforme al reglamento de propiedad intelectual de la institución; Ángel Luis Catagua González y José Alexi Mendoza Álava damos nuestro consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"



Ing. Ángel Luis
Catagua González



Ing. José Alexi
Mendoza Álava

RESUMEN

Los precios dinámicos consisten en la determinación del precio óptimo en un ambiente que pueden cambiar fácilmente, y donde los mismos pueden ajustarse a características propias del cliente y mercado. En este trabajo proponemos una metodología de personalización de precios mediante la aplicación de técnicas de aprendizaje automático, con la finalidad de ofrecer el precio correcto, al cliente correcto, en el momento correcto. Diferentes técnicas de aprendizaje automático son usadas para la clasificación de clientes por clúster, el pronóstico de las ventas y, por último, el cálculo de rangos de porcentajes de descuentos personalizados. Los mejores modelos son elegidos entre los candidatos a través de diferentes métricas de interés. El rendimiento general de la propuesta luce adecuado, donde se encontró que las características de mayor relevancia para explicar los descuentos son el ingreso por compra, la familia del producto, el clúster al cual pertenece el cliente y la temporalidad por medio del año y mes. Finalmente, la propuesta ha sido implementada a través de un aplicativo web, el cual se facilitará al usuario.

Palabras Clave: Precios Dinámicos, Pronósticos, Aprendizaje automático, Rangos de precios.

ABSTRACT

Dynamic pricing consists of determining the optimal price in an environment that can change quickly and it can be adjusted to the characteristics of the client and the market. In this work, we propose a price personalization methodology by applying machine learning techniques to offer the right price to the right customer at the right time. Different machine learning techniques are used to classify customers by cluster, forecast sales, and, finally, calculate ranges of percentages of personalized discounts. The best models are chosen among the candidates through different interest metrics. The general performance of the proposal looks adequate, where it was found that the most relevant characteristics to explain the discounts are the income per purchase, the family of the product, the cluster to which the client belongs, and the temporality through the year and month. Finally, the proposal has been implemented through a web application provided to the user.

Keywords: Dynamic pricing, Forecasting, Machine Learning, Price ranges.

ÍNDICE GENERAL

CAPÍTULO 1.....	8
1.PLANTEAMIENTO DE LA PROBLEMÁTICA	8
1.1 Descripción del problema	8
1.2 Justificación	10
1.3 Objetivos.....	11
1.3.1 Objetivo General	11
1.3.2 Objetivos Específicos	11
1.4 Metodología	11
1.5 Resultados Esperados.....	13
1.6 Dataset	13
CAPÍTULO 2.....	15
2.ESTADO DEL ARTE	15
2.1 Fundamentos del problema	15
2.2 Soluciones de analítica y aprendizaje relacionadas al problema.....	16
2.3 Fuentes de datos relacionadas al problema	17
2.4 Librerías y software a utilizar	18
2.4.1 Software	19
2.4.2 Lenguajes de programación.....	19
2.4.3 Librerías	20
CAPÍTULO 3.....	23
3.DISEÑO E IMPLEMENTACIÓN	23
3.1 Exploración y Validación de datos	23

3.2	Prototipos de algoritmos y modelos	25
3.2.1	Clúster de clientes.	25
3.2.2	Modelos para descuentos.....	26
3.2.3	Modelos para pronósticos de series temporales.....	26
3.3	Plataforma y prototipos de visualización.....	27
3.3.1	Módulos	27
3.3.2	Roles.....	27
3.3.3	Usuarios.....	28
3.4	Métricas y comunicación de resultados	28
CAPÍTULO 4.....		30
4.ANÁLISIS DE RESULTADOS.....		30
.....		30
4.1	Análisis estadístico descriptivo del problema.....	30
4.2	Resultados de Segmentación de clientes	31
4.3	Resultados de Rangos de precios dinámicos	34
4.4	Resultados de predicciones y evaluación de rangos de precios.....	38
4.5	Rangos de precios y evaluación de medidas.....	42
4.6	Análisis de beneficios del cálculo de precios dinámicos	44
CAPÍTULO 5.....		46
CONCLUSIONES Y RECOMENDACIONES.....		46
Conclusiones		46
Recomendaciones		46
BIBLIOGRAFÍA.....		47

ÍNDICE DE FIGURAS

Figura 1.1 Propuesta de técnicas para construcción de precios dinámicos.....	11
Figura 3.1 Propuesta de técnicas para construcción de precios dinámicos.....	23
Figura 3.2 Estructura de la solución propuesta.....	28
Figura 4.1 Consumo Totales de toneladas por familia durante el periodo enero de 2020 a junio de 2021	31
Figura 4.2 Clúster por número de SKU y Toneladas promedio.....	33
Figura 4.3 Gráfico de Cajas por clúster y familia	35
Figura 4.4 Series temporales de descuentos por clúster desde 2018	36
Tabla 4.3: Métricas obtenidas para los diferentes métodos ajustados para modelar los valores de descuentos por clústers.....	37
Figura 4.5 Series temporales de los clústers de clientes para los años 2018 a 2021	39
Figura 4.6 Participación de Familia por clúster	41
Figura 4.7 Rangos de precios para los clústers 0 y 3	43
Figura 4.8 Comparación de descuentos YTgo (Jul – Dic) 2021 vs. YTD (Ene - Jun) año 2021	44
Figura 4.9 Comparación de descuentos otorgados en los años 2020 y 2021 para el período de julio a diciembre	45

ÍNDICE DE TABLAS

Tabla 2.1 Clúster de Clientes desarrollado en Python	20
Tabla 2.2 Detalle de librerías de R utilizadas para rangos de precios	21
Tabla 2.3 Forecasting de Clientes desarrollada en R	21
Tabla 2.4 Visualización e Interfaz Gráfica desarrollada en JavaScript	22
Tabla 4.1: Promedio de Sku, Familias y Toneladas Consumidas por nodos de clientes	32
Tabla 4.2: Métricas obtenidas para los diferentes métodos ajustados para modelar los valores de descuentos por clústers.....	37
Tabla 4.3: Evaluación de Modelos para Forecasting de los clientes	40
Tabla 4.4: Tabla de ejemplo de modelos de Forecasting para familia	42

INTRODUCCIÓN

En la actualidad las empresas se encuentran con la compleja tarea de determinar el mejor precio para un bien o servicio que esté dispuesto a pagar el cliente. Esta tarea pasa del conocimiento de sus costos a la valorización de propuesta que se presenta al cliente. Para garantizar este precio óptimo, las empresas se han visto en la obligación de utilizar diferentes puntos de vistas y probar diferentes alternativas, incluso algunas poco tradicionales, pasando del regateo a las negociaciones por rendimiento del producto, o por volúmenes de compras. Las primeras empresas o industrias en trabajar con precios diferenciados por un mismo bien fueron las aerolíneas, sin embargo, otras industrias como el *retail* la llevaron a aplicaciones que incluían el estudio de la competencia, siendo muchas veces mal visto y acusado de precios discriminatorios. El propósito de este trabajo es abordar la problemática de precios dinámicos utilizando diferentes técnicas de aprendizaje automático. El objetivo va más allá de encontrar como solución precios fijos diferenciados, sino estimar un rango de opciones a evaluar en función al pronóstico de venta y familia de productos a ser adquiridos. Este documento consta de cuatro capítulos, más la sección final de conclusiones y recomendaciones. El primero se enfoca en el planteamiento de la problemática describiendo el cómo, y dejando claro la necesidad del estudio de precios dinámicos; el segundo capítulo se enfoca en dar una vista del tratamiento del mismo problema desde el abordaje de otros autores. Por otro lado, el tercer capítulo tiene como objetivo el diseño e implementación de la propuesta a la problemática escrita en el capítulo uno. Finalmente, el último capítulo describe los resultados obtenidos y las medidas de evaluación.

CAPÍTULO 1

1. PLANTEAMIENTO DE LA PROBLEMÁTICA

1.1 Descripción del problema

La Industria 4.0, data del año 2010 ante una iniciativa alemana, la cual consiste en la digitalización de los procesos industriales por medio de la interacción del aprendizaje automático e inteligencia artificial con los dispositivos IoT (*internet of things*) y la optimización de recursos enfocada en nuevas estrategias de hacer negocios y como predecir/decidir sobre ellos. La Industria 4.0 tiene como objetivo garantizar de forma ingeniosa la satisfacción del cliente y la personalización de sus servicios; asimismo, la predicción de nuevos servicios, o en su defecto, alinear al cliente a un nuevo enfoque de sus necesidades [1].

Empresas *startup* como las de comercio electrónico y retail han adoptado tecnología que les ha permitido resolver la compleja tarea de determinar de manera correcta los precios para cobrar a un cliente un producto o un servicio, esta tarea implica más que conocimiento de la estructura de costos de la empresa y la disponibilidad de suministros que esta maneja, si no también otro aspecto como la valorización del productor, entre otros.

Dentro de los desafíos claves para garantizar la satisfacción del cliente y optimizar la ganancia de la compañía, se encuentra la decisión que toma la empresa en cuanto a los precios de sus productos, la optimización de estos y la administración del ingreso en un entorno totalmente cambiante compitiendo con empresas que tienen apetitos de riesgos altos.

La fijación de precios dinámicos es una estrategia flexible en la que el precio del producto está determinado por las condiciones actuales del mercado [4]. Se busca establecer una política de precios óptima considerando los costos, que los clientes están dispuestos a comprar, así como también las tendencias del mercado.

Los precios dinámicos son conceptualmente, el hecho de ofrecer bienes a diferentes precios que varían según la demanda del cliente [5]; este arte también es conocido como discriminación de precios a nivel individual [5] o gestión de ingresos [6]. La

personalización de precios por medio de segmentación de clientes y evaluación de futuras compras es un fenómeno muy extendido en la industria de *e-commerce*, *retail*, incluso se ha estudiado su aplicación en proyectos de boletos de aviones, consumo de energía eléctrica, entre otros [7], [8], [9], [10], [11].

Las nuevas tecnologías permiten de manera eficiente el estudio de patrones de consumo en los clientes, las herramientas de apoyo a la toma de decisiones emergen conjuntamente, y la experiencia en el sector de retail está otorgando una mejora de competitividad, esta combinación está otorgando a las empresas mayores márgenes de utilidad. A su vez esto ayuda a los clientes a elegir la oferta justa y a la industria trabajar en la eficiencia de sus recursos, respectivamente [12].

La empresa VitaPRO se encuentra dedicada al desarrollo de la industria de la acuicultura, cuenta con siete lustros de estudios y acciones para consolidar su objetivo de “nutrir el mañana” a través de estrategias y soluciones alimenticias sustentables que generen valor en todos los aspectos a la sociedad [13].

VitaPRO es parte de ALICORP; multinacional con presencia en más de 30 países con liderazgo en 150 marcas propias, aplicadas en 4 líneas de negocios, a saber [14]: acuicultura, consumo masivo, estrategia negocio por negocio (conocida en términos de mercado como “B2B”), y molienda. Ubicada en nuestro país en la provincia del Guayas, cantón Yaguachi, a 33 kilómetros de la cabecera provincial, Guayaquil; además, esta se encuentra presente en varios países de América Latina. En el sur en países como Perú y Chile, América Central en Panamá, Costa Rica, Honduras, Nicaragua y Guatemala, mientras que en América del Norte tiene presencia en México.

El horizonte en el corto plazo que se ha trazado VitaPRO es asentarse cómo un consorcio bajo los estándares de Industria 4.0 con lo cual lograría consolidarse, aún más, como una compañía sólida y resiliente [15]. Dentro de las oportunidades de mejora encontradas en la empresa se encuentra el cambio de su modelo de *pricing* buscando ofrecer al cliente correcto, el precio correcto en el tiempo correcto.

Dentro de los desafíos claves para garantizar la satisfacción del cliente y optimizar la ganancia de la compañía, se encuentra la decisión que toma la empresa en cuanto a los precios de sus productos y servicios, la optimización de estos y la administración del ingreso en un entorno totalmente cambiante.

1.2 Justificación

Empresas con enfoque basado en datos y experiencia en precios dinámicos como industrias de aerolíneas y hoteleras han tenido mucho éxito en la aplicación de este modelo de negocios, al examinar la gestión de ingreso incluso han encontrados patrones de comportamientos en clientes que les ha permitido realizar ajustes y maximizar utilidades. Por otro lado, las empresas privadas de esta metodología encuentran poco conocimiento en los perfiles de consumo de sus clientes y al ajustar sus precios de manera homogéneas encuentran altas tasas de deserción hacia los competidores.

