



**FACULTAD DE POSTGRADO  
TRABAJO FINAL DE GRADUACIÓN**

**PROPUESTA DE IMPLEMENTACIÓN DE TÉCNICAS DE  
CLUSTERIZACIÓN PARA LA RECOMENDACIÓN  
PERSONALIZADA DE PRODUCTOS PARA UNA  
PLATAFORMA DE E-COMMERCE EN HONDURAS**

**SUSTENTADO POR:**

**KENETH MAURICIO PÉREZ BRIZUELA  
ANDY JAVIER REYES ERAZO**

**PREVIA INVESTIDURA AL TÍTULO DE**

**MÁSTER EN  
ANALÍTICA DE NEGOCIOS**

**TEGUCIGALPA, FRANCISCO MORAZÁN, HONDURAS, C.A.**

**ENERO, 2025**

**UNIVERSIDAD TECNOLÓGICA CENTROAMERICANA**

**UNITEC**

**FACULTAD DE POSTGRADO**

**AUTORIDADES UNIVERSITARIAS**

**PRESIDENTE EJECUTIVO /  
RECTORA  
ROSALPINA RODRÍGUEZ**

**SECRETARIO GENERAL /  
PRORRECTOR  
ROGER MARTÍNEZ MIRALDA**

**VICERRECTOR ACADÉMICO NACIONAL  
JAVIER ABRAHAM SALGADO LEZAMA**

**DECANA FACULTAD DE POSTGRADO  
ANA DEL CARMEN RETTALLY**

**PROPUESTA DE IMPLEMENTACIÓN DE TÉCNICAS DE  
CLUSTERIZACIÓN PARA LA RECOMENDACIÓN  
PERSONALIZADA DE PRODUCTOS PARA UNA  
PLATAFORMA DE E-COMMERCE EN HONDURAS**

**TRABAJO PRESENTADO EN CUMPLIMIENTO DE LOS  
REQUISITOS EXIGIDOS PARA OPTAR AL TÍTULO DE  
MÁSTER EN**

**ANÁLITICA DE NEGOCIOS**

**ASESOR**

**JOSÉ RODOLFO SORTO**

**MIEMBROS DE LA TERNA:**

**MARIO ALBERTO GALLO  
JESÚS RICARDO RODRÍGUEZ  
GERARDO LUJANO**

# **DERECHOS DE AUTOR**

© Copyright 2024  
Keneth Mauricio Pérez Brizuela  
Andy Javier Reyes Erazo

Todos los derechos son reservados.



## **FACULTAD DE POSTGRADO**

# **PROPUESTA DE IMPLEMENTACIÓN DE TÉCNICAS DE CLUSTERIZACIÓN PARA LA RECOMENDACIÓN PERSONALIZADA DE PRODUCTOS PARA UNA PLATAFORMA DE E-COMMERCE EN HONDURAS**

**Keneth Mauricio Pérez  
Andy Javier Reyes Erazo**

### **Resumen**

Esta tesis tiene como propósito mejorar la personalización de las recomendaciones de productos en la plataforma LynxPay en Honduras, donde el comercio electrónico está en crecimiento, pero enfrenta desafíos para personalizar la experiencia de compra. El objetivo es segmentar usuarios en grupos homogéneos mediante algoritmos de clusterización como K-Means y DBSCAN permitiendo una mejor implementación de técnicas de recomendación, basándose en su comportamiento. Para ello, se recopilaron y analizaron datos de la plataforma, se aplicaron los algoritmos y se evaluó su efectividad mediante métricas como la tasa de conversión y la retención de clientes. Se analizó la viabilidad económica considerando el costo-beneficio. Finalmente, se espera que los resultados de este estudio contribuyan a optimizar las estrategias de marketing y ventas en el comercio electrónico hondureño.

**Palabras claves: (Clusterización, comercio electrónico, recomendaciones personalizadas, LynxPay, Honduras.)**



**GRADUATE SCHOOL**

**PROPOSAL FOR THE IMPLEMENTATION OF CLUSTERING  
TECHNIQUES FOR THE PERSONALIZED  
RECOMMENDATION OF PRODUCTS FOR AN E-  
COMMERCE PLATFORM IN HONDURAS**

**Keneth Mauricio Pérez  
Andy Javier Reyes Erazo**

**Abstract**

This thesis aims to improve the personalization of product recommendations on the LynxPay platform in Honduras, where e-commerce is growing, but faces challenges to personalize the shopping experience. The objective is to segment users into homogeneous groups using clustering algorithms such as K-Means and DBSCAN, allowing a better implementation of recommendation techniques, based on their behavior. To do this, data from the platform was collected and analyzed, algorithms were applied, and its effectiveness was evaluated using metrics such as conversion rate and customer retention. Economic viability was analyzed considering cost-benefit. Finally, it is expected that the results of this study will contribute to optimizing marketing and sales strategies in Honduran e-commerce.

**Keywords: (Clustering, e-commerce, personalized recommendations, LynxPay, Honduras.)**

## **DEDICATORIA**

A nuestros queridos padres, este logro es el resultado de su amor incondicional, esfuerzo y constante apoyo. Nos enseñaron desde pequeños el valor del trabajo duro, la perseverancia y la integridad, brindándonos las herramientas para enfrentar los retos de la vida. Gracias por creer en nosotros y ser la guía en cada paso de nuestro camino. Su dedicación, sacrificios y palabras de aliento nos han permitido alcanzar esta meta. Sin su apoyo inquebrantable, este proyecto no habría sido posible. Les debemos no solo el éxito académico, sino también los principios y valores que nos han llevado a él. Con inmensa gratitud, les dedicamos esta tesis, nuestra mayor fuente de inspiración. Este logro es tanto suyo como nuestro.

## **AGRADECIMIENTO**

En primer lugar, agradecemos a Dios por la fortaleza y bendiciones que nos permitieron culminar esta etapa. A nuestros padres, nuestro mayor pilar, por su amor incondicional, paciencia y sacrificio; este trabajo es un reflejo de sus enseñanzas. Agradecemos a nuestro asesor, Ing. José Rodolfo Sorto, por su valiosa orientación y apoyo constante; su experiencia fue fundamental en este proceso. A nuestros compañeros de equipo, por el esfuerzo y dedicación para superar los desafíos. Finalmente, a nuestros amigos y seres queridos, por su apoyo y comprensión en los momentos difíciles. Su aliento nos impulsó a seguir adelante.

# ÍNDICE DE CONTENIDO

DEDICATORIA .....	ix
AGRADECIMIENTO .....	x
ÍNDICE DE CONTENIDO .....	xi
ÍNDICE DE FIGURAS.....	xvi
ÍNDICE DE TABLAS .....	xvii
CAPÍTULO I. PLANTEAMIENTO DE LA INVESTIGACIÓN .....	1
1.1 INTRODUCCIÓN .....	1
1.2 ANTECEDENTES DEL PROBLEMA .....	2
1.3 DEFINICIÓN DEL PROBLEMA .....	3
1.3.1 ENUNCIADO DEL PROBLEMA .....	3
1.3.2 FORMULACIÓN DEL PROBLEMA.....	4
1.3.3 PREGUNTAS DE INVESTIGACIÓN.....	5
1.4 OBJETIVOS DEL PROYECTO.....	6
1.4.1 OBJETIVO GENERAL.....	6
1.4.2 OBJETIVOS ESPECIFICOS.....	6
1.5 JUSTIFICACIÓN.....	7
CAPITULO II. MARCO TEÓRICO .....	9
2.1 ANÁLISIS DE LA SITUACIÓN ACTUAL .....	9
2.1.1 ANÁLISIS DEL MACROENTORNO.....	9
2.1.2 ANÁLISIS DEL MICROENTORNO .....	12
2.1.3 ANÁLISIS INTERNO.....	15
2.2 CONCEPTUALIZACIÓN.....	18
2.2.1 VARIABLE DEPENDIENTE .....	18
2.2.2 VARIABLES INDEPENDIENTES .....	19
2.2.3 COMERCIO ELECTRÓNICO.....	20
2.2.4 TASA DE CONVERSIÓN EN COMERCIO ELECTRÓNICO.....	20
2.2.5 SEGMENTACIÓN DE CLIENTES.....	21
2.2.6 APRENDIZAJE NO-SUPERVISADO .....	21

2.2.7	CLUSTERING.....	22
2.2.8	K-MEANS .....	22
2.2.9	DBSCAN .....	23
2.2.10	PERSONALIZACIÓN EN EL COMERCIO ELECTRONICO .....	23
2.2.11	SISTEMAS DE RECOMENDACIÓN.....	24
2.2.12	FILTRADO BASADO EN CONTENIDO.....	24
2.2.13	FILTRADO COLABORATIVO .....	25
2.2.14	ALGORITMOS HÍBRIDOS .....	25
2.2.15	CARRITO DE COMPRA.....	25
2.2.16	RECOMENDADOR DE PRODUCTOS.....	26
2.2.17	INICIO EN FRIO.....	26
2.2.18	ARQUITECTURA MULTI-INQUILINO.....	26
2.2.19	SCIKIT-LEARN (SKLEARN).....	26
2.3	TEORÍAS DE SUSTENTO .....	27
2.3.1	BASES TEÓRICAS.....	27
2.3.2	TEORÍA DE LA SEGMENTACIÓN.....	27
2.3.3	TEORÍA DE LA COLABORACION DE FILTRADO .....	28
2.3.4	TEORÍA DE LA MINERÍA DE DATOS .....	31
2.3.5	INSTRUMENTOS UTILIZADOS POR LOS INVESTIGADORES .....	32
2.4	MARCO LEGAL .....	34
2.4.1	MARCO LEGAL INTERNACIONAL .....	34
2.4.2	PROTECCIÓN DE DATOS EN HONDURAS .....	37
2.4.3	MARCO LEGAL NACIONAL.....	38
CAPÍTULO III. METODOLOGÍA .....		42
3.1	CONGRUENCIA METODOLÓGICA.....	42
3.1.1	MATRIZ METODOLÓGICA .....	42
3.1.2	ESQUEMA DE VARIABLES DE ESTUDIO.....	44
3.1.3	OPERACIONALIZACIÓN DE LAS VARIABLES.....	45
3.1.4	HIPÓTESIS.....	47
3.2	ENFOQUE Y MÉTODOS .....	47
3.2.1	ENFOQUE.....	48

3.2.2	ALCANCE.....	48
3.2.3	DISEÑO .....	49
3.2.4	MÉTODOS .....	49
3.2.5	INSTRUMENTOS.....	50
3.3	DISEÑO DE LA INVESTIGACIÓN .....	50
3.3.1	POBLACIÓN.....	51
3.3.2	MUESTRA .....	54
3.3.3	TÉCNICA DE MUESTREO .....	56
3.3.4	CORTE DE LA INVESTIGACIÓN.....	58
3.4	TÉCNICAS, HERRAMIENTAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS .....	60
3.4.1	TÉCNICAS.....	60
3.4.2	MINERÍA DE DATOS.....	60
3.4.3	FASES DEL PROCESO DE LA MINERÍA DE DATOS .....	61
3.4.4	MODELOS DE PREDICCIÓN.....	61
3.4.5	HERRAMIENTAS .....	62
3.4.5.1	PUTTY .....	62
3.4.5.2	DBEAVER.....	62
3.4.5.3	PYTHON.....	62
3.4.5.4	VSCODE.....	63
3.4.5.5	SKLEARN .....	63
3.4.5.6	MATPLOTLIB.....	63
3.4.6	PROCEDIMIENTO .....	63
3.4.7	INSTRUMENTOS UTILIZADOS.....	65
3.4.7.1	SCRIPTS EN PYTHON PARA EL PROCESO ETL (EXTRACCIÓN, TRANSFORMACIÓN Y CARGA).....	65
3.4.7.2	ALGORITMOS DE AGRUPACIÓN Y MODELOS DE APRENDIZAJE AUTOMÁTICO (K-MEANS Y DBSCAN).....	66
3.4.7.3	CONSULTAS SQL PARA LA EXTRACCIÓN DE DATOS .....	67
3.4.7.4	VISUALIZACIONES EN MATPLOTLIB .....	68
3.4.7.5	MODELOS DE RECOMENDACIÓN BASADOS EN FILTRACIÓN	

COLABORATIVA .....	69
3.5 FUENTES DE INFORMACIÓN.....	70
3.5.1 FUENTES PRIMARIAS .....	70
3.5.2 FUENTES SECUNDARIAS .....	70
CAPÍTULO IV. RESULTADOS Y ANÁLISIS .....	72
4.1 INFORME DE PROCESO DE RECOLECCIÓN DE DATOS.....	72
4.2 RESULTADOS Y ANÁLISIS DE LAS TÉCNICAS APLICADAS .....	72
4.2.1 SITUACIÓN ACTUAL.....	73
4.2.2 FACTORES SOCIODEMOGRÁFICOS.....	77
4.2.3 ANÁLISIS RFM.....	80
4.2.4 MODELOS PREDICTIVOS .....	84
4.2.5 SISTEMA DE RECOMENDACIÓN .....	96
4.2.6 ANÁLISIS COSTO-BENEFICIO DEL ALGORITMO HÍBRIDO DE RECOMENDACIÓN .....	99
4.3 COMPROBACIÓN DE HIPÓTESIS .....	100
CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES.....	101
5.1 CONCLUSIONES .....	101
5.2 RECOMENDACIONES .....	102
CAPITULO VI. APLICABILIDAD.....	103
6.1 NOMBRE DE LA PROPUESTA .....	103
6.2 JUSTIFICACIÓN DE LA PROPUESTA .....	103
6.3 ALCANCE DE LA PROPUESTA .....	103
6.3.1 OBJETIVO GENERAL DE LA PROPUESTA .....	104
6.3.2 OBJETIVOS ESPECÍFICOS DE LA PROPUESTA .....	104
6.4 DESCRIPCIÓN Y DESARROLLO .....	104
6.4.1 DESCRIPCIÓN .....	104
6.4.2 DESARROLLO .....	104
6.5 MEDIDAS DE CONTROL. ....	112
6.6 CRONOGRAMA DE IMPLEMENTACIÓN Y PRESUPUESTO. ....	114
6.6.1 ACTIVIDADES DE IMPLEMENTACIÓN DE PROPUESTA .....	114
6.6.2 COSTOS DE LA IMPLEMENTACIÓN DE LA PROPUESTA .....	115

6.7 CONCORANCIA DE LOS SEGMENTOS DE LA TESIS CON LA PROPUESTA.	117
BIBLIOGRAFÍA .....	122
ANEXOS .....	124
Anexo 1 .....	124
Anexo 2 .....	125
Anexo 3 .....	125
Anexo 4 .....	126
Anexo 5 .....	126

## ÍNDICE DE FIGURAS

Figura 1 Esquema de estructura multi-inquilino Fuente: Elaboración Propia .....	18
Figura 2. Datos de entrenamiento sin categorizar para aprendizaje automático. ....	21
Figura 3. Agrupación de información. ....	22
Figura 4. Resumen de esquema de las variables de estudio. ....	44
Figura 5 Resumen de esquema de las variables de estudio .....	48
Figura 6. Número de ítems adquiridos por inquilino .....	73
Figura 7 Histórico de ítems vendidos por inquilino.....	74
Figura 8 Ventas totales inquilino Ferretería .....	76
Figura 9 Ventas totales inquilino Maquillaje.....	77
Figura 10 Compras por departamento en inquilino Ferretería .....	78
Figura 11 Compras por departamento inquilino Maquillaje .....	79
Figura 12 Distribución de valores faltantes en campo departamento .....	80
Figura 13 Histograma de recencia inquilino Ferretería .....	81
Figura 14 Histograma de recencia inquilino Maquillaje .....	81
Figura 15 Frecuencia de compra .....	82
Figura 16 Distribución de compras inquilino Maquillaje .....	83
Figura 17 Distribución de compras inquilino Ferretería .....	84
Figura 18 Código para generar características (Recency).....	85
Figura 19 Código para generar características (Frecuencia) .....	85
Figura 20 Código para generar características de tendencia de compra .....	86
Figura 21 Código para verificar valores atípicos.....	86
Figura 22 Matriz de correlación de variables en inquilino Ferretería.....	87
Figura 23 Resumen de rendimiento de modelo Kmeans con diferente configuración de centroides.....	88
Figura 24 Código elaboración de clúster en datos .....	89
Figura 25 PCA para visualización en tres dimensiones.....	90
Figura 26 Distribución de clientes en cada clúster .....	91
Figura 27 Análisis clúster 0.....	92
Figura 28 Análisis Clúster 1 .....	93
Figura 29 Análisis clúster 2.....	94
Figura 30 Puntuación Silueta entre parámetros .....	94
Figura 31 Clústeres obtenidos utilizando algoritmo DBSCAN .....	95
Figura 32 Precisión de sistema de recomendación híbrido inquilino Ferretería .....	96
Figura 33 Precisión de sistema de recomendación híbrido inquilino Maquillaje .....	97
Figura 34 Diagrama de Gantt de actividades de implementación de propuesta .....	115

## ÍNDICE DE TABLAS

Tabla 1. Matriz de congruencia metodológica.....	43
Tabla 2. Tabla de Operacionalización de variables independientes.....	45
Continuación Tabla 3. Tabla de Operacionalización de variables independientes.....	46
Tabla 4. Tabla de actividades de la investigación.....	50
Continuación Tabla 5. Tabla de actividades de la investigación.....	51
Tabla 6 Esquema de tabla transaccional.....	75
Tabla 7 Puntuación de métricas de evaluación.....	91
Tabla 8 Métricas de puntuación DBSCAN.....	95
Tabla 9 Recomendaciones a usuario 3467, inquilino Maquillaje.....	97
Tabla 10 Recomendaciones a usuario 3501, inquilino maquillaje.....	98
Tabla 11 Recomendaciones usuario 3120, inquilino ferretería.....	98
Tabla 12 Plan de capacitación de propuesta.....	111
Tabla 13 Actividades implementación de propuesta.....	114
Tabla 14 Costos de Propuesta.....	115
Tabla 15 Matriz de Concordancia.....	117

# **CAPÍTULO I. PLANTEAMIENTO DE LA INVESTIGACIÓN**

En este capítulo se establece que estructura tiene la investigación, se describe el problema y antecedentes para facilitar su análisis, prosiguiendo con las preguntas de investigación y objetivos generales y específicos que definen las metas del presente estudio, incluyendo finalmente la justificación del problema que esclarece la importancia del proyecto a continuación.

## **1.1 INTRODUCCIÓN**

LynxPay es una plataforma creada por LynxLabs empresa nacional hondureña, enfocada en la creación de herramientas de carácter tecnológico cuenta con más de 7 años en la industria, su plataforma busca ayudar a los comerciantes locales, desde tiendas de productos textiles hasta establecimientos gastronómicos. Ofreciendo la posibilidad de adaptarse a las necesidades de sus clientes.

La plataforma LynxPay cuenta con clientes que utilizan su plataforma y generan datos que podrían ser relevantes, generando un gran interés en su extracción y evaluación para encontrar mejoras posibles en sus procesos y especialmente en el tema tratado por esta investigación en una mejor recomendación de productos para la plataforma.

En la era digital actual, el comercio electrónico se ha convertido en un pilar fundamental para la expansión y el éxito de numerosas empresas alrededor del mundo. Plataformas como Shopify y WordPress han facilitado enormemente el proceso de establecer una presencia en línea, permitiendo a los negocios alcanzar a un público global de manera eficiente. Sin embargo, el simple hecho de contar con una tienda en línea no garantiza el éxito, ya que la clave radica en comprender a fondo el comportamiento y las necesidades de los clientes, y ofrecerles una experiencia de compra personalizada. La personalización en el comercio electrónico se ha convertido en un factor crucial para aumentar la satisfacción del cliente, la fidelización y las ventas.

Duy-Nghia Nguyen (2024) encontró que las recomendaciones personalizadas pueden incrementar significativamente el Valor Promedio por Orden (AOV) hasta en un 369%. Este indicador, definido por Chaffey (2016) como el valor promedio por transacción en un canal específico, es crucial para evaluar la rentabilidad de las estrategias de comercio electrónico. En este contexto, LynxLabs ha desarrollado una plataforma innovadora denominada LynxPay, la cual no solo ofrece funcionalidades de comercio electrónico, sino que también recopila una gran

cantidad de datos valiosos sobre los usuarios y sus interacciones con las tiendas alojadas en dicha plataforma. Estos datos incluyen información detallada sobre los carritos de compra y productos favoritos, como otros.

En este contexto, la clusterización de clientes se presenta como una técnica eficaz para la segmentación y la personalización. Esta técnica consiste en agrupar clientes con características similares en base a su comportamiento de compra, preferencias y datos demográficos. Se considera que la implementación de un sistema de recomendación basado en la clusterización podría tener un potencial de mejora en métricas relacionadas con el consumidor y el valor que encuentre en la plataforma.

El objetivo principal de este proyecto de investigación es validar la implementación de un sistema de clusterización que permita identificar grupos de clientes con características y comportamientos similares en cada una de las tiendas alojadas en la plataforma LynxPay, con el fin de ofrecer recomendaciones de productos personalizadas. Para ello, se analizarán los datos de la base de datos de LynxPay y se aplicarán modelos de clusterización como K-MEANS, DBSCAN entre otros para segmentar a los clientes de cada tienda. La investigación se llevó a cabo mediante un enfoque cuantitativo.

Se espera que, con esta investigación, LynxPay pueda validar si la implementación de este sistema de recomendación basado en clusterización es sustentable en su retorno de inversión y contribuir a mejorar la experiencia de usuario en dicha plataforma.

## **1.2 ANTECEDENTES DEL PROBLEMA**

Estudios recientes como el realizado por Brilliance (2024) han demostrado el impacto significativo de las recomendaciones personalizadas de productos en el aumento de las ventas y los ingresos en plataformas de comercio electrónico. Serrano (2023) afirma que "las recomendaciones de productos representan hasta el 31% de los ingresos de los sitios de comercio electrónico".

En marzo del 2024 se realizó un estudio con título "A personalized product recommendation model in e-commerce based on retrieval strategy", su principal objetivo fue utilizar diferentes modelos de machine learning para implementar un sistema de recomendación aplicando no solo algoritmos de filtrado colectivo y de popularidad, si no, también potenciando su

rendimiento con ciertos modelos. Este estudio concluyó que, a pesar de utilizar redes neuronales para realizar esta actividad, el modelo LightGBM supera las redes neuronales y a un menor costo computacional, haciendo énfasis en la importancia de la selección del modelo ideal para ciertas actividades. (Duy-Nghia Nguyen, 2024)

A su vez en el estudio llamado “Research on Segmenting E-Commerce Customer through an Improved K-Medoids Clustering Algorithm” como principal objetivo establecieron el uso de modelos de clusterización para segmentación de clientes dentro de una plataforma de comercio electrónico con el fin de categorizar los clientes, para su posterior identificación con fines de selección de estrategias de mercadeo. Utilizando valores como el comportamiento de compra y añadiendo algoritmos especializados de segmentación, pudieron mejorar la categorización de clientes de modelos anteriormente utilizados, pudiendo categorizar 37,376 clientes. (Zengyuan Wu, 2022)

Y, por último, en el artículo “Review of Clustering-Based Recommender Systems” el principal objetivo fue realizar un análisis de modelos de agrupación o clusterización aplicado a sistemas de recomendación. Haciendo énfasis, en los principales algoritmos utilizados como los de filtrado colaborativo (CF), basados en el contenido (CBF).

Esto concluyó con que la implementación de aproximaciones híbridas de las implementaciones puede resultar más efectivo en el entorno aplicado por los mismos investigadores, sin embargo, hacen hincapié en la imprecisa mejora de métodos que utilizan aprendizaje automático en modelos de recomendación.

Por último, hacen notar un problema en el ámbito de negocio, sobre como tomar en cuenta los cambios constantes en las preferencias de los usuarios, mencionando que los sistemas de recomendación tradicionales no toman en cuenta el factor temporal, por lo cual cambios en el comportamiento de compra del usuario podría repercutir al final en peores recomendaciones. (Beregovskaya Irina, 2021)

## **1.3 DEFINICIÓN DEL PROBLEMA**

### **1.3.1 ENUNCIADO DEL PROBLEMA**

El comercio electrónico en Honduras se encuentra en una etapa de crecimiento acelerado, impulsado por la digitalización y la adopción de nuevas tecnologías. Plataformas como LynxPay,

desarrollada por LynxLabs, facilitan a las empresas la creación y gestión de tiendas en línea. Sin embargo, a pesar de las funcionalidades que ofrecen estas plataformas, persisten desafíos significativos en la optimización de la experiencia del usuario y la personalización de las ofertas.

La falta de herramientas para segmentar clientes, identificar patrones de comportamiento y ofrecer recomendaciones personalizadas limita la capacidad de las empresas para brindar experiencias de compra satisfactorias y aprovechar al máximo el potencial de la plataforma. Esto se traduce en bajas tasas de conversión, altas tasas de abandono del carrito de compra y pérdida de oportunidades de venta.

Estudios recientes (Duy-Nghia Nguyen, 2024) y (Bariliance, 2024) demuestran el impacto positivo de las recomendaciones personalizadas en el aumento de las ventas, los ingresos y la fidelización de clientes en el comercio electrónico. La implementación de un sistema de recomendación basado en la clusterización de clientes, que agrupa usuarios con características similares, se presenta como una solución prometedora para abordar este problema.

### 1.3.2 FORMULACIÓN DEL PROBLEMA

El uso de recomendaciones personalizadas ha sido un hito importante para mejorar la experiencia de usuario en plataformas web, buscar que el usuario obtenga datos personalizados mejora muchas métricas tales como AOV para comercios electrónicos que mide el valor que el usuario promedio de la orden se ven incrementadas potencialmente en más de 369% (Duy-Nghia Nguyen, 2024)

Desde los inicios del comercio electrónico, la personalización ha sido crucial para aumentar el valor del usuario y, por consiguiente, el retorno de las plataformas. Un ejemplo es Amazon, donde las recomendaciones personalizadas representan el 35% de las ventas (Linden G, 2003). Si bien LynxPay posee una gran cantidad de datos sobre el comportamiento de sus usuarios, actualmente carece de un método estructurado para personalizar las recomendaciones de productos. Más allá de recomendaciones rudimentarias basadas en categorías similares, existe la oportunidad de utilizar algoritmos de aprendizaje automático para impulsar recomendaciones más efectivas basadas en los comportamientos de compra individuales.

Según el medio Digital Commerce (2024) la media en latino américa de vendedores con plataformas de trecientas plataformas de comercio electrónico fue de \$89 (ochenta y nueve dólares

americanos), esto contrasta con datos actuales de LynxPay que rondan el número a \$40 (cuarenta dólares americanos) o aproximadamente a conversión el día 17 de nov. de 24 a L1,013.00 (mil trece lempiras), existiendo una oportunidad de aplicar técnicas de recomendación para aumentar el valor que podemos obtener de los clientes.

Esta presente investigación pretende implementar un modelo, que utilizando clusterización y técnicas de recomendación, permita a LynxPay maximizar el valor de sus clientes en la plataforma, utilizando técnicas de minería de datos que a su vez mejoren a los algoritmos de aprendizaje automático.

Con esta esperada resolución se busca ayudar a la plataforma LynxPay se beneficiará en términos de efectividad y mejora en KPIs relacionados a la estadía del usuario en la plataforma maximizando su valor.

Basado en todo la anterior, se formula el problema de investigación con la siguiente pregunta: ¿Cómo la implementación de un sistema de recomendación basado en clusterización en la plataforma LynxPay puede mejorar la experiencia de usuario y aumentar la precisión de las recomendaciones contexto del comercio electrónico en Honduras?

### 1.3.3 PREGUNTAS DE INVESTIGACIÓN

- a. ¿Cuáles son las variables clave (valores demográficos, geográficos, historial de transacciones, productos en el carrito, etc.) que mejor predicen el comportamiento de compra de los usuarios en LynxPay?
- b. ¿Qué algoritmo de clusterización (K-means, DBSCAN u otros) ofrece la mejor segmentación de clientes en LynxPay, considerando la naturaleza de los datos y las variables de comportamiento clave?
- c. ¿Determinar la combinación de modelo de clustering y técnica de recomendación que maximice la precisión (medida con Precision@10) en la recomendación de productos a usuarios de la plataforma?
- d. ¿Cuál es el costo-beneficio de implementar modelos de clusterización en LynxPay, considerando los recursos computacionales, el tiempo de procesamiento y el potencial aumento en las ventas?
- e. ¿Qué estrategias de mejora y optimización se pueden implementar en LynxPay, con base en

los resultados de la validación de los modelos de clusterización y recomendación, para maximizar la personalización y la experiencia del usuario?

## **1.4 OBJETIVOS DEL PROYECTO**

### **1.4.1 OBJETIVO GENERAL**

El objetivo principal de este proyecto de tesis es validar la viabilidad y efectividad de las técnicas de clusterización como herramienta para la segmentación de clientes en la plataforma de comercio electrónico LynxPay, con el propósito de generar recomendaciones de productos personalizadas que optimicen la experiencia del usuario, aumenten las tasas de conversión y mejoren la competitividad de las tiendas en línea en Honduras.

### **1.4.2 OBJETIVOS ESPECIFICOS**

1. Recopilar y analizar de forma exhaustiva los datos disponibles en la plataforma LynxPay, tales como transacciones realizadas y productos añadidos al carrito, con el objetivo de identificar las principales variables que influyen el comportamiento de compra de los usuarios.
2. Aplicar diferentes algoritmos de clusterización, como K-MEANS y DBSCAN, con el propósito de segmentar la base de clientes en grupos homogéneos basados en variables clave de comportamiento, como la frecuencia de compra, el tipo de productos adquiridos y las preferencias de navegación.
3. Desarrollar y validar modelos de recomendación personalizados para los segmentos identificados, analizando el impacto de estos modelos en las métricas clave de rendimiento, tales como la precisión de las recomendaciones.
4. Explorar la viabilidad económica de implementar modelos de clusterización en LynxPay mediante un análisis de costo-beneficio, tomando en cuenta los recursos computacionales, el tiempo de procesamiento de los datos y el impacto esperado en las ventas de las tiendas participantes.
5. Proponer estrategias de mejora basadas en los resultados obtenidos durante la validación de los modelos, sugiriendo posibles optimizaciones para la implementación futura de los sistemas de recomendación en LynxPay, con miras a incrementar la personalización y mejorar la experiencia del usuario.

## 1.5 JUSTIFICACIÓN

En un contexto global donde el comercio electrónico ha transformado las dinámicas tradicionales de compra y venta, la personalización se ha convertido en un factor decisivo para la satisfacción del cliente y la maximización de las ventas. Estudios recientes Flores Bautista, Sánchez Ayala, Jiménez DeLucio, & Rojo Cisneros (2023), sugieren que las recomendaciones personalizadas no solo aumentan la fidelización de los clientes, sino que también tienen el potencial de incrementar significativamente las ventas y el valor promedio por transacción.

En el caso de Honduras, el comercio electrónico ha crecido de forma acelerada en los últimos años Berrios Lagos, Osorio Molina, & Medina Casco (2022), con plataformas como LynxPay que ofrecen una solución integral para que las empresas que gestionen sus tiendas en línea con ellos. Sin embargo, la falta de personalización en la experiencia del usuario sigue siendo un reto crítico para muchas de estas tiendas, lo que se refleja en tasas de conversión relativamente bajas y un elevado abandono de carritos de compra.

La importancia de esta investigación radica en su contribución al desarrollo de soluciones tecnológicas adaptadas al contexto del comercio electrónico en Honduras, un mercado en crecimiento que necesita optimizar su competitividad. Mediante la validación de técnicas de clusterización, este estudio permitirá a LynxPay y a las empresas que utilizan su plataforma identificar patrones de comportamiento en sus clientes, segmentar a los usuarios en grupos más precisos y ofrecer recomendaciones de productos que se alineen mejor con sus preferencias y necesidades.

El impacto práctico de esta investigación es significativo, ya que los resultados esperados incluyen una mejora en las tasas de conversión, un incremento en las ventas y una mayor retención de clientes. Además, se generarán insumos valiosos para las empresas que operan en el comercio electrónico en Honduras en la plataforma LynxPay, permitiendo que estas optimicen sus estrategias de marketing y ventas en función de datos empíricos y análisis avanzados.

Desde una perspectiva teórica, esta investigación aportará nuevo conocimiento al campo de la personalización en el comercio electrónico, validando el uso de modelos de clusterización en plataformas emergentes. Además, la investigación podría servir como referencia para futuros estudios que busquen optimizar la personalización en contextos similares, contribuyendo al desarrollo de tecnologías más avanzadas de recomendación de productos.

Por último, el valor metodológico de este proyecto reside en la implementación de un enfoque cuantitativo para evaluar el desempeño de los algoritmos de recomendación y el impacto económico de su implementación. Esta aproximación permitirá ofrecer recomendaciones basadas en resultados medibles y empíricos, lo que hará que los resultados del estudio sean aplicables y escalables a nivel empresarial.