Dentro de los beneficios de tener precios dinámicos se resalta la gestión flexible para fijar un precio de producto o servicio, en función de diversos factores como compra del mercado, límites de precios y estacionalidad en ventas. Una buena estrategia permite reajustar los precios rápidamente y a escala mientras se comprenden los efectos de sus cambios.

La tecnificación de esta estrategia permite pasar a las empresas del regateo que antiguamente se utilizaba, a las propuestas que trabajan con compras en aumento de volumen y cumplimientos de metas que les permite acceder a mejores descuentos, incluso en el mercado minorista permite a sus vendedores a corregir los escenarios de fijar un precio a largo plazo y fijarlo demasiado bajo renunciando a un margen innecesario o también corregir el cobrar demasiado y evitar la deserción de clientes.

La construcción de estos rangos de precios dinámicos permitirá a los vendedores tener una estrategia más sensible con el cliente, crear un aplicativo para evaluación

de rangos de precios dará facilidad a la negociación y entendimiento al comprador para asegurar a quien más valora el producto la oportunidad de comprarlo.

1.3 Objetivos

1.3.1 Objetivo General

Desarrollar una metodología de personalización de precios mediante la aplicación de técnicas de aprendizaje automático con la finalidad de ofrecer el precio correcto, al cliente correcto, en el momento correcto.

1.3.2 Objetivos Específicos

- Crear segmentaciones de clientes basado en la selección de atributos relevantes y su comportamiento de compra
- Determinar rangos de precios por cada segmento de clientes mediante técnicas de aprendizaje supervisado
- Predecir el nivel de compras por cliente para la evaluación de rangos de precios

1.4 Metodología

La propuesta considera la combinación de tres técnicas diferentes: segmentación de clientes, rangos de precios adecuados por cada segmentación, y tercero, predicción de volúmenes para evaluación del nuevo esquema de precios.

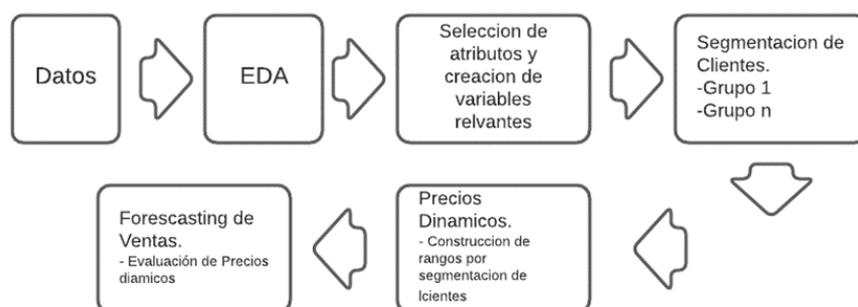


Figura 1.1 Propuesta de técnicas para construcción de precios dinámicos

Datos & Análisis Exploratorio de Datos (EDA)

Como primer paso se recolectarán los datos, donde luego se integrarán 4 conjuntos de datos con información histórica sobre el volumen y precios, entre otras características explicadas posteriormente en la sección 7. Como segundo paso se realizará el análisis exploratorio de datos. Estadística descriptiva a través de tablas de frecuencias y diagramas de cajas (*boxplots*), y Estadística Inferencial para prueba de hipótesis de, relación para variables numéricas, independencia para datos categóricos y pruebas ANOVA y Kruskal-Wallis para comparaciones múltiples de medias.

Selección de atributos

En esta etapa se evalúan los atributos relevantes antes de realizar la segmentación.

Se crearán las variables auxiliares:

- Número de diferentes SKUs que consume un cliente
- Porcentaje de participación de productos por la clasificación de familia del SKU
- Porcentaje de participación de productos por la clasificación de tipo de rentabilidad (dietas alto desempeño, desempeño pleno y bajo costo)

Se realizará análisis de datos multivariantes, específicamente, métodos de reducción de dimensiones para seleccionar los atributos de mayor relevancia, para encontrar similitudes entre los diversos tipos de clientes.

Segmentación de clientes

La segmentación de clientes se realiza sobre los atributos seleccionados. Algoritmos de agrupación no supervisada (*K-means*, por ejemplo) se utilizarán para encontrar similitudes entre los usuarios.

Precios dinámicos

En base a la segmentación de clientes, se realizará un modelo que determine los rangos de precios por cada uno de ellos; se utilizarán técnicas de aprendizaje automático para encontrar los rangos apropiados por cada segmento. El rango de precios para los grupos se representa a través del poder adquisitivo del cliente, el cual se medirá de acuerdo con los históricos de compras y a su potencial de crecimiento.

Forescasting de Ventas

Se realizarán proyecciones de ventas por clientes usando técnicas estadísticas clásicas como las series temporales o técnicas más relacionadas al ML como los modelos aditivos generalizados.

Calculadora de Pricing

Por último, se construirá un aplicativo web que provea el rango de precios dinámicos a ofrecer a un determinado cliente, de acuerdo con su perfil de compra histórica, volumen de compra y tipos de producto.

1.5 Resultados Esperados

Dentro de los resultados esperados se desea lo siguiente:

Entregables de modelación y script

- Segmentación de clientes basado en atributos relevantes
- Modelo con rangos de precios por cada segmento de clientes
- Algoritmo de predicción de compra de clientes
- Esquema de precios dinámicos con detalle de segmentos y rangos de precios

Otros

- Aplicativo web con calculadora de precios basado en el esquema de precios para evaluación de clientes en prospección.

1.6 Dataset

Para fines de la investigación, se utilizará un conjunto de datos de un mercado vendedor de alimento balanceado para camarón, con las siguientes características:

- Detalle de consumo histórico de toneladas de alimento balanceado de clientes por cada producto desde el año 2020 a mayo del 2021. El data set contiene 68 mil transacciones, más de 30 clientes, y cerca de 300 productos diferentes
- Detalle del precio de los productos desde 2020 a mayo 2021 por país

- Detalle de los descuentos otorgados a los clientes por volumen de compra y por acción de retenciones y/o acción de aumento de volumen
- Detalle de maestro de cliente con nivel de agrupación nodo

CAPÍTULO 2

2. ESTADO DEL ARTE

2.1 Fundamentos del problema

La fijación de precios puede ser definida como el ajuste situacional de los precios para un marco de condiciones que varía en el tiempo. Dentro de las empresas del sector de servicios se encuentra con mayor frecuencia nuevas estrategias de precios que surgen como respuesta a la problemática de competitividad. Si bien en el pasado era posible fijar precios exclusivamente sobre la base de los costos atribuibles (costos de producción), hoy estos se orientan a las estrategias de los competidores directos (precios del mercado), y en particular, del lado de la demanda de los clientes [13]; por lo tanto, los precios ya no suelen ser estáticos, ni dependientes únicamente de sus costos de producción durante todo el ciclo de vida de la empresa. El hecho de que las empresas ajusten sus precios a las condiciones cambiantes en un periodo de tiempo corto se lo considera como factor fundamental de éxito. Las actividades de investigación relacionales a esta problemática se asignan a las áreas de marketing, economía y comercio electrónico.

Los precios dinámicos pueden ser considerados como la combinación dos campos de investigación: el aprendizaje estadístico, aplicado específicamente al problema de estimación de funciones de demanda, y la optimización de precios. En esta sección, describimos brevemente los fundamentos históricos a partir de los cuales la fijación de precios y el aprendizaje dinámico han surgido, señalando algunas referencias claves.

El primero en ser reconocido en la utilización de formulación matemática para describir la relación precio-demanda de productos, y posteriormente resolver el problema de determinar el precio de venta óptimo, fue Cournot [16]. Otros científicos como Fisher también abordaron una explicación con aplicaciones de método matemáticos [17], siendo estos enfoques matemáticos para resolver problemas económicos bastante novedosos en esa época.

Para poder realizar los primeros pasos de aplicación del conocimiento teórico-matemático sobre la teoría de precios óptimo, es necesario tener una estimación de la función de demanda, principalmente, y luego abordar el problema desde dos puntos de vistas:

- Modelos donde la función de demanda cambia dinámicamente con el tiempo
- Modelos donde la función de demanda es estática pero los precios están sujetos al nivel de inventario

Dentro de los objetivos comunes encontrados en la literatura sobre precios dinámicos y aprendizaje, se tiene el desarrollar una política de precios que considere la incertidumbre propia entre la relación de precio y demanda, i.e., la dinámica de los precios está completamente causada por el hecho que la empresa conoce la relación precio-demanda.

2.2 Soluciones de analítica y aprendizaje relacionadas al problema

Dentro de las soluciones analíticas para el problema de precios dinámicos encontramos también las orientadas por el aprendizaje automático. El enfoque de estas se ha concentrado en diseñar un modelo realista para los mercados electrónicos con aplicación de técnicas que buscan medir el desempeño con las políticas tradicionales. Existen un amplio catálogo de soluciones para llegar a dinamizar precios e innumerables métricas de rendimiento, por lo que alcanzar una comparación puede resultar ser una tarea bastante difícil de explorar. Por esto, describiremos las soluciones que los autores consideran tienen mayor relevancia, y que se relacionan directamente no solo con el problema planteado, sino también con el giro del negocio de la industria de alimento balanceado para camarones.

Técnicamente, la definición de precios dinámicos se la conoce como la discriminación de precios, que consiste en vender un mismo producto o servicio a diferente precio considerando distintos grupos de clientes [18]; sin embargo, para los productos de primera necesidad, este esquema se ha

considerado discriminatorio desde el punto de vista legal, e.g., discriminatorio a grupos vulnerables como el género, la orientación sexual, entre otros. Para este estudio en particular no se ha considerado asumir este riesgo potencial legal y se tratará el ejercicio desde el punto de vista técnico y su solución basada en datos.

La literatura concerniente a la solución del problema de precios, apunta a dos principales abordajes para el cálculo de los precios dinámicos: basados en grupos y basados en el tiempo. Para el primero se utilizan métodos estadísticos y algoritmos de aprendizaje automático para sectorizar los grupos de clientes y luego ofrecer diferentes precios a los mismos, considerando predicciones a mayor predisposición de pago según un tipo de variable de interés (demográfica, ubicación, etc.). El segundo enfoque relaciona directamente la variable tiempo al comportamiento de compra, haciendo un precio más atractivo cuando los clientes tienen menos predisposición de compra.

Para mitigar el impacto negativo de la discriminación de precios dinámicos, se manejará en este estudio la estrategia efectiva de trabajar con un precio base para todos los clientes y un efecto diferenciador por los descuentos. Dentro de las soluciones encontradas detallamos a Rategenie [19], que provee una calculadora de precios dinámicos que trabaja con las variables de grupos clientes y tiempo.

2.3 Fuentes de datos relacionadas al problema

Dentro de las aplicaciones más conocidas en precios dinámicos encontramos las trabajadas desde el comercio electrónico en sector del retail. Una muestra de esto, es el data set *Marketing: Electronic Products and Pricing Data* disponible en Kaggle. Este conjunto de datos contiene Más de 7,000 productos electrónicos con información de precios en 10 campos únicos proporcionados por la base de datos de productos de Datafiniti, en el cual se plantea una solución desde el punto de vista de la demanda, y basado en el

tiempo. Otro ejemplo es el conjunto de datos públicos brasileños de comercio electrónico de Olist [20] el cual contiene información de pedidos y compras desde el año 2016 hasta el año 2018. En este se pueden observar pedidos desde diferentes dimensiones: desde el estado del pedido, precio, pago, desempeño del flete, ubicación del cliente, atributos del producto y finalmente las reseñas escritas por los clientes.

Por otro lado, [21] ofrece una guía para el modelaje del precio de viviendas usando técnicas de aprendizaje supervisado. Los resultados abordados en este cuaderno, si bien resultan estar bien ajustados a los datos, se enfrenta a la problemática de establecer el precio solo por variables de series temporales, por lo que pierde visión de negocio al no poder ser replicado en otro ejercicio.

Finalmente encontramos el data set de modelo de precios [22], el cual optimiza la asignación de precios por medio de clústers de productos, donde su principal objetivo es determinar segmentos en los cuales se ha podido bajar las tarifas sin afectar la rentabilidad. Dentro de este cuaderno se ha abordado la creación de nuevas características y clusterización; técnicas similares a las propuestas en nuestro trabajo, pero que no tiene en consideración la parte de pronósticos y la evaluación de los precios luego de encontrar el mejor modelo.

2.4 Librerías y software a utilizar

Se abordará una solución de precios dinámicos con tres acercamientos: clúster, rango de precios y forecasting; estos últimos nos ayudarán a resolver los objetivos planteados en el Capítulo 1. Se utilizarán dos lenguajes de programación para el tratamiento de datos y desarrollo de algoritmos (Python y R), y otros dos para el desarrollo de la aplicación web (Php y JavaScript). A continuación, describimos los softwares y lenguajes de programación usados para la implementación y distribución de nuestra propuesta

2.4.1 Software

Jupyter Lab: interfaz de usuario basada en web de próxima generación para Project Jupyter. La versión utilizada en este proyecto es la 3.1.10 [23].

R Studio: entorno de desarrollo integrado para R y Python, con una consola, un editor de resaltado de sintaxis que admite la ejecución directa de código y herramientas para el trazado, el historial, la depuración y la gestión del espacio de trabajo [24].

Visual Studio Code: editor de código fuente ligero pero potente que se ejecuta en su escritorio y está disponible para Windows, macOS y Linux. Viene con soporte integrado para JavaScript, TypeScript y Node.js. Tiene un rico ecosistema de extensiones para otros lenguajes (como C++, C#, Java, Python, PHP, Go) y tiempos de ejecución (como .NET y Unity) [25].

Apache Http Server: popular servidor web multiplataforma de fuente abierta que, según los números, es el servidor web más popular que existe. Es activamente mantenido por Apache Software Foundation. Algunas empresas de alto perfil que utilizan Apache incluyen a Cisco, IBM, Salesforce, General Electric, Adobe, VMware, Xerox, LinkedIn, Facebook, Hewlett-Packard, AT & T, Siemens, eBay, y muchas más [26].

SQL Server (Express Edition): SQL Server 2019 Express es una edición gratuita de SQL Server ideal para el desarrollo y la producción de aplicaciones de escritorio, aplicaciones web y pequeñas aplicaciones de servidor [27].

2.4.2 Lenguajes de programación

PHP: lenguaje de código abierto muy popular y ampliamente usado para el desarrollo web. Puede ser incrustado en HTML. Será utilizado para realizar toda la lógica de negocio de la aplicación web [28].

JAVASCRIPT: Lenguaje de secuencias de comandos que permite crear contenido de actualización dinámica, controlar multimedia, animar imágenes y mucho más. Permite implementar funciones complejas en páginas web. [29].

PYTHON: Lenguaje de programación que permite trabajar rápidamente e integrar los sistemas de forma más eficaz. Será utilizado para realizar la segmentación de clientes y de familias de productos mediante aprendizaje automático [30].

R: Entorno de software libre para gráficos y análisis estadístico. Se compila y ejecuta en una amplia variedad de plataformas como UNIX, Windows y MacOS. Usado principalmente para determinar los rangos de precios y el forecasting de las series de tiempo [31].

2.4.3 Librerías

Tabla 2.1 Clúster de Clientes desarrollado en Python

Librerías	Módulo	Breve Descripción
Pandas		Herramienta de manipulación y análisis de datos de código abierto, rápida, potente, flexible y fácil de usar [32].
Numpy		Librería especializada en el cálculo numérico y el análisis de datos, especialmente para datos en grandes volúmenes [33].
Matplotlib	pyplot	Biblioteca para la generación de gráficos a partir de datos contenidos en listas o arrays con extensión matemática NumPy. Proporciona una API, pylab, diseñada para emular a la de MATLAB [34].
Seaborn		Biblioteca de visualización de datos basada en matplotlib. Proporciona una interfaz de alto nivel para generar gráficos estadísticos atractivos [35].
Mpl_toolkits	mplot3d	Herramientas básicas de trazado en 3D (dispersión, surf, línea, malla). No es la biblioteca 3D más rápida o con mayor número de funciones, sin embargo se envía con Matplotlib, siendo una solución más liviana para algunos casos de uso [36].
Statsmodels	api	Cálculos, modelos estadísticos y métodos transversales.
Sklearn	sklearn.preprocessing Método: StandardScaler	Preprocesamiento y Normalización. Estandarización de variables.
	sklearn.decomposition Método: PCA	Descomposiciones de matrices. Estadística multivariante, específicamente, análisis de componentes principales (PCA).

	sklearn.model_selection Método: train_test_split	Selección de modelo. División de matrices en subconjuntos de prueba y entrenamiento.
	sklearn.metrics Método: silhouette_score	Métricas de rendimiento, sea en ajuste o poder predictivo. Cálculo del coeficiente de silueta medio de todas las muestras.
	sklearn.cluster Método: Kmeans	Agrupación usando el método K-Means para encontrar cluster.

Tabla 2.2 Detalle de librerías de R utilizadas para rangos de precios

Librerías	Módulo	Breve Descripción
readxl		Lectura e importación de archivos con formato Excel.
writexls		Escribe un marco de datos en un archivo xlsx, con una o varias hojas de trabajo.
tidyverse	readr	Proporciona una forma rápida y sencilla de leer datos rectangulares (como csv, tsv y fwf). Está diseñado para analizar de manera flexible muchos tipos de datos que se encuentran en la naturaleza
	dplyr	Gramática de manipulación de datos. Proporciona un conjunto consistente de verbos que resuelven los desafíos de manipulación de datos más comunes.
	stringr	Ofrece un conjunto cohesivo de funciones diseñadas para hacer que trabajar con cadenas sea sencillo.
	ggplot2	Sistema para crear gráficos de forma declarativa, basado en "The Grammar of Graphics".
lubridate		Contiene funciones útiles para el análisis de fechas en R, útil entre otras cosas para preparar nuestros datos antes de un estudio de forecasting..

Tabla 2.3 Forecasting de Clientes desarrollada en R

Librerías	Módulo	Breve Descripción
readxl		Lectura e importación de archivos con formato Excel.
writexls		Escribe un marco de datos en un archivo xlsx, con una o varias hojas de trabajo.
tidyverse	readr	Proporciona una forma rápida y sencilla de leer datos rectangulares (como csv, tsv y fwf). Está diseñado para analizar de manera flexible muchos tipos de datos que se encuentran en la naturaleza.
	dplyr	Gramática de manipulación de datos. Proporciona un conjunto consistente de verbos que resuelven los desafíos de manipulación de datos más comunes.
	stringr	Ofrece un conjunto cohesivo de funciones diseñadas para hacer que trabajar con cadenas sea sencillo.
	ggplot2	Sistema para crear gráficos de forma declarativa, basado en "The Grammar of Graphics".
	tidymodels	Colección de paquetes para modelado y aprendizaje automático utilizando principios tidyverse [41].
	tsibble	Proporciona una clase 'tbl_ts' (el 'tsibble') para datos temporales en un formato orientado a datos y modelos. El 'tsibble' proporciona

		herramientas para manipular y analizar fácilmente datos temporales, como llenar espacios de tiempo y agregar períodos de calendario.
lubridate		Contiene funciones útiles para el análisis de fechas en R, útil entre otras cosas para preparar nuestros datos antes de un estudio de Forecasting.
chron		Crea objetos cronológicos. Representa fechas y horas del día.
fabletools		Proporciona herramientas, ayudantes y estructuras de datos para desarrollar modelos y funciones de series de tiempo para 'Fable' y paquetes de extensión. Admite una interfaz coherente y ordenada para el modelado y análisis de series de tiempo.
fpp2		Carga los datos y librerías necesarias para la reproducción de los ejemplos y ejercicios utilizados en el libro Forecasting: Principles and Practice (2ª edición) de Rob J Hyndman y George Athanasopoulos [42].
manipulate		Funciones de trazado interactivo para su uso dentro de RStudio. Acepta una expresión de trazado y un conjunto de controles (por ejemplo, control deslizante, selector, casilla de verificación o botón) que se utilizan para cambiar dinámicamente los valores dentro de la expresión y el respectivo gráfico de forma automática.
Metrics		La precisión se define como la proporción de elementos reales que son iguales a los correspondientes elementos en predicho.
modeltime		Marco de predicción de series de tiempo para el ecosistema 'tidymodels'. Los modelos incluyen ARIMA, suavizado exponencial y modelos de series de tiempo adicionales de otros paquetes.
nortest		Pruebas de hipótesis para el supuesto de normalidad de los residuos.
seasonal		Interfaz fácil de usar para X-13-ARIMA-SEATS, el software de ajuste estacional de la Oficina del Censo de EE.UU. Ofrece acceso completo a casi todas las opciones y salidas de X-13, incluidos X-11 y SEATS. Permite la búsqueda automática de modelos ARIMA.
xgboost		es una biblioteca optimizada de aumento de gradiente distribuida diseñada para ser altamente eficiente, flexible y portátil. Implementa algoritmos de aprendizaje automático bajo el marco Gradient Boosting [43]
xts		Proporcionar el manejo uniforme de las diferentes clases de datos de tipo DateTime en R.

Tabla 2.4 Visualización e Interfaz Gráfica desarrollada en JavaScript

Librerías	Módulo	Breve Descripción
Bootstrap		Es un Kit de herramientas de código abierto para front-end.
Highcharts		Sirve para crear gráficos interactivos para proyectos web y móviles.

CAPÍTULO 3

3. DISEÑO E IMPLEMENTACIÓN

3.1 Exploración y Validación de datos

Como un primer paso, se trabajó en la construcción de características relevantes que permitan agrupar, en un primer acercamiento, a clientes que presenten cierta similitud en su comportamiento de compra. En la Figura 3.1 se muestra el diagrama de solución planteado para resolver el problema de precios dinámicos en cinco etapas.

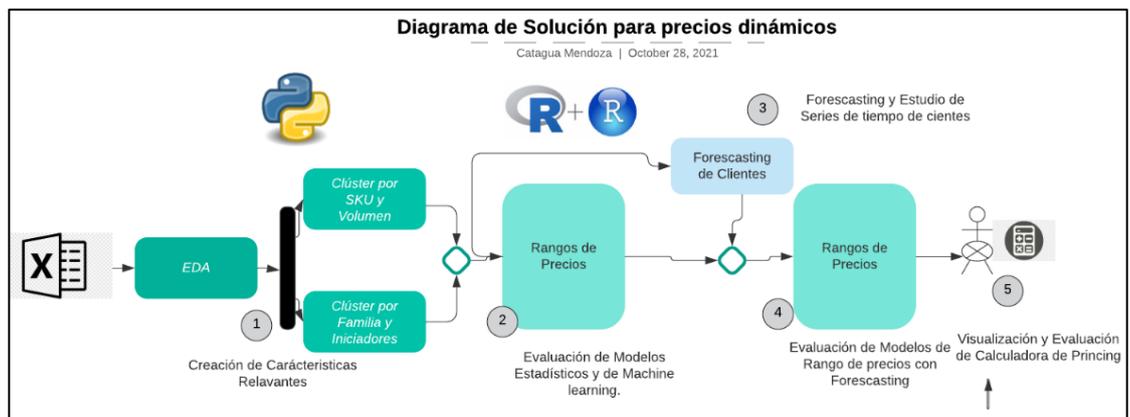


Figura 3.1 Propuesta de técnicas para construcción de precios dinámicos

Para esto, se realizó un análisis exploratorio de datos, donde para la creación de las características relevantes en la segmentación (etapa 1), se consideró la clasificación empírica de los grupos de clientes de acuerdo a la importancia de estos dentro de la comercialización de productos, además del conocimiento del negocio de los autores.

La creación de los clústers se realizó en el software Python. Dos tipos de agrupaciones fueron consideradas usando los indicadores a , b , c y d , definidos por

a : número de SKUs diferentes que consumen dentro de un mes

b : total de toneladas métricas (tms) que consume dentro de un mes

$c(i)$: porcentaje de tms de producto de la familia i comprado por mes

$d(j)$: porcentaje de tms de producto de la etapa j comprado por mes,

donde $i = \{1,2,\dots,6\}$ (Classic, Finalis, Katal, Origin, T rap, Otros) y $j = \{1,2\}$ (Iniciador, Engorde).

La primera agrupaci3n se realiz3 de acuerdo con las caracter sticas a y b, mientras que la segunda fue de acuerdo con las caracter sticas a, c y d. Luego, se combinaron las dos segmentaciones de clientes y se procedi3 a revisar la homogeneidad en los consumos de productos.