## CAPITULO II. MARCO TEÓRICO

### 2.1 ANÁLISIS DE LA SITUACIÓN ACTUAL

#### 2.1.1 ANÁLISIS DEL MACROENTORNO

##### 2.1.1.1 TENDENCIAS GLOBALES EN SISTEMAS DE RECOMENDACIÓN

En los últimos años, el volumen de información generado y disponible ha experimentado un crecimiento exponencial, impulsado en gran medida por la masificación de los teléfonos inteligentes y el desarrollo de tecnologías como los sensores y dispositivos conectados, comúnmente conocidos como Internet de las Cosas (IOT, por sus siglas en inglés). Este avance ha dado lugar a un fenómeno sin precedentes, en el que la cantidad de datos generados supera ampliamente las capacidades de procesamiento y análisis tradicionales. (Badrul Sarwar, 2021). En este contexto, surge el concepto de Big Data, que integra un conjunto de tecnologías y metodologías dedicadas a la recolección, almacenamiento y análisis de grandes volúmenes de información en tiempo real. Su propósito principal es interpretar y extraer conocimiento útil de estos datos masivos, transformando de manera significativa todos los sectores de la economía y modificando profundamente cómo las empresas gestionan sus recursos, interactúan con sus clientes y toman decisiones estratégicas.

El impacto del Big Data en el ámbito empresarial ha sido particularmente notable. Para muchas organizaciones, la capacidad de manejar y analizar grandes conjuntos de datos ha representado una ventaja competitiva sin precedentes. Esta tecnología permite identificar patrones, anticipar tendencias y optimizar los procesos internos, lo que se traduce en una mayor productividad, una toma de decisiones más precisa y la minimización de riesgos operativos. Este enfoque de análisis ha generado valor económico tangible, posicionando a las empresas que aprovechan el Big Data en un lugar de liderazgo dentro de sus respectivos sectores. (Haefner, Wincent, Parida, & Gassmann, 2021).

Tendencias Globales en Sistemas de Recomendación y Comercio Electrónico El comercio electrónico ha experimentado un auge a nivel mundial, con un crecimiento del 12-15% anual en las ventas totales, según E-Commerce Statistics of 2024, Forbes (2024). Las tendencias indican que el mercado global alcanzará los \$7.4 billones de dólares en 2025, con un impacto significativo de la personalización y las recomendaciones basadas en datos. Empresas líderes como Amazon,

Alibaba y Shopify han adoptado sistemas avanzados de inteligencia artificial (IA) para anticipar comportamientos de compra y mejorar sus tasas de conversión.

En el ámbito de los sistemas de recomendación, el uso de modelos híbridos y redes neuronales profundas ha ganado popularidad. Estas tecnologías permiten superar problemas tradicionales como el inicio en frío y el aislamiento de datos, maximizando la precisión en las recomendaciones. Además, la incorporación de análisis de comportamiento en tiempo real y personalización dinámica ha permitido una mayor retención de clientes.

Comparativa Regional En Asia, Alibaba utiliza técnicas avanzadas de clusterización para personalizar experiencias de compra, lo que ha resultado en un aumento del 20% en la retención de clientes. Por otro lado, en Europa, plataformas como Zalando han integrado IA para segmentar clientes según su estilo y preferencias, aumentando el valor promedio por orden (AOV) en un 15%.

Estas estrategias son relevantes para replicar en regiones emergentes como Centroamérica, donde el comercio electrónico aún está en desarrollo, pero con un crecimiento significativo en países como Costa Rica y Guatemala.

#### 2.1.1.2 TENDENCIAS IMPACTO DE SEGMENTACIÓN EN COMERCIO ELECTRÓNICO INTERNACIONAL

Un caso especialmente relevante del impacto del Big Data es el sector del comercio electrónico, donde empresas líderes como Amazon y Alibaba han experimentado un cambio fundamental en sus modelos de negocio mediante la implementación de estrategias basadas en Big Data. (Valencia Arias, y otros, 2024). Una de las aplicaciones más destacadas de esta tecnología en el comercio electrónico es el uso de modelos de segmentación avanzados, que se implementan con técnicas de clusterización (o agrupamiento) para generar recomendaciones personalizadas. La clusterización permite a estas empresas segmentar a sus usuarios en grupos con características y comportamientos similares, basándose en variables como el historial de compras y preferencias de productos.

Esta técnica, esencial para el desarrollo de sistemas de recomendación, permite a empresas como Amazon y Alibaba ofrecer sugerencias de productos altamente personalizadas que se adaptan de manera precisa a los intereses y necesidades de cada usuario. Al aplicar algoritmos de machine learning y clustering, las empresas pueden analizar los datos de sus clientes para predecir

qué productos podrían interesar a cada uno, optimizando así la experiencia de compra y aumentando las probabilidades de conversión y fidelización. En el caso de Amazon, por ejemplo, la segmentación de usuarios basada en clustering se traduce en una experiencia de usuario más personalizada y en un catálogo de recomendaciones ajustado al perfil de cada cliente. Alibaba, por su parte, utiliza técnicas similares para entender y anticipar las demandas de su amplia base de usuarios, lo cual le permite ajustar sus estrategias de marketing y su oferta de productos de acuerdo con las preferencias específicas de cada segmento de clientes.

La aplicación de estos modelos de segmentación y técnicas de clusterización ha permitido a las empresas de comercio electrónico no solo optimizar la satisfacción del cliente mediante recomendaciones personalizadas, sino también mejorar el rendimiento de cada uno de los eslabones de su cadena de valor. Desde la gestión de proveedores hasta la personalización de la experiencia del cliente en tiempo real, el uso estratégico de Big Data, en combinación con técnicas de machine learning y clusterización, representa un recurso fundamental en la obtención de ventajas competitivas. (Salah, Hamada, Taj Eddin, & Kasem, 2023). En este sentido, tanto Amazon como Alibaba son ejemplos de éxito en el aprovechamiento de la información para optimizar sus operaciones, mejorar la eficiencia y rentabilidad, y consolidarse como líderes en un mercado cada vez más competitivo. Estos casos demuestran cómo el Big Data y los modelos de segmentación avanzados son esenciales para impulsar la innovación y fortalecer la relación con el cliente en el ámbito del comercio electrónico.

#### 2.1.1.3 MEJORES PRÁCTICAS EN LA PERSONALIZACIÓN DE EXPERIENCIA DE USUARIO

Una de las aplicaciones más destacadas de esta tecnología en el comercio electrónico es el uso de modelos de segmentación avanzados, entre los que destaca el método RFM (Recency, Frequency, Monetary) junto con técnicas de clusterización para generar recomendaciones personalizadas. El método RFM permite clasificar y segmentar a los clientes en función de tres criterios clave:

- Recency (Recencia): cuánto tiempo ha pasado desde la última compra del cliente.
- Frequency (Frecuencia): la cantidad de veces que el cliente ha realizado una compra en un periodo determinado.

- Monetary (Monetario): el valor total gastado por el cliente en la empresa.

Estos tres criterios ofrecen una visión integral del comportamiento del cliente, y permiten a las empresas identificar y priorizar a los segmentos con mayor potencial. En el comercio electrónico, el análisis RFM, combinado con algoritmos de machine learning y técnicas de clusterización, permite que empresas de comercio electrónico generen recomendaciones personalizadas al segmentar a sus usuarios en grupos con características y comportamientos similares, basados en la frecuencia de sus compras, la cantidad de dinero que gastan y el tiempo que pasa entre cada transacción (Smaili & Hachimi, 2023).

## 2.1.2 ANÁLISIS DEL MICROENTORNO

### 2.1.2.1 COMERCIO ELECTRÓNICO EN LATINOAMÉRICA

En los últimos años, el comercio electrónico en Latinoamérica ha experimentado un crecimiento acelerado, impulsado por el incremento en el acceso a internet, la masificación de los dispositivos móviles y la expansión de soluciones de pago digital. Según Statista (2024), la región mostró un crecimiento del 25% anual en el comercio electrónico, siendo Brasil, México y Argentina los mercados más destacados. Este aumento se debe, en gran medida, a cambios en los hábitos de consumo, donde los usuarios prefieren realizar compras desde la comodidad de sus hogares. Sin embargo, este panorama también presenta desafíos, como la necesidad de mejorar la personalización en la experiencia de usuario, un área donde los sistemas de recomendación han demostrado ser fundamentales. Empresas como MercadoLibre y Rappi han implementado con éxito estas tecnologías, optimizando tanto la satisfacción del cliente como las tasas de conversión.

**Panorama General del Comercio Electrónico** El comercio electrónico en Centroamérica ha experimentado un crecimiento acelerado en los últimos años, impulsado por la mejora en la infraestructura digital y la adopción de pagos electrónicos. Según un informe de Statista (2024), las ventas en línea en la región crecieron un 23% anual en promedio, con Honduras, Costa Rica y Guatemala liderando el crecimiento. Aun así, el uso de sistemas de recomendación avanzados sigue siendo limitado en la región.

**Honduras:** La penetración de internet alcanzó un 65% en 2024, y el comercio electrónico representa un 4% del total de ventas al por menor, con plataformas locales como LynxPay destacándose en el mercado. Sin embargo, la personalización sigue siendo un desafío importante.

Costa Rica: Este país tiene una de las infraestructuras digitales más desarrolladas en la región, con una penetración de internet del 79%. Plataformas como Tienda Monge han implementado sistemas básicos de recomendación, mejorando las tasas de conversión en un 10%.

Guatemala: Aunque la penetración de internet es menor (54%), las plataformas de comercio electrónico como Elektra han mostrado interés en tecnologías de segmentación de clientes, comenzando a explorar modelos de clusterización.

#### 2.1.2.2 ADOPCIÓN DE SISTEMAS DE RECOMENDACIÓN EN HONDURAS

En Honduras, el comercio electrónico se encuentra en una etapa de desarrollo, con un crecimiento significativo en la última década gracias a iniciativas como LynxPay. A pesar de este progreso, el país enfrenta retos importantes en términos de infraestructura tecnológica, educación digital y confianza del consumidor en las plataformas en línea. Actualmente, la adopción de sistemas de recomendación es limitada, y muchas empresas recurren a soluciones básicas basadas en categorías predefinidas en lugar de algoritmos avanzados. En 2018, el AOV medio para los minoristas de América Latina 300 fue de 89 dólares, lo que resalta el impacto positivo que las herramientas de personalización pueden tener en el aumento del valor promedio de las transacciones. La implementación de sistemas de recomendación personalizados a través de técnicas de clusterización, como K-Means y DBSCAN, podría ser una estrategia diferenciadora para empresas locales, permitiéndoles competir en igualdad de condiciones con plataformas globales.

#### 2.1.2.3 ADOPCIÓN DE SISTEMAS DE RECOMENDACIÓN EN LATINOAMÉRICA

A nivel regional, la adopción de sistemas de recomendación ha avanzado considerablemente en los últimos años. Países como Brasil, México y Chile lideran la implementación de estas tecnologías, especialmente en sectores como el comercio minorista y las plataformas de streaming. Empresas como Falabella, Amazon México y Spotify Latinoamérica han demostrado que el uso de algoritmos de aprendizaje automático y técnicas de clusterización no solo mejora la experiencia del usuario, sino que también incrementa significativamente el valor promedio por transacción (Chaffey, 2016). Sin embargo, aún existen barreras en la región, como la falta de inversión en infraestructura tecnológica y el acceso desigual a internet. Para países con economías emergentes, la implementación de estos sistemas se presenta como una oportunidad

estratégica para fortalecer su competitividad en el mercado global. En este contexto, Honduras puede beneficiarse de las lecciones aprendidas en otros países de la región para acelerar la adopción de estas tecnologías y posicionarse como un actor clave en el comercio electrónico regional.

Panorama General del Comercio Electrónico en Latinoamérica se ha posicionado como una de las regiones de mayor crecimiento en comercio electrónico a nivel mundial. Según el Latin America Ecommerce Report (2024) de Statista, las ventas en línea crecieron un 25% anual, con un valor proyectado de \$160 mil millones de dólares para 2025. Países como Brasil, México y Argentina lideran el mercado, representando más del 70% de las transacciones regionales. A pesar de estos números, los sistemas de recomendación avanzados todavía no son comunes fuera de los grandes actores globales.

**Brasil:** Es el mercado más grande de comercio electrónico en la región, con empresas como Mercado Livre utilizando IA y técnicas de clusterización para personalizar recomendaciones, lo que ha incrementado el valor promedio por orden en un 15%.

**México:** Con una penetración de internet del 74%, empresas como Amazon México han implementado recomendaciones personalizadas basadas en algoritmos híbridos, logrando reducir la tasa de abandono del carrito en un 20%.

**Argentina:** La plataforma Tienda Nube reportó un aumento del 35% en las ventas al integrar sistemas de recomendación que segmentan a los clientes por comportamiento.

Otros países en Latinoamérica, como Colombia, Chile y Perú, están creciendo rápidamente en este rubro, impulsados por la expansión del acceso a internet y la mejora en la infraestructura de pagos electrónicos.

**Retos específicos de Centroamérica:** Limitada infraestructura tecnológica en países con menor penetración de internet (Nicaragua, Honduras).

Baja confianza en sistemas de pagos en línea, aunque se están popularizando métodos de pago como billeteras digitales y transferencias bancarias instantáneas.

Oportunidades de Segmentación y Recomendación en la Región.

**Técnicas Avanzadas de Clusterización:** La implementación de algoritmos como DBSCAN y K-Means puede mejorar la personalización en sectores como moda, tecnología y productos de

consumo masivo.

Recomendaciones Contextuales: Integrar datos demográficos y transaccionales permite ajustar las recomendaciones al contexto local, como estaciones climáticas o festividades.

Casos de éxito replicables: Mercado Libre (Latinoamérica): Ha optimizado su modelo de filtrado colaborativo para aumentar la retención de usuarios en un 25%.

Rappi (Colombia y otros países): Utiliza datos de comportamiento en tiempo real para segmentar clientes según sus hábitos de consumo.

Fortalecimiento del Comercio Electrónico Local: Las plataformas regionales pueden aprender de gigantes como Amazon y Mercado Libre para incorporar sistemas híbridos que combinen recomendaciones basadas en contenido y comportamiento. Statista (2024)

### 2.1.3 ANÁLISIS INTERNO

#### 2.1.3.1 DIAGNÓSTICO DE LA SITUACIÓN DE LA PLATAFORMA CON RELACIÓN AL USO DE MODELOS DE SEGMENTACIÓN

LynxPay, plataforma objetivo para el análisis de esta investigación, cuenta con muchas características que posibilitan al emprendedor realizar su introducción al comercio electrónico, entre ellas permite:

- Pago electrónico haciendo uso de tarjetas de crédito y débito.
- Integración con cuenta bancaria nacional.
- Personalización de tienda.

#### 2.1.3.2 DIAGNÓSTICO DE LA SITUACIÓN

Actualmente LynxPay carece de funcionalidades específicas para la segmentación de clientes y la recomendación de productos. Según LynxLabs, esta ausencia se debe a la limitación de recursos humanos, que han sido priorizados para fortalecer la seguridad en las transacciones.

La falta de estas herramientas puede impactar en la capacidad de los emprendedores que utilizan LynxPay para:

- Personalizar la experiencia de compra.
- Optimizar las estrategias de marketing.
- Incrementar las ventas.

### 2.1.3.3 ESTRATEGIA Y PLANES FUTUROS DE LYNXPAY

A pesar de las limitaciones actuales, LynxLabs reconoce la importancia de la segmentación y la recomendación de productos para el éxito de las tiendas online (O. Sánchez, comunicación personal, 30 de octubre de 2024). Ambas funcionalidades se encuentran en su hoja de ruta de desarrollo, aunque su implementación se ha pospuesto debido a la priorización de la seguridad transaccional.

### 2.1.3.4 OPORTUNIDADES DE MEJORA A TRAVÉZ DE LA PERSONALIZACIÓN Y SEGMENTACIÓN

En el entorno actual especialmente enfocado en el e-commerce, se cuenta con abundancia de información y opciones disponibles puede resultar para los usuarios abrumadora. Minimizar las opciones al usuario y que al mismo tiempo resulten de valor es crucial para mejorar la experiencia de usuario y por consiguiente aumentar el valor percibido por los mismos en transacciones. Para ello los sistemas de recomendación surgen como herramienta fundamental para guiar al usuario hacia nuestros productos y específicamente a los que mejor se adecuen a sus requerimientos, basándonos en previas transacciones.

Por ello, los sistemas de recomendación en el comercio electrónico se han vuelto crucial para el éxito de compañías como Amazon, donde se reportó que el 35% de sus ingresos provenían directamente de las recomendaciones hechas por su sistema.

Para Ricci, Rokach y Shapira (2011) los sistemas de recomendación desempeñan un papel fundamental en sitios web como Amazon, Netflix y YouTube, permitiendo a los usuarios encontrar contenido o servicios de manera eficiente.

Los beneficios van desde:

- Mejora de experiencia de usuario: Los usuarios encuentran fácilmente lo que buscan y descubren nuevas opciones relevantes. (Duy-Nghia Nguyen, 2024)
- Aumento de la retención de clientes: La personalización crea una experiencia más atractiva y relevante para los usuarios. (Chaffey, 2016)
- Mayores tasas de conversión: Las recomendaciones personalizadas aumentaran los incentivos de compra e interacción con el contenido. (Duy-Nghia Nguyen, 2024)

### 2.1.3.5 RETOS EN LA IMPLEMENTACIÓN DE SISTEMAS DE RECOMENDACIÓN

Según LinkedIn (2024), LynxLabs es una compañía establecida en Tegucigalpa, Honduras fundada en 2017, cuenta con aproximadamente de 11 a 50 empleados dedicada al desarrollo de software a medida, que vio la oportunidad de generar una plataforma de comercio electrónico a la cual llamó “LynxPay”, en la cual compañías honduras han confiado su información y propuesta para gestionar su sitio web y por consiguiente permitirles tener un nuevo medio para llegar a sus clientes.

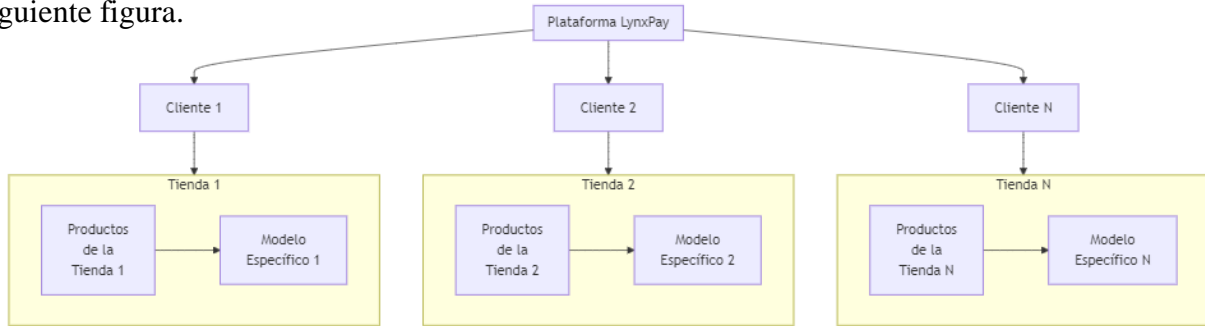
Podemos dividir los retos en operativas y técnicas, la primera resulta especialmente importante debido al recurso humano necesario para la generación e implementación de estas herramientas y modelos. LynxLabs cuenta con recurso humano especializado en desarrollo, implementaciones como esta pueden requerir de estudios más específicos, dado esto el primer reto es el recurso humano necesario. LynxLabs ha enfocado sus recursos a la seguridad de la plataforma como su primer objetivo, la asignación de personal a la implementación de un sistema de recomendación pasa a ser una prioridad más baja. (O. Sánchez, comunicación personal, 30 de octubre de 2024)

Entre los retos técnicos para la implementación de un sistema de recomendación encontramos:

- Aislamiento de datos: La arquitectura multi-inquilino en LynxPay implica que los datos de cada cliente están separados. Esto puede dificultar la generación de recomendaciones precisas, especialmente para el filtrado colaborativo, que requiere acceso a datos de múltiples usuarios para identificar patrones y similitudes.
- Inicio en frío: Este problema hace hincapié en la dificultad de generar recomendaciones para usuarios nuevos o productos recién agregados, debido a la falta de información histórica. (Duy-Nghia Nguyen, 2024)
- Escasa cantidad de datos: Relacionado al último reto, para emprendedores la cantidad de datos de interacción de los usuarios puede resultar en recomendaciones aleatorias y poco precisas. (Duy-Nghia Nguyen, 2024)
- Privacidad de datos: Es crucial garantizar la privacidad de los datos de los usuarios

al recopilar y utilizar información para la generación de recomendaciones.

Especialmente la arquitectura multi-inquilino presentada por la plataforma, véase la siguiente figura.



**Figura 1 Esquema de estructura multi-inquilino**

Fuente: Elaboración Propia

Representa un reto para ser específicos la creación de modelos para cada cliente inquilino en la plataforma, debido a que se desean realizar recomendaciones específicas de cada negocio, por lo cual se considera importante para esta implementación el filtrado a priori de la información para cada tienda.

## 2.2 CONCEPTUALIZACIÓN

### 2.2.1 VARIABLE DEPENDIENTE

#### 2.2.1.1 PRECISIÓN DE RECOMENDACIONES

Con el fin de verificar y la efectividad del sistema de recomendación híbrido, se realizan los experimentos de algoritmo único y algoritmo híbrido respectivamente para comparar la precisión de los algoritmos. La precisión es un factor muy importante como indicador para evaluar el rendimiento de los sistemas de recomendación. La precisión se define de la siguiente manera:

$$Precision = \frac{sum(R(u) \cap T(u))}{sum(R(u))} \quad (1)$$

Donde:

- $R(u)$ : Representa la lista de recomendaciones hechas al usuario  $u$  basadas en su comportamiento en el conjunto de entrenamiento. Es decir, es el conjunto de ítems que el sistema le recomienda al usuario.
- $T(u)$ : Es el conjunto de ítems que el usuario realmente ha valorado o seleccionado

en el conjunto de prueba. Este conjunto representa lo que el usuario realmente considera relevante.

Explicación:

La precisión mide la proporción de recomendaciones correctas que el sistema ha hecho en relación con todas las recomendaciones realizadas. En otras palabras, la precisión evalúa qué tan exacto es el sistema al recomendar ítems relevantes para el usuario.

- Intersección  $R(u) \cap T(u)$ : La intersección entre  $R(u)$  y  $T(u)$  representa los ítems que han sido recomendados y que realmente son de interés para el usuario según el conjunto de prueba. Este es el conjunto de recomendaciones correctas o relevantes.
- Suma  $\text{sum}(R(u))$ : La suma de  $R(u)$  es el total de ítems recomendados por el sistema.

Interpretación: La precisión se calcula dividiendo el número de ítems relevantes recomendados (los elementos en  $R(u) \cap T(u)$ ) entre el total de ítems recomendados ( $\text{sum}(R(u))$ ). Si el valor de precisión es alto, significa que el sistema es efectivo recomendando ítems que el usuario realmente considera relevantes. Yonghong Tian (2019).

## 2.2.2 VARIABLES INDEPENDIENTES

### 2.2.2.1 SEGMENTACIÓN TRANSACCIONAL

La segmentación transaccional se basa en el análisis de datos relacionados con las transacciones realizadas por los consumidores, como el monto total gastado, la frecuencia de compra, el tipo de producto adquirido y los métodos de pago utilizados. Este enfoque permite identificar patrones de gasto y comportamiento de los consumidores para desarrollar estrategias personalizadas de marketing, según autores como, Kotler, Philip y Gary (2017).

### 2.2.2.2 SEGMENTACIÓN GEOGRAFICA

La segmentación geográfica clasifica a los consumidores en función de su ubicación, considerando factores como país, región, ciudad e incluso clima. En el contexto de e-commerce, este enfoque permite adaptar la oferta de productos y servicios a las necesidades y preferencias locales, lo que mejora la relevancia del producto para el usuario final, Kotler y Keller (2017). La segmentación geográfica es útil para identificar patrones de compra y preferencias de consumo que varían entre diferentes regiones, ayudando a personalizar las recomendaciones y estrategias

de distribución.

### 2.2.2.3 SEGMENTACIÓN DEMOGRÁFICA

La segmentación demográfica clasifica a los consumidores en función de características personales como edad, género, nivel educativo, estado civil, ingreso económico, ocupación y tamaño del hogar. Según Kotler y Keller (2017), este enfoque permite adaptar estrategias de marketing a las características básicas y necesidades generales de los consumidores.

### 2.2.3 COMERCIO ELECTRÓNICO

El comercio electrónico, también conocido como e-commerce, se refiere a la compra y venta de bienes y servicios a través de medios electrónicos, como internet y otras redes informáticas. En el contexto del marketing digital, el comercio electrónico juega un papel fundamental en la estrategia de ventas y distribución de las empresas. (Chaffey, 2016)

Laudon y Traver (2020) definen el comercio electrónico como "el uso de Internet y la web para realizar transacciones de negocios; más formalmente, el intercambio digital de bienes, servicios e información" (p.45). Esta definición enfatiza no solo las transacciones comerciales, sino también el intercambio de información como un componente esencial del comercio electrónico. (Laudon & Traver, 2020, p. 45).

En términos de modelos de negocio, el comercio electrónico se puede clasificar en varias categorías, como B2B (business-to-business), B2C (business-to-consumer), C2C (consumer-to-consumer) y C2B (consumer-to-business), cada uno con sus propias características y requerimientos estratégicos. (Chaffey, 2016)

### 2.2.4 TASA DE CONVERSIÓN EN COMERCIO ELECTRÓNICO

La tasa de conversión en comercio electrónico, según Chaffey (2016), se refiere a la métrica que evalúa la eficacia de un sitio web en convertir visitantes en clientes reales. Esta métrica es esencial para medir el rendimiento de las estrategias de marketing digital y la experiencia del usuario en un entorno online. (p. 460).

Chaffey (2016) explica que la tasa de conversión se calcula dividiendo el número de conversiones (por ejemplo, compras realizadas) entre el número total de visitantes o usuarios activos en el sitio web. Esta métrica proporciona información valiosa sobre la efectividad de las acciones de marketing y la capacidad del sitio para persuadir a los

visitantes a realizar una acción deseada, como una compra.

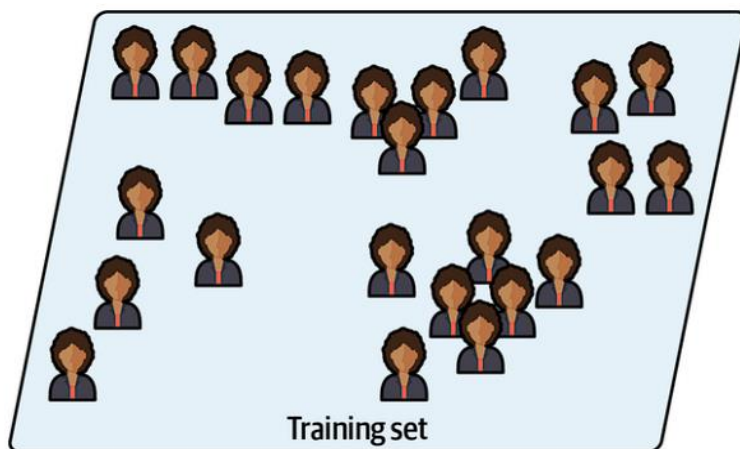
## 2.2.5 SEGMENTACIÓN DE CLIENTES

La segmentación de clientes es un proceso analítico clave en el comercio electrónico y el marketing, donde se agrupan los clientes en subconjuntos según características, comportamientos y preferencias similares. Esta técnica permite a las empresas personalizar sus estrategias y ofertas, mejorando la eficiencia del marketing, aumentando la satisfacción del cliente y optimizando los recursos empresariales. (Zhou, 2004)

El Machine Learning y las técnicas de análisis de datos juegan un papel crucial en la segmentación de clientes, permitiendo el manejo de grandes volúmenes de datos y la identificación de patrones complejos que no serían apreciables a simple vista. Hennig (2015, pág. 234) destaca que la utilización de algoritmos de clusterización, como K-means o análisis jerárquico, facilita la creación de grupos de clientes con características homogéneas, lo cual es fundamental para la personalización de servicios y productos en plataformas de comercio electrónico (p. 234).

## 2.2.6 APRENDIZAJE NO-SUPERVISADO

El aprendizaje no supervisado es en pocas palabras un modelo que es entrenado con información sin categorizar, aprende sin maestro.

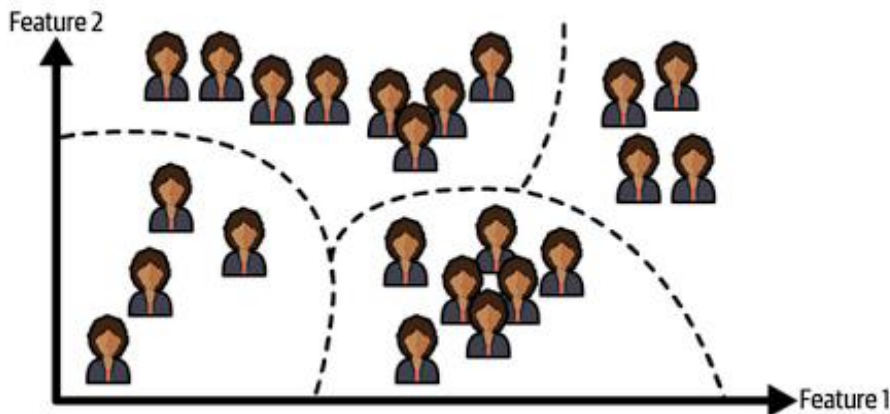


**Figura 2. Datos de entrenamiento sin categorizar para aprendizaje automático.**

Fuente: (Aurélien, 2023)

Ejemplo de esto podría resultar en la detección de grupos similares de visitantes en un blog en específico nótese figura 2 y 3, en el cual basado en características relevantes en datos obtenidos de los usuarios podría realizarse un análisis de grupos que resulte en agrupaciones inherentes de

visitantes, posiblemente encontrando patrones de comportamiento.



**Figura 3. Agrupación de información.**

Fuente: (Aurélien, 2023)

### 2.2.7 CLUSTERING

Géron (2023) menciona que el objetivo principal de este concepto es agrupar instancias similares juntas en grupos. Confirma que es una gran herramienta para análisis de datos, segmentación de clientes, sistemas de recomendación, segmentación de imágenes, y más.

Para Ricci y Shapira (2011) también se le conoce como aprendizaje no supervisado, consiste principalmente en la asignación de instancias a grupos, para que las instancias en los grupos compartan similitudes que como objetivo final nos permita descubrir de manera natural grupos que existen en la información.

Con esto, podremos establecer grupos dentro de los datos de cada tienda dentro de LynxPay, encontrando similitudes en comportamiento de compra de los clientes, llegando a poder ofrecer productos que sean acorde a sus necesidades posiblemente derivado de comportamiento de otros pares.

### 2.2.8 K-MEANS

K-means es un algoritmo de agrupamiento que divide un conjunto de datos en  $k$  grupos distintos, buscando que los elementos dentro de cada grupo sean lo más similares posible entre sí y lo más diferentes posible a los elementos de otros grupos. Para lograrlo, el algoritmo asigna cada elemento al grupo cuyo centroide (punto central) está más cerca, utilizando una medida de distancia. Este proceso se repite iterativamente, ajustando los centroides y la asignación de los

elementos a los grupos hasta que se estabilizan. (Ricci F. R., 2011)

Ricci y Shapira (2011) también menciona que K-means a pesar de ser un algoritmo simple y eficiente, presenta algunas limitaciones, dado que se requiere conocer a priori la cantidad de grupos (k) que se desean formar y esto puede cambiar notablemente la posición de centroides.

A su vez Géron (2023) confirma que k-means no es perfecto, el algoritmo debe ejecutarse una gran cantidad de ocasiones para evitar soluciones no óptimas, además de especificar un número de nodos.

### 2.2.9 DBSCAN

Para Gerón (2023) el modelo DBSCAN (density-based spatial clustering of applications with noise), en su traducción “Agrupación espacial basado en densidad espacial para aplicaciones con ruido” define agrupaciones como regiones continuas de alta intensidad, haciendo simplícidamente lo siguiente:

- Por cada instancia, el algoritmo cuenta cuantas instancias están localizadas en una pequeña región  $\epsilon$  o (épsilon)
- Si existe una cierta cantidad de muestras en esta región es considerada una instancia principal.
- Todas las instancias en el entorno de la instancia principal pertenecen al mismo grupo.
- Cualquier instancia que no se considere una instancia principal y no tiene una en su entorno, es considerada una anomalía.

### 2.2.10 PERSONALIZACIÓN EN EL COMERCIO ELECTRONICO

Ricci y Shapira (2011) menciona que la personalización en el comercio electrónico se refiere a la adaptación de la experiencia de compra a las necesidades y preferencias individuales de cada cliente. Esto implica ofrecer productos, servicios y contenido relevantes para cada usuario, basándose en su comportamiento, historial de compras y datos demográficos. La personalización puede mejorar significativamente la satisfacción del cliente, aumentar la fidelización y las ventas.