Estos cl sters anteriormente calculados, fueron usados para la creaci3n de rango de precios, con el fin de asignar el descuento de acuerdo con los atributos de compra del cliente. Se realiz3 una primera limpieza de los valores hist3ricos de descuentos, considerando las siguientes pol ticas. Si el descuento del mes n crece en un 0,5% en relaci3n con el promedio de descuentos de los tres meses anteriores, y el ingreso del mes n decrece en un 10% en comparaci3n con el promedio del ingreso de los tres meses anteriores, entonces no se considera dentro del modelamiento, puesto que la acci3n de menor descuento no ejerce una mejora en el consumo del cliente. Por otro lado, si el descuento del mes n decrece en un 0,5% (en relaci3n con el promedio de descuentos de los tres meses anteriores), y el ingreso del mes n crece en un 10% (en comparaci3n con el promedio del ingreso de los tres meses anteriores), entonces tampoco estos se consideran dentro del modelamiento, puesto que la acci3n de crecimiento de ingreso no es determinada por el aumento de descuento, sino que responde a otros factores no estudiados en este trabajo. Los puntos de 0,5% en descuento y 10% en ingreso, responden al giro de negociaciones de la empresa.

Una vez con los valores de descuento depurados, se realiz3 el modelado por medio de an lisis supervisado y tambi3n se realiz3 una modelizaci3n usando m3todos estad sticos cl sicos; espec ficamente, un modelo de efectos mixtos descrito en la siguiente secci3n.

Los pronósticos fueron realizados para el pronóstico de clientes y pronóstico de proporciones de compra de las familias de productos. Para el modelado de los pronósticos de clientes, se consideraron 10 pronósticos diferentes combinados con transformaciones de datos. Para los pronósticos de proporciones de productos, se consideraron cuatro modelos para calcular la serie de datos, que luego pasarán a ser sumados y sacar el porcentaje total.

Una vez elegidos los modelos ganadores para la predicción de series temporales y la modelización de descuentos, se realizaron pruebas para revisar el rango de cálculo sugerido por el modelo, de acuerdo con sus pronósticos de venta y participación por familia. Mediante una aplicación web los usuarios pueden tener acceso a una plataforma de rangos de precios, la cual contiene tres módulos: el primero con la descripción de clúster, el segundo centrado en la descripción de descuentos y el tercero, el cuál muestra la calculadora para evaluar los descuentos de un nuevo cliente potencial

3.2 Prototipos de algoritmos y modelos

En la solución planteada se presentan tres frentes de algoritmos o modelos: algoritmos utilizados para clasificación, algoritmos y modelos utilizados para el cálculo de descuentos y modelos para el desarrollo de *Forescasting*.

3.2.1 Clúster de clientes.

- K-means: algoritmo de aprendizaje no supervisado más utilizado para determinación de clústers. Dentro de sus características, está la escalabilidad con la cantidad de datos y su actualización de posición del centroide a la media. Se lo utilizó con los datos normalizados y se calculó el número óptimo de acuerdo con el coeficiente de silueta.

3.2.2 Modelos para descuentos.

A continuación, se presentan los distintos métodos para modelar los descuentos. Todos los modelos pertenecen a la categoría de algoritmos supervisados de Machine Learning.

- MARS: Regresión spline adaptativa multivariada. A diferencia del modelo GAM, este es un método no paramétrico. Además, usa tensores, o funciones bisagra para modelar las no linealidades.
- KNN: K vecinos más cercanos. Puede ser usado de forma no supervisada.
- LM: Regresión lineal clásica. Errores normales, independientes y homocedásticos.
- CART: Árboles de decisión y clasificación. Usa reglas basadas en covariables para decidir el valor o grupo de una nueva observación.
- Random Forest: Modelo formado por un conjunto de árboles de decisiones individuales.
- SVM Radial y SVM Polinómico: Máquina de soporte vectorial con kernel radial y polinómico, respectivamente.
- Red Neuronal simple: Considera una simple neurona.
- GAM: Modelo aditivo generalizado. Es un modelo de regresión semiparamétrico el cual permite modelar no linealidades a través de la inclusión de funciones suaves, e.g., splines, wavelets, etc.
- MLEM: Modelo Lineal Efectos Mixtos. Extiende a los modelos de regresión clásicos al poder modelar la correlación entre individuos u observaciones que pertenecen a un mismo grupo. Errores independientes entre grupos, pero dependientes dentro de estos.

Los modelos LM, GAM y MLEM son modelos estadísticos clásicos que pueden ser considerados como parte del Statistical Learning.

3.2.3 Modelos para pronósticos de series temporales

- Naïve. Modelo básico. Consiste en replicar el valor observado más reciente.

- Snaive. Llamado de método ingenuo. Básico en predicción. Usado como punto de referencia para revisar posibles modelos más complejos.
- ARIMA. Modelos que contienen integrada una parte autoregresiva (AR) y de media móvil (MA).
- Drift. Modelo derivado del ARIMA que contiene un paseo aleatorio.
- STL + Naïve. Descompone una serie de tiempo en términos de su estacionalidad y tendencia.
- Holt-Winters: Contiene una ampliación perfeccionada del enfoque de la suavización exponencial.

Para los modelos anteriores, será considerado usar transformaciones de Box-Cox para normalizar la distribución de los datos, cuando sea necesario.

3.3 Plataforma y prototipos de visualización

La plataforma de visualización contiene módulos, roles, usuarios:

3.3.1 Módulos

Primer módulo: Corresponde al modelo descriptivo del clúster de clientes.

Segundo módulo: Corresponde al módulo para revisar la descripción de los descuentos.

Tercer módulo: Corresponde a la calculadora de descuentos.

3.3.2 Roles

Rol Admin: Este rol permite realizar el control de usuarios (creación, edición o inactivación) y reasignar los roles a los usuarios.

Rol Consultor: Este rol permite tener acceso a los módulos antes mencionados.

3.3.3 Usuarios

admin: Este usuario tiene asignado el rol Admin, por lo tanto, es el administrador de la aplicación web.

demo: Este usuario tiene asignado el rol Consultor, por lo tanto, es el que podrá visualizar los modelos y la calculadora de descuentos para los precios dinámicos.

El acceso a la aplicación web es mediante usuario y contraseña, estos se obtienen desde un motor de base de datos la cual se describe en la sección 2.4.1.

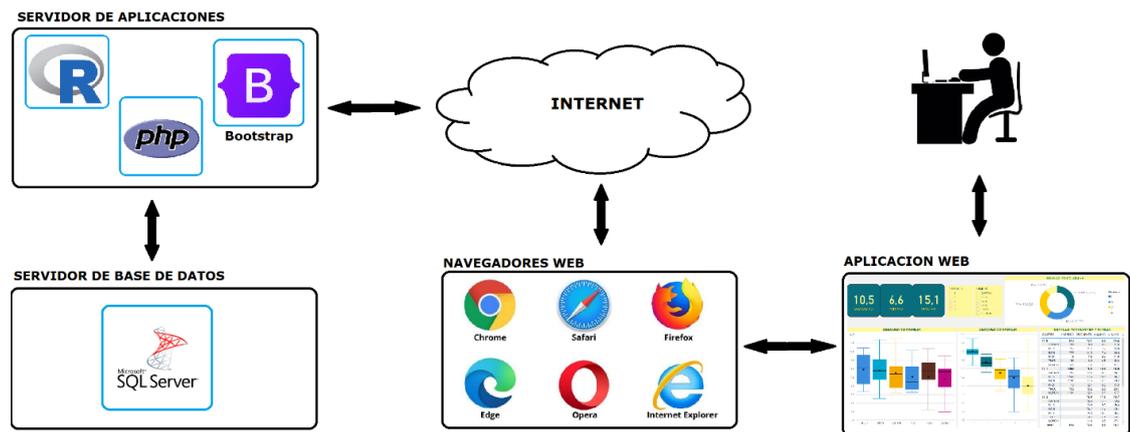


Figura 3.2 Estructura de la solución propuesta

3.4 Métricas y comunicación de resultados

Las métricas a ser utilizadas para determinar el rendimiento del modelo de rango de precios con los pronósticos son:

$$\% \Delta DaD = \frac{\% \text{ descuento mes } n}{\% \text{ descuento mes } n - 1} - 1$$

$$\% \Delta IaI = \frac{\text{Valor Ingreso mes } n}{\text{Valor Ingreso mes } n - 1} -$$

Estas medidas permitirán evidenciar si a medida que existe crecimiento en ingreso, el porcentaje de descuentos también crece o se mantiene por medio de gráficos evolutivos, con la cual se calculará clientes con demanda

creciente, y así poder evidenciar si el modelo de descuentos presenta valores extremos con respecto a meses anteriores reales.

$$\% \Delta \text{DYTGo} = \frac{\sum_6^{12} \text{descuento proyectado año 2021 (n)}}{\sum_1^6 \text{descuento real año 2021 (n)}} - 1$$

$$\% \Delta \text{DYTD} = \frac{\sum_6^{12} \text{descuento proyectado año 2021 (n)}}{\sum_6^{12} \text{descuento real año 2020 (n)}} - 1$$

Para el clúster se calcularon dos valores de los descuentos en el periodo de 6 meses en la predicción y se los comparó con los resultados del mismo intervalo de tiempo, pero en el año anterior, así como también con el semestre anterior, con el objetivo de evidenciar el rendimiento del modelo y la estabilidad de sus descuentos.

Se han considerado también las medidas tradicionales dentro del negocio como los montos totales de descuentos por cliente, clúster, familias, etc.

La comunicación de resultados está planificada en tres fases detalladas a continuación:

Fase 1: pruebas alpha. Se comunicarán los resultados a especialistas del negocio y colegas para revisar si el modelo implementado tiene coherencia con lo revisado en la parte descriptiva.

Fase 2: pruebas beta. Se comunicará a las gerencias y partes interesadas (e.g., como directores de proyectos), en la cual se centrará en realizar pruebas en la calculadora de descuentos para clientes, evaluando el modelo incluso con nuevos datos.

Fase 3: socialización. Se apertura la calculadora de *pricing* para los usuarios. Se realizará una reunión de lanzamiento y la entrega de material de apoyo como manuales, entre otros.

CAPÍTULO 4

4. ANÁLISIS DE RESULTADOS

4.1 Análisis estadístico descriptivo del problema

Se realizó el análisis descriptivo de los datos al data set principal ya revisado, el cual contiene 183 mil registros con 34 variables. En primera instancia, se consideraron 43 clientes para formar parte de los 5 grupos de clústers resultantes. Solo 28 clientes fueron usados para entrenar el modelo de rangos de precios, donde los restantes 15 fueron excluidos por criterios del negocio; estos fueron, aquellos clientes pequeños con compras irregulares y aquellos desertores que no presentaron compras en el año 2021.

Para un mejor ordenamiento de los análisis se dividieron en base a los tres objetivos principales.

Clúster: se realizaron tres scripts en Python. El primero para el análisis exploratorio de datos; el segundo para encontrar los clústers detallados por número promedio mensual de SKUs comprados en el último año, y el tercero para los clústers por familia.

Forescasting: se realizaron tres scripts en R. El primero correspondiente a limpieza de los datos y gráficos exploratorios; el segundo para el pronóstico de la serie temporal de consumo de toneladas por cliente; y el tercero, para las series temporales de consumo de los clientes por familia.

Precios Dinámicos: se realizaron tres scripts en R. El primero para el análisis exploratorio de datos enfocado en los descuentos; el segundo, para los modelos evaluados; y el tercero, quien calcula las métricas de evaluación.

Los datos han sido enmascarados para protección y seguridad de la empresa, por lo que se describen las familias de productos como personajes de ciencia ficción determinados por los autores. De igual forma, se utilizó un factor aleatorio para

alterar de manera consistente los datos, para así no divulgar los valores reales, sino poder interpretar su modelamiento. Los pseudónimos para las familias de productos son: Iron, Thor, Captain, Nick, Falcon, Hulk, Ultron y Otros.