En la perspectiva de mercadeo digital, Chaffey (2016) menciona que la personalización envuelve la entrega personalizada de contenido a individuos, a través de la web, correo electrónico

o cualquier otra tecnología de contacto.

### 2.2.11 SISTEMAS DE RECOMENDACIÓN

Los sistemas de recomendación en el comercio electrónico son herramientas tecnológicas avanzadas diseñadas para predecir y sugerir productos o servicios a los usuarios basándose en sus preferencias, comportamientos anteriores y datos de otros usuarios con gustos similares. Estos sistemas juegan un papel crucial en la personalización de la experiencia de compra online, ayudando a los usuarios a descubrir productos que probablemente les interesen, mejorando la satisfacción del cliente y aumentando la eficiencia de las ventas.

En el Hennig (2015, pág. 117) discuten cómo los métodos de clusterización pueden ser utilizados para mejorar la precisión de los sistemas de recomendación. Al agrupar a los usuarios según sus preferencias, es posible ofrecer recomendaciones más precisas y personalizadas.

Algunas técnicas de filtrado son:

- Filtrado basado en contenido
- Filtrado colaborativo
- Agrupación K-Means.

### 2.2.12 FILTRADO BASADO EN CONTENIDO

Ricci y Shapira (2011) este algoritmo se enfoca en recomendar productos que son similares a los que el usuario le gustó en el pasado. La similitud de los productos es calculado basado en características asociadas.

En alto nivel los sistemas de filtrado basado en contenido necesitan ejecutar técnicas adecuadas para representar los productos y producir un perfil de usuario adecuado, el proceso de recomendación se realiza en tres pasos:

**Analizar contenido:** Cuando la información no es estructurada como texto plano, se requiere algún tipo de preprocesamiento para extraer información relevante.

**Aprendizaje de perfil:** Recolecta información representativa de las preferencias del usuario y trata de generalizar la información para poder construir un perfil de cliente.

Este paso explora el perfil del usuario para sugerir productos con relevancia que encajen

en el perfil del usuario.

### 2.2.13 FILTRADO COLABORATIVO

Para Chaffey (2016) el filtrado colaborativo es un método de recomendación que sugiere productos a los usuarios basándose en las preferencias de otros usuarios con gustos similares. Este enfoque se basa en la idea de que, si dos usuarios han coincidido en sus preferencias en el pasado, es probable que también coincidan en el futuro. Para ello, se analizan las valoraciones o el historial de compras de los usuarios, y se identifican aquellos con patrones similares. Cuanto mayor sea la cantidad de información disponible sobre los usuarios, más precisas serán las recomendaciones. Un ejemplo de este método es el sistema de recomendación de Amazon, que sugiere productos basándose en las compras de otros usuarios.

### 2.2.14 ALGORITMOS HÍBRIDOS

Los sistemas de recomendación híbridos combinan diferentes técnicas, como el filtrado colaborativo y el basado en contenido, para aprovechar las ventajas de cada una y superar sus limitaciones. Esto permite, por ejemplo, recomendar artículos nuevos que no tienen valoraciones previas, o adaptar las recomendaciones al contexto del usuario, como la época del año o la ocasión.

Existen diferentes enfoques para incorporar la información contextual en los sistemas de recomendación, como el pre-filtrado, el post-filtrado y la modelización contextual. Estos métodos permiten ajustar las recomendaciones a las circunstancias específicas del usuario, mejorando la precisión y la satisfacción con las recomendaciones. (Ricci F. R., 2011)

### 2.2.15 CARRITO DE COMPRA

El carrito de compra en el contexto del comercio electrónico es una herramienta virtual que permite a los usuarios recolectar artículos digitales antes de realizar una compra final. Funciona como un equivalente digital del carrito de compras físico utilizado en las tiendas tradicionales, permitiendo a los compradores navegar por un sitio web, seleccionar productos de interés y reservarlos temporalmente mientras continúan explorando otros productos. Al final del proceso de compra, el usuario puede revisar los artículos en el carrito, realizar cambios si es necesario, y proceder al pago. (Laudon & Traver, 2020).

### 2.2.16 RECOMENDADOR DE PRODUCTOS

Los sistemas de recomendación de productos son herramientas tecnológicas avanzadas utilizadas en el comercio electrónico para sugerir productos a los usuarios basándose en una variedad de factores, incluidos los comportamientos de compra previos, las preferencias y los intereses expresados, y los patrones de consumo de otros usuarios con perfiles similares. Estos sistemas utilizan algoritmos de aprendizaje automático y técnicas de filtrado para personalizar las experiencias de compra y aumentar la probabilidad de conversión al presentar a los usuarios artículos que probablemente sean de su interés.

El resultado según Ricci y Shapira (2011) los sitios de comercio electrónico utilizan sistemas de recomendación de productos para mejorar sus ingresos, mejorando las ventas cruzadas. (p. 289)

### 2.2.17 INICIO EN FRIO

El inicio en frío en los sistemas de recomendación hacen referencia especialmente a la dificultad que tienen modelos en ofrecer dichas recomendaciones por la falta de información del individuo, Ricci y Shapira (2011) mencionan que para adaptar las recomendaciones los sistemas necesitan conocer los gustos y preferencias de los usuarios, que se suele obtener basado en interacciones o valoraciones.

A su vez Nguyen, Van-Ho Nguyen, Trinh, Ho, Le (2024) menciona que el inicio en frío resulta ser un problema y que todo sistema de recomendación debería ser capaz de sugerir productos a nuevos usuarios basados en información demográfica.

### 2.2.18 ARQUITECTURA MULTI-INQUILINO

En la arquitectura multi-inquilino en el contexto de software como servicio en la que se basa LynxPay para su funcionamiento, se refiere a la creación de una sola aplicación o software para múltiples clientes o “inquilinos”. Cada inquilino tiene sus propios datos y configuraciones, que está totalmente aislados de sus demás inquilinos. Esto permite a proveedores de SaaS ofrecer servicios de forma más eficiente y económica. (Golding, 2024)

### 2.2.19 SCIKIT-LEARN (SKLEARN)

Scikit-Learn es una librería que implementa muchos algoritmos de aprendizaje de máquina eficientemente, lo hace un gran punto de partida para el aprendizaje de máquina, creado por David

Cournapeau en 2007. (Aurélien, 2023)

Algunos modelos disponibles en la librería van desde simples regresiones lineales, clasificación hasta agrupaciones y reducción de dimensionalidad. (Scikit Learn, 2024).

## 2.3 TEORÍAS DE SUSTENTO

### 2.3.1 BASES TEÓRICAS

Llevando a cabo un análisis de la literatura disponible, lo cual permitió identificar y recopilar una variedad de teorías orientadas a la segmentación de clientes. Estas teorías proporcionan distintos enfoques y metodologías para clasificar y agrupar a los consumidores en función de características específicas, con el objetivo de optimizar las estrategias de marketing y personalización en las tiendas en línea. Entre las teorías identificadas se destacan aquellas que buscan mejorar el conocimiento del cliente, maximizando la efectividad de las recomendaciones personalizadas y adaptando los servicios y productos a las necesidades y preferencias de distintos grupos de usuarios. A continuación, se presentan algunas de estas teorías relevantes para la segmentación de clientes.

### 2.3.2 TEORÍA DE LA SEGMENTACIÓN

En la bibliografía estudiada, la segmentación es el proceso de dividir un mercado amplio y diverso en grupos más pequeños de consumidores que comparten características, necesidades o comportamientos similares. Este concepto fue desarrollado a partir de la idea de que es más efectivo y rentable para las empresas enfocar sus esfuerzos de marketing en subgrupos específicos de consumidores en lugar de dirigirse a todo el mercado de forma homogénea. Según Kotler y Armstrong (2017), la segmentación permite diseñar estrategias de marketing más personalizadas que logran satisfacer mejor las necesidades de cada segmento, incrementando así la eficacia de las campañas de marketing y las posibilidades de éxito comercial.

#### 2.3.2.1 BASES DE SEGMENTACIÓN DE MERCADOS

Para segmentar un mercado, existen diversas bases o criterios que permiten clasificar a los consumidores. Algunos de los enfoques más comunes incluyen:

- **Segmentación demográfica:** que clasifica a los consumidores según variables como la edad, el género, el ingreso, la educación y el ciclo de vida familiar.

- **Segmentación geográfica:** que divide a los consumidores según su ubicación geográfica, como país, región, ciudad o incluso barrio.
- **Segmentación psicográfica:** que considera aspectos relacionados con el estilo de vida, los valores, las actitudes y la personalidad de los consumidores.
- **Segmentación conductual:** que se enfoca en el comportamiento de compra, uso y lealtad del consumidor hacia ciertos productos o marcas.

Estas bases permiten que las empresas identifiquen y clasifiquen a los clientes de manera más precisa, ofreciendo productos y servicios adaptados a las particularidades de cada segmento Wedel & Kamakura (2000), centrándonos nosotros para esta investigación en la segmentación conductual debido a que necesitaremos el comportamiento de compra, uso y lealtad del consumidor dentro de las tiendas en línea en la plataforma.

#### 2.3.2.2 SEGMENTACIÓN Y MODELOS DE VALOR DEL CLIENTE

La teoría de la segmentación también se basa en modelos de valor del cliente, los cuales evalúan la rentabilidad potencial de cada cliente o segmento. Modelos como el RFM (Recency, Frequency, Monetary) y el Análisis de Ciclo de Vida del Cliente (CLV, por sus siglas en inglés) ayudan a segmentar a los clientes según su valor económico y sus hábitos de compra, permitiendo que las empresas dirijan sus esfuerzos hacia los segmentos más valiosos. Estos modelos enfatizan que no todos los clientes generan el mismo valor para la empresa, y que es estratégico destinar recursos a los clientes más rentables Reinartz & Kumar (2000).

La segmentación ha evolucionado hacia enfoques más personalizados gracias al avance del Big Data y las técnicas de machine learning, que permiten analizar grandes volúmenes de datos para identificar patrones específicos en el comportamiento de los clientes. Este enfoque, conocido como segmentación personalizada, permite a las empresas adaptar sus productos, servicios y comunicaciones de manera más detallada. Según Rust y Verhoef (2005), la personalización basada en segmentación puede mejorar la experiencia del cliente y, al mismo tiempo, aumentar la lealtad y la retención de los consumidores.

#### 2.3.3 TEORÍA DE LA COLABORACION DE FILTRADO

La Teoría de la Colaboración de Filtrado se refiere a los métodos y modelos utilizados para hacer recomendaciones automáticas a los usuarios, especialmente en plataformas digitales como

la que trabajaremos en esta investigación, mediante el análisis de patrones de comportamiento y preferencias. Esta teoría es fundamental en sistemas de recomendación, ya que permite sugerir productos, servicios o contenidos a usuarios basándose en datos de otros usuarios con comportamientos o intereses similares.

La colaboración de filtrado se originó en el ámbito de la recuperación de información y la inteligencia artificial, y fue aplicada por primera vez en la década de 1990 para mejorar los sistemas de recomendación. El principio detrás de esta teoría es que los usuarios que han mostrado comportamientos o preferencias similares en el pasado probablemente seguirán patrones semejantes en el futuro. Este enfoque se basa en el supuesto de que los usuarios pueden beneficiarse de los datos generados por otros usuarios, creando así una red colaborativa de preferencias. Según Schafer, Konstan, y Riedl (2001), el filtrado colaborativo es uno de los métodos más efectivos para proporcionar recomendaciones personalizadas.

#### 2.3.3.1 TIPOS DE FILTRADO COLABORATIVO

Según los estudios realizados que analizamos para esta investigación existen dos tipos principales de modelos de filtrado colaborativo:

- **Filtrado colaborativo basado en el usuario:** Este método sugiere elementos a un usuario específico en función de lo que han preferido otros usuarios con patrones de comportamiento similares. Por ejemplo, si dos usuarios tienen historiales de compra similares, los artículos comprados por uno de ellos pueden ser recomendados al otro.
- **Filtrado colaborativo basado en el ítem:** En lugar de buscar similitudes entre usuarios, este método se centra en encontrar similitudes entre ítems (productos, películas, etc.) según las preferencias de los usuarios. Si varios usuarios que compraron un libro también adquirieron otro título específico, entonces el sistema recomendará ese segundo libro a nuevos usuarios interesados en el primero.

Estos enfoques nos permiten crear recomendaciones precisas y relevantes, y son ampliamente utilizados en plataformas como Amazon, Netflix y Spotify, Badrul Sarwar (2021).

#### 2.3.3.2 TEORÍA DE LOS ALGORITMOS DE VECINDAD Y MATRIZ DE SIMILARIDAD

La teoría detrás de los algoritmos de filtrado colaborativo se basa en la idea de vecindad y

la matriz de similaridad. En el filtrado colaborativo basado en el usuario, se utiliza una matriz que mide la similitud entre usuarios (o ítems) para identificar "vecinos" cercanos en términos de preferencias y comportamientos. Esta teoría matemática se fundamenta en métricas como el coseno de similitud, la correlación de Pearson o la distancia euclidiana, que cuantifican qué tan cercanos están dos usuarios o ítems en un espacio de preferencias. Según Koren, Bell y Volinsky (2009), estos modelos de vecindad han sido cruciales para mejorar la precisión y relevancia de los sistemas de recomendación.

### 2.3.3.3 MODELOS DE DESCOMPOSICIÓN DE MATRICES (FACTORIZACIÓN DE MATRICES)

Otro enfoque teórico importante en la colaboración de filtrado es el modelo de descomposición de matrices, también conocido como factorización de matrices. Este enfoque es común en sistemas de recomendación modernos y fue ampliamente popularizado por la incursión de Netflix. En este modelo, las interacciones entre usuarios e ítems se representan en una matriz, que luego se descompone en factores latentes, es decir, características ocultas que explican las similitudes y diferencias entre usuarios e ítems. Esta técnica permite que el sistema prediga las preferencias de los usuarios para ítems que aún no han visto, lo cual es fundamental para mejorar la experiencia de usuario Koren, Bell y Volinsky (2009).

### 2.3.3.4 FILTRADO COLABORATIVO Y APRENDIZAJE AUTOMÁTICO

Con el avance del Big Data y el aprendizaje automático, el filtrado colaborativo ha evolucionado para incorporar algoritmos más complejos, como el machine learning. Los sistemas de recomendación actuales utilizan modelos de aprendizaje profundo y redes neuronales para analizar grandes cantidades de datos y mejorar la precisión de las recomendaciones. Este enfoque permite identificar patrones de comportamiento complejos y personalizar las recomendaciones a un nivel más detallado y dinámico. Según Ricci (2016), los modelos basados en aprendizaje automático ofrecen mejores resultados al captar patrones no lineales y adaptarse rápidamente a cambios en las preferencias de los usuarios.

### 2.3.3.5 FILTRADO COLABORATIVO HÍBRIDO

La teoría de colaboración de filtrado también ha evolucionado hacia enfoques híbridos, que combinan el filtrado colaborativo con otros métodos, como el filtrado basado en contenido. Este

tipo de sistemas integran diferentes fuentes de información, tales como datos demográficos y de contenido, para mejorar la precisión y relevancia de las recomendaciones. Los sistemas de filtrado híbrido pueden adaptarse a contextos en los que el filtrado colaborativo puro se ve limitado, como en el caso del problema de inicio frío (cold start), en el que no se dispone de suficiente información sobre nuevos usuarios o ítems. Los sistemas híbridos ayudan a superar las limitaciones individuales de los modelos de filtrado, creando recomendaciones más robustas y adaptables.

La colaboración de filtrado se ha convertido en una herramienta fundamental en diversas industrias, como el comercio electrónico, el entretenimiento y las redes sociales. Sin embargo, también presenta desafíos, como el problema de cómo se esparce la matriz, en el que la falta de datos suficientes sobre ciertos usuarios o ítems dificulta la generación de recomendaciones precisas. Otra limitación es el problema de inicio frío para nuevos usuarios o ítems, que carecen de suficiente historial de datos. A pesar de estos desafíos, la teoría de la colaboración de filtrado sigue siendo una base sólida para los sistemas de recomendación y es complementada continuamente por nuevas tecnologías y enfoques.

#### 2.3.4 TEORÍA DE LA MINERÍA DE DATOS

La minería de datos es una disciplina que combina técnicas de estadística, aprendizaje automático e inteligencia artificial para extraer patrones significativos y conocimiento útil a partir de grandes volúmenes de datos. Según Han, Kamber y Pei (2012), la minería de datos es el proceso de descubrir información previamente desconocida, válida y potencialmente útil a partir de bases de datos. Este enfoque es particularmente relevante en el contexto del comercio electrónico, donde los datos generados por los usuarios son voluminosos, diversos y complejos.

##### **Componentes Clave de la Minería de Datos:**

- **Preparación de Datos:** Consiste en limpiar, transformar y seleccionar datos relevantes para su análisis. Esto asegura que los datos sean consistentes, precisos y aptos para los modelos que se utilizarán.
- **Modelado y Análisis:** Incluye técnicas como la segmentación de clientes mediante algoritmos de clusterización (K-Means, DBSCAN) y la construcción de modelos predictivos para identificar patrones y tendencias relevantes.
- **Evaluación y Visualización:** Permite validar los resultados obtenidos y

representarlos de manera que puedan ser interpretados fácilmente, como gráficos o reportes que respaldan la toma de decisiones estratégicas.

### **Aplicación en Sistemas de Recomendación:**

En sistemas de recomendación, la minería de datos se utiliza para analizar el comportamiento del usuario, como transacciones, historial de navegación y preferencias. Esta información permite identificar relaciones significativas entre los datos, facilitando la personalización de recomendaciones. Técnicas de clustering y clasificación son especialmente útiles en este contexto, ya que ayudan a segmentar usuarios en grupos homogéneos y predecir sus preferencias futuras.

### **Relación con la Clusterización y el Comercio Electrónico:**

Para Ricci, Rokach y Shapira (2011), la minería de datos es un componente central en la implementación de sistemas de recomendación basados en clusterización. Estas técnicas permiten agrupar a los clientes en clústeres con características similares, optimizando la personalización y aumentando las tasas de conversión en plataformas de comercio electrónico.

### **Relevancia en el Contexto de LynxPay:**

En el caso de la plataforma LynxPay, la minería de datos es indispensable para extraer conocimiento a partir de los datos transaccionales y demográficos de los usuarios. A través de su implementación, es posible identificar patrones de comportamiento y diseñar estrategias efectivas de recomendación, mejorando la experiencia del usuario y potenciando el valor de cada transacción.

## **2.3.5 INSTRUMENTOS UTILIZADOS POR LOS INVESTIGADORES**

### **2.3.5.1 PUTTY**

PuTTY es un cliente de terminal que permite conectarse a servidores remotos mediante protocolos como SSH (Secure Shell). En minería de datos, se utiliza para acceder a servidores donde están alojados grandes volúmenes de datos o donde se ejecutan modelos de recomendación en entornos más robustos, como clústeres de servidores o sistemas en la nube. (Tatham, 2024)

Ejemplo: Conectarse a un servidor remoto para ejecutar scripts de Python o consultar bases de datos utilizadas en el proceso de minería.

### 2.3.5.2 DBEAVER

DBeaver es una herramienta de gestión de bases de datos que permite conectar, explorar y gestionar bases de datos relacionales y no relacionales. Es crucial para realizar consultas y extraer datos necesarios para alimentar modelos de recomendación. (Dbeaver, 2024)

Ejemplo: Utilizar DBeaver para realizar consultas SQL avanzadas que obtengan datos relevantes de una base de datos, como historiales de compras o preferencias de usuarios, que luego se procesarán en Python.

### 2.3.5.3 PYTHON

Python es un lenguaje de programación ampliamente utilizado en ciencia de datos y aprendizaje automático. Es esencial para construir, entrenar y evaluar modelos de recomendación gracias a su versatilidad y la disponibilidad de bibliotecas especializadas. (Python Org, 2024)

Ejemplo: Escribir un script en Python que implemente un algoritmo de clustering (como K-Means) para segmentar usuarios con características similares en un sistema de recomendación.

### 2.3.5.4 VISUAL STUDIO CODE (VSCODE)

VSCode es un editor de código que facilita el desarrollo de scripts y aplicaciones en diversos lenguajes, incluido Python. Su integración con extensiones y sistemas de control de versiones como Git lo hace ideal para trabajar en proyectos colaborativos de minería de datos.

Ejemplo: Utilizar VSCode para escribir y depurar código en Python, incluyendo la implementación de algoritmos de recomendación o scripts para procesar grandes volúmenes de datos. (Visual Studio Code, 2024)

### 2.3.5.5 SCIKIT-LEARN (SKLEARN)

Sklearn es una biblioteca de Python que ofrece herramientas para aprendizaje automático, incluyendo algoritmos para sistemas de recomendación, clasificación, regresión y clustering. Es ideal para experimentar con distintos enfoques antes de implementar un modelo final. (Scikit Learn, 2024)

Ejemplo: Usar Sklearn para implementar un modelo de filtrado colaborativo basado en usuarios o productos y evaluar su precisión con métricas como Precision@10 o Recall.

#### 2.3.5.6 MATPLOTLIB

Matplotlib es una biblioteca de Python para crear visualizaciones de datos. En minería de datos y sistemas de recomendación, se utiliza para explorar datos, analizar patrones y presentar resultados. (Hunter, 2024)

Ejemplo: Crear gráficos para visualizar la segmentación de usuarios basada en clústeres o analizar la distribución de datos de productos más comprados.

#### 2.3.5.7 METODOLOGÍA PRACTICA DEL FLUJO

- Conexión y extracción de datos: Utilizar PuTTY para conectarse al servidor donde se almacena la base de datos y DBeaver para realizar consultas SQL y extraer información relevante.
- Procesamiento y modelado: Desarrollar scripts en VSCode usando Python y bibliotecas como Sklearn para construir modelos de recomendación basados en los datos extraídos.
- Evaluación y visualización: Utilizar Matplotlib para graficar métricas de rendimiento del modelo y presentar resultados clave.

Con este conjunto de herramientas se nos permite trabajar eficientemente en todo el ciclo de desarrollo de modelos de recomendación, desde la extracción de datos hasta la evaluación del modelo.

## 2.4 MARCO LEGAL

### 2.4.1 MARCO LEGAL INTERNACIONAL

#### 2.4.1.1 REGLAMENTO GENERAL DE PROTECCIÓN DE DATOS (GDPR) – UNIÓN EUROPEA

El GDPR, en vigor desde mayo de 2018, es considerado el estándar internacional en la protección de datos personales. A pesar de ser una regulación europea, ha influido significativamente en la creación de leyes similares en otras regiones debido a su alcance extraterritorial. Aspectos Clave:

Consentimiento explícito: Las empresas deben obtener un consentimiento claro y específico de los usuarios antes de recopilar y procesar datos.

Derecho al olvido: Los usuarios pueden solicitar la eliminación de sus datos personales en cualquier momento.

Transferencias internacionales de datos: Regula la transferencia de datos personales fuera de la UE, exigiendo garantías adecuadas en países receptores.

Sanciones: Las empresas que incumplen pueden enfrentar multas de hasta el 4% de sus ingresos anuales globales o €20 millones, lo que sea mayor.

Relevancia para Latinoamérica: Países como Brasil (Ley General de Protección de Datos, LGPD) y Argentina (Ley 25.326) han adaptado elementos del GDPR en sus legislaciones locales, marcando una tendencia que Honduras podría seguir.

#### 2.4.1.2 LEY MODELO DE COMERCIO ELECTRÓNICO DE LA UNCITRAL

Adoptada en 1996 por la Comisión de las Naciones Unidas para el Derecho Mercantil Internacional (UNCITRAL), esta ley proporciona principios generales para regular las transacciones electrónicas a nivel global. Puntos destacados:

Reconoce la validez jurídica de documentos y firmas electrónicas.

Establece criterios para la autenticación de las partes involucradas en transacciones digitales.

Fomenta la equivalencia funcional entre documentos físicos y digitales.

Impacto regional: Países como México y Colombia han adoptado disposiciones basadas en esta ley para dar seguridad jurídica al comercio electrónico. En Honduras, aunque existe una Ley de Comercio Electrónico, podría beneficiarse al alinearse más estrechamente con estas disposiciones.

#### 2.4.1.3 CONVENIO DE BUDAPEST SOBRE CIBERDELINCUENCIA

Este tratado internacional, adoptado en 2001 y respaldado por el Consejo de Europa, es el principal instrumento legal para combatir delitos informáticos, garantizando la cooperación internacional. Principales disposiciones:

Criminalización de actos como el acceso no autorizado, la interceptación ilícita y el fraude informático.

Establecimiento de procedimientos para recolectar evidencia digital en investigaciones penales.

Promoción de la cooperación entre países para enfrentar cibercriminos transnacionales.

Adopción en Latinoamérica: Aunque pocos países de la región lo han ratificado, como República Dominicana y Chile, su implementación ofrece un marco robusto para proteger plataformas digitales como LynxPay de ataques cibernéticos.

#### 2.4.1.4 ACUERDO DE ASOCIACIÓN TRANSPACÍFICO (TPP) Y SU EVOLUCIÓN (CPTPP)

El TPP, reformulado como el Acuerdo Integral y Progresista de Asociación Transpacífico (CPTPP) en 2018, incluye disposiciones específicas sobre comercio digital y protección de datos.

Aspectos relevantes:

Prohíbe exigir a las empresas la localización de servidores en un país específico, promoviendo la interoperabilidad digital.

Establece principios para la transferencia internacional de datos, respetando la privacidad de los usuarios.

Promueve la eliminación de barreras arancelarias en bienes y servicios digitales.

Impacto potencial: Si Honduras busca integrarse a tratados de libre comercio con enfoque digital, como el CPTPP, deberá garantizar estándares internacionales en protección de datos y comercio electrónico.

#### 2.4.1.5 DECLARACIÓN DE PRINCIPIOS DE LA OCDE SOBRE PRIVACIDAD Y FLUJOS TRANSFRONTERIZOS DE DATOS PERSONALES

Adoptada en 1980 y actualizada en 2013, esta declaración establece lineamientos para la protección de la privacidad en el procesamiento de datos personales, particularmente en el contexto de transferencias transfronterizas. Principios clave:

Limitación en la recolección de datos: Solo deben recopilarse los datos necesarios para un propósito específico.

Calidad de los datos: Los datos deben ser precisos, completos y estar actualizados.

Responsabilidad: Las empresas deben ser responsables del cumplimiento de estos principios en sus operaciones globales.

Relevancia regional: Los principios de la OCDE han servido como base para marcos legales en países de América Latina, como Chile y México, y pueden ser una guía para el desarrollo normativo en Honduras.

#### 2.4.1.6 NORMAS DE PROTECCIÓN AL CONSUMIDOR EN EL COMERCIO ELECTRÓNICO

La Conferencia de las Naciones Unidas sobre Comercio y Desarrollo (UNCTAD) ha promovido recomendaciones sobre derechos del consumidor en el entorno digital, que han sido adoptadas por varios países. Derechos destacados:

Acceso a información clara sobre términos y condiciones de compra.

Derecho a la devolución de productos comprados en línea dentro de un plazo razonable.

Transparencia en los costos asociados a la compra (impuestos, tarifas de envío).

Casos de adopción: En Latinoamérica, Brasil y Colombia han incorporado estas recomendaciones en sus leyes locales, ofreciendo mayor protección a los consumidores y fomentando la confianza en el comercio electrónico.

#### 2.4.1.7 ACUERDO DE LIBRE COMERCIO ENTRE ESTADOS UNIDOS, MÉXICO Y CANADÁ (USMCA)

Este acuerdo, en vigor desde 2020, incluye disposiciones innovadoras sobre comercio digital que pueden servir como modelo para países de América Latina. Características principales:

Protección de los datos personales de los usuarios en plataformas digitales.

Prohibición de imponer aranceles a productos digitales, como aplicaciones y software.

Fomento de la colaboración entre las partes para mejorar la seguridad cibernética.

#### 2.4.2 PROTECCIÓN DE DATOS EN HONDURAS

En Honduras, la protección de datos personales y la privacidad de la información constituyen áreas de creciente importancia, reflejando las tendencias globales hacia la digitalización y el incremento en la utilización de tecnologías de información. La Constitución de

Honduras establece salvaguardas para la protección del honor, la intimidad personal y familiar, y la propia imagen en su artículo 76, así como el derecho a la inviolabilidad y al secreto de las comunicaciones en su artículo 100. Adicionalmente, el artículo 182 reconoce el Habeas Data como una garantía constitucional para la protección de datos personales.

A pesar de estos fundamentos constitucionales, Honduras aún enfrenta desafíos en la implementación práctica de un marco legislativo robusto para la protección de datos personales. Desde abril de 2018, el Congreso Nacional ha mantenido en suspenso la aprobación del proyecto de Ley de Protección de Datos Personales. Este proyecto es vital dada la ausencia de legislación específica en la materia y propone innovaciones en definiciones y mecanismos para la protección de datos. No obstante, algunas áreas de la propuesta legislativa requieren clarificación para evitar interpretaciones erróneas o la aplicación arbitraria de la ley. (Paz, 2022).

La situación actual refleja la necesidad urgente de avanzar en la aprobación e implementación de una legislación que proteja de manera efectiva los datos personales en Honduras. Tal legislación no solo protegería los derechos de privacidad de los individuos, sino que también establecería un marco de confianza esencial para el desarrollo de la economía digital en el país.

## 2.4.3 MARCO LEGAL NACIONAL

### 2.4.3.1 LEY DEL COMERCIO ELECTRÓNICO (DECRETO NO. 149-2020)

En respuesta a la necesidad de mantener activo el sector económico durante situaciones de emergencia, Honduras ha tomado medidas para simplificar administrativamente la implementación de mecanismos de comercio y firma electrónicos. La Ley sobre firmas electrónicas ha sido reformada para permitir el uso de firmas electrónicas en una amplia gama de casos, reconociendo y otorgando efectos legales equivalentes a las firmas electrónicas avanzadas y a otros medios de identificación confiables basados en tecnologías como PKI, firma biométrica, sistemas en la nube y doble factor.

La ley también reconoce la validez de las firmas electrónicas y certificados emitidos dentro y fuera de Honduras, siempre que tengan un nivel de confiabilidad similar. Esto facilita el reconocimiento transfronterizo de tales firmas y certificados, fomentando así el comercio electrónico internacional. También establece que entidades del sector público o privado pueden

designar responsables para certificar autorizaciones, agilizando las operaciones electrónicas.

Un aspecto relevante es la autorización para que todo tipo de actos, contratos y negocios jurídicos puedan celebrarse por medios electrónicos, siempre que se pueda demostrar fehacientemente la voluntad de las partes. Además, se ha habilitado a personas jurídicas y entidades del Estado a celebrar reuniones de órganos de gobierno y supervisión mediante medios electrónicos, validando las decisiones tomadas en dichas reuniones. (Zelaya, 2020).

Esta ley establece las bases legales para las transacciones realizadas mediante medios electrónicos y busca fomentar la confianza de consumidores y empresas en el entorno digital. Aspectos clave:

**Reconocimiento de firmas electrónicas:** Se establece la validez jurídica de las firmas electrónicas y avanzadas, que son equivalentes a las manuscritas, siempre que cumplan con los requisitos técnicos de autenticidad.

**Validez de documentos electrónicos:** Garantiza que los documentos generados en formato digital tengan el mismo peso legal que los físicos, siempre que se cumplan los estándares técnicos.

**Obligaciones de las plataformas:** Las empresas deben informar claramente sobre los términos y condiciones de sus servicios, costos asociados y políticas de devolución.

**Protección de datos personales:** Aunque no desarrolla ampliamente este tema, se menciona la obligación de garantizar la privacidad en las transacciones electrónicas.

**Relevancia para LynxPay:** Al implementar sistemas de recomendación, la plataforma debe cumplir con las disposiciones sobre transparencia y notificación a los usuarios sobre el uso de sus datos para personalizar recomendaciones.

#### 2.4.3.2 LEY DE PROTECCIÓN DE DATOS PERSONALES (DECRETO NO. 25-2022)

Esta ley regula el tratamiento de los datos personales, estableciendo principios fundamentales que deben cumplir las empresas para proteger la privacidad de los usuarios. Principios y disposiciones clave:

**Consentimiento informado:** Las empresas deben obtener el consentimiento explícito de los usuarios antes de recopilar y procesar sus datos personales.

Finalidad específica: Los datos recopilados solo pueden ser utilizados para el propósito autorizado por el usuario.

Derecho de acceso, rectificación y cancelación (ARCO): Los usuarios tienen derecho a acceder a sus datos, corregir errores y solicitar su eliminación.

Transferencia de datos: Establece restricciones para la transferencia de datos personales a terceros, exigiendo garantías de seguridad.

Sanciones: Se imponen multas y sanciones administrativas para quienes incumplan las disposiciones de la ley.

Impacto en LynxPay: Deberá implementar políticas claras sobre el manejo de datos, como la creación de términos de uso que expliquen cómo se recopilan, almacenan y utilizan los datos para los sistemas de recomendación.