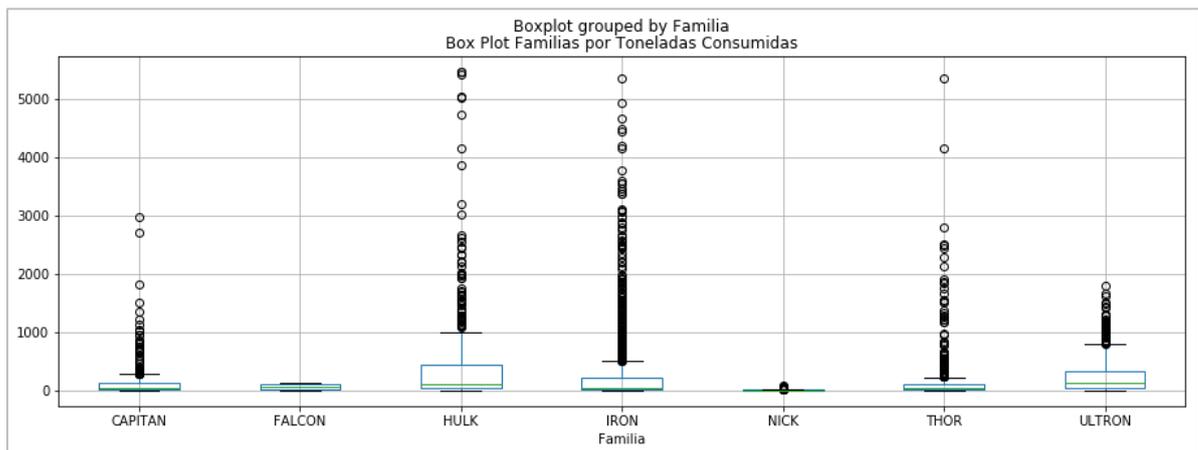


Figura 4.1 Consumo Totales de toneladas por familia durante el periodo enero de 2020 a junio de 2021

Dentro de los resultados encontrados, se puede apreciar a través de los diagramas de cajas en el gráfico 1 que las familias de Iron, Hulk y Thor son las que presentan mayores consumos que el resto. Estas familias han tenido consumos de más de 5 mil toneladas por cliente dentro del mes. Por otro lado, las familias con consumos más bajos son la Falcon y la Nick, las que presentan en su diagrama de caja valores por debajo de las mil toneladas.

4.2 Resultados de Segmentación de clientes

Se calcularon las variables descritas en capítulo 3 para la determinación de los clústers. La primera corrida fue realizada considerando el número promedio de SKUs consumidos por mes, y la suma de toneladas de producto. En la tabla 1 se puede observar el consumo promedio de toneladas de los clientes, junto al número de SKUs diferentes comprados mensuales en su top 10. De estos, resaltamos que el cliente Q, es el mayor cliente con compras promedio de 19 mil toneladas y un promedio de 20 SKUs de consumo diferentes por mes. El segundo mayor es el cliente AX con 18 mil toneladas de consumo y 19 SKUs promedio. Luego de estos

dos, siguen clientes con consumo menor a las 6 mil toneladas promedio, pero con un número de SKUs altos, como lo son los clientes Y y Z, con 6,7 mil toneladas y 16 SKUs, y 3,3 mil toneladas y 18 SKUs promedio, respectivamente.

Tabla 4.1: Promedio de SKUs, Familias y Toneladas Consumidas por nodos de clientes

Cod_nodo	Des_nodo	MeanSKU	MeanFamilia	MeanToneladas
9982211	Cliente_Q	20.00	4.75	19634.99
9982366	Cliente_AX	18.92	5.58	18660.69
9982225	Cliente_Y	16.42	5.25	6732.37
9982363	Cliente_AV	6.67	2.58	6479.15
9982364	Cliente_AW	17.42	5.67	4956.16
9982226	Cliente_Z	18.08	5.58	3386.67
9982216	Cliente_V	4.42	1.42	3137.71
9982362	Cliente_AU	17.08	5.75	2762.88
9982360	Cliente_AS	5.58	3.17	1282.99
9982215	Cliente_U	2.25	1.42	1110.83

Los clústers fueron evaluados utilizando el método K-means, donde para determinar el número óptimo de clústers, se probó un rango de dos a diez; además se utilizó el índice de silueta como medida de optimización. Bajo estos escenarios, el número óptimo de clústers fue cinco, con un índice de silueta de 0.56 (obsérvese gráfica en los anexos). Con este número óptimo se procedió a calcular los clústers detallados a continuación. De los 5 grupos mostrados en el gráfico 2 detallamos:

Clúster 0: Consumo promedio de 2 SKUs y 284 toneladas promedio al mes

Clúster 1: Consumo promedio de 17 SKUs y 4.4 mil toneladas promedio al mes

Clúster 2: Consumo promedio de 19 SKUs y 19 mil toneladas promedio al mes

Clúster 3: Consumo promedio de 9 SKUs y 2,2 mil toneladas promedio al mes

Clúster 4: Consumo promedio de 4 SKUs y 543 toneladas promedio al mes

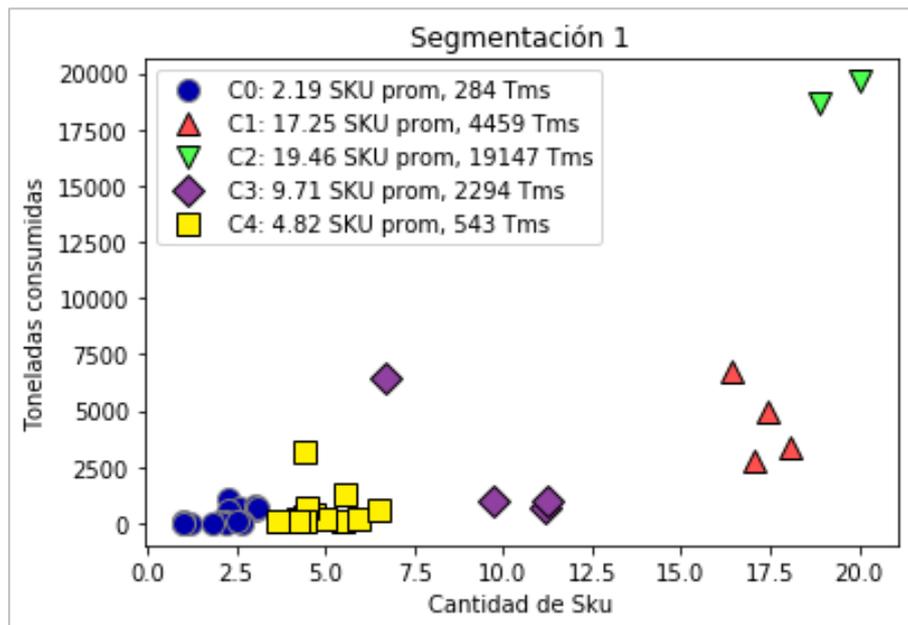


Figura 4.2 Clúster por número de SKU y Toneladas promedio

Dentro del clúster 2 encontramos dos clientes que son los de mayor consumo de alimento. El clúster 1 contiene 4 clientes con consumos de más de 4 mil toneladas y que consumen varios productos. Por otro lado, el clúster 3 contiene clientes más pequeños. Los otros dos clústers son clientes con consumo promedio menor a mil toneladas. Si bien se podría agrupar a estos últimos, para el estudio de precios los autores decidieron dejarlos agrupados para su estudio particular de precios. Fueron realizadas otras corridas de clúster con los algoritmos espectral y DBSCAN, donde en el primero se obtuvieron los mismos grupos de clasificación y similitud en la clasificación por cliente, mientras que con el segundo se obtuvieron solo dos clústers. Luego de una exhaustiva revisión, estos últimos fueron descartados por su falta de alineación al proyecto.

Una segunda corrida de análisis de clústers fue realizada con K-means, en las cuales se quería evidenciar la agrupación de los clientes dependiendo el consumo interno de las familias. En principio, se utilizó componentes principales para la reducción de dimensiones de estas variables y proceder a graficar. El número óptimo de clúster resultó siete, los cuales se detallan a continuación.

Tabla 4.2 Resultados de clusterización por familia de productos y toneladas de consumo

Cluster	IRON	HULK	CAPITAN	NICK	...	ULTRON	THOR	FASE 1	Toneladas
0	0.80	0.07	0.01	0.00	...	0.03	0.09	0.14	13772.33
1	0.43	0.02	0.42	0.00	...	0.01	0.09	0.14	5429.44
2	0.34	0.25	0.01	0.06	...	0.30	0.04	0.28	24404.48
3	1.00	0.00	0.00	0.00	...	0.00	0.00	1.00	148.48
4	0.35	0.00	0.23	0.00	...	0.00	0.42	0.33	1976.96
5	0.13	0.08	0.00	0.00	...	0.78	0.01	0.13	2590.72
6	0.21	0.72	0.02	0.00	...	0.02	0.03	0.10	2634.24

En el clúster 0, se encuentran los clientes que en su preferencia consumen la familia Iron; en el clúster 1 los clientes utilizan Iron y Captain; los del clúster 2 utilizan Iron y Ultron; el clúster 4 están los clientes que utilizan Iron, Captain y Thor; principalmente, en el clúster 6, se encuentran los clientes que utilizan Iron y Hulk. Adicionalmente, se obtuvo un escenario plausible con 3 clústers, pero después de un estudio este fue descartado a favor de la inicial con siete. Vale mencionar que el primer clúster fue utilizado como principal segmentación, mientras que el segundo fue utilizado para poder describir los datos de manera más efectiva.

4.3 Resultados de Rangos de precios dinámicos

Dentro del estudio de exploración para los rangos de precios, descuentos mayores al 40% fueron descartados pues, como nos fue corroborado por la gerencia, estos se debían a errores de registro en el sistema de precios de la compañía. Otros análisis fueron realizados para la depuración de casos donde los valores de descuentos no obedecían a las políticas de la empresa. Por ejemplo, se excluyeron los aumentos de descuentos que no respondían a un aumento de consumo mensual, así como también los descuentos que no respondían a una estrategia de crecimiento con volumen.

En un primer acercamiento, se consideraron pruebas dentro de los modelos con la variable etapa, la cual hace referencia a la etapa en el que el alimento es requerido

por el camarón, siendo sus valores etapa 1 y 2, mas esta variable no respondió favorablemente a las métricas de evaluación, por lo que se decidió excluirla del análisis, aliviando el número de variables de modelo.

Para la corrección de valores atípicos, se incluyó una regla de validación de los descuentos por familia, dentro de su rango de variabilidad estudiada. En el gráfico 3 se puede observar el comportamiento de las familias, de acuerdo a lo segmentación mostrada en el Gráfico 2. Se puede ver que los clientes de los clústers 1 y 2 presentan descuentos más altos que el resto, siendo los del clúster 2 los que presentan menor variabilidad. El clúster 4 presenta varios valores atípicos que han pasado los filtros antes descritos. Por otro lado, el clúster 0 se encuentran descuentos con mayor rango Inter cuartil dentro de las familias. El clúster 3 presenta descuentos menores en comparación con los clústers 1 y 2, sin embargo, su variabilidad esta es menor que la de los demás, a excepción de la familia Hulk.

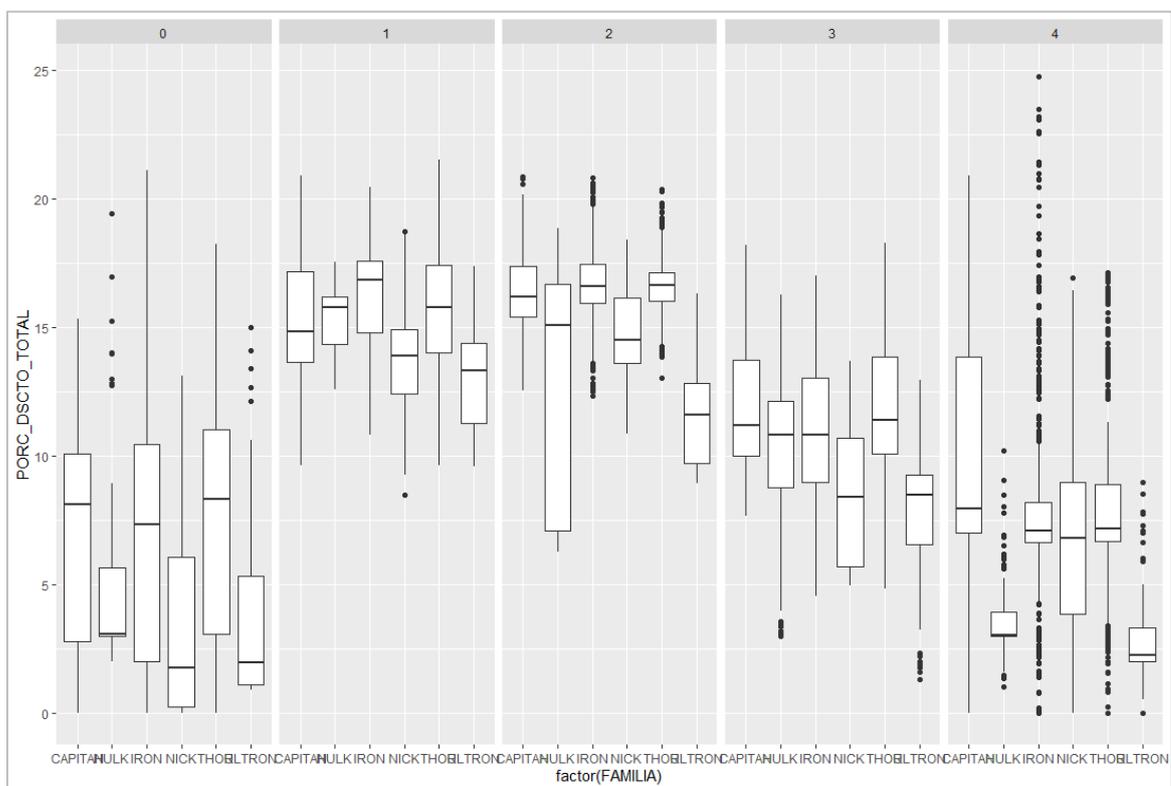


Figura 4.3 Gráfico de Cajas por clúster y familia

Adicionalmente es de interés estudiar el comportamiento del otorgamiento de los

descuentos como función del tiempo, por lo que las series temporales para cada clúster fueron calculadas. Estas se presentan en el gráfico 4. De este último gráfico, se aprecia que los clientes del clúster 0 reciben niveles de descuentos más bajos que el resto, entre el 5% y el 10%, pero con una tendencia al alza en el 2021. Con respecto al clúster 1 y 2, vemos que los clientes reciben los máximos niveles de descuentos, alrededor del 15% al 20%, donde para el clúster se puede apreciar una ligera tendencia decreciente en los últimos dos años, a diferencia del clúster 2, el cual presenta un crecimiento constante pero moderado. Clientes del clúster 3 han recibido descuentos promedios similares, alrededor del 10%, durante los últimos tres años. Finalmente, el clúster 4 recibe los segundos niveles de descuentos más bajos, alrededor del 5 al 10%, y cierta tendencia al alta postpandemia.