#### 2.4.3.3 LEY DE PROTECCIÓN AL CONSUMIDOR (DECRETO NO. 24-2008)

Esta ley protege los derechos de los consumidores en todas las transacciones comerciales, incluyendo aquellas realizadas en plataformas digitales. Disposiciones relevantes:

Derecho a la información: Los consumidores deben recibir información completa, clara y veraz sobre los productos o servicios ofrecidos.

Política de devoluciones: Los usuarios tienen derecho a devolver productos adquiridos en línea dentro de un período razonable.

Resolución de conflictos: Se garantiza el acceso a mecanismos alternativos de resolución de conflictos, como mediación y arbitraje, en disputas relacionadas con el comercio electrónico.

Publicidad engañosa: Prohíbe la difusión de información falsa o que induzca a error a los consumidores.

Implicaciones para LynxPay: Deberá garantizar que los comerciantes que utilicen su plataforma cumplan con estas disposiciones, brindando a los consumidores protección en sus transacciones.

#### 2.4.3.4 CÓDIGO PENAL (REFORMAS 2021)

El Código Penal hondureño incluye disposiciones específicas sobre delitos relacionados

con el comercio electrónico y la protección de datos, fundamentales para garantizar la seguridad de plataformas digitales.

Delitos relacionados con la ciberseguridad:

Acceso no autorizado a sistemas informáticos: Penaliza la intrusión en sistemas digitales sin consentimiento, con penas de 4 a 8 años de prisión.

Uso indebido de datos personales: Criminaliza la recopilación, venta o distribución de datos personales sin autorización, con penas que oscilan entre 2 y 6 años de prisión.

Suplantación de identidad digital: Castiga a quienes utilicen datos personales para hacerse pasar por otra persona, con penas de 3 a 5 años de prisión.

Fraude informático: Incluye la manipulación de sistemas o datos con fines fraudulentos, con sanciones de 5 a 10 años de prisión.

Relevancia para LynxPay: La plataforma debe implementar medidas de seguridad robustas para proteger los datos de los usuarios y evitar ser blanco de ataques cibernéticos. También debe garantizar que los algoritmos de recomendación no vulneren la privacidad de los usuarios.

#### 2.4.3.5 OTRAS LEYES COMPLEMENTARIAS

Ley sobre Firmas Electrónicas (Decreto No. 131-2018): Regula el uso y reconocimiento legal de las firmas electrónicas, estableciendo estándares de autenticidad y validez.

Relevancia: Permite a LynxPay establecer contratos y acuerdos con sus usuarios de forma digital.

Ley de Transparencia y Acceso a la Información Pública (Decreto No. 170-2006): Asegura que las empresas y entidades públicas sean transparentes en el manejo de datos personales.

Implicación: Exige a LynxPay informar a sus usuarios sobre el uso y protección de sus datos.

Ley del Instituto Nacional de Telecomunicaciones (Decreto No. 185-1995): Regula la infraestructura digital, incluyendo las redes de telecomunicaciones utilizadas para el comercio electrónico.

## **CAPÍTULO III. METODOLOGÍA**

### **3.1 CONGRUENCIA METODOLÓGICA**

Congruencia En este capítulo se presenta el marco teórico, el cual constituye la base esencial sobre la que se sustenta toda la investigación. Además, se explicará en detalle la congruencia metodológica, describiendo cómo se alinean los diferentes componentes del estudio para garantizar la coherencia y validez de los resultados. Se abordará el enfoque de la investigación, especificando si es cualitativo, cuantitativo o mixto, y se detallarán los métodos utilizados para la recolección y análisis de datos.

Asimismo, se describirá el diseño de la investigación, incluyendo el tipo de estudio, las etapas del proceso investigativo y las estrategias implementadas para llevar a cabo el análisis. Se determinará la población objetivo y se seleccionará una muestra representativa que permita obtener resultados significativos y generalizables.

Se explicarán las técnicas e instrumentos aplicados para la recolección de datos, y se justificará su elección en función de los objetivos del estudio. Finalmente, se enumerarán las fuentes de información utilizadas, que pueden incluir bases de datos, artículos científicos, libros y otros recursos relevantes que proporcionaron el fundamento teórico y empírico necesario para la investigación.

#### **3.1.1 MATRIZ METODOLÓGICA**

La herramienta de matriz metodológica es fundamental para optimizar el tiempo y los esfuerzos dedicados a la investigación. Esta herramienta permite organizar de manera eficiente los objetivos específicos del estudio, relacionándolos con las variables dependientes e independientes, así como con sus respectivas dimensiones. De esta forma, se establece una secuencia lógica que facilita el análisis y la interpretación de los datos. La matriz metodológica no solo ayuda a estructurar el proceso investigativo, sino que también asegura que cada componente del estudio esté alineado con los objetivos planteados. Esto incluye la identificación de las variables clave y la definición de sus dimensiones, lo que permite una comprensión más profunda y detallada del fenómeno en estudio. Se procederá a desarrollar la matriz metodológica, la cual se presenta en la tabla 2. Esta tabla ilustrará cómo se interrelacionan los diferentes elementos del estudio, proporcionando una visión clara y ordenada del enfoque metodológico adoptado.

**Tabla 1. Matriz de congruencia metodológica.**

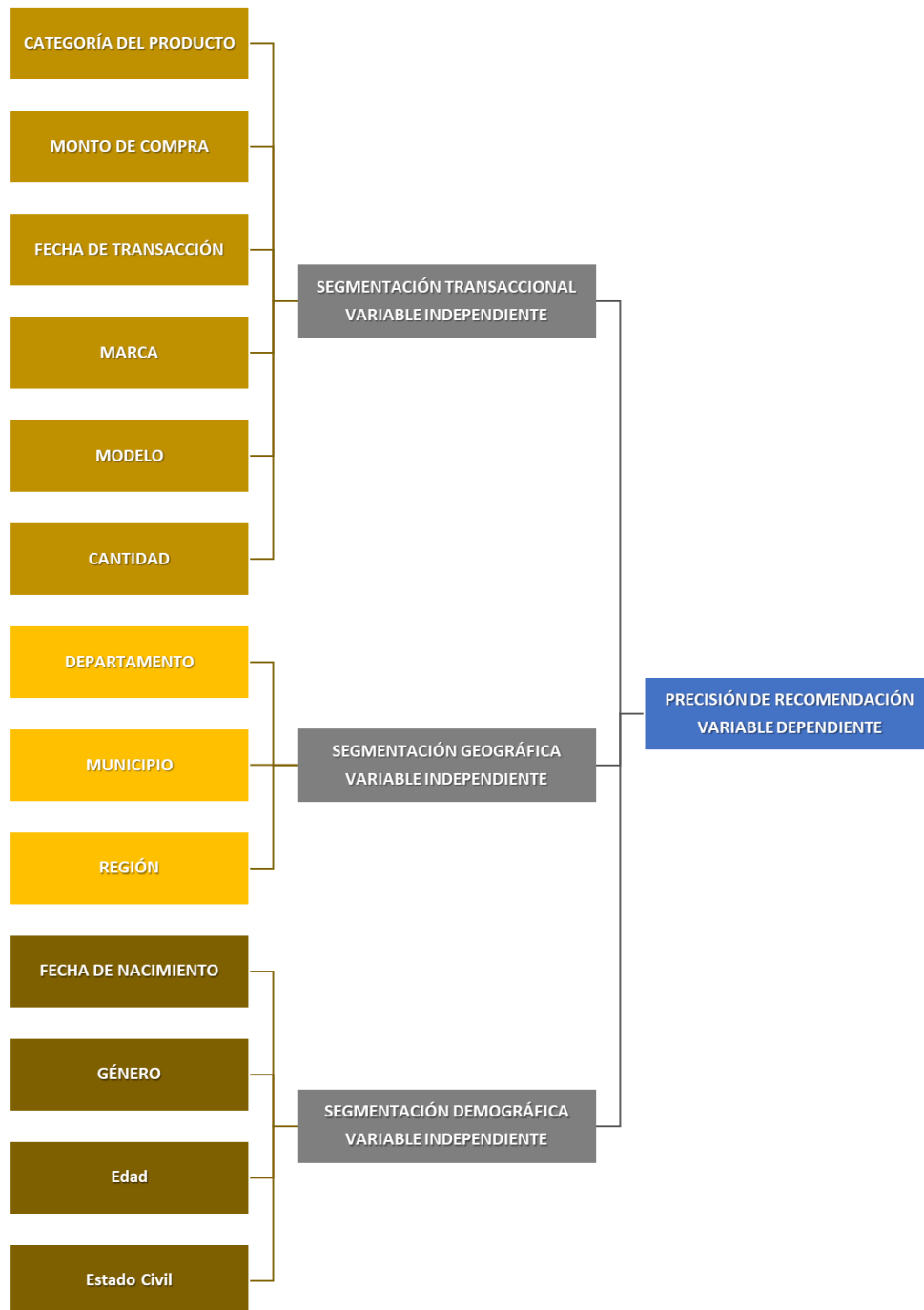
<b>Título: Propuesta de Implementación de Técnicas de Clusterización para la Recomendación Personalizada de Productos para una Plataforma de e-Commerce en Honduras</b>				
<b>Problema</b>	<b>Objetivo General</b>	<b>Objetivos Especifico</b>	<b>Variable Independiente</b>	<b>Variable Dependiente</b>
¿Cómo la implementación de un sistema de recomendación basado en clusterización en la plataforma LynxPay puede mejorar la experiencia de usuario y aumentar las tasas de conversión en el contexto del comercio electrónico en Honduras?	El objetivo principal de este proyecto de tesis es validar la viabilidad y efectividad de las técnicas de clusterización como herramienta para la segmentación de clientes en la plataforma de comercio electrónico LynxPay, con el propósito de generar recomendaciones de productos personalizadas que optimicen la experiencia del usuario, aumenten las tasas de conversión y mejoren la competitividad de las tiendas en línea en Honduras.	Recopilar y analizar de forma exhaustiva los datos disponibles en la plataforma LynxPay, tales como transacciones realizadas y productos añadidos al carrito, con el objetivo de identificar las principales variables que influyen en el comportamiento de compra de los usuarios.	Segmentación Transaccional	Precisión de Recomendación
		Aplicar diferentes algoritmos de clusterización, como K-MEANS y DBSCAN, con el propósito de segmentar la base de clientes en grupos homogéneos basados en variables clave de comportamiento, como la frecuencia de compra y el tipo de productos adquiridos.		
		Desarrollar y validar modelos de recomendación personalizados para los segmentos identificados, analizando el impacto de estos modelos en las métricas clave de rendimiento, tales como la precisión de las recomendaciones.	Segmentación Geográfica	
		Explorar la viabilidad económica de implementar modelos de clusterización en LynxPay mediante un análisis de costo-beneficio, tomando en cuenta los recursos computacionales, el tiempo de procesamiento de los datos y el impacto esperado en las ventas de las tiendas participantes.	Segmentación Demográfica	
		Proponer estrategias de mejora basadas en los resultados obtenidos durante la validación de los modelos, sugiriendo posibles optimizaciones para la implementación futura de los sistemas de recomendación en LynxPay, con miras a incrementar la personalización y mejorar la experiencia del usuario.		

Fuente: Elaboración propia

### 3.1.2 ESQUEMA DE VARIABLES DE ESTUDIO

En esta sección se describen las variables en función de la relación o conexión que existe entre ellas. La Figura 4, que se presenta a continuación, ofrece un resumen de las relaciones entre las variables utilizadas en esta investigación.

**Figura 4. Resumen de esquema de las variables de estudio.**



Fuente: Elaboración propia

### 3.1.3 OPERACIONALIZACIÓN DE LAS VARIABLES

La operacionalización de las variables implica descomponer los elementos que forman la estructura de la hipótesis, enfocándose especialmente en las variables. Este proceso se completa cuando las variables se dividen en dimensiones, las cuales se traducen en indicadores que permiten su observación y medición directa. La operacionalización es crucial porque a través de ella se identifican y definen los aspectos y elementos que se desean cuantificar, conocer y registrar, con el objetivo de llegar a conclusiones precisas. (Espinoza Freire, 2019)

A continuación, se detalla una tabla con la operacionalización de las variables a utilizar en la presente investigación:

**Tabla 2. Tabla de Operacionalización de variables independientes**

Variable	Definición Conceptual	Definición Operacional	Dimensiones	Ítems
Segmentación Transaccional	La segmentación transaccional se basa en el análisis de datos relacionados con las transacciones realizadas por los consumidores, como el monto total gastado, la frecuencia de compra, el tipo de producto adquirido y los métodos de pago utilizados. Este enfoque permite identificar patrones de gasto y comportamiento de los consumidores para desarrollar estrategias personalizadas de marketing, según autores como, Kotler, Philip y Gary (2017).	En el sistema de recomendación, la segmentación transaccional se implementará mediante el análisis de las características y el historial de transacciones de los usuarios. Por ejemplo, el sistema puede clasificar a los clientes en segmentos como compradores frecuentes, ocasionales o de alto valor según el monto acumulado gastado, sugiriendo promociones y productos específicos para cada grupo.	Categoría del Producto	Iluminación Ferretería Equipos Electricidad Deportes Construcción Automotriz Jardinería Oficina Pintura Seguridad Tornillería
			Monto de Compra	Lempiras de 1.00 a n
			Fecha de Transacción	Días del calendario Gregoriano
			Marca	Texto
			Modelo	texto
			Cantidad	Valores enteros de 1 a n

Fuente: Elaboración propia

**Continuación Tabla 3. Tabla de Operacionalización de variables independientes**

Variable	Definición Conceptual	Definición Operacional	Dimensiones	Ítems
Segmentación Geográfica	La segmentación geográfica clasifica a los consumidores en función de su ubicación, considerando factores como país, región, ciudad e incluso clima. En el contexto de e-commerce, este enfoque permite adaptar la oferta de productos y servicios a las necesidades y preferencias locales, lo que mejora la relevancia del producto para el usuario final, Kotler y Keller (2017). La segmentación geográfica es útil para identificar patrones de compra y preferencias de consumo que varían entre diferentes regiones, ayudando a personalizar las recomendaciones y estrategias de distribución.	La segmentación geográfica en la plataforma e-commerce permitirá identificar patrones de compra según la región de los usuarios en Honduras. Esto ayudará a que el sistema de recomendación sugiera productos que sean populares en ciertas zonas o ajustados a las características y demandas específicas de cada área. Por ejemplo, productos más apropiados para el clima o las costumbres locales de cada región.	Municipio	La Ceiba ... Yorito (298 en total)
			Departamento	Atlántida Choluteca Colón Comayagua Copán Cortés El Paraíso Francisco Morazán Gracias a Dios Intibucá Islas de la Bahía La Paz Lempira Ocotepeque Olancho Santa Bárbara Valle Yoro
			Región	Zona Norte Zona Sur Zona Oriente Zona Occidente
Segmentación Demográfica	La segmentación demográfica clasifica a los consumidores en función de características personales como edad, género, nivel educativo, estado civil, ingreso económico, ocupación y tamaño del hogar. Según Kotler y Keller (2017), este enfoque permite adaptar estrategias de marketing a las características básicas y necesidades generales de los consumidores.	En el sistema de recomendación, la segmentación demográfica se utilizará para adaptar las ofertas de productos según las características personales de los usuarios. Por ejemplo, podría sugerir productos de juguetes y ropa infantil para usuarios con hijos pequeños o herramientas de bricolaje para adultos jóvenes con intereses específicos.	Fecha de Nacimiento	Día/Mes/Año (DD/MM/AAAA)
			Genero	Masculino, Femenino, Otro
			Edad	Rango de edad del usuario (ej. 18-24, 25-34, etc.).
			Estado Civil	Soltero/a, casado/a, otros.

Fuente: Elaboración propia

### 3.1.4 HIPÓTESIS

Una hipótesis de investigación es una afirmación provisional que sugiere una posible relación entre dos o más variables. Se formula con el objetivo de ser comprobada o refutada a través de la investigación científica<sup>1</sup>. Es una suposición educada basada en observaciones previas y conocimientos teóricos, que guía el diseño y la ejecución del estudio.

Las hipótesis de investigación se entienden como afirmaciones provisionales sobre las posibles relaciones entre dos o más variables. (Hernández Sampieri, Collado, & Baptista Lucio, 2014)

Hipótesis nula ( $H_0$ ) sobre el problema planteado en esta investigación:

El sistema de recomendación que utiliza modelos de agrupación no mejora el rendimiento en comparación con la implementación de algoritmos básicos de filtrado colaborativo. Esto implica que, al aplicar técnicas de clusterización, no se observa una mejora significativa en la precisión y relevancia de las recomendaciones en la plataforma de e-commerce.

Hipótesis de investigación ( $H_1$ ) planteada es la siguiente:

El sistema de recomendación que utiliza modelos de agrupación mejora el rendimiento en comparación con la implementación de algoritmos básicos de filtrado colaborativo. Esto significa que, mediante la aplicación de técnicas de clusterización, se logra una mejora significativa en la precisión y relevancia de las recomendaciones, proporcionando sugerencias más útiles y personalizadas para los usuarios de la plataforma de e-commerce.

Estas hipótesis se fundamentan en los objetivos y la problemática identificada en la investigación, así como en el análisis de los métodos de clusterización aplicados (como K-Means y DBSCAN) para lograr una mejor segmentación y personalización.

## 3.2 ENFOQUE Y MÉTODOS

El enfoque de la investigación se refiere a la perspectiva general que orienta el diseño del estudio, definiendo su naturaleza y alcance. En contraste, los métodos de investigación incluyen las técnicas específicas empleadas para la recolección y el análisis de datos en el contexto del estudio. Hernández Sampieri, Collado, & Baptista Lucio, (2014).

La Figura 5 resume las definiciones contenidas en esta sección.

<b>Enfoque</b>	<b>Cuantitativo</b>
<b>Alcance</b>	<b>Correlacional</b>
<b>Tipo de Diseño</b>	<b>No experimental / Longitudinal / Retrospectivo</b>
<b>Método</b>	<b>Minería de datos / Modelos de agrupación</b>
<b>Instrumentos</b>	<b>Python y DBeaver</b>

**Figura 5 Resumen de esquema de las variables de estudio**

Fuente: Elaboración propia

### 3.2.1 ENFOQUE

Este estudio utiliza una metodología cuantitativa, enfocada en el análisis numérico de los datos para obtener información precisa y cuantificable. Este enfoque es ideal para los objetivos de la investigación, que buscan identificar los factores que afectan la precisión de los modelos de recomendación.

### 3.2.2 ALCANCE

El alcance se define como el conjunto de restricciones o límites que establecen los parámetros de un proyecto de investigación. En otras palabras, el alcance determina qué características, factores, población objetivo o temas específicos serán incluidos en el estudio y cuáles quedarán excluidos. Hernández Sampieri, Collado, & Baptista Lucio, (2014).

Considerando lo anterior, podemos afirmar que esta investigación tiene un alcance correlacional, cuyo objetivo es identificar los factores que influyen en las compras de los clientes a través del análisis de datos. Esto permitirá proponer el desarrollo de un algoritmo que mejore la precisión de los modelos de recomendación para las tiendas de e-commerce en Honduras.

### 3.2.3 DISEÑO

Los diseños de investigación transeccional o transversal recopilan datos en un único momento. Su objetivo es describir variables y analizar su impacto y relaciones en ese instante específico. Este enfoque puede compararse con una “fotografía” que captura una situación en un punto determinado del tiempo. Hernández Sampieri, Collado, & Baptista Lucio, (2014).

La investigación es de tipo no experimental, transversal y retrospectiva, ya que no se manipulan deliberadamente las variables, sino que se observan en su contexto natural. Este enfoque permite analizar las variables tal como se presentan en el entorno real, sin intervención del investigador. Además, la investigación es transversal, lo que significa que la recolección de datos se realiza en un único momento en el tiempo. Este método es adecuado para la explicación del problema, ya que solo requiere una medición de datos en un punto específico.

La evaluación de las variaciones en la predicción de la probabilidad de recompra se lleva a cabo mediante una sola recolección de datos. Para ello, se utiliza el conjunto de información obtenido de la base de datos de la plataforma LynxPay. Esta base de datos proporciona un registro detallado de las transacciones y comportamientos de los usuarios, lo que permite realizar un análisis exhaustivo de los factores que influyen en la precisión de recomendación. Al utilizar datos históricos y actuales, se puede obtener una visión completa y precisa de las tendencias y patrones de comportamiento de los clientes.

### 3.2.4 MÉTODOS

Las técnicas de minería de datos que respaldan los análisis de datos pueden aplicarse con dos objetivos principales: describir el conjunto de datos objetivo o predecir resultados utilizando algoritmos de machine learning.

Estos métodos se emplean para organizar y filtrar datos, revelando la información más útil, desde la detección de fraudes hasta el análisis de comportamientos de usuarios, identificación de cuellos de botella e incluso violaciones de seguridad.

Para los fines de esta investigación, se está utilizando la minería de datos para realizar un análisis exploratorio de datos (EDA) y aplicar modelos predictivos como K-Means, DBSCAN y técnicas híbridas de Clusterización + Modelos Predictivos, Segmentación Basada en Optimización. Estos métodos permiten explorar en profundidad los patrones presentes en el

conjunto de datos proporcionado, con el objetivo de identificar con mayor precisión los modelos de recomendación.

### 3.2.5 INSTRUMENTOS

En este estudio, se empleará un conjunto de datos proporcionado por la Plataforma LynxPay, que abarca las transacciones realizadas entre los años 2021 y 2024. El data set será extraído utilizando la herramienta DBeaver como cliente de base de datos y Putty para conectarse al servidor. Este conjunto de datos contiene información crucial que permitirá realizar la investigación de manera efectiva, proporcionando una base sólida para evaluar la precisión de las recomendaciones generadas por el modelo.

### 3.3 DISEÑO DE LA INVESTIGACIÓN

Una vez que se ha definido claramente el planteamiento del problema se ha establecido el alcance inicial de la investigación y se han formulado las hipótesis (o se ha decidido no hacerlo debido a la naturaleza del estudio), el investigador debe considerar cómo responder de manera práctica y concreta a las preguntas de investigación y alcanzar los objetivos propuestos. Esto requiere seleccionar o desarrollar uno o más diseños de investigación y aplicarlos al contexto específico del estudio. El término “diseño” se refiere al plan o estrategia diseñada para obtener la información necesaria para responder al planteamiento del problema. Hernández Sampieri, Collado, & Baptista Lucio, (2014).

**Tabla 4. Tabla de actividades de la investigación.**

Diseño de la Investigación			
Actividades		Tiempo	Responsable
<b>Fase 1</b>	<b>Preparación de Instrumentos</b>	1 semana	Keneth Mauricio Pérez Brizuela Andy Javier Reyes Erazo
1.1	Elección de software a utilizar.		
1.2	Técnicas para extracción de datos.		
<b>Fase 2</b>	<b>Recopilación de Datos</b>	1 semana	Andy Javier Reyes Erazo
2.1	Limpieza de datos.		
2.2	Creación de dashboard de situación actual.		
<b>Fase 3</b>	<b>Análisis de Datos</b>	3 semanas	Keneth Mauricio Pérez Brizuela
3.1	Analizar datos obtenidos.		
3.2	Crear modelo de recomendación en Python.		

Fuente: Elaboración propia

### Continuación Tabla 5. Tabla de actividades de la investigación.

<b>Fase 4</b>	<b>Interpretación y Documentación</b>	2 semanas	Keneth Mauricio Pérez Brizuela
4.1	Análisis de resultados obtenidos.		
4.2	Despliegue de modelo de recomendación.		
<b>Fase 5</b>	<b>Conclusiones y Presentación</b>	1 semana	Andy Javier Reyes Erazo
5.1	Exportar datos y graficas.		
5.2	Informe de conclusiones.		

Fuente: Elaboración propia

#### 3.3.1 POBLACIÓN

El objetivo principal de esta investigación es obtener información relevante sobre las características de una población definida como el conjunto total de elementos que comparten características comunes, constituyendo el universo de análisis. La población es un concepto esencial en cualquier investigación, ya que permite abordar de manera integral el problema de estudio, facilitando el entendimiento de patrones, tendencias y dinámicas.

En términos generales, la población puede referirse a un grupo de individuos de la misma especie dentro de un área geográfica específica y un periodo de tiempo determinado. Este enfoque es aplicable en diversas disciplinas, como la demografía, la biología, la sociología y la geografía. En el contexto humano, la población incluye aspectos como el número total de personas, su distribución geográfica, y características demográficas como edad, género, etnia, nivel educativo y otros factores dinámicos, como tasas de natalidad, mortalidad y migraciones. Estas características son clave para comprender las dinámicas y particularidades del grupo analizado.

Para este estudio, la población incluye todos los registros de usuarios y clientes disponibles en la base de datos de la plataforma LynxPay, siempre y cuando cumplan con los siguientes criterios establecidos para garantizar que la información recolectada sea relevante, precisa y alineada con los objetivos de la investigación:

#### **Criterios de Inclusión:**

- Clientes registrados en tiendas en línea alojadas en LynxPay.
- Clientes que hayan realizado al menos una compra documentada.
- Clientes cuyas compras hayan sido efectuadas en una de las tiendas activas dentro de la plataforma LynxPay.
- Transacciones cuya facturación haya ocurrido desde el año 2021 en adelante.

### **Características de los Datos Utilizados:**

- Cantidad de registros: Se identificaron 4,842 registros únicos en la base de datos, provenientes de 2 comerciantes activos en la plataforma LynxPay.
- Precisión de la información: Los datos disponibles son completos, verificados y representan el total de las interacciones relevantes dentro del periodo definido.

### **Razonamiento para Utilizar El Total de la Población**

En este estudio, se tomó la decisión de trabajar con la totalidad de los datos disponibles, es decir, con toda la población identificada, debido a las siguientes razones:

- Tamaño manejable de la población: La población total consta de 4,842 registros, una cantidad suficientemente grande para realizar análisis significativos y suficientemente pequeña como para ser procesada en su totalidad mediante herramientas tecnológicas como Python. Este tamaño permite trabajar con todos los datos disponibles sin comprometer la capacidad computacional ni la calidad del análisis.
- Calidad y completitud de los datos: Los registros disponibles en la base de datos son completos, lo que asegura que cada cliente y transacción incluidos en el estudio aporten información valiosa para alcanzar los objetivos planteados. No existen limitaciones significativas que justifiquen la selección de una muestra, ya que trabajar con toda la población elimina cualquier riesgo de exclusión de variaciones relevantes en los datos.
- Reducción de sesgos: Al analizar toda la población, se evita el sesgo de selección que podría ocurrir al trabajar con una muestra. Esto significa que se considerarán todas las variaciones posibles dentro del conjunto de datos, incluyendo patrones de comportamiento menos comunes que podrían ser ignorados en un análisis basado en una muestra. Este enfoque asegura una representación completa y precisa del comportamiento de los usuarios y comerciantes en la plataforma.
- Explotación integral de los datos: La base de datos utilizada en el estudio fue segmentada mediante filtros específicos aplicados con Python, permitiendo identificar únicamente los registros que cumplen con los criterios de inclusión.

Trabajar con toda la población permite realizar análisis detallados y exhaustivos, maximizando la extracción de información útil y relevante para el desarrollo de sistemas de clusterización y recomendaciones personalizadas.

- Validez de los resultados: Utilizar toda la población asegura que los resultados obtenidos no dependan de estimaciones estadísticas o generalizaciones, sino que representen directamente la realidad del conjunto de datos. Esto aumenta la validez interna y externa de los hallazgos, lo que significa que los resultados son tanto precisos para la población estudiada como aplicables en contextos similares.

### **Metodología de Recolección de Datos**

La obtención de los datos para este estudio se realizó mediante el análisis de la base de datos de LynxPay, utilizando herramientas tecnológicas avanzadas como Python para filtrar y segmentar los registros. Este enfoque permitió:

- Identificar los registros que cumplen estrictamente con los criterios de inclusión.
- Asegurar la calidad de los datos mediante procesos de limpieza y validación.
- Facilitar la extracción de características clave, como datos demográficos, patrones de compra y comportamiento de los clientes.
- El enfoque tecnológico adoptado no solo garantizó la precisión en la selección de la población, sino que también hizo posible trabajar con grandes volúmenes de datos de manera eficiente.

### **Ventajas del Enfoque Censal**

Trabajar con toda la población disponible en la base de datos presenta varias ventajas significativas para esta investigación:

- Representación Completa: Al incluir a todos los clientes y transacciones que cumplen con los criterios establecidos, se asegura que todos los patrones de comportamiento estén representados en el análisis.
- Mayor Precisión en los Resultados: No es necesario realizar inferencias estadísticas, ya que los resultados se basan en la totalidad de los datos reales.

- Capacidad de Identificar Subgrupos: Este enfoque permite segmentar la población en subgrupos específicos, como clientes frecuentes, clientes ocasionales y diferentes tipos de comerciantes, para un análisis más detallado.
- Eliminación de Riesgos de Exclusión: Al no depender de una muestra, no se excluyen posibles registros atípicos que podrían ser valiosos para el análisis.

### 3.3.2 MUESTRA

La representatividad de una muestra permite que los resultados obtenidos puedan generalizarse, primero a la población accesible, que incluye a los individuos de la población objetivo que están disponibles para participar en el estudio, y luego a la población objetivo en general. Para que una muestra sea representativa, es fundamental que haya sido seleccionada de forma aleatoria, lo que asegura que todos los integrantes de la población accesible y objetivo hayan tenido la misma probabilidad de ser elegidos. Este proceso corresponde al uso de técnicas de muestreo probabilístico. Además, el tamaño de la muestra debe ser suficiente para reflejar numéricamente a la población de origen en cuanto a la distribución de la variable que se está analizando. De esta manera, se puede garantizar que los resultados sean aplicables al grupo más amplio del que se extrajo la muestra. (Otzen & Manterola, 2017).

Dado que se tiene acceso completo a todos los registros disponibles en la base de datos para esta investigación, no fue necesario diseñar una muestra específica. En lugar de ello, se utilizó la totalidad de los datos disponibles, lo que permitió un análisis exhaustivo y detallado de la población en estudio. Esta metodología asegura que todas las variaciones y características presentes en la base de datos sean consideradas, proporcionando una visión integral y precisa de los elementos investigados. Al no limitarse a una muestra, se eliminan posibles sesgos de selección y se maximiza la validez de los resultados obtenidos.

En esta investigación, se tomó la decisión de trabajar con la totalidad de la población disponible, es decir, con todos los registros identificados en la base de datos de LynxPay que cumplen con los criterios de inclusión previamente definidos. Esta decisión se fundamenta en varios factores clave relacionados con la naturaleza y el volumen de los datos disponibles, así como con los objetivos específicos del estudio.

**Justificación Principal: Tamaño y Disponibilidad de los Datos**

Acceso completo a los datos: Una de las principales razones para no emplear una muestra en este estudio es que se tiene acceso directo y completo a toda la base de datos de LynxPay, que incluye 4,842 registros únicos de clientes y transacciones realizadas desde 2021.

Este acceso completo elimina la necesidad de seleccionar un subconjunto de datos para analizar, permitiendo en cambio realizar un análisis censal de toda la población.

Tamaño manejable de la población: Aunque los 4,842 registros representan un volumen considerable de datos, este tamaño sigue siendo perfectamente manejable utilizando herramientas tecnológicas como Python. El procesamiento de estos datos no implica una carga computacional excesiva y permite realizar análisis exhaustivos sin restricciones operativas.

Calidad y estructura de los datos: Los registros disponibles son completos y verificables, con información estructurada sobre variables clave como características demográficas, comportamiento de compra y transacciones realizadas. Esto asegura que toda la información sea utilizable para el análisis, sin necesidad de excluir registros o recurrir a estimaciones estadísticas.

Ventajas de Utilizar Toda la Población: Trabajar con la totalidad de los datos disponibles presenta múltiples ventajas que justifican la decisión de no utilizar una muestra:

- Representación completa: Al analizar todos los registros disponibles, se incluye la totalidad de las variaciones y características presentes en la población, lo que asegura que los resultados sean completamente representativos y no estén influenciados por posibles errores o sesgos asociados al proceso de selección de una muestra.
- Eliminación de sesgos de selección: El uso de una muestra siempre implica un riesgo inherente de exclusión de patrones menos comunes o atípicos que podrían ser relevantes para los objetivos del estudio. Al trabajar con toda la población, este riesgo se elimina por completo.
- Mayor precisión en los resultados: No es necesario realizar inferencias estadísticas para extrapolar los resultados a la población total, ya que se basa directamente en la totalidad de los datos disponibles.
- Explotación integral de los datos: Este enfoque permite maximizar el uso de los datos disponibles, aprovechando toda la información relevante para realizar análisis

más detallados, como la segmentación por clústeres y la identificación de patrones de comportamiento específicos.

- Flexibilidad para análisis específicos: Al tener acceso a toda la población, se pueden realizar análisis detallados de subgrupos específicos (por ejemplo, clientes frecuentes, clientes ocasionales o diferentes categorías de productos), lo que facilita la personalización de las recomendaciones.

### 3.3.3 TÉCNICA DE MUESTREO

Las muestras se pueden clasificar en dos tipos: probabilísticas y no probabilísticas. En el caso de las probabilísticas, se utilizan métodos que permiten conocer la probabilidad de que cada individuo sea incluido en la muestra, lo que se logra a través de una selección aleatoria. Esto asegura que la muestra sea representativa de la población objetivo. Por otro lado, las muestras no probabilísticas se seleccionan basándose en ciertos criterios o características que los investigadores consideran relevantes para el estudio en ese momento. Sin embargo, estas muestras suelen ser menos confiables, válidas y reproducibles, ya que no se fundamentan en principios probabilísticos. Esto implica que no garantizan que cada sujeto elegido represente adecuadamente a la población objetivo. Por tanto, el uso de técnicas no probabilísticas puede limitar la capacidad de generalizar los resultados del estudio a la población de interés de manera precisa. (Otzen & Manterola, 2017).

Dado que se tiene acceso completo a todos los registros disponibles en la base de datos para esta investigación, no será necesario diseñar una técnica de muestreo específica. En lugar de seleccionar una muestra representativa, se utilizará la totalidad de los datos disponibles. Esto permitirá un análisis exhaustivo y detallado de la población en estudio, asegurando que todas las variaciones y características presentes en la base de datos sean consideradas. Al no limitarse a una muestra, se eliminan posibles sesgos de selección y se maximiza la validez y precisión de los resultados obtenidos. Esta metodología proporciona una visión integral y completa de los elementos investigados, lo que es fundamental para abordar de manera efectiva el problema de investigación planteado.

¿Qué Tipo de Muestra Usar y Por Qué? Aunque en este estudio se decidió trabajar con toda la población, en un contexto donde no fuera posible analizar todos los registros, sería necesario utilizar una muestra representativa de los datos. A continuación, se describen los tipos de muestra que podrían haberse utilizado y sus posibles aplicaciones en esta investigación:

### 3.3.3.1 MUESTREO PROBABILÍSTICO

El muestreo probabilístico es ideal cuando se busca una representación estadística precisa de la población y se cuenta con suficiente información para seleccionar la muestra de manera aleatoria. Algunos métodos específicos incluyen:

- Muestreo aleatorio simple: En este método, cada registro en la base de datos tiene la misma probabilidad de ser seleccionado. Es fácil de implementar y asegura que todos los elementos tengan igualdad de oportunidades.
- Muestreo estratificado: Este método divide la población en subgrupos o estratos (por ejemplo, edad, frecuencia de compra, género, región geográfica) y luego selecciona una muestra proporcional de cada estrato. Asegura que cada subgrupo esté representado en la muestra.
- Muestreo sistemático: Consiste en seleccionar cada  $n$ -ésimo registro de una lista ordenada de la población. Es eficiente y fácil de implementar.

### 3.3.3.2 MUESTREO NO PROBABILÍSTICO

Este tipo de muestreo es útil cuando no se tiene acceso completo a la población o cuando se busca analizar un subconjunto específico de interés. Algunos métodos relevantes incluyen:

- Muestreo por conveniencia: Selecciona registros fácilmente accesibles o disponibles. Puede introducir sesgos significativos, ya que no garantiza la representatividad.
- Muestreo por cuotas: Se selecciona una muestra que cumpla con ciertas cuotas predefinidas (por ejemplo, porcentaje de clientes frecuentes y ocasionales). Asegura que se incluyan proporciones específicas de ciertos subgrupos.
- Muestreo intencional o dirigido: Selecciona registros que se consideran representativos según el criterio del investigador. Es subjetivo y puede introducir sesgos significativos.

### 3.3.3.3 ENFOQUE CENSAL

En este estudio, el enfoque censal fue preferido por las siguientes razones:

- Acceso completo: Dado que se cuenta con todos los registros disponibles, no hay necesidad de realizar un muestreo.
- Tamaño manejable: Los 4,842 registros son procesables sin limitaciones técnicas.
- Mayor validez: Los resultados reflejan directamente la realidad de la población, sin depender de estimaciones o inferencias.

### 3.3.4 CORTE DE LA INVESTIGACIÓN

El corte de la investigación se refiere al periodo de tiempo en el cual se recopilan y analizan los datos que componen el estudio. Este corte temporal es fundamental para garantizar que los resultados reflejen las dinámicas, tendencias y comportamientos más relevantes de la población en estudio, permitiendo una interpretación precisa y contextualizada de los datos.

#### 3.3.4.1 PERIODO DE ANÁLISIS

El periodo de análisis definido para este estudio abarca desde el 1 de enero de 2021 hasta el 31 de diciembre de 2024, lo que equivale a un total de 4 años consecutivos. Este periodo fue seleccionado considerando los siguientes factores:

- Relevancia temporal: Elegir un periodo de análisis amplio permite capturar tendencias de largo plazo y variaciones temporales en el comportamiento de los usuarios y comerciantes de la plataforma LynxPay.

Esto incluye la identificación de patrones recurrentes o estacionales (por ejemplo, compras durante eventos como Black Friday, Navidad o el inicio del año).

- Disponibilidad de datos: El acceso a registros completos desde el año 2021 garantiza que el análisis se base en datos actualizados y completos, lo cual es crucial para validar la efectividad de los sistemas de clusterización y recomendaciones personalizadas.
- Periodo representativo: El periodo de 4 años incluye suficientes interacciones y transacciones como para identificar cambios significativos en las dinámicas del comercio electrónico, ya sea por la evolución del mercado, eventos económicos externos o tendencias digitales.

### 3.3.4.2 JUSTIFICACIÓN DEL PERIODO DE ANÁLISIS

El periodo seleccionado cumple con varias características esenciales para que los resultados de la investigación sean relevantes y válidos:

- Captura de variaciones estacionales: En el comercio electrónico, los picos de actividad suelen estar vinculados a eventos específicos, como campañas de descuentos (Black Friday, Cyber Monday) o temporadas festivas (Navidad, Día de la Madre). Estos periodos generan un volumen significativo de transacciones que reflejan el comportamiento intensivo de los consumidores.

Al abarcar 4 años, el análisis incluye suficientes eventos estacionales para identificar patrones regulares en el comportamiento de compra.

- Análisis de tendencias a largo plazo: Este rango temporal permite observar cómo evolucionan las dinámicas de compra y las interacciones en LynxPay a lo largo del tiempo. Por ejemplo:
  - Cambios en las preferencias de los usuarios.
  - Incremento o disminución en la frecuencia de compra.
  - Introducción de nuevas categorías de productos o servicios por parte de los comerciantes.
  - También permite analizar el impacto de factores externos (como la pandemia de COVID-19) en el comercio electrónico.
- Enfoque en datos recientes: Elegir un rango temporal relativamente reciente asegura que los resultados sean aplicables al estado actual de la plataforma LynxPay y puedan guiar decisiones estratégicas futuras.
- Suficiente volumen de datos: Durante este periodo, la plataforma generó un volumen considerable de registros (4,842 transacciones únicas), lo que proporciona un conjunto de datos robusto para realizar análisis significativos sin comprometer la capacidad de procesamiento.

## **3.4 TÉCNICAS, HERRAMIENTAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS**

### **3.4.1 TÉCNICAS**

Las técnicas de investigación son métodos sistemáticos para recolectar, analizar e interpretar datos, y entre las más comunes se encuentran la investigación bibliográfica, que revisa material existente; los métodos cualitativos, que exploran percepciones y comportamientos a través de datos no numéricos; los métodos cuantitativos, que utilizan datos numéricos y estadísticas para probar hipótesis; la revisión sistemática, que analiza todos los estudios relevantes sobre un tema específico; y los estudios de caso, que investigan en profundidad uno o pocos casos específicos.

En esta investigación, no se emplean técnicas tradicionales de recolección de datos como encuestas o entrevistas. Esto se debe a que la información necesaria es proporcionada directamente por la empresa, lo que permite un acceso más preciso y detallado a los datos relevantes. Para llevar a cabo el proceso de Extracción, Transformación y Carga (ETL), se utilizan herramientas especializadas como DBeaver y Python. Estas herramientas son fundamentales para la extracción de los datos correspondientes al período comprendido entre los años 2021 y 2024. DBeaver facilita la conexión y gestión de bases de datos, mientras que Python, con su amplia gama de bibliotecas y capacidades de programación, permite la manipulación y transformación de los datos de manera eficiente. Este enfoque asegura que los datos sean procesados y preparados adecuadamente para su análisis posterior, garantizando la integridad y calidad de la información utilizada en la investigación.

### **3.4.2 MINERÍA DE DATOS**

En nuestra investigación, implementamos la minería de datos utilizando Putty para conectarnos de manera segura a servidores remotos y acceder a los datos almacenados. Utilizamos DBeaver para explorar y consultar las bases de datos, lo que facilitó la gestión y extracción de la información necesaria. Python y VSCode fueron empleados conjuntamente para escribir y depurar scripts de minería de datos, aprovechando las capacidades de programación de Python y las herramientas de desarrollo de VSCode. Aplicamos Sklearn para implementar algoritmos de aprendizaje automático y construir modelos predictivos basados en los datos extraídos. Finalmente, utilizamos Matplotlib para crear visualizaciones detalladas y comprensibles, lo que

facilitó la interpretación y comunicación de los resultados del análisis de datos.

### 3.4.3 FASES DEL PROCESO DE LA MINERÍA DE DATOS

La minería de datos se compone de varias fases o procesos esenciales, los cuales hemos implementado detalladamente en este estudio. A continuación, se describen cada una de estas fases para proporcionar una comprensión completa del enfoque y la metodología aplicada.

- **Acceso a Datos:** Utilizamos Putty para conectarnos de manera segura a los servidores remotos de LynxPay, donde se almacenan los datos transaccionales y de usuario necesarios para nuestro análisis.
- **Gestión y Exploración de Datos:** Empleamos DBeaver para explorar y consultar las bases de datos de LynxPay. Esta herramienta nos permite gestionar eficientemente la información y extraer los datos relevantes para el desarrollo del modelo de recomendación.
- **Desarrollo de Scripts:** Utilizamos Python y VSCode para escribir y depurar scripts de minería de datos. Python, con su amplia gama de bibliotecas, nos permite realizar transformaciones de datos y preparar el conjunto de datos para el análisis. VSCode facilita el proceso de desarrollo gracias a sus características avanzadas de edición y depuración.
- **Análisis de Datos y Modelado:** Aplicamos Sklearn (Scikit-learn) para implementar algoritmos de aprendizaje automático que nos permiten construir el modelo de recomendación. Utilizamos técnicas como la filtración colaborativa y el análisis de similitud para predecir las preferencias de los usuarios de LynxPay.
- **Visualización de Resultados:** Utilizamos Matplotlib para crear visualizaciones detalladas y comprensibles de los datos y los resultados del modelo de recomendación. Estas visualizaciones nos ayudan a interpretar los patrones y tendencias en los datos, y a comunicar los hallazgos de manera efectiva a los interesados.

### 3.4.4 MODELOS DE PREDICCIÓN

En este trabajo de investigación se probaron los siguientes modelos de agrupación:

- K-means
- DBSCAN

### 3.4.5 HERRAMIENTAS

En esta investigación el proceso de minería de datos para desarrollar un modelo de recomendación incluye la conexión segura a servidores remotos con Putty, la gestión y exploración de bases de datos con DBeaver, el desarrollo de scripts con Python y VSCode, el análisis y modelado de datos con Sklearn, y la visualización de resultados con Matplotlib. Este enfoque integral nos permite crear un modelo de recomendación robusto y eficiente que mejora la experiencia del usuario en nuestra plataforma.

#### 3.4.5.1 PUTTY

Putty es una herramienta de terminal que utilizamos para conectarnos de manera segura a los servidores remotos de LynxPay a través de protocolos como SSH. Esto nos permitió acceder a los datos almacenados en los servidores y ejecutar scripts de análisis de datos de manera eficiente. Putty es esencial para administrar y ejecutar procesos de minería de datos en un entorno remoto. (Tatham, 2024)

#### 3.4.5.2 DBEAVER

DBeaver es una herramienta de administración de bases de datos que soporta una amplia variedad de sistemas de gestión de bases de datos (DBMS). En nuestro proceso, utilizamos DBeaver para explorar, consultar y gestionar las bases de datos de LynxPay. Esta herramienta nos permitió realizar consultas SQL para extraer datos relevantes, facilitando la gestión y preparación de la información necesaria para el análisis. (Dbeaver, 2024)

#### 3.4.5.3 PYTHON

Python es un lenguaje de programación muy popular en el campo de la minería de datos debido a su simplicidad y la gran cantidad de bibliotecas disponibles. Utilizamos Python para escribir scripts que automatizan el proceso de extracción, transformación y carga (ETL) de datos. Además, Python se empleó para realizar análisis estadísticos y aplicar algoritmos de minería de datos, lo que es crucial para el desarrollo del modelo de recomendación. (Python Org, 2024)

#### 3.4.5.4 VSCODE

VSCoDe (Visual Studio Code) es un editor de código fuente que soporta múltiples lenguajes de programación y tiene una gran cantidad de extensiones. Utilizamos VSCoDe para escribir y depurar scripts en Python. Sus características avanzadas, como el resaltado de sintaxis, la integración con control de versiones y las extensiones específicas para Python, hicieron que el desarrollo de scripts fuera más eficiente y manejable. (Visual Studio Code, 2024)

#### 3.4.5.5 SKLEARN

Sklearn (Scikit-learn) es una biblioteca de Python que proporciona herramientas simples y eficientes para el análisis de datos y la minería de datos. En LynxPay, utilizamos Sklearn para implementar algoritmos de aprendizaje automático que nos permitieron construir y evaluar modelos predictivos. Estos modelos son fundamentales para el sistema de recomendación, ya que ayudan a predecir las preferencias de los usuarios basándose en sus comportamientos y datos históricos. (Scikit Learn, 2024)

#### 3.4.5.6 MATPLOTLIB

Matplotlib es una biblioteca de Python para la creación de gráficos y visualizaciones. Utilizamos Matplotlib para visualizar los datos y los resultados del análisis de manera clara y comprensible. Esta herramienta nos permitió crear gráficos de líneas, barras, histogramas y dispersión, entre otros, lo que facilitó la interpretación y comunicación de los hallazgos de manera efectiva. (Hunter, 2024)

### 3.4.6 PROCEDIMIENTO

Texto. En nuestra investigación en LynxPay, el proceso de minería de datos para desarrollar un modelo de recomendación incluyó la conexión segura a servidores remotos con Putty, la gestión y exploración de bases de datos con DBEaver, el desarrollo y depuración de scripts con Python y VSCoDe, el análisis y modelado de datos con Sklearn, y la visualización de resultados con Matplotlib. Este enfoque integral nos permitió crear un modelo de recomendación robusto y eficiente, mejorando la experiencia del usuario en nuestra plataforma.

#### Paso 1: Conexión a Servidores Remotos

- Abrimos Putty: Iniciamos Putty en nuestras computadoras.

- Configuramos Conexión SSH: Introdujimos la dirección IP del servidor remoto de LynxPay y seleccionamos el protocolo SSH.
- Accedimos al Servidor: Ingresamos nuestras credenciales de acceso para conectarnos de manera segura al servidor remoto.
- Navegamos al Directorio de Datos: Utilizamos comandos de terminal para navegar al directorio donde se almacenaban los datos necesarios.

#### Paso 2: Exploración y Gestión de Bases de Datos

- Abrimos DBeaver: Iniciamos DBeaver en nuestras computadoras.
- Configuramos Conexión a la Base de Datos: Configuramos una nueva conexión a la base de datos de LynxPay introduciendo las credenciales y detalles del servidor.
- Exploramos la Base de Datos: Navegamos por las tablas y esquemas de la base de datos para familiarizarnos con la estructura de los datos.
- Realizamos Consultas SQL: Ejecutamos consultas SQL para extraer los datos relevantes que luego fueron procesados y analizados.

#### Paso 3: Desarrollo de Scripts

- Abrimos VSCode: Iniciamos Visual Studio Code en nuestras computadoras.
- Configuramos Entorno de Python: Aseguramos que Python estuviera instalado y configurado correctamente en VSCode.
- Escribimos Scripts de ETL: Creamos y editamos scripts en Python para automatizar el proceso de extracción, transformación y carga (ETL) de los datos.
- Depuramos Scripts: Utilizamos las herramientas de depuración de VSCode para identificar y corregir errores en los scripts.

#### Paso 4: Análisis de Datos y Modelado

- Importamos Bibliotecas en Python: Aseguramos tener instaladas las bibliotecas necesarias como Sklearn y Matplotlib.
- Preparamos los Datos: Utilizamos Python para limpiar y preparar los datos

extraídos para el análisis.

- Implementamos Algoritmos de Sklearn: Aplicamos algoritmos de aprendizaje automático de Sklearn para construir y evaluar modelos predictivos.
- Entrenamos el Modelo: Entrenamos el modelo de recomendación utilizando los datos preparados.
- Evaluamos el Modelo: Evaluamos el rendimiento del modelo utilizando métricas adecuadas y ajustamos los parámetros según fue necesario.

#### Paso 5: Visualización de Resultados

- Creamos Gráficos con Matplotlib: Utilizamos Matplotlib para crear visualizaciones de los datos y los resultados del modelo.
- Interpretamos Visualizaciones: Analizamos los gráficos para interpretar los patrones y tendencias en los datos.
- Comunicamos Resultados: Preparamos informes visuales que resumieron los hallazgos y las recomendaciones basadas en el análisis de datos.

### 3.4.7 INSTRUMENTOS UTILIZADOS

En el contexto de esta investigación, los instrumentos son los mecanismos específicamente diseñados para recolectar, procesar y estructurar los datos de manera sistemática, con el objetivo de analizarlos y obtener resultados relevantes. Estos instrumentos van más allá de las herramientas utilizadas, ya que están diseñados para cumplir una función específica en el marco del proceso de minería de datos y desarrollo del sistema de recomendaciones en LynxPay. A continuación, se describen los instrumentos desarrollados y su papel dentro del estudio:

#### 3.4.7.1 SCRIPTS EN PYTHON PARA EL PROCESO ETL (EXTRACCIÓN, TRANSFORMACIÓN Y CARGA)

Los scripts desarrollados en Python automatizan el proceso ETL, una etapa clave en el manejo de datos. Este instrumento fue diseñado para garantizar que los datos extraídos de las bases de datos de LynxPay cumplan con los criterios de inclusión del estudio y estén preparados para el análisis.

- Propósito:
  - Automatizar la extracción de datos relevantes desde las bases de datos mediante consultas estructuradas.
  - Transformar los datos para corregir errores, eliminar duplicados y estructurarlos en un formato adecuado para el análisis.
  - Cargar los datos procesados en estructuras que puedan ser utilizadas directamente por los algoritmos de minería de datos.
- Aplicaciones en LynxPay:
  - Extracción de información transaccional de los usuarios, como historial de compras y productos añadidos al carrito.
  - Procesamiento de datos demográficos de los clientes (edad, género, ubicación geográfica).
  - Estandarización de los datos para que sean compatibles con los modelos de aprendizaje automático implementados en Sklearn.
- Características del Instrumento:
  - Bibliotecas utilizadas: Pandas (para manipulación de datos), Numpy (para cálculos matemáticos) y SQLAlchemy (para conectar y ejecutar consultas en la base de datos).
  - Ventajas: Reducción de errores manuales, mayor consistencia en los datos y eficiencia en el manejo de grandes volúmenes de información.

#### 3.4.7.2 ALGORITMOS DE AGRUPACIÓN Y MODELOS DE APRENDIZAJE AUTOMÁTICO (K-MEANS Y DBSCAN)

Los algoritmos implementados en Sklearn se consideran instrumentos fundamentales, ya que permiten analizar y segmentar los datos mediante técnicas de aprendizaje no supervisado. Estos modelos identifican patrones y agrupaciones dentro de los datos transaccionales y demográficos.

- Propósito:
  - Segmentar a los clientes de LynxPay en clústeres homogéneos basados en su comportamiento de compra y características demográficas.
  - Identificar patrones ocultos en los datos que ayuden a personalizar las recomendaciones y mejorar la experiencia del usuario.
- Aplicaciones en LynxPay:
  - K-means: Utilizado para agrupar clientes en función de variables como la frecuencia de compra, el valor promedio por transacción y el tipo de productos comprados. Este modelo facilita la identificación de segmentos como "compradores frecuentes", "compradores ocasionales" y "clientes con alta inversión".
  - DBSCAN: Empleado para detectar grupos de clientes con patrones menos estructurados o fuera de lo común, como clientes que realizan compras durante eventos específicos o de manera esporádica.
- Características del Instrumento:
  - Parámetros ajustados: Se optimizaron parámetros como el número de clústeres (para K-means) y el radio de vecindad (para DBSCAN) mediante validación cruzada.
  - Resultados generados: Identificación de segmentos específicos que sirven como base para personalizar las estrategias de recomendación.

### 3.4.7.3 CONSULTAS SQL PARA LA EXTRACCIÓN DE DATOS

Las consultas SQL diseñadas específicamente para este estudio se consideran instrumentos, ya que permiten filtrar y extraer la información relevante desde las bases de datos de LynxPay. Estas consultas son esenciales para seleccionar únicamente los registros que cumplen con los criterios de inclusión definidos.

- Propósito:
  - Acceder y extraer información específica de las bases de datos de LynxPay.

- Filtrar datos irrelevantes o duplicados, asegurando que solo se utilicen registros válidos en el análisis.
- Aplicaciones en LynxPay:
  - Selección de registros de transacciones realizadas entre 2021 y 2024.
  - Extracción de datos demográficos asociados a los usuarios que hayan realizado al menos una compra.
  - Filtrado de información por categorías de productos, regiones geográficas o montos de transacción.
- Características del Instrumento:
  - Consultas dinámicas: Utilizan filtros condicionales para segmentar los datos de manera específica.
  - Resultados obtenidos: Creación de subconjuntos de datos listos para ser procesados por los scripts de Python en el proceso ETL.

#### 3.4.7.4 VISUALIZACIONES EN MATPLOTLIB

Los gráficos y visualizaciones creados con Matplotlib se consideran instrumentos en esta investigación, ya que facilitan la interpretación de los resultados obtenidos a partir del análisis de datos.

- Propósito:
  - Representar visualmente los patrones y tendencias identificados en los datos.
  - Comunicar de manera clara y comprensible los hallazgos a los interesados en la investigación.
- Aplicaciones en LynxPay:
  - Creación de gráficos de barras, histogramas y diagramas de dispersión para analizar la frecuencia de compra, la distribución de clientes por clústeres y el rendimiento del modelo de recomendación.
  - Identificación de outliers o anomalías en los datos mediante

representaciones gráficas.

- Características del Instrumento:
  - Flexibilidad: Posibilidad de personalizar los gráficos para incluir etiquetas, colores y estilos que mejoren la presentación.
  - Ventajas: Facilitan la toma de decisiones al permitir que los hallazgos sean fácilmente comprensibles.

#### 3.4.7.5 MODELOS DE RECOMENDACIÓN BASADOS EN FILTRACIÓN COLABORATIVA

Los modelos predictivos construidos mediante filtración colaborativa se consideran instrumentos, ya que permiten predecir las preferencias de los usuarios en función de su historial de interacciones y datos demográficos.

- Propósito:
  - Desarrollar un sistema de recomendaciones personalizadas que mejore la experiencia del usuario en LynxPay.
  - Predecir productos relevantes para cada cliente basándose en patrones observados en clientes similares.
- Aplicaciones en LynxPay:
  - Recomendación de productos relacionados para compradores frecuentes.
  - Identificación de tendencias emergentes basadas en el comportamiento agregado de los usuarios.
- Características del Instrumento:
  - Entrenamiento: Los modelos se entrenaron utilizando conjuntos de datos divididos en datos de entrenamiento y prueba para evaluar su desempeño.
  - Evaluación: Se utilizaron métricas como la precisión y el recall para medir la efectividad de las recomendaciones generadas.

### **3.5 FUENTES DE INFORMACIÓN**

Las fuentes de información son todos los recursos o materiales de donde se obtiene información relevante para desarrollar una investigación o proyecto. Pueden incluir datos, ideas, o conocimientos, ya sea provenientes de documentos, personas, bases de datos, libros, artículos, páginas web o cualquier medio que aporte contenido útil y confiable.

#### **3.5.1 FUENTES PRIMARIAS**

Las fuentes primarias se refieren a datos originales o información obtenida directamente por el investigador de las personas u organizaciones relacionadas con el tema de estudio. Estas fuentes son fundamentales para ofrecer un sustento empírico a la investigación.

En este proyecto, las fuentes primarias incluyen los datos proporcionados por la plataforma LynxPay, tales como historial de transacciones y productos añadidos al carrito de compra. Estos datos son clave para la aplicación de técnicas de clusterización y el desarrollo de recomendaciones personalizadas.

En este proyecto, las fuentes primarias incluyen:

- La base de datos proporcionada por LynxPay, la cual contiene información relevante como historial de transacciones y productos añadidos al carrito de compra. Esta información es fundamental para aplicar técnicas de clusterización y desarrollar sistemas de recomendación personalizados.
- Entrevistas y reuniones con los representantes de LynxPay, donde se obtuvieron detalles específicos sobre las necesidades y objetivos del sistema de recomendación.

#### **3.5.2 FUENTES SECUNDARIAS**

Las fuentes secundarias son aquellas que reúnen, sintetizan y analizan información previamente publicada en fuentes primarias. Estas fuentes son útiles para enriquecer el marco teórico y respaldar el análisis de la investigación.

Para esta tesis, se consultaron diversas fuentes secundarias, entre ellas:

- Libros tales como Machine Learning Yearning de Andrew Ng y Data Mining: Concepts and Techniques de Han, Kamber y Pei.

- Artículos académicos sobre sistemas de recomendación y algoritmos de clusterización, como el publicado por Duy-Nghia Nguyen (2024) y Zengyuan Wu (2022).
  - Tesis y publicaciones previas relacionadas con el uso de inteligencia artificial y minería de datos en comercio electrónico.
1. Entradas de blogs especializados y páginas web como Trade.gov y LynxPay, que aportaron información contextual sobre el comercio electrónico en Honduras.

## **CAPÍTULO IV. RESULTADOS Y ANÁLISIS**

Este capítulo se centra en examinar y comprender los resultados de la investigación. Para ello, se comparan los resultados con los objetivos iniciales y las preguntas de investigación planteadas al principio del estudio. Además, se analizan los datos en un contexto más amplio para comprender mejor su significado. El propósito principal de este capítulo es mostrar los resultados de manera clara y organizada.

### **4.1 INFORME DE PROCESO DE RECOLECCIÓN DE DATOS**

Para llevar a cabo este análisis, LynxLabs otorgó acceso a una réplica de su base de datos transaccional, actualizada hasta el 4 de noviembre de 2024. Se estableció una conexión segura mediante túnel SSH a través de PuTTY confirmando el funcionamiento del túnel SSH, por consiguiente, esto permitió acceder al servidor Postgres con Dbeaver y extraer la información relevante.

Los datos históricos se exportaron en formato CSV para su posterior análisis. Este análisis se realizó utilizando Jupyter Notebooks en Visual Studio Code, lo cual facilitó la exploración y el procesamiento de la información.

En resumen, el acceso a la réplica de la base de datos y el uso de herramientas como PuTTY, Dbeaver y Jupyter Notebooks permitieron un análisis exhaustivo y eficiente de los datos proporcionados por LynxLabs.

Se debe hacer notar que los campos demográficos no se encontraron en el proceso de recolección de datos, considerados básicos para la segmentación de clientes, por lo tanto, se incluyen en las recomendaciones, para que la plataforma pueda obtener mejores resultados, con la finalidad de enriquecer los análisis e incluso modelos necesarios para una mejor segmentación.

### **4.2 RESULTADOS Y ANÁLISIS DE LAS TÉCNICAS APLICADAS**

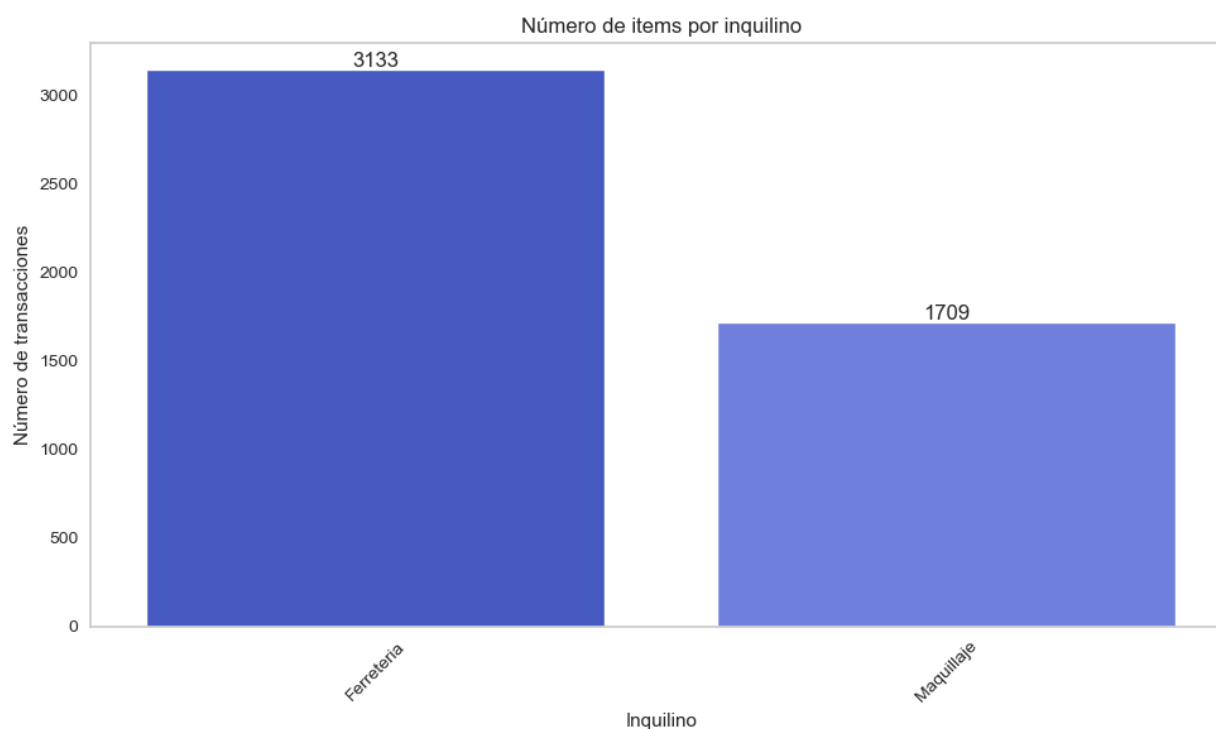
En este capítulo se ofrecen los descubrimientos hechos a base de los análisis de los datos recolectados para el estudio, con el objetivo principal de realizar descubrimientos fundamentados, así como recomendaciones para abordar el problema de investigación.

## 4.2.1 SITUACIÓN ACTUAL

En esta sección se muestra un análisis exploratorio de los datos obtenidos por inquilino dentro de la plataforma LynxPay, datos que remontan a los inicios de cada cliente, estos aproximadamente desde 2021 a 2024, los datos de 2024 solo abarcan hasta noviembre.

### 4.2.1.1 ANÁLISIS DE VENTAS POR INQUILINO

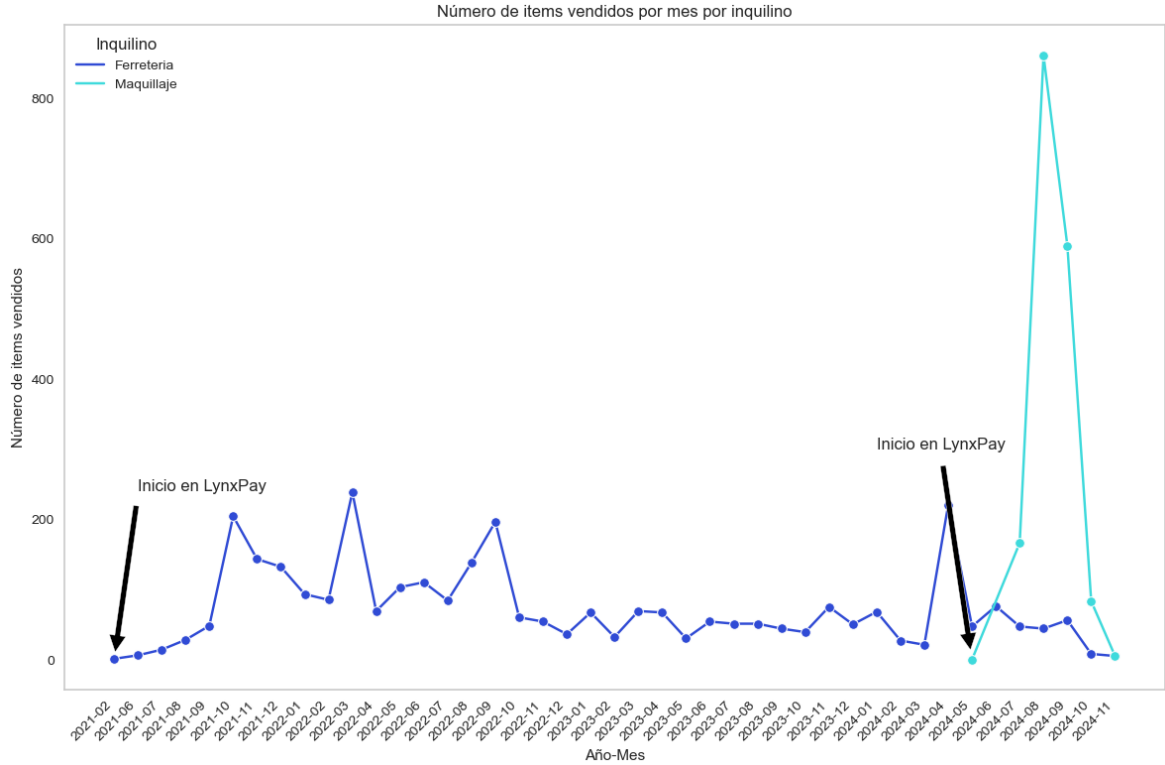
En la figura 6 podemos visualizar la cantidad de datos por cada inquilino activo presente en la plataforma LynxPay, El inquilino "Ferretería" registra un número significativamente mayor de ítems, con un total de 3133. Por otro lado, el inquilino "Maquillaje" presenta 1709 ítems, esta diferencia sugiere que la actividad o el volumen de datos asociados al inquilino "Ferretería" es considerablemente mayor en comparación con el inquilino "Maquillaje".