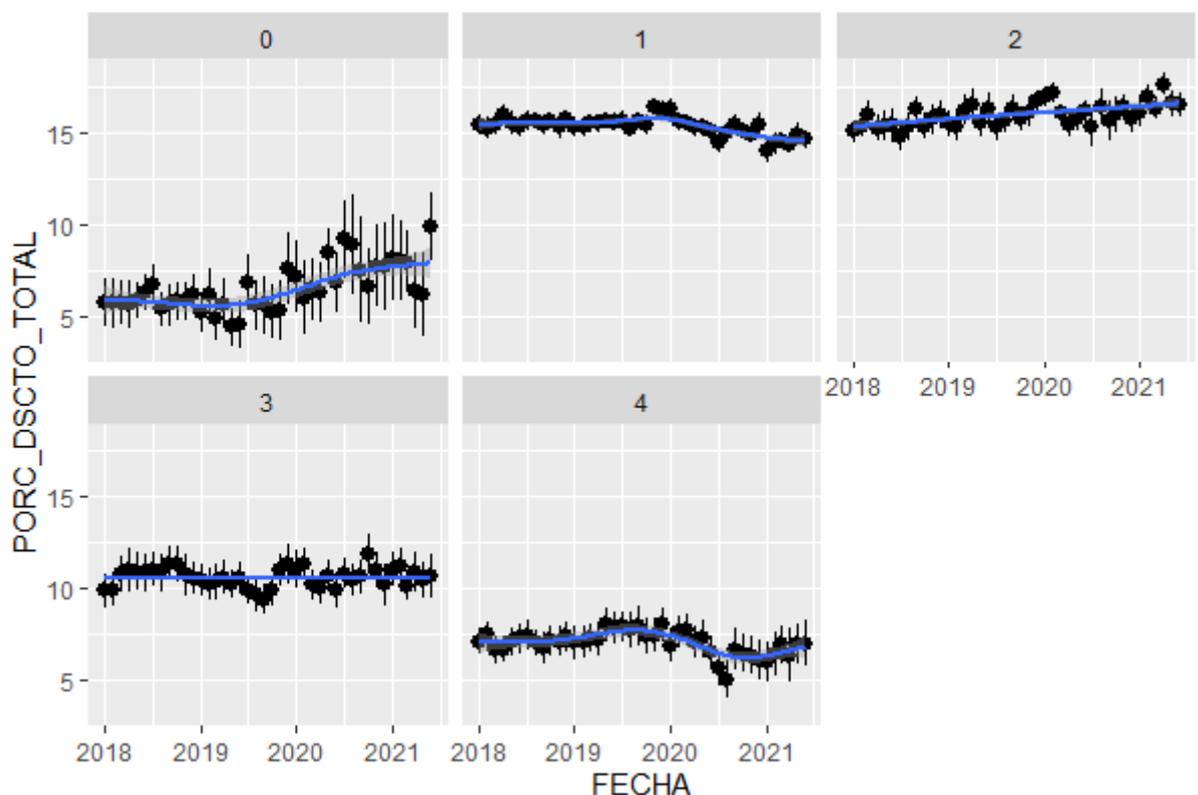


Figura 4.4 Series temporales de descuentos por clúster desde 2018

Luego de esta exploración, se procedió a la evaluación de los modelos descritos en el capítulo anterior, realizándose una partición del data set en 75% para los datos de entrenamiento, y 25% para los datos de prueba. Para el grupo de los modelos

de machine Learning, los tres mejores modelos ajustados de acuerdo a la medida del RMSE fueron, de mejor a peor: el modelo de máquinas de soporte vectorial con método radial, con un RMSE de 2,16 y un coeficiente R2 de 77%, el modelo de máquinas de soporte vectorial con método polinómico, con un RMSE de 2.18 y un R2 de 76%, y el MARS con un RMSE de 2.31 y un coeficiente R2 de 73%. Si bien en estos resultados se busca llegar el mejor ajuste posible, los autores consideraron que la explicación de los modelos trabajados agrega valor al negocio y que puede ser explicado con complejidad media.

	RMSE	Rsquared	MAE
res_gam1	3.200577	0.49252301	2.510882
res_gam2	3.211108	0.48927623	2.514702
res_gam3	2.521915	0.68464897	1.868495
res_mars	2.312816	0.73456786	1.712942
res_knn	2.573544	0.67088271	1.934139
res_kknn	2.227747	0.75355945	1.655463
res_lm	2.444141	0.70372483	1.846847
res_rpart	3.350865	0.44194033	2.603144
res_rf	2.079209	0.78621343	1.533376
res_svm1	2.043443	0.78667958	1.420300
res_svm2	2.184743	0.76727711	1.558204
res_nnet	11.487357	0.04935794	10.580894

Tabla 4.3: Métricas obtenidas para los diferentes métodos ajustados para modelar los valores de descuentos por clústers

El modelo ganador explica el descuento en función del ingreso por familia, la temporalidad (mes y año), y el clúster al cual pertenece el cliente. Cabe recalcar que la variable ingreso se la trabajó con una transformación logarítmica para mejorar la linealidad del modelo y la simetría de los datos. Este modelo se utilizará para otorgar un sugerido de descuento en la aplicación que se trabajó.

Finalmente, se exploraron modelos sencillos de Deep Learning (e.g., redes neuronales) y otros modelos que incluyen transformaciones polinómicas, combinaciones de variables e inclusión de nuevas características, sin embargo, el

rendimiento de estos no justificaba la complejidad de explicación de los mismos, por lo que no se muestran dentro de los resultados de la Tabla 2.

Ajuste de modelos estadísticos clásicos

En contraste con los modelos anteriores, se probaron también modelos tradicionales estadísticos teniendo en cuenta los principales supuestos a cumplir, así como características de los datos como la temporalidad, y posible correlación entre clientes de un mismo clúster. A continuación, se describen el mejor modelo estadístico encontrado, junto a una descripción de las características del mismo.

Se desarrollaron modelos GAM (Generalized Additive Model) para incorporar posibles tendencias no lineales de los predictores, donde se probaron diferentes funciones bases para modelar la no linealidad como Splines, polinomios, entre otros. El mejor ajuste fue usando el método *gamSpline* disponible a través de *caret*, donde se utilizó el parámetro *random* para incluir un efecto aleatorio y así modelar el comportamiento entre clústers y también su interacción con el monto.

4.4 Resultados de predicciones y evaluación de rangos de precios

Las predicciones fueron abordadas en dos grupos: las predicciones de las series temporales de demanda en toneladas y las correspondientes a la participación de las familias de productos. Una vez obtenida las predicciones, se realizó el cálculo del precio promedio por familia para finalmente encontrar el valor del ingreso por compra y el respectivo descuento.

En la gráfica 5 se muestran las series temporales de las toneladas consumidas por cada clúster para los años 2018 a 2021. El análisis de series temporales fue realizado a detalle por cada cliente, sin embargo, con el objetivo de tener una herramienta de calculadora de precio dinámica, i.e., que sea capaz actualizarse de acuerdo a la nueva data, se lo calculó por medio de la estrategia *train-test* junto a una medida de precisión de interés. Si bien la temporalidad de las series no se puede apreciar de manera marcada en los clústers, dentro de los clientes se

revisaron posibles tendencias y estacionalidad con la evaluación de varios modelos.

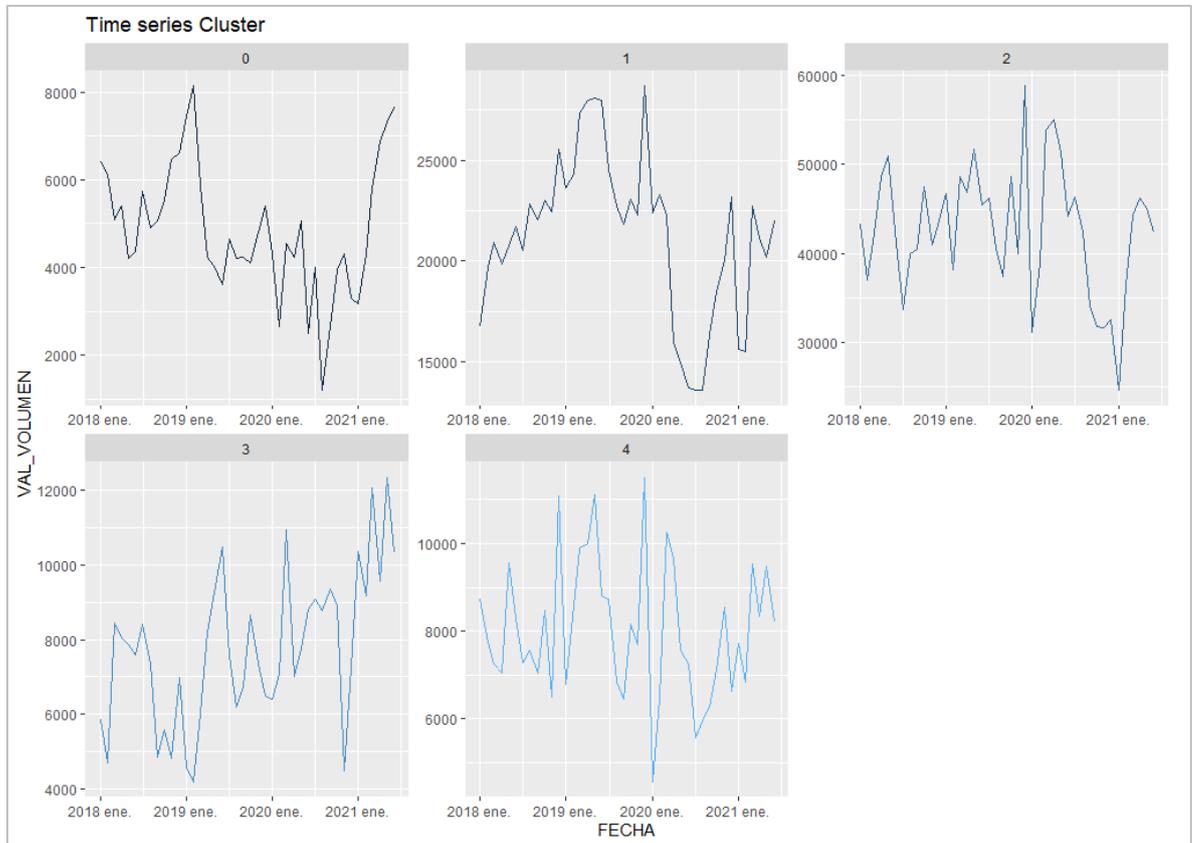


Figura 4.5 Series temporales de los clústers de clientes para los años 2018 a 2021

Fueron considerados 10 modelos por cada serie, tres de los cuales incluyen transformaciones de Box-Cox. En el data set para la evaluación del modelo, después de encontrar el modelo ganador, se realizaron proyecciones a una ventana de tiempo de 18 meses.