**Figura 6. Número de ítems adquiridos por inquilino**

Fuente: Elaboración propia

Sin embargo, sin el contexto histórico estos datos podrían parecer muy reveladores, véase la figura 7, en el cual vemos que el inquilino enfocado en la industria de ferretería tiene más tiempo dentro de la plataforma, por lo tanto, sus datos ciertamente se han acumulado durante una mayor cantidad de tiempo.



**Figura 7 Histórico de ítems vendidos por inquilino**

Fuente: Elaboración propia

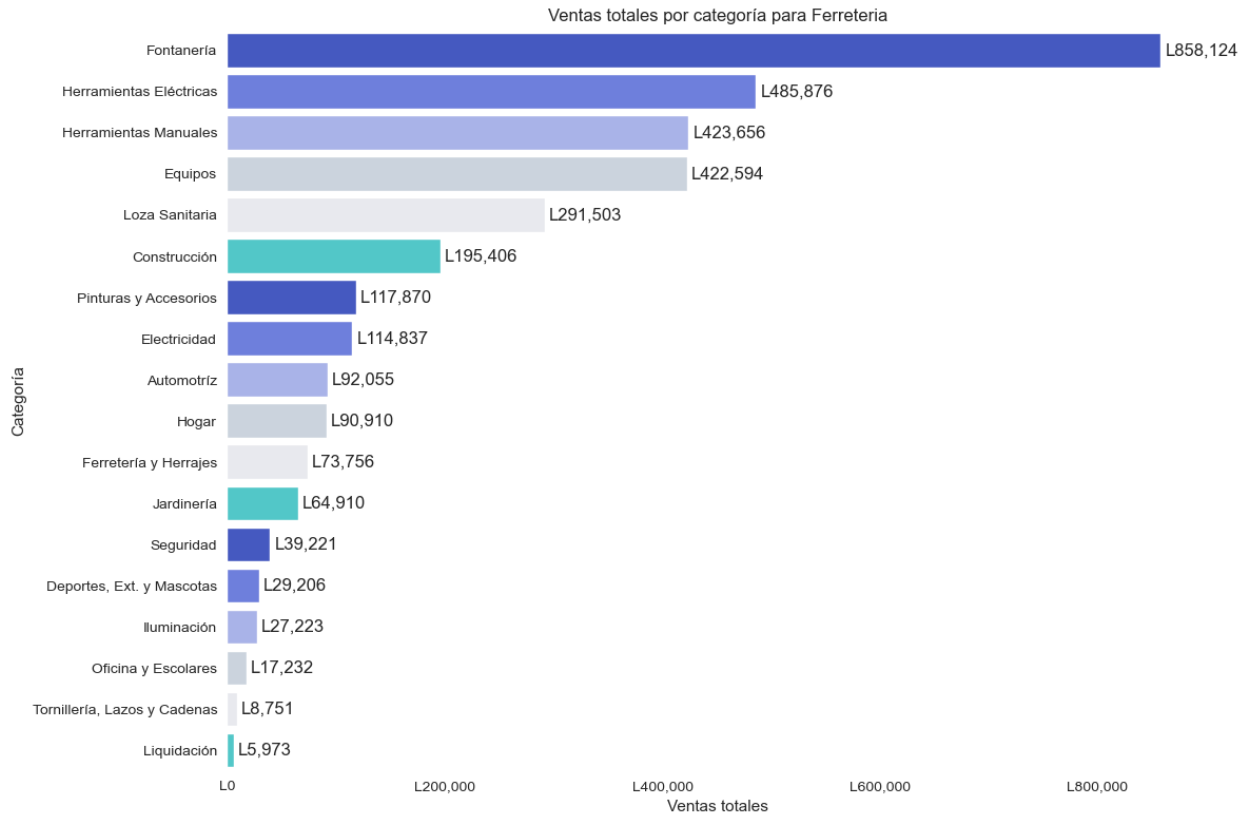
Nótese que en este contexto ítems se refiere a el listado de productos comprados por un usuario, sin tomar en cuenta la cantidad, es decir, un usuario dentro del inquilino Ferretería al realizar una compra de dos productos distintos, tendría 2 ítems dentro de esta fuente de datos, véase en la tabla 4 donde se muestra el esquema de esta tabla.

**Tabla 6 Esquema de tabla transaccional**

Industria Inquilino	Región	Nombre del producto	Categoría	Descripción corta	Nombre de la variante	Precio	Cantidad	Fecha Transacción	Total	Marca	Modelo
Maquillaje	N/A	CARMEX mini lip balm promoción	Promociones	Llévate 2 carmex mini lip balm por solo L150.	CARMEX mini lip balm promoción	130.44	1	21/08/2024	130.44	N/A	N/A
Ferretería	N/A	Termocuple para Calentador 18" 410576/115	Electricidad	Termocuple para Calentado	Termocuple para Calentador 18"	195.65	1	07/03/2022	195.65	Do It Best	410576
Maquillaje	N/A	elf Hydrating Camo Concealer	Maquillaje	Corrector líquido de cobertura media con acabado hidratante.	Medium Neutral	304.35	1	20/08/2024	304.35	N/A	N/A
Ferretería	N/A	Malla de Fibra de Vidrio para Reparar Paredes P8096	Construcción	Malla de Fibra de Vidrio para Reparar Paredes P8096	Malla de Fibra de Vidrio para Reparar Paredes	91.3	1	02/03/2022	91.3	Prime Line	N/A
Ferretería	N/A	Promoción Inodoro + Lavamanos Casablanca Bone CASTEL	Loza Sanitaria	Promoción Inodoro + Lavamanos Casablanca Bone CASTEL	PROM INOD+LAV CASABLANCA BONE PED CASTEL	4477.39	3	24/06/2024	13432.17	Castel	N/A

Fuente: Elaboración propia.

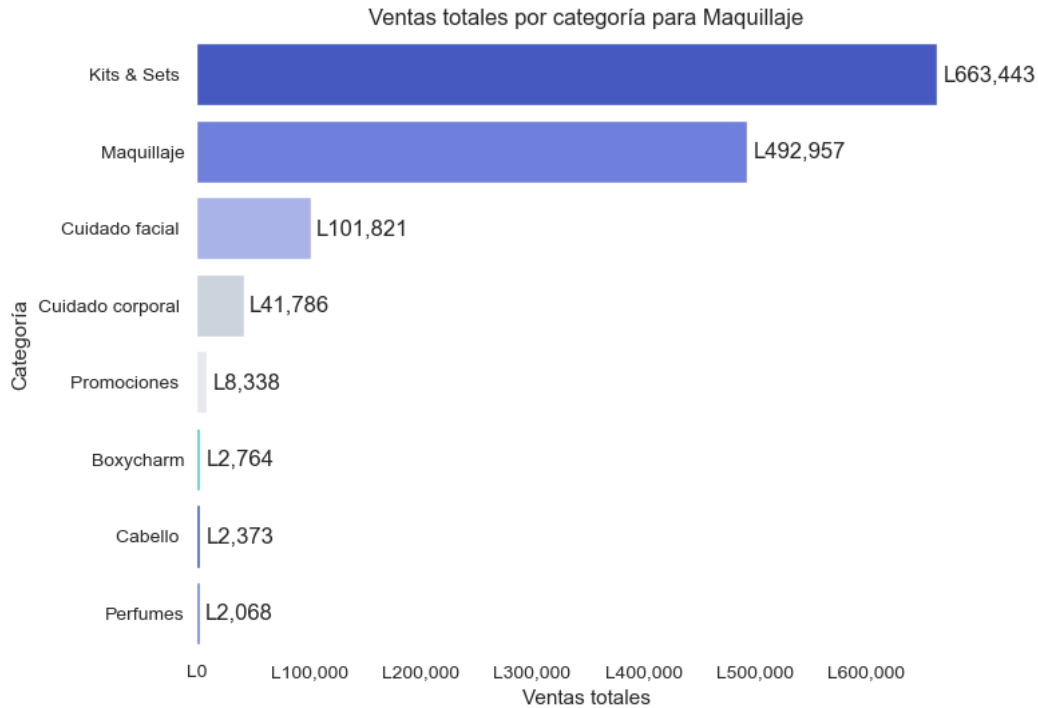
El siguiente gráfico proporciona una vista de las ventas generadas para la plataforma para cada uno de los inquilinos presentes en la misma, dado la cantidad de tiempo, podemos notar diferencias notables en las sumas.



**Figura 8 Ventas totales inquilino Ferreteria**

Fuente: Elaboración propia.

En esta, el inquilino dedicado al rubro de la ferreteria presenta ventas notablemente mayores en fontanería, posteriormente se encuentran presentes tres categorías muy equiparables, siendo herramientas eléctricas, herramientas manuales y equipos los que se encuentran con sumas cercanas al medio millón de lempiras, dando información valiosa de que categorías de productos son más buscados por los clientes de este inquilino.



**Figura 9 Ventas totales inquilino Maquillaje**

Fuente: Elaboración propia

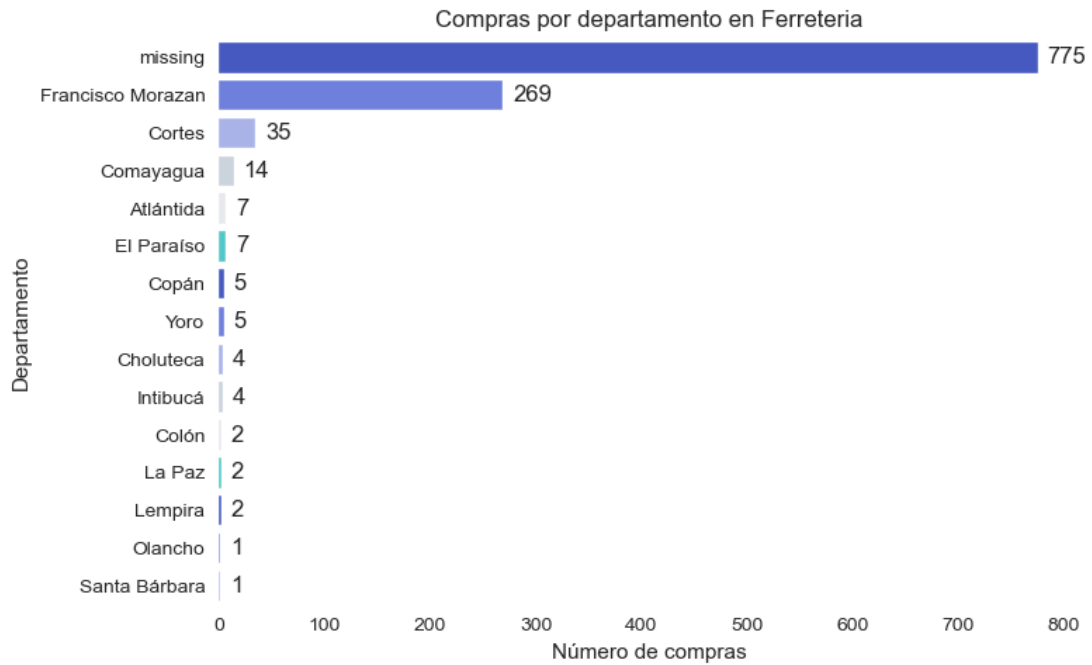
En la figura número 9 podemos visualizar como la categoría Kit & Sets que hace referencia a productos que se venden en conjunto presentan un notable repunte en las sumas vendidas para este inquilino, seguidamente de maquillaje que es el rubro principal al que se dedica.

Ambas figuras representan fielmente el comportamiento de los usuarios dentro de la plataforma en cada inquilino, que fueron de gran ayuda para el modelo.

#### 4.2.2 FACTORES SOCIODEMOGRÁFICOS

##### 4.2.2.1 ANÁLISIS GEOGRÁFICO

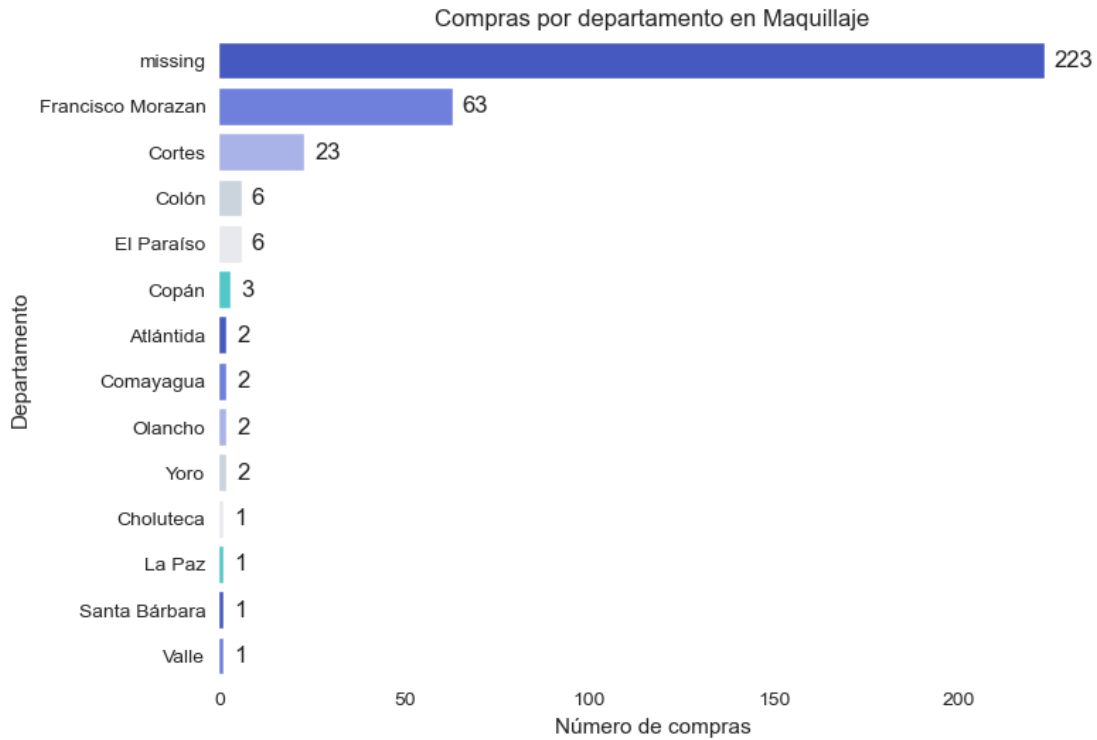
Se puede visualizar en la siguiente figura la cantidad de ventas realizadas en cada uno de los inquilinos por zona geográfica registrada en la dirección del usuario.



**Figura 10 Compras por departamento en inquilino Ferreteria**

Fuente: Elaboración propia

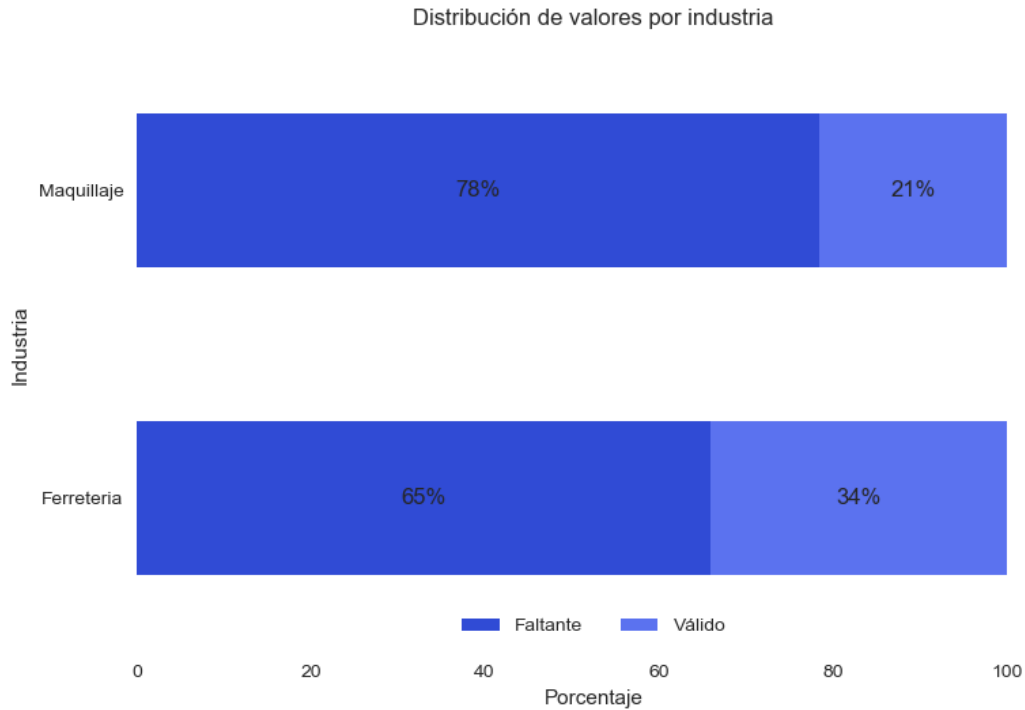
Como se puede visualizar existen valores faltantes notables en los datos, seguidos por Francisco Morazán como uno de los departamentos donde más se efectúan transacciones en la plataforma.



**Figura 11 Compras por departamento inquilino Maquillaje**

Fuente: Elaboración propia

En ambos casos, una cantidad importante de los datos presentan valores nulos o vacíos en campos geográficos, resultado de ventas no finalizadas dentro de la plataforma, para fines investigativos se hará uso de la información transaccional para visualización de comportamientos de compra, evitando el uso de datos geográficos que sesguen el resultado debido a datos faltantes.



**Figura 12 Distribución de valores faltantes en campo departamento**

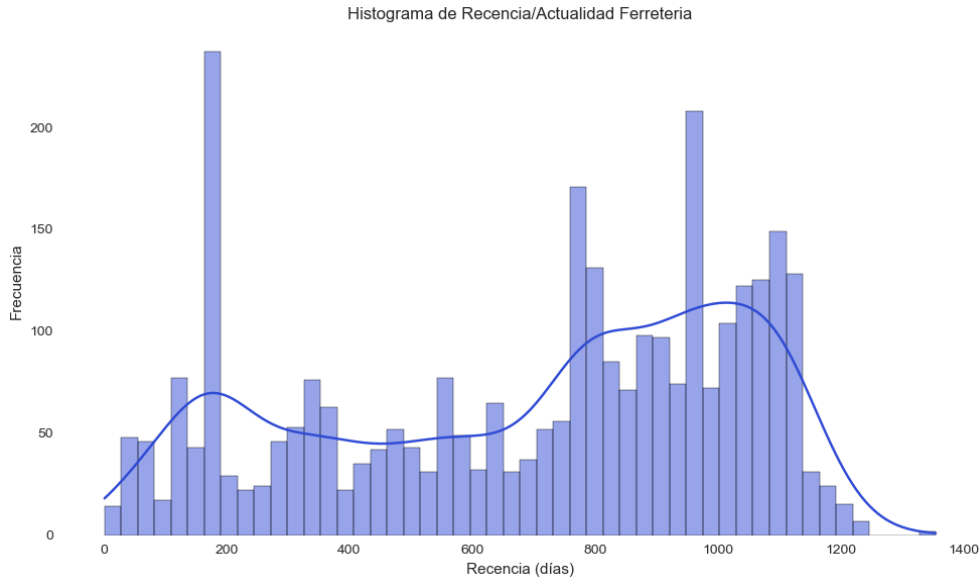
Fuente: Elaboración propia

En la figura 12 se hace visible la proporción de transacciones u ordenes dentro de los datos que cuentan con un departamento asignado dentro de los valores del cliente. Aseverando la necesidad de realizar una decisión acerca del uso de esta característica dentro del análisis.

#### 4.2.3 ANÁLISIS RFM

##### 4.2.3.1 ACTUALIDAD (RECENCY)

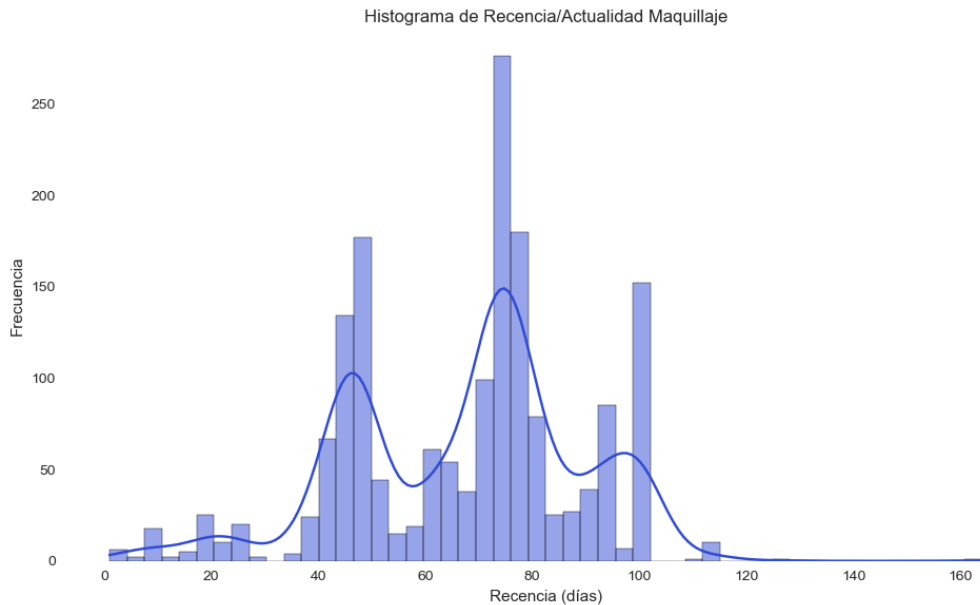
En el siguiente análisis como se menciona en la conceptualización, se evalúa la diferencia de las fechas de compra de cada uno de los clientes, con el fin de conocer la última compra hecha, y cuantos días de diferencia existe entre esa compra y la actualidad, para el fin del análisis, la fecha final, se decide utilizando la última transacción en la fuente de datos para cada industria presente.



**Figura 13 Histograma de recencia inquilino Ferretería**

Fuente: Elaboración propia

La figura 13 presenta una frecuencia que se mantiene constante dentro de los primeros 600 días, es decir aproximadamente 20 meses o 1 año y 8 meses, para luego obtener una mayor cantidad de usuarios que utilizaron la plataforma en sus inicios, denotando el auge del uso de plataformas web dada las limitaciones de movilidad en los primeros años luego de pandemia.



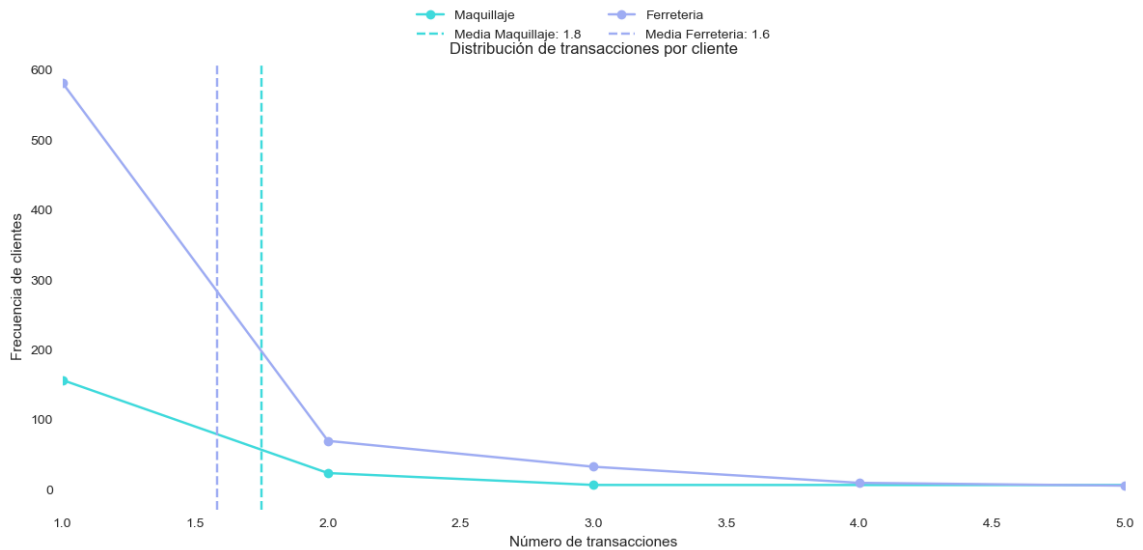
**Figura 14 Histograma de recencia inquilino Maquillaje**

Fuente: Elaboración propia

En la figura 14, la visualización del inquilino muestra distribución no simétrica, con una ligera asimetría hacia la derecha, mostrando que hay una mayor concentración de clientes con valores de recencia más bajos, mostrando en pocas palabras que una menor cantidad de personas han comprado maquillaje recientemente.

#### 4.2.3.2 FRECUENCIA (FREQUENCY)

En el ámbito de frecuencia los datos se vuelven más cercanos entre ambos inquilinos, visualizando la figura 15 podemos comprender que existe una similitud entre ambos, la media de frecuencia ronda los 1.8 transacciones por inquilino en industria de maquillaje y 1.6 en la industria de la ferretería.

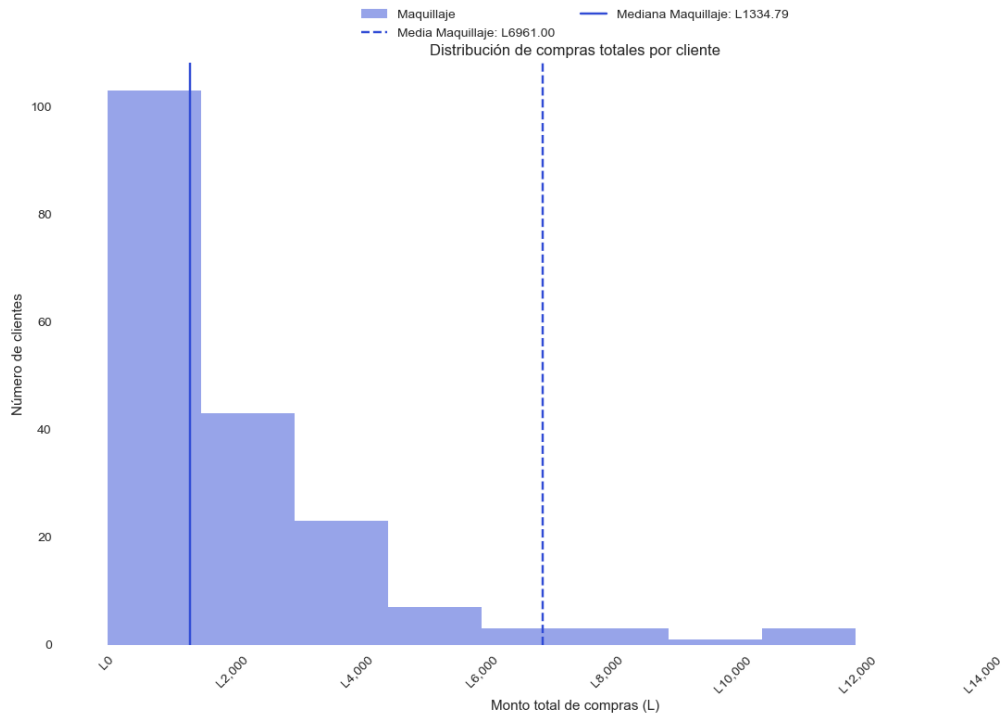


**Figura 15 Frecuencia de compra**

Fuente: Elaboración propia

#### 4.2.3.3 VALOR (MONETARY)

En esta siguiente sección tenemos como fin visualizar la distribución de compras en la plataforma por cada uno de los inquilinos, agrupándolos y sumando la frecuencia de aparición en 5 intervalos.

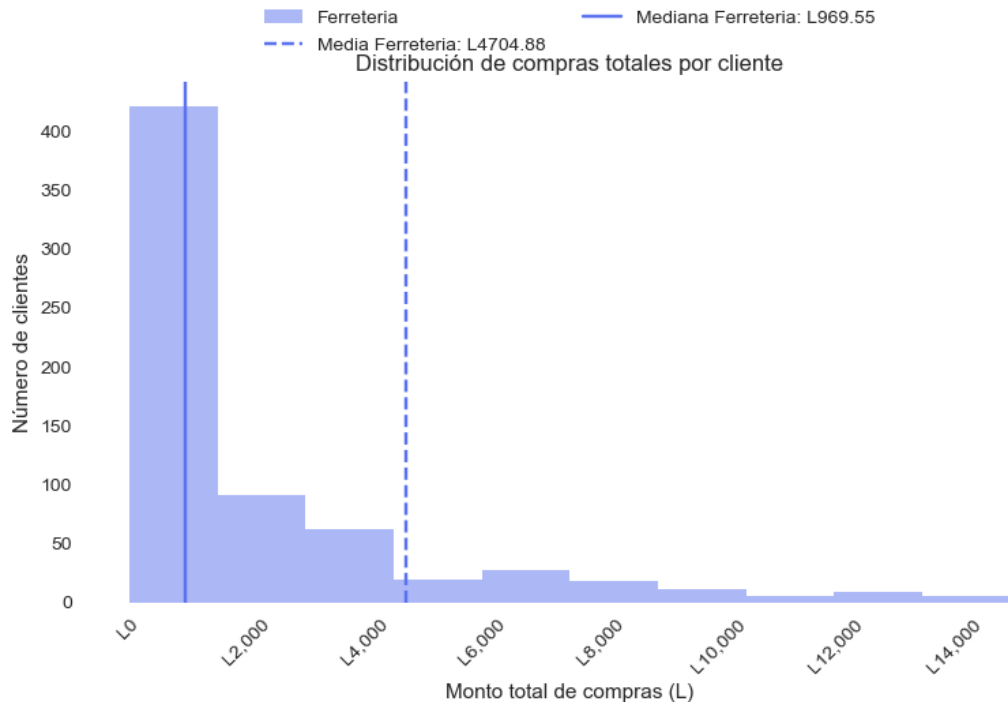


**Figura 16 Distribución de compras inquilino Maquillaje**

Fuente: Elaboración propia

Podemos observar una media de L6,961 lempiras por transacción, esto debido a valores atípicos que mueven la media a un valor alto, se añade la mediana que muestra que estas transacciones rondan en realidad a valores cercanos a los L1,334 lempiras para el inquilino maquillaje.

A su vez, el inquilino Ferretería, muestra valores que son afectados por valores atípicos, la media se encuentra en los L4,704; mientras que la mediana ronda los L969.



**Figura 17 Distribución de compras inquilino Ferreteria**

Fuente: Elaboración propia

#### 4.2.4 MODELOS PREDICTIVOS

Para este análisis se hizo uso explícitamente de lenguajes como Python, comúnmente conocido por el uso extendido para machine learning, con fin de mejorar el rendimiento del modelo utilizando los datos de LynxPay fue crucial realizar ingeniería de características, permitiendo obtener más información relevante dentro de la misma fuente, dada las anteriores visualizaciones se consideró pertinente evitar el uso de características geográficas, así como demográficas dato al considerable número de valores que en dichas columnas provenían de manera nula o vacía o a su vez, no existentes dentro de la fuente, como se hizo hincapié en los últimos en la sección 4.1.

Para esta fase se logra consolidar columnas tales como:

- Días desde la última compra (Recency).
- Cantidad de productos comprados (Frecuencia).
- Gasto promedio (Monetary)
- Cantidad de productos únicos comprados

```

this_df = this_df.loc[(this_df['category'].str.contains('50') == False)]
this_df['valor_total_pedido'] = this_df['price'] * this_df['quantity']
this_df['categoria'] = this_df['category'].str.lower().replace(' ', '_', regex=True)

# Calculate the number of days since the last purchase for each customer (Recency)
this_df.rename(columns={'created_at': 'fecha_factura'}, inplace=True)
this_df['fecha_factura'] = pd.to_datetime(this_df['fecha_factura'])
this_df['fecha_factura'] = this_df['fecha_factura'].dt.date
fecha_mas_reciente_por_cliente = this_df.groupby('customer_id')['fecha_factura'].max().reset_index()
fecha_maxima_en_df = pd.to_datetime(this_df['fecha_factura'].max())
fecha_mas_reciente_por_cliente['dias_desde_ultima_compra'] = (fecha_maxima_en_df - pd.to_datetime(fecha_mas_reciente_por_cliente['fecha_factura'])).dt.days
fecha_mas_reciente_por_cliente.drop(columns=['fecha_factura'], inplace=True)

valor_total_cliente = this_df.groupby('customer_id')['valor_total_pedido'].sum().reset_index()
valor_total_cliente.columns = ['customer_id', 'valor_total']

carritos_por_cliente = this_df.groupby('customer_id')['cart_id'].nunique().reset_index()
carritos_por_cliente.columns = ['customer_id', 'conteo_carritos']

# Customer purchases count (Frequency)
compras_cliente = this_df.groupby('customer_id')['cart_id'].count().reset_index()
compras_cliente.columns = ['customer_id', 'total_transacciones']
productos_comprados_cliente = this_df.groupby('customer_id')['quantity'].sum().reset_index()
productos_comprados_cliente.columns = ['customer_id', 'total_productos_comprados']

# Calculate the mean purchase value for each customer (Monetary)
valor_total_compra = valor_total_cliente.groupby('customer_id')['valor_total'].sum().reset_index()
valor_total_compra.columns = ['customer_id', 'gasto_total']
valor_promedio_transaccion = valor_total_compra.merge(compras_cliente, on='customer_id')
valor_promedio_transaccion['promedio_gasto_total'] = valor_promedio_transaccion['gasto_total'] / valor_promedio_transaccion['total_transacciones']

# Products diversity
productos_unicos_por_cliente = this_df.groupby('customer_id')['product_name'].nunique().reset_index()
productos_unicos_por_cliente.columns = ['customer_id', 'productos_unicos_comprados']
✓ 0.0s

```

**Figura 18 Código para generar características (Recency)**

Fuente: Elaboración propia

Para valores de características de tiempo de la columna “Fecha de creación” las siguientes columnas:

- Promedio de días entre compras
- Hora favorita de compra
- Día favorito de la semana

```

this_df['fecha_factura'] = pd.to_datetime(this_df['fecha_factura'])
this_df['day_of_week'] = this_df['fecha_factura'].dt.dayofweek
this_df['hour_of_day'] = this_df['fecha_factura'].dt.hour

# Days between purchases
days_between_purchases = this_df.groupby('customer_id')['fecha_factura'].apply(lambda x: (x.diff().dropna()).apply(lambda y: y.days))
avg_days_between_purchases = days_between_purchases.groupby('customer_id').mean().reset_index()
avg_days_between_purchases.columns = ['customer_id', 'avg_days_between_purchases']

# Favorite day of the week
favorite_day_of_week = this_df.groupby(['customer_id', 'day_of_week']).size().reset_index(name='count')
favorite_day_of_week = favorite_day_of_week.loc[favorite_day_of_week.groupby('customer_id')['count'].idxmax()]['customer_id', 'day_of_week']

# Favorite hour of the day
favorite_hour_of_day = this_df.groupby(['customer_id', 'hour_of_day']).size().reset_index(name='count')
favorite_hour_of_day = favorite_hour_of_day.loc[favorite_hour_of_day.groupby('customer_id')['count'].idxmax()]['customer_id', 'hour_of_day']
favorite_hour_of_day.columns = ['customer_id', 'favorite_hour_of_day']
✓ 0.0s

```

**Figura 19 Código para generar características (Frecuencia)**

Fuente: Elaboración propia

Para analizar los patrones de compra de los usuarios dentro de la fuente de datos, se generaron las siguientes características:

- Promedio de compras mensual

- Tendencia de compra (aumento, decadencia, ninguna)

```

from scipy.stats import linregress
this_df['year'] = this_df['fecha_factura'].dt.year
this_df['month'] = this_df['fecha_factura'].dt.month

# Monthly spending
monthly_spending = this_df.groupby(['customer_id', 'year', 'month'])['valor_total_pedido'].sum().reset_index()

# Check seasonal buying behavior
seasonal_buying_pattern = monthly_spending.groupby('customer_id')['valor_total_pedido'].agg(['mean', 'std']).reset_index()
seasonal_buying_pattern.rename(columns={'mean': 'gasto_promedio_mensual', 'std': 'std_gasto_mensual'}, inplace=True)
seasonal_buying_pattern['std_gasto_mensual'].fillna(0, inplace=True)

def calculate_trend(spend_data):
    # If there are more than one data points, we calculate the trend using linear regression
    if len(spend_data) > 1:
        x = np.arange(len(spend_data))
        slope, _, _, _ = linregress(x, spend_data)
        return slope
    # If there is only one data point, no trend can be calculated, hence we return 0
    else:
        return 0

trend_monthly_spend = monthly_spending.groupby('customer_id')['valor_total_pedido'].apply(calculate_trend).reset_index()
trend_monthly_spend.columns = ['customer_id', 'tendencia_mensual_gasto']

```

**Figura 20 Código para generar características de tendencia de compra**

Fuente: Elaboración propia

A estos valores anteriormente mencionados, añade la frecuencia de compra de cada categoría presente en la fuente de datos para cada inquilino, dado que cada inquilino mantiene diferentes categorías, estas mismas columnas pueden variar.

Posteriormente dentro del procedimiento se eligió utilizar “Isolation Forest” como algoritmo de detección de valores atípicos, es decir aquellos casos dentro de los análisis que parecían salirse de la norma, luego se realizó una normalización de los datos que permitieran a los algoritmos encontrar datos dentro de un rango en específico conocido, finalmente se realizó la división de valores de entrenamiento y prueba, con margen 80/20 respectivamente.

```

from sklearn.ensemble import IsolationForest
model = IsolationForest(contamination=0.05, random_state=42)
pivot_table['outlier_score'] = model.fit_predict(pivot_table.iloc[:, 1:].to_numpy())
pivot_table['is_outlier'] = pd.Series(pivot_table['outlier_score']).replace({1: False, -1: True})
pivot_table['is_outlier'].value_counts()

```

```

[102] ✓ 0.1s
...
is_outlier
False    118
True      7
Name: count, dtype: int64

```

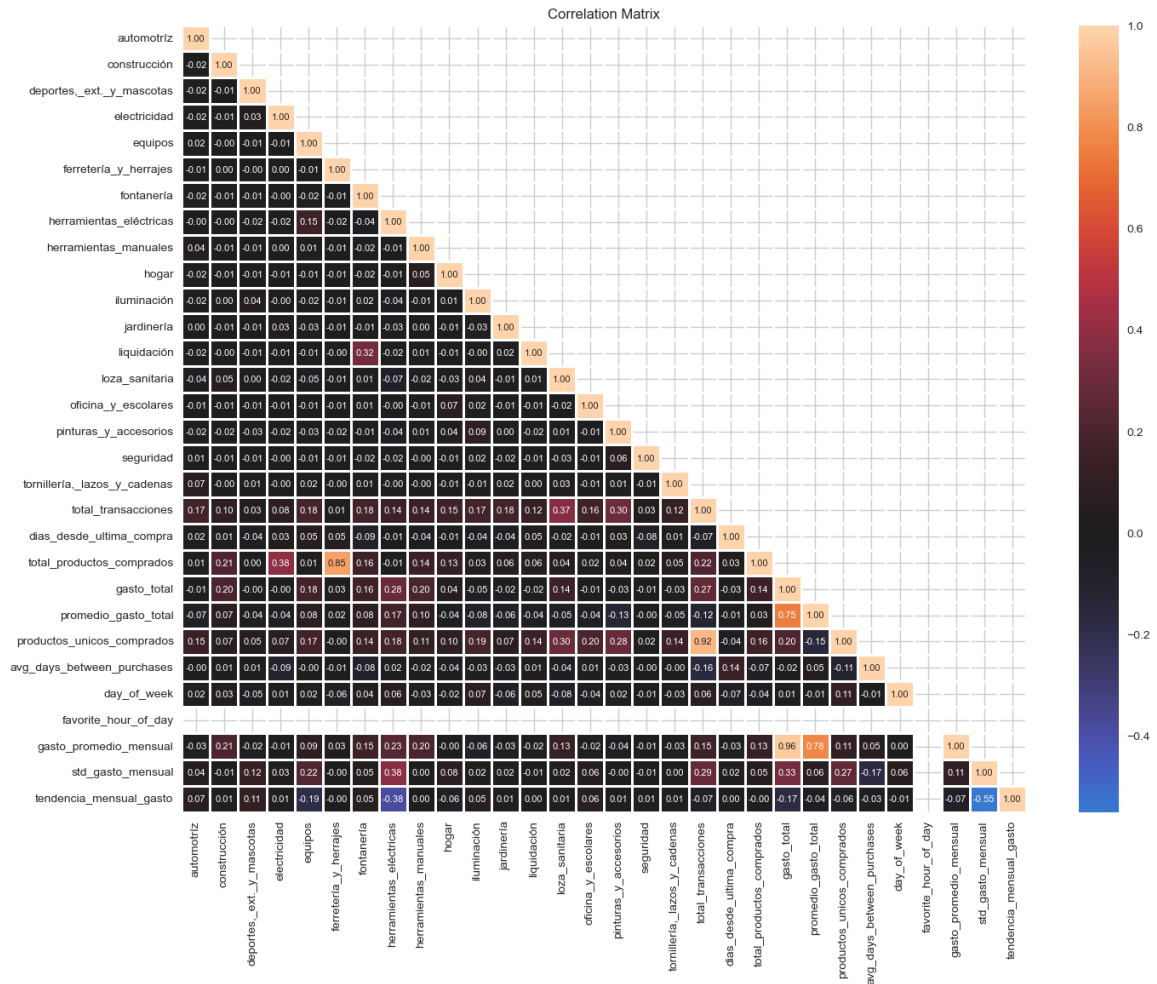
**Figura 21 Código para verificar valores atípicos**

Fuente: Elaboración propia

Como métricas de rendimiento para los modelos se utilizó el coeficiente Silhouette, como afirma Aurélien Géron (2023) el coeficiente de Silhouette puede rondar entre -1 y +1, los valores

cercanos a +1 significa que la instancia se encuentra bien agrupada, mientras que los cercanos a 0, significa que se encuentra cerca de una agrupación cercana, finalmente los cercanos a -1 significa que fueron asignados a una agrupación incorrecta.

Esta métrica nos permitirá conocer que algoritmo agrupa mejor los datos, para posteriormente realizar la implementación de técnicas de recomendación en cada grupo, mejorando el rendimiento de estas. (Duy-Nghia Nguyen, 2024)



**Figura 22 Matriz de correlación de variables en inquilino Ferretería**

Fuente: Elaboración propia

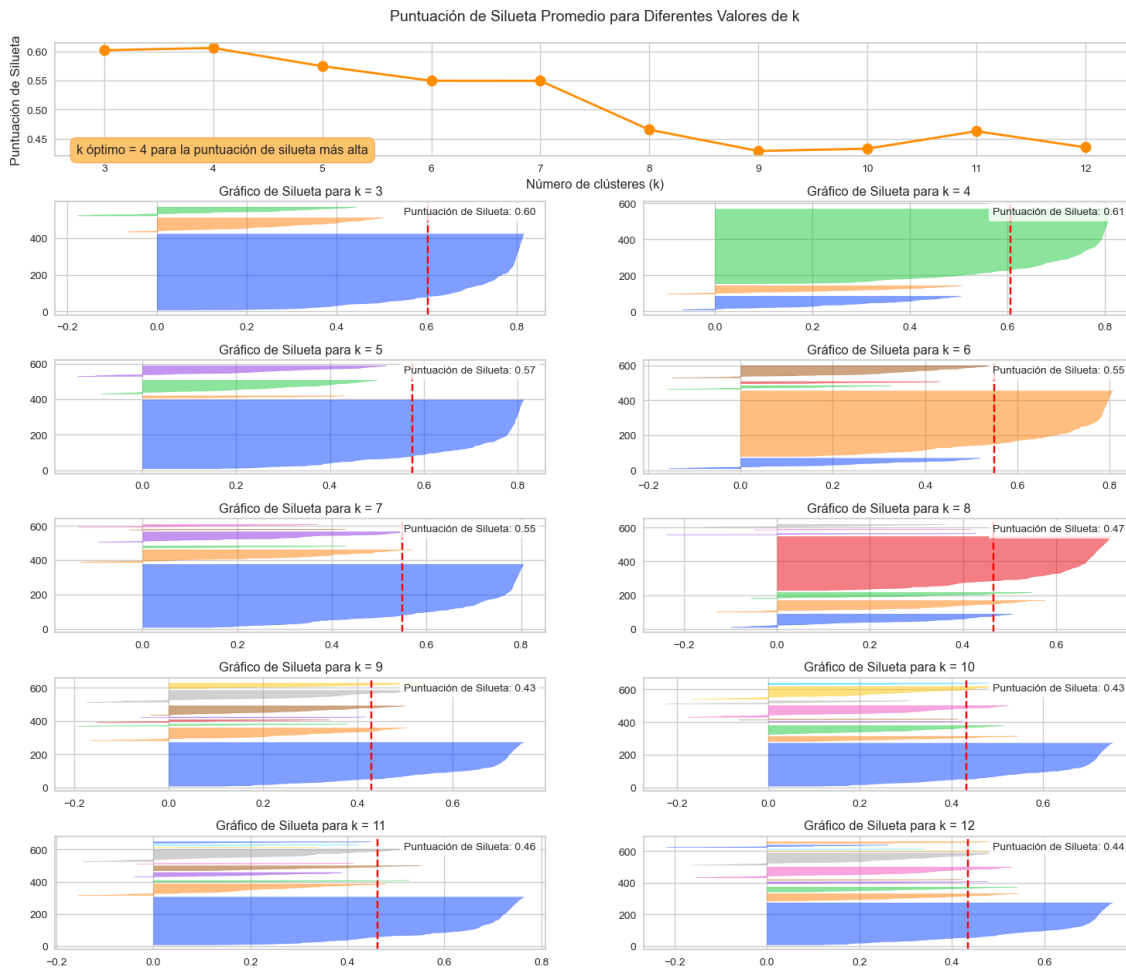
Existen ciertas correlaciones fuertes en los datos, por ejemplo, la variable “Cantidad de productos comprados” y “Cantidad de productos únicos” tienen una relación positiva, los clientes tienen a comprar mayor variedad de productos.

Correlaciones como “Cantidad de transacciones” y “Cantidad de productos comprados”

indica también que quienes realizan más transacciones, tienden a comprar una mayor cantidad de productos.

La mayoría de las categorías dentro de esta fuente, muestran poca relación, indicando que son relativamente independientes, dando a entender que, si un cliente compra artículos deportivos, no tiene mayor o menos probabilidad de comprar artículos de ferretería, en el caso de este inquilino.

#### 4.2.4.1 KMEANS



**Figura 23** Resumen de rendimiento de modelo Kmeans con diferente configuración de centroides

Fuente: Elaboración propia

Para el análisis de la figura 19 se interpretan diferentes criterios con el fin de conocer la mejor configuración de centroides para el análisis, desde el ancho de las siluetas, puntuación y

grosor del clúster.

Para el punto número uno, encontramos que se observan anchos relativamente buenos en la mayoría de los datos, sin embargo, existe notoriamente que existe un grupo predominante entre los datos, visualmente en el ancho de las siluetas entre tres y cuatro clústeres no se encuentra diferencia, pero, la configuración de tres clústeres aunque ofrece una menor puntuación que la configuración de cuatro, se observa que el nuevo grupo incluiría muy pocos valores relevantes para ser categorizado como grupo.

```
from collections import Counter
# Apply KMeans clustering using the optimal k
kmeans = KMeans(n_clusters=3, init='k-means++', n_init=10, max_iter=100, random_state=0)
kmeans.fit(pca_ds)

# Get the frequency of each cluster
cluster_frequencies = Counter(kmeans.labels_)

unique_labels = np.unique(kmeans.labels_)

# Step 2: Create dynamic mapping based on actual clusters
label_mapping = {}
target_order = [1, 0, 2] # Your desired order

for i, label in enumerate(unique_labels):
    if i < len(target_order):
        label_mapping[label] = target_order[i]
    else:
        label_mapping[label] = i # Fallback for extra clusters

# Step 3: Apply mapping
new_labels = np.array([label_mapping[label] for label in kmeans.labels_])

# Step 4: Add to pivot table
pivot_table['cluster'] = new_labels

pca_ds['cluster'] = new_labels
```

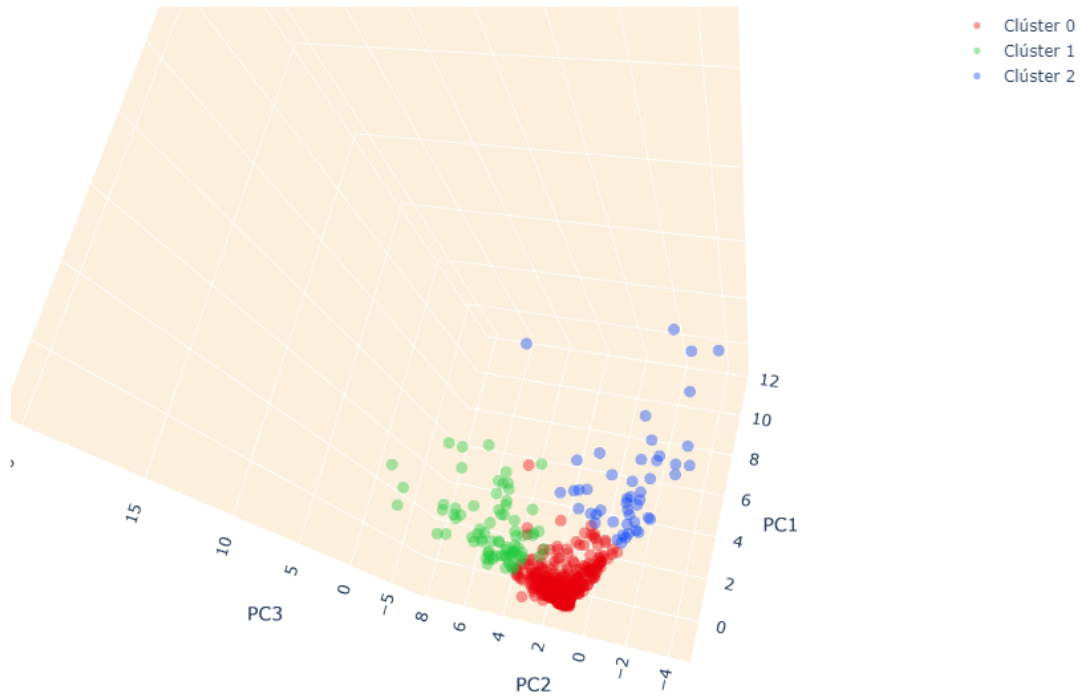
✓ 0.0s

**Figura 24** Código elaboración de clúster en datos

Fuente: Elaboración propia

#### 4.2.4.1.1 PCA

Visualización de Clústers en 3D



**Figura 25 PCA para visualización en tres dimensiones**

Fuente: Elaboración propia

Como se observa, existe una acumulación notable del clúster número 0 (rojo), presentando características cercanas de comportamiento, seguidamente del clúster 1 (verde), en valores medios de PC2 y PC3, finalmente el clúster 3 (azul) es el más pequeño, se ubica en valores altos de PC1 y bajos de PC2 y PC3.

Se distingue una relativa separación de los grupos con cierta superposición entre el clúster 0 y clúster 1.



**Figura 26 Distribución de clientes en cada clúster**

Fuente: Elaboración propia

Existe como en la figura 20, una distribución desigual y en esta figura 21 se hace notable aún más, el clúster 0 domina con un 76.57% de los clientes, mientras que los clústeres 1 y 2 tienen un 14.39% y 9.04% respectivamente, la concentración de clientes en el clúster 0 sugiere que este grupo puede representar a la mayoría de los clientes "típicos" o "promedio", mientras que los clústeres 1 y 2 podrían representar grupos más pequeños con características distintivas.

**Tabla 7 Puntuación de métricas de evaluación**

Métrica	Valor
Número de observaciones	542
Puntuación de Silueta	0.58
Puntuación de Calinski Harabasz	282
Puntuación de Davies Bouldin	0.86

Fuente: Elaboración propia

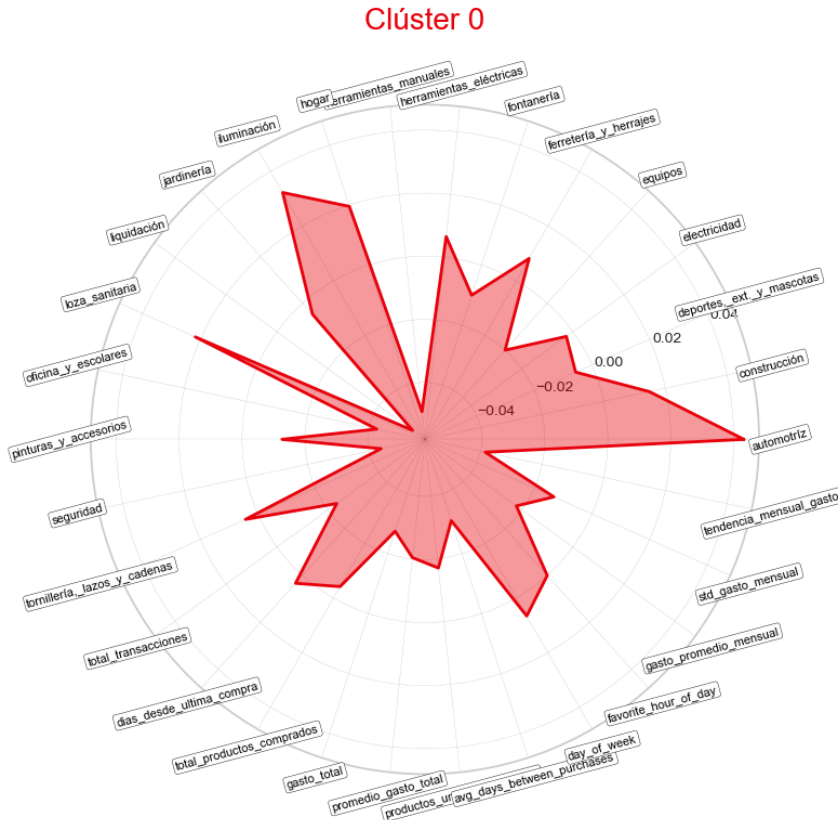
En general, la puntuación de Silueta del modelo kmeans es de 0.58 y se considera razonablemente bueno, lo que sugiere que los clústeres están relativamente bien separados. Esto significa que, en general, los objetos dentro de cada clúster son más similares entre sí que a los objetos de otros clústeres.

La puntuación Calinski Harabasz un valor de 282 sugiere una separación decente entre los clústeres, aunque sin ser excepcionalmente alta, cuanto mayor sea el valor, mejor definida estará la estructura de los clústeres, indicando una mayor separación entre clústeres por lo tanto una mayor cohesión dentro de cada clúster.

Y finalmente en la puntuación de Davies-Bouldin indica una estructura de clústeres

aceptable, aunque hay margen de mejora, cuanto menor sea el valor, mejor será la separación entre los clústeres.

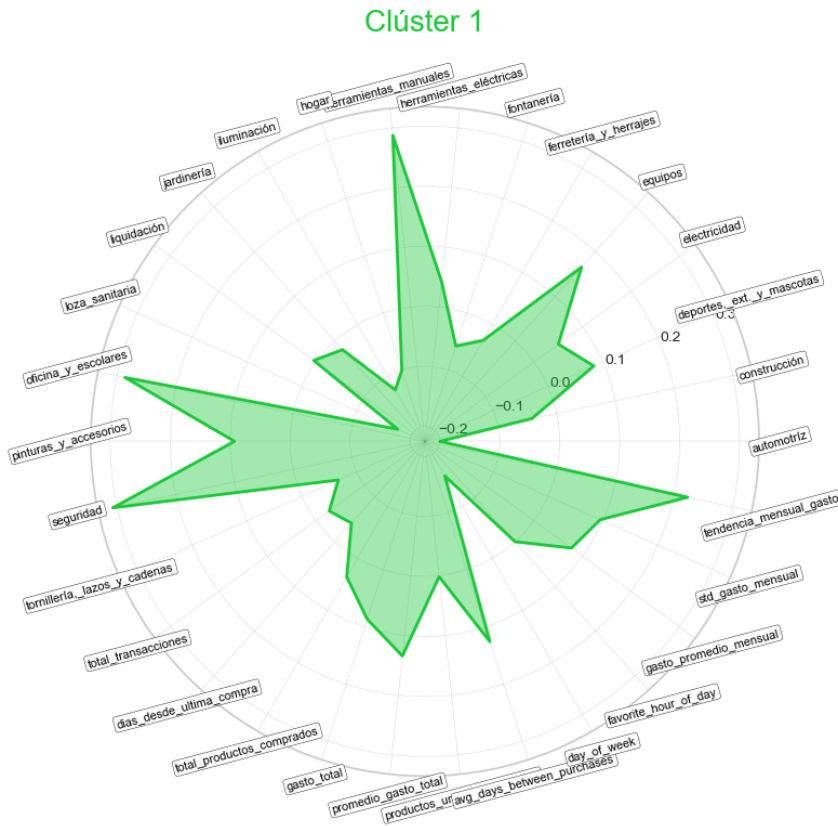
#### 4.2.4.1.2 ANÁLISIS DE CLÚSTERES



**Figura 27 Análisis clúster 0**

Fuente: Elaboración propia

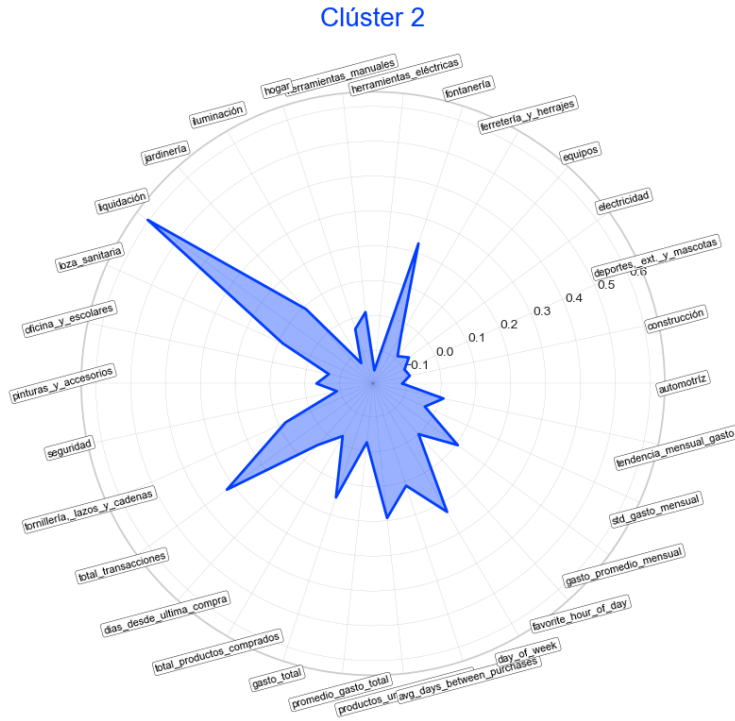
El análisis del Clúster 0, representado mediante un gráfico de radar, revela un patrón distintivo en el comportamiento de compra de este grupo. Se observa un gasto significativamente superior a la media en categorías como hogar, iluminación, automotriz, herramientas eléctricas y fontanería, lo que sugiere un marcado interés en productos de bricolaje y mejoras del hogar. Este patrón se complementa con un gasto moderadamente alto en ferretería y herrajes, y un comportamiento de compra cercano o inferior a la media en el resto de las categorías. Estos hallazgos permiten inferir que el Clúster 0 podría estar compuesto por individuos que realizan reparaciones o proyectos en sus hogares, o incluso profesionales del sector de la construcción.



**Figura 28 Análisis Clúster 1**

Fuente: Elaboración propia

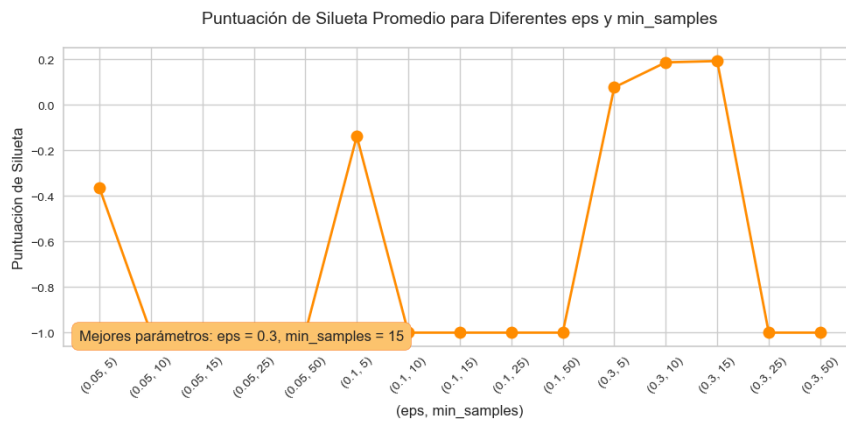
Se puede definir el perfil del Clúster 1 como un grupo de clientes con un comportamiento de compra diversificado, que invierte en una variedad de productos para el hogar, el trabajo, el ocio y la seguridad. Su gasto promedio total es alto, lo que indica un buen nivel adquisitivo, pero no se concentra en una categoría específica, sino que se distribuye en diferentes áreas de interés.



**Figura 29 Análisis clúster 2**  
Fuente: Elaboración propia

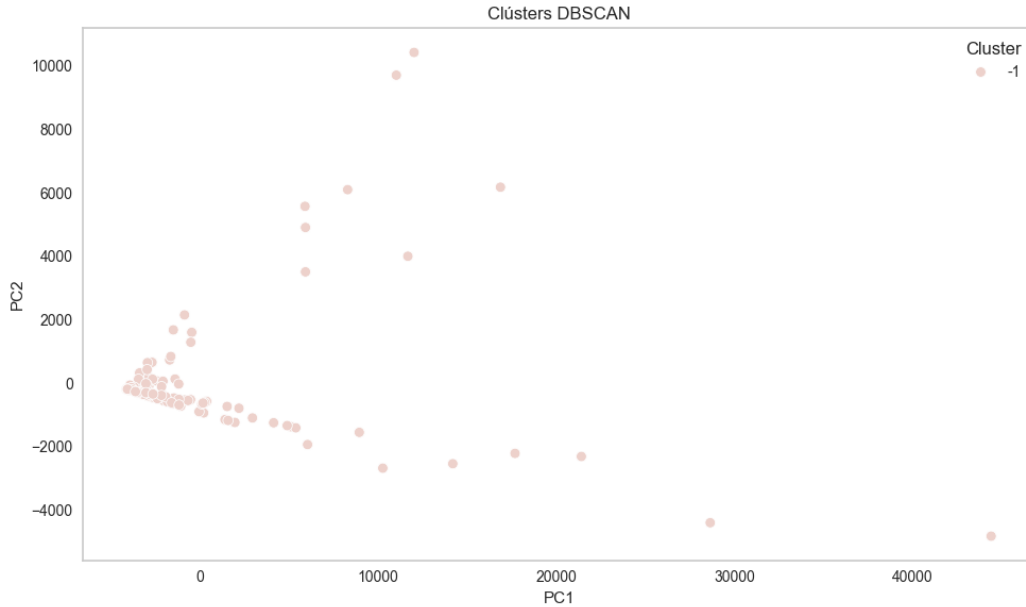
Este clúster parece representar un grupo de clientes "cazadores de ofertas" con un interés particular en productos de fontanería. Podrían ser personas que aprovechan las liquidaciones para adquirir productos para sus hogares o para su trabajo, o incluso pequeños comerciantes que buscan abastecerse a precios bajos.

#### 4.2.4.2 DBSCAN



**Figura 30 Puntuación Silueta entre parámetros**  
Fuente: Elaboración propia

Se puede apreciar en la figura 25, que variando los parámetros épsilon y número de muestras entre un rango entre 0.05 a 0.3 y muestras de 5 a 50, encontramos que los parámetros que mejor promedio de puntuación de Silueta obtiene es la configuración (0.3,15).



**Figura 31 Clústeres obtenidos utilizando algoritmo DBSCAN**

Fuente: Elaboración propia

Para el algoritmo DBSCAN cualquier valor categorizado con un valor negativo, es decir -1, se considera ruido dentro del análisis, en este sentido, DBSCAN parece no ser una opción ideal para el objetivo de este análisis, para confirmar lo anteriormente dicho, se visualizaron la puntuación de las demás métricas.