En la tabla 3 se muestran los resultados de los diferentes modelos aplicados en términos de la métrica RMSE, ordenados de mejor a peor. De esta última se puede apreciar que el modelo ganador para la proyección fue el modelo Holt-Winter multiplicativo, mientras que el segundo mejor modelo fue el *drift* multiplicativo. Vale mencionar que todos los clientes fueron evaluados de manera independiente con esta metodología.

Tabla 4.4: Evaluación de Modelos para Forecasting de los clientes

	Nombre_Modelo	RMSE
Test set6	Modl_holt_win_multiplicativo	822.9636
Test set3	Modl_drift	982.5683
Test set2	Modl_arima	1007.7469
Test set5	Modl_holt_win_aditivo_pro	1015.1772
Test set	Modl_naive	1045.8992
Test set4	Modl_stl_naive	1263.6098
Test set1	Modl_snaive	1902.4980
Test set8	Modl_arima_Box	2253.0632
Test set9	Modl_stl_naive_Box	2254.5874
Test set7	Modl_snaive_Box	2255.2074

Para el cálculo de la participación de familia (entiéndase como la distribución del volumen de comprar en términos de cada familia de productos), se proyectaron cada uno de los volúmenes por familia de forma independiente. Dado que el total de toneladas de consumo promedio por clúster es una cantidad conocida la cual ya ha sido proyectada anteriormente, se procedió a calcular los porcentajes de participación por familia, simplemente dividiendo las toneladas de compra por familia para el gran total.

En el grafico 6 se puede apreciar la participación por familia para cada clúster dentro del mismo período de tiempo de estudio. Para todos los clústers se puede ver una tendencia general de disminución de participación de la familia Captain, un aumento de la familia Thor, y finalmente la participación de la familia Ultron a partir del año 2020.

En particular, los clústers 0 y 4, son predominantes en la compra de la familia Captain, mientras que los clústers del 1 al 3, son principales compradores de la familia Nick. Con respecto a la familia Hulk, los clientes de los clústers 1 y 2 son sus mayores compradores, seguidos por los de los clústers 3 y 4 quienes son compradores medianos del producto, y finalmente el clúster 0 que casi no lo compra. El clúster 1 presenta tendencias estables en los años de estudio, donde

se ve una reducción en la participación de la familia Iron y un aumento en la participación de la familia Hulk y Ultron. Podemos finalizar mencionando que en el clúster 3 se puede evidenciar una participación de familias bastante constante.



Figura 4.6 Participación de Familia por clúster

Los pronósticos de familia fueron abordados usando cinco modelos: ARIMA, ETS, regresión lineal múltiple, Snaïve y una combinación de ARIMA-naïve. En la Tabla 4-4 se pueden apreciar los diferentes modelos ajustados junto a sus medidas de precisión por cliente. Estas predicciones fueron abordadas utilizando series temporales independientes por familia y clúster.

Una vez elegidos los mejores modelos por cliente, se realizó la evaluación de los descuentos para estos pronósticos de ingreso.

Tabla 4.5: Tabla de ejemplo de modelos de Forecasting para familia

.model	Cod_Nom	.type	ME	RMSE	MAE	MPE	MAPE
model_arima	9982225_Cliente_Y	Test	553.6000	684.9029	590.7200	23.870819	26.53137
model_average	9982225_Cliente_Y	Test	1064.3200	1173.8943	1064.3200	49.310322	49.31032
model_lm	9982225_Cliente_Y	Test	1080.8415	1179.9869	1080.8415	50.432205	50.43220
model_snaive	9982225_Cliente_Y	Test	1575.0400	1683.4969	1575.0400	74.749826	74.74983
model_arima	9982226_Cliente_Z	Test	732.4973	998.0149	894.3235	28.009220	53.29457
model_average	9982226_Cliente_Z	Test	1017.4486	1218.1510	1019.0018	51.256934	51.49961
model_lm	9982226_Cliente_Z	Test	1063.1348	1234.0246	1063.1348	56.526129	56.52613
model_snaive	9982226_Cliente_Z	Test	1302.4000	1463.5103	1302.4000	74.504647	74.50465
model_lm	9982362_Cliente_AU	Test	545.8252	673.7974	562.0859	36.386426	38.94766
model_arima	9982362_Cliente_AU	Test	556.0229	700.6601	596.1522	36.039802	42.36056
model_average	9982362_Cliente_AU	Test	778.4915	869.6545	778.4915	58.556893	58.55689
model_snaive	9982362_Cliente_AU	Test	1000.9600	1062.0391	1000.9600	81.073984	81.07398
model_average	9982364_Cliente_AW	Test	262.7473	713.5032	602.8590	7.793964	26.88203
model_lm	9982364_Cliente_AW	Test	558.7674	717.0852	558.7674	24.074418	24.07442
model_arima	9982364_Cliente_AW	Test	-940.1054	1119.2720	940.1054	-51.612794	51.61279
model_snaive	9982364_Cliente_AW	Test	1465.6000	1632.7814	1465.6000	67.200722	67.20072

4.5 Rangos de precios y evaluación de medidas

Es de interés no solo considerar una estimación puntual de los descuentos, los que se podrían mal interpretar como fijos dentro de las negociaciones con los clientes, sino obtener estimaciones intervalares, i.e., intervalos de descuentos. Con este fin, se utilizó una metodología basada en el remuestreo, para así poder obtener información sobre la distribución de los descuentos para cada escenario. En particular, se utilizaron 200 muestras Bootstrap donde el modelo ganador SVM fue ajustado, para así obtener una estimación de la variabilidad de los precios, bajo las diferentes condiciones de compra de cada cliente.

En la ilustración 2 se pueden observar los valores actuales de descuento, así como los rangos de descuentos obtenidos (intervalos de predicción del 90%), esto para los clústers 0 y 3 en los paneles de la izquierda y derecha, respectivamente. Esta ilustración es resultado de la calculadora de descuentos. Como se puede evidenciar en ambos casos, los rangos de descuentos contienen los valores puntuales que ha otorgado el modelo. Por otro lado, en el gráfico del clúster 3 se puede evidenciar que

los intervalos de predicción son ligeramente más amplios que los del clúster 0 para valores de descuento entre 10% y 12%, valores en los que se tiene la mayor cantidad de observaciones y variabilidad de los descuentos.

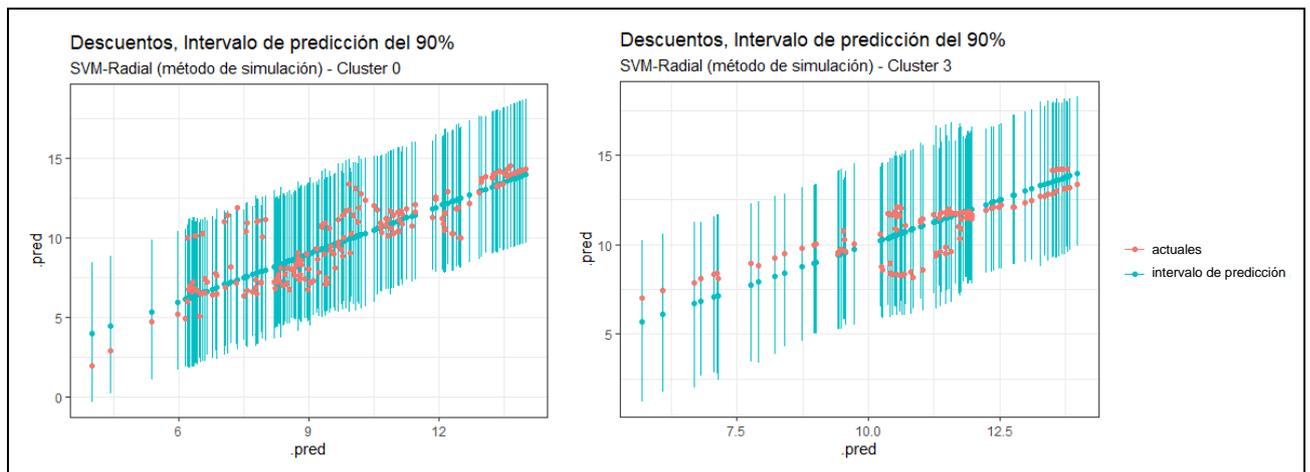


Figura 4.7 Rangos de precios para los clústers 0 y 3

En la sección 3.4 se propuso como métrica la comparativa YTD, que consiste en comparar los descuentos reales otorgados en los meses de enero a junio del año 2021, versus lo estimado por el modelo descuentos para la segunda mitad del año, es decir, en el período de julio a diciembre del año 2021. En el grafico 7 se pueden diagramas de cajas por clúster correspondientes a los descuentos reales otorgados en el primer semestre del año (YTD) en color verde, junto a los descuentos calculados por el modelo para el segundo semestre del año (YTgo) en color naranja.

De este último gráfico podemos comentar que, para los valores proyectados por el modelo, se evidencia la corrección de valores que se podrían considerar atípicos en el periodo inicial en todos los clústers, a excepción del primero. Además, en clúster 2 se puede apreciar que los descuentos pronosticados tienen menor rango intercuartil que los del primer semestre del año real; si bien para el último cuartil del año los pronósticos evidencian un mayor rango en descuentos que los otorgados realmente, se puede observar la corrección de valores atípicos evidenciados en el primer semestre del año del mismo período.

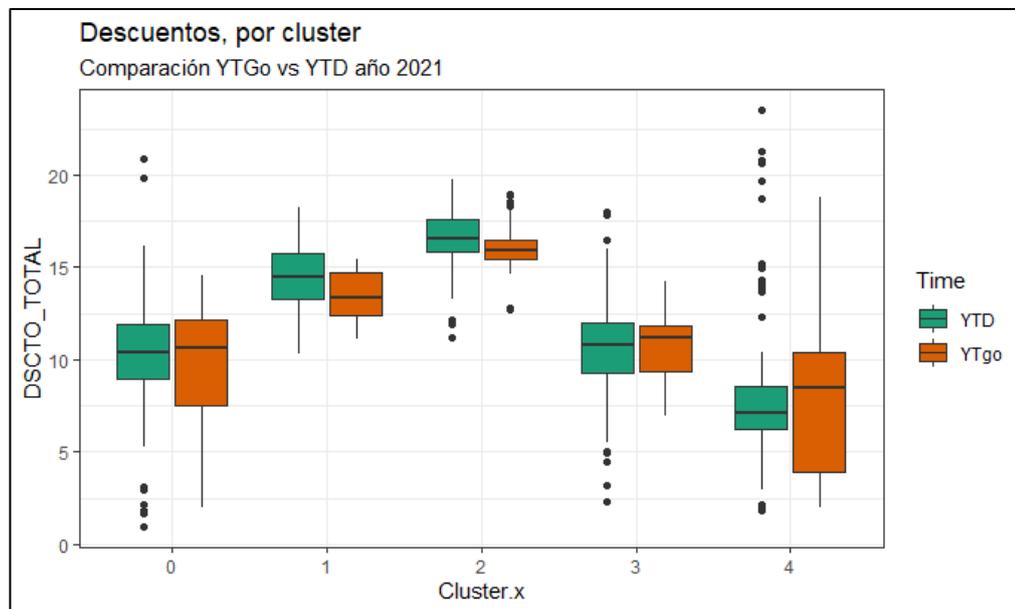


Figura 4.8 Comparación de descuentos YTgo (Jul – Dic) 2021 vs. YTD (Ene - Jun) año 2021

Con respecto al segundo período del año (julio a diciembre), se realizó una comparación entre los valores reales de descuento otorgados en el año 2020, con los pronosticados a través de la propuesta para el año 2021, la cuál se puede ver resumida en la figura 8. De los diagramas de cajas se puede observar que, en general, se espera en el año 2021 que exista una mayor variabilidad en el otorgamiento de descuentos. Respecto a los valores medianos de descuentos por clústers, en el clúster 1 se puede apreciar un evidente incremento para el año 2021, mientras que un efecto negativo es observado para el clúster 4.

4.6 Análisis de beneficios del cálculo de precios dinámicos

Dentro de las limitantes encontradas con respecto al modelamiento de descuentos se evidenció la falta de registros de las políticas de descuentos por parte de la empresa, por lo que, basado en los hallazgos del presente trabajo, ciertas directrices son recomendadas en las secciones de conclusiones y recomendaciones.

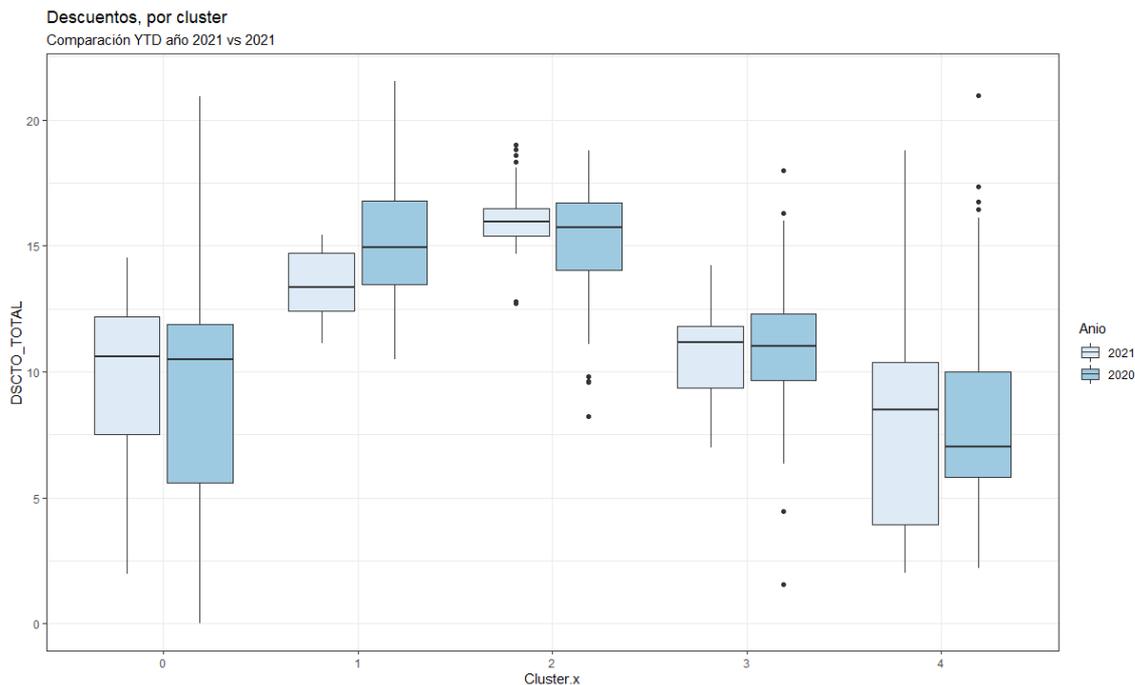


Figura 4.9 Comparación de descuentos otorgados en los años 2020 y 2021 para el período de julio a diciembre

A continuación, se detallan tres beneficios relevantes del presente trabajo:

1. Se estima que al mes existen más de 25 consultas de políticas de descuentos para cambios de volúmenes en clientes. Estas consultas se demoran en responder cerca de 45 minutos de acuerdo a entrevistas levantadas. Así, con la calculadora de descuentos, se dispone un ahorro estimado de 18 horas de trabajo en la jornada laboral.
2. Dentro de otros beneficios, encontramos los pronósticos a medida de rangos, sin dejar de lado la estimación puntual compuesta por el pronóstico de la serie de ingreso y modelo de ajuste del descuento.
3. Al ofrecer un rango de precios sugerido, el cual no incluye valores que podrían ser considerados como valores de descuentos atípicos basado en transacciones anteriores, el uso de la calculadora promovería la selección de un valor adecuado, eliminando situaciones como, e.g., a quienes no cumplen su acuerdo mínimo de compra. La eliminación de este último caso, representaría un ingreso estimado de un 0.5% más de valor neto facturado.

CAPÍTULO 5

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

- Se identificaron 5 grupos principales en la segmentación de clientes, los cuales responden a la lógica del negocio. Por otro lado, la segunda segmentación permite entender la distribución de compra respecto a la familia de productos.
- El modelo ganador dentro del ajuste de los modelos de rangos de precios fue el SVM, donde las características que mejor explican la variabilidad de los descuentos son ingreso, familia de producto, año, mes y el clúster al que pertenece el cliente.
- Dentro de los modelos tradicionales, se observó un gran desempeño por parte del modelo de efectos mixtos, que si bien es superado en término de poder predictivo por algoritmos de Machine Learning, su estructura paramétrica multinivel, permite una explicación más clara de las características y sus efectos en el modelo.
- Si bien se realizó un estudio exhaustivo de series temporales para el pronóstico de ventas, se debe tener en cuenta que existe un efecto causado por el COVID, el cual no fue considerado en el modelo.
- La calculadora de precios, disponible a través de un aplicativo web, provee el análisis del presente trabajo en diferentes módulos, incluyendo el pronóstico del ingreso y el cálculo de intervalo de descuentos sugeridos.

Recomendaciones

- En un futuro, es recomendable agregar al data set el registro de las acciones tácticas para retención o crecimiento de los clientes para obtener rangos de precios más precisos.
- Se recomienda que, una vez que se cuente con los datos de los años 2021 y 2022 de forma completa, se puede ajustar la data real del 2020 para poder comparar la proyección y así cuantificar el efecto COVID.
- Actualizar el análisis de segmentación por familia de producto para todo el año 2021, donde otras características podrían ganar importancia en el modelo.

BIBLIOGRAFÍA

[1] Team MixTrategy. (2020, April 13). ¿Qué es la Industria 4.0 o cuarta revolución industrial? [Online]. Disponible en: www.mixtrategy.com/la-industria-4-0-cuarta-revolucion-industrial/

[2] Esther Camuñas. (2017, July 12). Estrategias de fijación de precios dinámicos basados en el cliente. UNIR. [Online]. Disponible en: <https://www.unir.net/empresa/revista/estrategias-de-fijacion-de-precios-dinamicos-basados-en-el-cliente/>

[3] Asesoría LogicBUS. (2021, March). La Industria 4.0 a través de sus características y cómo funciona. Sección informativa. [Online]. Disponible en: www.logicbus.com.mx/que-es-la-industria-4-0.php

[4] CHI Software. (2020, January 27). Machine learning in dynamic pricing: CHI Software's experience CHI Software. [Online]. Disponible en: <https://chisoftware.medium.com/machine-learning-in-dynamic-pricing-chi-softwares-experience-ddcd4a322279>

[5] Y. Narahari, C. V. L. Raju, K. Ravikumar & Sourabh Shah. *Dynamic pricing models for electronic business*. Sadhana 30, 231–256 (2005).

[6] Hemant Warudkar. (2019, December 17). Developing machine learning models for dynamic pricing. [Online]. Disponible en: <https://expressanalytics.com/blog/developing-machine-learning-models-for-dynamic-pricing/>

[7] Gupta, R., & Pathak, C. "A machine learning framework for predicting purchase by online customers based on dynamic pricing". *Procedia Computer Science*, 36, 599-605. 2014

- [8] Alexandr Galkin. (August, 2017 24). Strategic Pricing in Retail with Machine Learning Cube.js. [Online]. Disponible en: <https://blog.statsbot.co/strategic-pricing-d1adcc2e0fd6>
- [9] Gan, L., Wang, H., & Yang, Z. "Machine learning solutions to challenges in finance: An application to the pricing of financial products". *Technological Forecasting and Social Change*, 153, 119928. 2020
- [10] Rafiei, M. H., & Adeli, H. "A novel machine learning model for estimation of sale prices of real estate units". *Journal of Construction Engineering and Management*, 142(2), 04015066. 2016
- [11] Koolen, D., Sadat-Razavi, N., & Ketter, W. "Machine learning for identifying demand patterns of home energy management systems with dynamic electricity pricing". *Applied Sciences*, 7(11), 1160. 2017
- [12] C. F. Ivaşcu. "Option pricing using Machine Learning". *Expert Systems with Applications*, vol. 163, pp. 0957-4174, January 2021
- [13] Redacción (Economía). (2021, March 12). VitaPRO ha invertido más de \$ 35 millones en ampliar capacidad de plantas en Ecuador y Perú. [Online]. Disponible en: <https://www.eluniverso.com/noticias/economia/vitapro-ha-invertido-mas-de-35-millones-en-ampliar-capacidad-de-plantas-en-ecuador-y-peru-nota/>
- [14] AliCORP. (2021, March 16). Nosotros Somos AliCORP. [Online]. Disponible en: www.alicorp.com.pe/ecu/es/nosotros/
- [15] NICOVITA. (2021, March 15). Somos Nicovita, una Marca de VitaPRO. [Online]. Disponible en: <https://nicovita.com/somos-nicovita/>
- [16] A.A. Cournot, *Researches into the Mathematical Principles of the Theory of Wealth*. Translated in 1897 by N.T. Bacon, with a Bibliography of Mathematical Economics by Irving Fisher, The Macmillan Company, New York, 1838

[17] I. Fisher, Cournot and mathematical economics, *Quart. J. Econ.* 12 (2) (1898) 119–138

[18] Campbell, P. (2021, August 24). How does dynamic pricing work? examples, strategies, and models. ProfitWell. [Online]. Disponible en: <https://www.priceintelligently.com/blog/bid/198355/how-to-implement-a-dynamic-pricing-strategy-without-the-pr-backlash>

[19] Rategenie. (2021, August 28). The Perfect Price, Everyday. [Online]. Disponible en: <https://www.rategenie.io>

[20] Kaggle Inc. (2021, August 28). Brazilian E-Commerce Public Dataset by Olist. [Online]. Disponible en: <https://www.kaggle.com/olistbr/brazilian-ecommerce/>

[21] Kaggle Inc. (2020, April 17). A Detailed Regression Guide with House-pricing. [Online]. Disponible en: <https://www.kaggle.com/masumrumi/a-detailed-regression-guide-with-house-pricing>

[22] Kaggle Inc. (2021, August 28). Pricing Model. [Online]. Disponible en: <https://www.kaggle.com/sumit9/pricing-model/tasks>

[23] Project Jupyter. (2021, August 28). JupyterLab Documentation. [Online]. Disponible en: <https://jupyterlab.readthedocs.io/en/stable/>

[24] RStudio, PBC. (2021, August 29). How we think about Data Science. [Online]. Disponible en: <https://www.rstudio.com/>

[25] Microsoft Corporation. (2021, August 29). Getting Started. Visual Studio Code in Action. [Online]. Disponible en: <https://code.visualstudio.com/docs>

[26] Kinsta Inc. (2020, January 7). ¿Qué Es Apache Web Server? Una Mirada Básica a lo que Es y Cómo Funciona. [Online]. Disponible en: <https://kinsta.com/es/base-de-conocimiento/que-es-apache/>

[27] Microsoft Corporation. (2021, August 30). Prueba SQL Server [Online]. Disponible en: <https://www.microsoft.com/es-es/sql-server/sql-server-downloads>

[28] The PHP Group. (2021, August 30). What is PHP? [Online]. Disponible en: <https://www.php.net/manual/es/intro-what-is.php>

[29] Mozilla Foundation. (2021, August 30). What is JavaScript? [Online]. Disponible en: https://developer.mozilla.org/es/docs/Learn/JavaScript/First_steps/What_is_JavaScript

[30] Python Software Foundation. (2021, August 28). Python. [Online]. Disponible en: <https://www.python.org/>

[31] The R Foundation. (2021, August 31). The R Project for Statistical Computing. [Online]. Disponible en: <https://www.r-project.org/>

[32] Pandas. (2021, August 31). Pandas. [Online]. Disponible en: <https://pandas.pydata.org/>

[33] NumPy. (2021, August 31). The fundamental package for scientific computing with Python. [Online]. Disponible en: <https://numpy.org/>

[34] The Matplotlib development team. (2021, September 1). Matplotlib: Visualization with Python. [Online]. Disponible en: <https://matplotlib.org/>

[35] Michael Waskom. (2021, September 1). Seaborn: Statistical data visualization. [Online]. Disponible en: <https://seaborn.pydata.org/>

[36] The Matplotlib development team. (2021, September 1). Toolkits. [Online].
Disponibile en: https://matplotlib.org/2.2.2/mpl_toolkits/index.html

[41] Tidymodels development team. (2021, October 26). Tidymodels, responsive sites with Bootstrap. [Online]. Disponibile en: <https://www.tidymodels.org/>

[37] Anaconda Inc. (2021, September 1). Anaconda Individual Edition. By data scientists, for data scientists. [Online]. Disponibile en: <https://www.anaconda.com/products/individual>

[38] NumFOCUS. (2021, September 1). Sponsored Projects. [Online]. Disponibile en: <https://numfocus.org/sponsored-projects>

[39] RStudio, PBC, Appl. Note Package base version 4.1.0.

[40] Bootstrap. (2021, August 28). Build fast, responsive sites with Bootstrap. [Online]. Disponibile en: <https://getbootstrap.com/>

[42] Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.

[43] XGBoost development team. (2021, October 26). XGBoost, responsive sites with XGBoost. [Online]. Disponibile en: <https://xgboost.readthedocs.io/en/latest/>