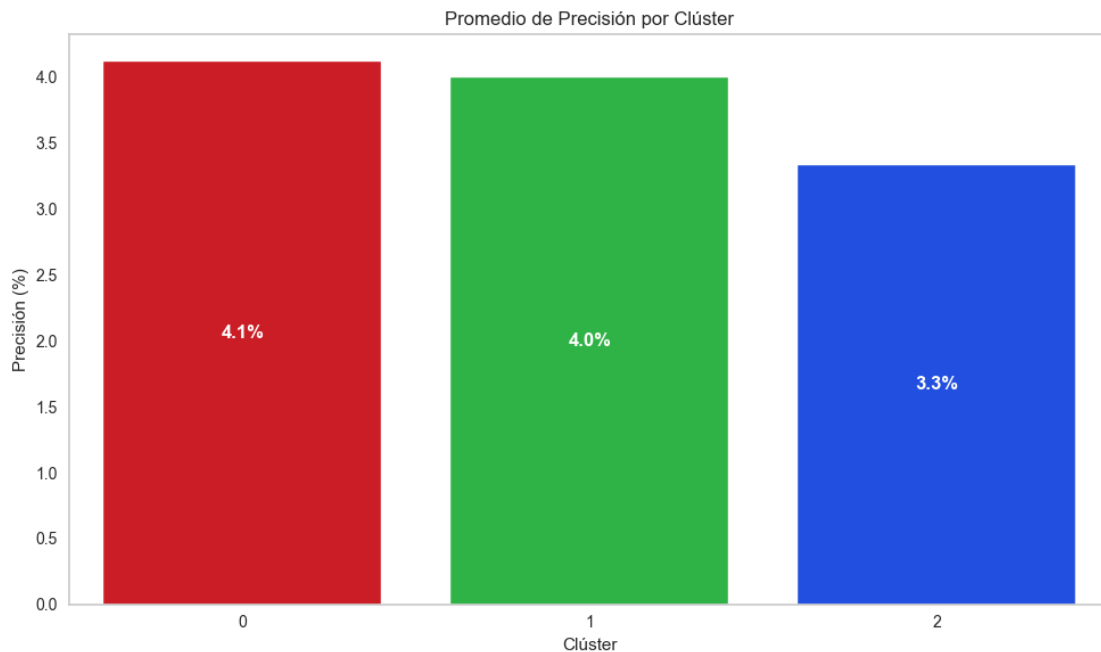
**Tabla 8 Métricas de puntuación DBSCAN**

Métrica	Valor
Número de observaciones	542
Puntuación de Silueta	0.19
Puntuación de Calinski Harabasz	60.60
Puntuación de Davies Bouldin	1.44

Fuente: Elaboración propia

Como se observa en la tabla 6, se visualiza una desmejora en las métricas, dado que como menciona Aurélien Géron (2023) el algoritmo DBSCAN funciona bien si los clústeres están bien separados por regiones de baja densidad. Por lo tanto, se descarta el uso de DBSCAN como método de agrupación.

#### 4.2.5 SISTEMA DE RECOMENDACIÓN



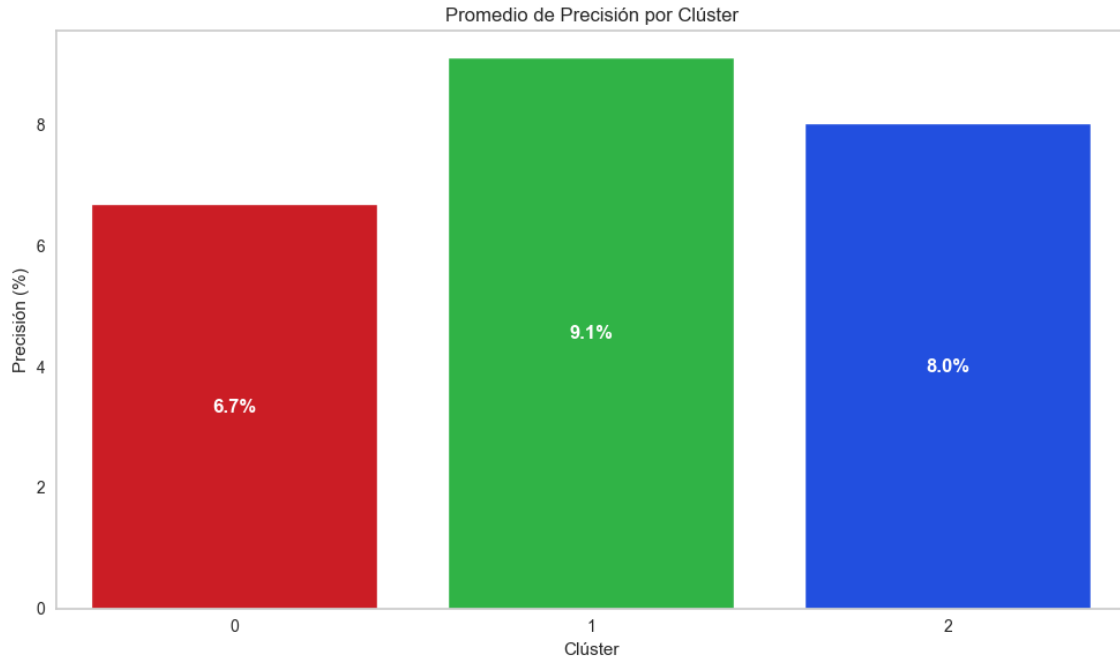
**Figura 32 Precisión de sistema de recomendación híbrido inquilino Ferretería**

Fuente: Elaboración propia

Se puede observar que utilizando el algoritmo de recomendación que implementa la segmentación de clientes, como a su vez utilizando técnicas de filtrado colaborativo y basado en contenido, se pueden ofrecer recomendaciones relevantes a los usuarios.

A su vez, se realizó la comparación de estas técnicas de manera independiente, y aunque no despreciable, estas consiguieron en promedio 3.26% de acierto en las recomendaciones, pero utilizando clústeres, se aumentó a 4.52% las recomendaciones.

Basado en este análisis, se puede constatar que, si al usuario se le recomiendan 100 productos dentro de su estancia dentro de la plataforma, 5 de ellos, podrían resultar en una compra.



**Figura 33 Precisión de sistema de recomendación híbrido inquilino Maquillaje**

Fuente: Elaboración propia

El resultado del inquilino maquillaje presenta una leve mejora en las recomendaciones, puntos claves de este inquilino, es que presenta una menor cantidad de clientes, pero con características más regulares y menos categorías de producto.

A su vez, se ve incrementado el rendimiento de los sistemas de recomendación en este caso, de 3.06% de precisión solo utilizando métodos colaborativos y de filtrado, comparado con la implementación híbrida de modelos de clústeres y métodos tradicionales, a un 8.37%.

Ejemplo de las recomendaciones antes mencionadas, es el usuario dentro del inquilino antes mencionado.

El usuario con id 3467 compró un producto “Judy Doll Ice Watery Lip Gloss”, se le recomendaron una cantidad de 10 productos.

**Tabla 9 Recomendaciones a usuario 3467, inquilino Maquillaje**

Recomendaciones	Resultado
Judy Doll Ice Watery Lip Gloss	
Maybelline Lash Sensational Sky High Mascara	
Judy Doll Watery Lip Gloss	Conversión
elf Glow Reviver Lip Oil	
Rare Beauty Soft Pinch Liquid Blush	
Laura Mercier Flawless To Go Mini Setting Duo	
Skin1004 Madagascar Centella Travel Kit	

Recomendaciones	Resultado
elf Power Grip Primer	
elf Power Grip Primer + 4% Niacinamide	
elf Stay All Night Blue Light Micro-Setting Mist	

Fuente: Elaboración propio

Lo que resultó en este caso en específico en una precisión del 10%.

En otro caso surgieron las siguientes recomendaciones, en el cual no tenía ningún historial de compra.

**Tabla 10 Recomendaciones a usuario 3501, inquilino maquillaje**

Recomendaciones	Resultado
Rare Beauty Perfect Stroke Volumizing Mascara	
Gucci Flora Gorgeous Gardenia Set	
Ulta Beauty Glow Getter	
Maybelline Lash Sensational Sky High Mascara	Conversión
Loreal True Match Hyaluronic Tinted Serum	
Nyx Thick it Stick it! Thickening Brow Gel Mascara	
Eucerin Pigment Control Sunscreen	
elf Power Grip Primer	
elf Stay All Night Blue Light Micro-Setting Mist	Conversión
Briogeo B. Well Organic + Australian 100%	

Fuente: Elaboración propia

Para el inquilino Ferretería, se muestran los siguientes ejemplos.

El siguiente usuario tuvo como historial de compra el producto “Botas de hule #12 (44)”.

**Tabla 11 Recomendaciones usuario 3120, inquilino ferretería**

Recomendaciones	Resultado
Botas de Hule #10 (42) 1003	
Botas de Hule #12 (44)	Conversión
Botas de Hule Baja #43 7 GATOS	
Tomacorriente Blanco Doble Polarizado EAGLE	
Accesorios para Colgar Cuadros	
Cuerda Elástica con Gancho 2 Piezas	
Barra de Seguridad para Inodoro	
Tinaco 750 Litros Beige Tricapa ROTOPLAS	
Lavadora a Presión 2000PSI 1.7GPM AR BLUE	
Tinaco 1,100 Litros Negro Bicapa ROTOPLAS	

Fuente: Elaboración propia

Surgiendo una precisión del 10% en este usuario particular, al que se le recomendó el mismo producto que compró, esta regla establecida en el código recomienda a los usuarios un mismo producto ya adquirido, por si este, estuviese interesado nuevamente en adquirirlo si fue de

su agrado.

#### 4.2.6 ANÁLISIS COSTO-BENEFICIO DEL ALGORITMO HÍBRIDO DE RECOMENDACIÓN

Se valida que, al momento de realizar esta investigación, no existen costos en términos de licenciamiento, debido a que las herramientas utilizadas son de uso libre, a su vez, LynxPay cuenta con infraestructura para ejecutar el entrenamiento y puesta en producción de un modelo predictivo, dado que ya está presente la misma, el costo se vuelve marginal para la organización. Tentativamente podrían surgir costos para la optimización de dicho modelo una vez se puedan implementar nuevas variables que incrementen la segmentación de los grupos presentes dentro de cada inquilino, permitiendo reducir la escasez de datos y mejorando los resultados de las predicciones.

Sobre los beneficios que tendría la implementación de un modelo de recomendación potenciado con la segmentación efectiva de clientes, se estima en términos monetarios que las ventas podrían aumentar considerablemente. Según los porcentajes ofrecidos por Bariliance (2024), se puede aumentar las tasas de conversión en productos recomendados aproximadamente un 15.2%. En un caso de estudio, utilizando técnicas de recomendación, se pudo incrementar en un 11.7% las tasas de conversión de una tienda de obras de arte. Considerando estos porcentajes, y dado que el AOV de LynxPay ronda los 40\$, podríamos esperar un incremento igual o más alto en el AOV de las tiendas participantes.

Este aumento se ve respaldado por el hecho de que regionalmente el AOV ronda \$80, lo que indica un alto potencial de crecimiento del 100% en las compras por usuario, es decir un aumento sustancial en las ventas netas en cada inquilino. La implementación de un modelo de recomendación efectivo, que considere no solo características de comportamiento, sino también demográficas y geográficas, permitiría ofrecer productos más relevantes a cada cliente, incentivando así un mayor gasto promedio por compra.

### 4.3 COMPROBACIÓN DE HIPÓTESIS

Hipótesis nula ( $H_0$ ) sobre el problema planteado en esta investigación:

El sistema de recomendación que utiliza modelos de agrupación no mejora el rendimiento en comparación con la implementación de algoritmos básicos de filtrado colaborativo. Esto implica que, al aplicar técnicas de clusterización, no se observa una mejora significativa en la precisión y relevancia de las recomendaciones en la plataforma de e-commerce.

Hipótesis de investigación ( $H_1$ ) planteada es la siguiente:

El sistema de recomendación que utiliza modelos de agrupación mejora el rendimiento en comparación con la implementación de algoritmos básicos de filtrado colaborativo. Esto significa que, mediante la aplicación de técnicas de clusterización, se logra una mejora significativa en la precisión y relevancia de las recomendaciones, proporcionando sugerencias más útiles y personalizadas para los usuarios de la plataforma de e-commerce.

Durante la presentación de la métrica de precisión de los sistemas de recomendación aplicados a los diferentes grupos dentro encontrados gracias a los modelos de agrupación, se pudo encontrar diferencias pequeñas pero notables en la mejora de recomendaciones al usuario que terminaron en una interacción a dicho producto por parte del cliente, en este caso, compra de un producto recomendado, a comparación del uso de sistemas de recomendación sin modelo de agrupación. Por lo cual se rechaza la hipótesis nula de investigación y se acepta la hipótesis alternativa que afirma:

El sistema de recomendación que utiliza modelos de agrupación mejora el rendimiento de los sistemas de recomendación clásicos como filtrado colaborativo y basado en contenido.

## CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES

### 5.1 CONCLUSIONES

1. A pesar de las limitaciones en los datos disponibles, este estudio demostró el valor de la segmentación de clientes y la personalización de recomendaciones en el contexto de LynxPay. Los modelos desarrollados lograron mejorar la precisión de las sugerencias, sin embargo, la ausencia de datos demográficos y geográficos limitó la capacidad de captar preferencias más profundas y contextuales.
2. El sistema de recomendación basado en técnicas de clusterización implementado en la plataforma LynxPay demostró una mejora significativa en la precisión de las recomendaciones. Al utilizar métricas como Precision@10, se verificó que los algoritmos empleados, especialmente K-Means, pudo identificar patrones de comportamiento en los usuarios y segmentarlos adecuadamente. Esto permitió que las recomendaciones ofrecidas fueran más relevantes y alineadas con los intereses y necesidades individuales de los clientes.
3. La precisión alcanzada confirma que el uso de modelos híbridos y técnicas avanzadas de análisis de datos no solo incrementa la efectividad de las sugerencias, sino que también mejora la experiencia del usuario en la plataforma. Estos resultados validan la viabilidad de emplear algoritmos de aprendizaje automático en contextos de comercio electrónico con datos limitados, siempre que se seleccionen adecuadamente las métricas y los modelos.
4. La viabilidad económica de esta iniciativa se ve respaldada por la infraestructura tecnológica existente de la empresa y el potencial de generar ingresos adicionales. Los resultados obtenidos sugieren que la inversión en esta área no solo es rentable a corto plazo, sino que también posiciona a LynxPay como una plataforma más competitiva y atractiva para sus emprendedores.
5. Se generaron propuestas estratégicas para enriquecer los modelos de recomendación de LynxPay con el objetivo de ofrecer una experiencia de usuario más personalizada y relevante. La incorporación de datos demográficos, geográficos y de navegación más detallados permitirá una segmentación más precisa y una adaptación de las recomendaciones a las preferencias individuales y al contexto de cada usuario.

## 5.2 RECOMENDACIONES

1. LynxPay debería establecer procesos robustos para recolectar y limpiar datos relevantes de los usuarios, como comportamientos de navegación, historial de compras y preferencias explícitas. La calidad de los datos es crucial para mejorar la precisión de las recomendaciones.
2. Utilizar las agrupaciones generadas por algoritmos de clusterización como K-Means y DBSCAN para desarrollar estrategias de segmentación más específicas. Esto permitirá ofrecer recomendaciones que se ajusten mejor a los intereses y necesidades de cada grupo.
3. Incorporar sistemas híbridos que combinen filtrado colaborativo y basado en contenido. Esta integración puede abordar las limitaciones de un único enfoque y aumentar la precisión al ofrecer recomendaciones contextualizadas.
4. Evaluar diferentes plataformas y herramientas de clusterización, incluyendo soluciones de código abierto y servicios en la nube, para determinar la opción que ofrezca el mejor equilibrio entre costo, escalabilidad y rendimiento para las necesidades de LynxPay. Considerar factores como el volumen de datos, la complejidad de los algoritmos y la frecuencia de actualización de las recomendaciones.
5. Desarrollar funcionalidades que consideren variables contextuales, como la ubicación geográfica, la temporada o el momento del día, para ajustar dinámicamente las recomendaciones y aumentar su relevancia, así como, enriquecer los perfiles de los usuarios con datos externos, como tendencias del mercado o patrones de consumo globales, para ofrecer recomendaciones más informadas y precisas.

## **CAPITULO VI. APLICABILIDAD**

### **6.1 NOMBRE DE LA PROPUESTA**

Plan de implementación de sistema de recomendación híbrido basado en modelos de clusterización para plataforma LynxPay.

### **6.2 JUSTIFICACIÓN DE LA PROPUESTA**

El uso optimizado del tiempo del usuario en plataformas de comercio electrónico es esencial para aumentar el valor de los clientes en la misma, así como mejorar la experiencia de usuario al poder encontrar lo que necesitan lo más pronto posible, resulta más rentable que el usuario visualice productos que puedan ser de interés conociendo factores básicos, como historial de compras, así como, factores demográficos y geográficos.

A través del análisis predictivo se pudo establecer que la implementación de técnicas de recomendación se puede mejorar las recomendaciones a los clientes que ingresen a la plataforma, así como conocer el grupo de comportamiento de este y conocer patrones que permitan a cada uno de los inquilinos establecer propuestas de mercadeo acordes a los grupos objetivos.

En el análisis costo beneficio se establece los costos marginales que representa esta implementación para la plataforma LynxPay, lo que podría significar mejoras notables en el AOV por cliente, logrando alcanzar cifras normales dentro del mercado latinoamericano.

La mejora en las recomendaciones en LynxPay significaría un aumento y normalización de los estándares de la región y posiblemente una entrada al uso de las mejores prácticas globales que maximicen el valor que se puede recolectar de los clientes, así mismo la utilidad de la plataforma en ofrecer soluciones a emprendedores hondureños.

### **6.3 ALCANCE DE LA PROPUESTA**

El objetivo de la propuesta es brindar el conocimiento de las tecnologías y procedimientos utilizados para la elaboración de un sistema de recomendación híbrido por inquilino, que permita a cada uno de los usuarios poder observar con mayor facilidad productos que les puedan interesar basados en información transaccional, demográfica y finalmente geográfica.

A continuación, se describe el objetivo general, así como, específicos que se buscaron alcanzar con la aplicación de la propuesta.

### 6.3.1 OBJETIVO GENERAL DE LA PROPUESTA

Implementar en la plataforma LynxPay un sistema de recomendación híbrido que retorne un listado de productos a partir de datos transaccionales, demográficos, geográficos y de comportamiento de los clientes de cada inquilino en LynxPay.

### 6.3.2 OBJETIVOS ESPECÍFICOS DE LA PROPUESTA

- Visualizar los resultados de la investigación y sistema de recomendación híbrido, como la precisión de sus recomendaciones.
- Establecer si existe infraestructura de IT para ejecutar el sistema de recomendación.
- Capacitar al recurso humano desarrollando la plataforma, que permita el uso del sistema de recomendación.
- Integración del sistema de recomendación en la plataforma.

## 6.4 DESCRIPCIÓN Y DESARROLLO

En la siguiente sección se presentan las actividades a realizar para la implementación de la propuesta de incorporación del sistema de recomendación para la plataforma LynxPay, que brinde recomendaciones a cada uno de los inquilinos en la misma.

### 6.4.1 DESCRIPCIÓN

Para que la siguiente propuesta tenga el impacto esperado en LynxPay, es necesario realizar un modelo disponible para el sistema a través de una capa de presentación, en este caso específico se pretende utilizar REST, con el fin de una variable de entrada, brindar las recomendaciones al usuario.

Fase 1: Elaboración de cadena de procesos para entrenamiento programado de modelo de agrupación.

Fase 2: Elaboración de servidor REST

### 6.4.2 DESARROLLO

En la siguiente sección se describen los lineamientos y requerimientos para la implementación de la propuesta.

#### 6.4.2.1 SOCIALIZACIÓN DE LOS RESULTADOS DE INVESTIGACIÓN

Es fundamental compartir los resultados de esta investigación para que los involucrados en el desarrollo de modelos de recomendación, usen la información y sugerencias que ofrece el modelo de recomendación híbrida. Para lograrlo, se usarán estrategias y métodos específicos para comunicar los hallazgos de manera efectiva.

- Plataforma Interactiva de Visualización

Utilizando herramientas como Streamlit que permita la generación de un aplicativo de visualización de las métricas, generando información relevante a cada uno de los inquilinos de la plataforma, o la misma implementación de estos gráficos en la consola general de cada inquilino.

- Comunicación Interna Constante

Se mantendrá al personal informado sobre los avances y resultados del proyecto a través de boletines, correos electrónicos y otros canales internos.

- Presentaciones Ejecutivas

Se realizarán presentaciones concisas y visualmente atractivas para la alta gerencia e inquilinos. Estas presentaciones destacarán los objetivos, la metodología, los hallazgos principales y las recomendaciones estratégicas del proyecto.

- Informes Detallados

Se elaborarán informes exhaustivos que documenten el proceso de investigación, los resultados y las recomendaciones. Estos informes se distribuirán a los inquilinos y departamentos técnicos para asegurar una implementación coherente del modelo.

- Reuniones de Seguimiento

Se programarán reuniones periódicas con los equipos involucrados para monitorear el progreso de la implementación, ajustar estrategias y resolver dudas. Estas reuniones también servirán para obtener retroalimentación sobre el uso del sistema de recomendación.

#### 6.4.2.2 PASOS REQUERIDOS PARA CONSTRUIR Y DESPLEGAR EL SISTEMA DE RECOMENDACIÓN

Para la implementación del sistema de recomendación propuesto, la empresa actualmente

cuenta con la siguiente infraestructura de IT que permite el despliegue de estos sistemas, como el actualmente sugerido.

El equipo proporcionado por la organización cuenta con:

- 2064 GB de almacenamiento
- 32 GB de memoria RAM
- CPU de 16 núcleos

Dada las herramientas de libre acceso utilizadas para la implementación, la empresa no requiere de software específico.

Se recomienda la instalación de Python 3.11, con las siguientes librerías para los procedimientos a mencionar.

- Sklearn
- Pandas
- Numpy
- Seaborn
- Matplotlib

Para construir e implementar el sistema de recomendación se deben realizar los siguientes procedimientos:

1. Extracción de datos: Se genera la fuente de datos para el modelo de recomendación mediante una consulta directa a una base de datos. Esta consulta se ejecuta periódicamente para mantener los datos actualizados con las nuevas interacciones y preferencias de los usuarios.

- a) Se elabora una consulta (query) SQL para extraer los datos relevantes de la base de datos.
- b) Los resultados de la consulta se exportan a un archivo en formato CSV (valores separados por comas). Este archivo servirá como entrada para la etapa de modelado.

2. Análisis descriptivo de los datos: Se realiza un análisis exploratorio de los datos extraídos para comprender sus características, distribuciones y posibles valores atípicos. Este análisis ayuda a identificar patrones iniciales y a guiar la posterior preparación de los datos.

3. Preparación de los datos: Se utiliza un script en Python para llevar a cabo la limpieza, transformación y preprocesamiento de los datos.

El script en Python lee los datos desde el archivo CSV generado en la etapa de extracción.

- a) Se aplican diversas técnicas de limpieza, como el manejo de valores faltantes, la eliminación de duplicados y la corrección de errores.
- b) Se realizan transformaciones relevantes, como la normalización, la estandarización o la codificación de variables categóricas, según sea necesario para el algoritmo de clusterización.

4. División de datos estratificada para evaluación: Se realiza una división estratificada de los datos, enfocándose en clientes con múltiples compras para asegurar una evaluación representativa.

- a) Se identifican a los clientes con más de una compra.
- b) Para estos clientes, se realiza una división 70/30 de sus transacciones: el 70% de sus transacciones se utilizarán para entrenamiento (train) y el 30% restante para prueba (test). Esta división se hace a nivel de transacción por usuario, no a nivel de usuario. Es decir, un mismo usuario tendrá transacciones en el conjunto de entrenamiento y en el de prueba. Esto es crucial para evaluar la capacidad del modelo de generalizar a nuevas interacciones del mismo usuario.
- c) Esta estratificación asegura que la evaluación se realiza sobre datos que representan el comportamiento real de los usuarios más activos, que suelen ser los más importantes para el negocio.
- d) Los clientes con una sola compra se incluyen únicamente en el conjunto de entrenamiento, dado que no es posible realizar una división estratificada para ellos.

5. Modelado y clusterización: Se implementa el algoritmo de clusterización en Python.

Se utiliza una biblioteca de Python como scikit-learn para implementar el algoritmo de clusterización (e.g., K-Means, DBSCAN, etc.).

- a) Se entrena el modelo de clusterización con los datos preparados. Este proceso asigna cada usuario a un clúster basándose en sus características y preferencias.

- b) El modelo entrenado, que contiene la información de los clústeres y sus centroides (o características representativas), se guarda en un archivo en formato pickle. Este formato permite serializar el modelo para su posterior uso.

6. Implementación del servicio de recomendaciones: Se crea un servicio REST que utiliza el modelo de clusterización para generar recomendaciones personalizadas.

- a) El servicio REST carga el modelo guardado en formato pickle (.pkl).
- b) Cuando un usuario realiza una solicitud, el servicio procesa sus datos y utiliza el modelo para asignarlo a un clúster específico.
- c) Basándose en las características del clúster al que pertenece el usuario, el servicio genera recomendaciones de productos o servicios relevantes como se observa en el anexo 2.

#### 6.4.2.3 INTEGRACIÓN DE SISTEMA DE RECOMENDACIÓN EN LYNXPAY

Para integrar y socializar los resultados del sistema de recomendación basado en clusterización en las operaciones actuales de la organización, se deben seguir los siguientes pasos:

1. Integración del modelo: Se implementa la integración del modelo de clusterización con los sistemas de la organización.
  - a) El modelo entrenado (archivo .pkl) se integra a un servicio REST, como se definió anteriormente. Este servicio actúa como una interfaz entre el modelo y las aplicaciones que consumirán las recomendaciones.
  - b) Se documenta la API REST (con Open API, como se mostró en el ejemplo anterior) para que otros equipos y sistemas puedan integrarse fácilmente.
  - c) Se implementan mecanismos de registro (logs) en el servicio REST para monitorear su funcionamiento y detectar posibles errores.
2. Despliegue y escalabilidad: Se despliega el servicio REST en un entorno de producción robusto que garantice su disponibilidad y escalabilidad.
  - a) Se utiliza un contenedor como Docker y una plataforma de orquestación como Kubernetes (o similar) para facilitar el despliegue y la gestión del servicio.

- b) Se configuran mecanismos de monitoreo y alertas para detectar problemas de rendimiento o fallos en el servicio.
  - c) Se considera la escalabilidad horizontal del servicio para manejar un volumen creciente de solicitudes de recomendación.
3. Actualización y reentrenamiento del modelo: Se establece un proceso para la actualización y el reentrenamiento periódico del modelo de clusterización.
- a) Se define una frecuencia de reentrenamiento (por ejemplo, mensual, trimestral o según la evolución de los datos). Esta frecuencia dependerá de la volatilidad de los datos y de los cambios en las preferencias de los usuarios.
  - b) Se automatiza el proceso de extracción de datos, preparación, entrenamiento del modelo y despliegue del nuevo modelo en el servicio REST. Esto se puede lograr mediante scripts o herramientas de automatización de pipelines CI/CD (Integración Continua/Entrega Continua).
  - c) Se realiza un seguimiento del rendimiento del modelo a lo largo del tiempo utilizando métricas relevantes. Si el rendimiento se deteriora, se investigan las causas y se ajustan los parámetros del modelo o el proceso de preprocesamiento de datos.
4. Comunicación y socialización de resultados: Se comunican los resultados del modelo y las recomendaciones generadas a los diferentes inquilinos.
- a) Se generan informes periódicos que resumen el rendimiento del modelo, los principales clústeres identificados y ejemplos de recomendaciones generadas.
  - b) Se presentan los resultados a los equipos de marketing, etc., para que puedan utilizar las recomendaciones en sus estrategias y operaciones.
  - c) Se crea un canal de comunicación para recibir feedback de los diferentes departamentos sobre la utilidad y la precisión de las recomendaciones.
  - d) Se realiza una capacitación a los usuarios finales sobre el uso de la API de recomendaciones, los diferentes parámetros de entrada, y el formato de los datos de salida.

Este proceso integral asegura que el sistema de recomendación se integre de manera efectiva en las operaciones de LynxPay, se mantenga actualizado y proporcione valor a los diferentes equipos. La automatización y el monitoreo continuo son claves para garantizar la eficiencia y la confiabilidad del sistema. Además, una comunicación clara y efectiva con los diferentes inquilinos asegura la adopción y el aprovechamiento máximo de las recomendaciones generadas.

#### 6.4.2.4 CAPACITACIÓN PARA TRABAJAR CON EL MODELO

Para la propuesta de capacitación se hará uso de la técnica de estimación de tres puntos, es una herramienta esencial en la gestión de proyectos que permite obtener una visión más precisa y realista de la duración o coste de una tarea, especialmente cuando existe incertidumbre. A diferencia de las estimaciones de un solo punto, que se basan en una única cifra, este método considera tres escenarios posibles: optimista, pesimista y más probable . Al combinar estos tres valores, se obtiene una estimación más completa que tiene en cuenta la variabilidad e incertidumbre inherentes a cualquier proyecto. (Ortiz)

El proceso de estimación de tres puntos implica los siguientes pasos:

- 1- Identificar la tarea: Define claramente la tarea o actividad que se va a estimar.
- 2- Determinar las tres estimaciones: Para cada tarea, se deben obtener tres valores, y es crucial que el equipo de desarrollo participe activamente en este proceso para obtener estimaciones más acertadas :
  - Estimación optimista (O): Representa el mejor escenario posible, donde todo sale según lo planeado y no hay retrasos ni contratiempos.
  - Estimación pesimista (P): Considera el peor escenario posible, donde se presentan obstáculos, retrasos y complicaciones.
  - Estimación más probable (M): Es la estimación que se considera más realista, basada en la experiencia y el conocimiento del equipo. La utilización de datos históricos de proyectos anteriores puede mejorar significativamente la precisión de esta estimación .

**Tabla 12 Plan de capacitación de propuesta**

Tema	Contenido	Tiempo Optimista	Tiempo Más Probable	Tiempo Pesimista	Tiempo Estimado	Costo Aproximado (USD)
<b>Arquitectura del Sistema de Recomendación (Revisión)</b>	Breve repaso de la arquitectura, enfatizando el flujo de datos desde la solicitud a la API hasta la respuesta con las recomendaciones.	0.5 horas	1 hora	1.5 horas	1 hora	<b>\$50 - \$100</b>
<b>Esquemas de Respuesta de la API (en detalle)</b>	Descripción exhaustiva de los esquemas JSON de las respuestas de la API, incluyendo todos los campos, tipos de datos y posibles valores. Ejemplos concretos y casos de uso.	2.5 horas	3 horas	3.5 horas	3 horas	<b>\$150 - \$300</b>
<b>Métricas de Evaluación del Modelo de Clusterización</b>	Explicación detallada de la métrica del Coeficiente de Silueta (Silhouette Score). Cómo se calcula, qué representa, rangos de valores y cómo interpretar los resultados para evaluar la calidad del clustering.	1.5 horas	2 horas	2.5 horas	2 horas	<b>\$100 - \$200</b>
<b>Métricas de Evaluación de las Recomendaciones</b>	Explicación detallada de la métrica de Precisión@K. Cómo se calcula, qué representa, cómo elegir el valor de K y cómo interpretar los resultados para evaluar la calidad de las recomendaciones.	1.5 horas	2 horas	2.5 horas	2 horas	<b>\$100 - \$200</b>
<b>Introducción a otros Algoritmos de Recomendación</b>	Introducción conceptual a los algoritmos de Filtrado Colaborativo (basado en usuario y basado en ítem) y Filtrado Basado en Contenido.	2 horas	2.5 horas	3 horas	2.5 horas	<b>\$125 - \$250</b>
<b>Casos de Uso y Mejores Prácticas</b>	Ejemplos prácticos de cómo utilizar la API para diferentes escenarios.	2.5 horas	3 horas	3.5 horas	3 horas	<b>\$150 - \$300</b>

Tema	Contenido	Tiempo Optimista	Tiempo Más Probable	Tiempo Pesimista	Tiempo Estimado	Costo Aproximado (USD)
<b>Q&amp;A y Feedback</b>	Sesión de preguntas y respuestas para aclarar dudas y recibir feedback.	0.5 horas	1 hora	1.5 horas	1 hora	<b>\$50 - \$100</b>
<b>Total</b>					14.5 horas	<b>\$725 - \$1450</b>

Fuente: Elaboración propia

## 6.5 MEDIDAS DE CONTROL.

A continuación, se describen los indicadores clave de desempeño (KPIs) establecidos para evaluar el impacto y la efectividad del modelo de recomendación personalizado implementado en LynxPay. Estos indicadores permiten medir la mejora en las métricas clave del negocio y en la experiencia de las recomendaciones del usuario.

1. **Precisión del Modelo (PM):** Evalúa la proporción de recomendaciones relevantes ofrecidas por el sistema en relación con las interacciones reales de los usuarios.

Fórmula:

$$Precision = \frac{sum(R(u) \cap T(u))}{sum(R(u))} \quad (1)$$

Donde:

- **R(u):** Representa la lista de recomendaciones hechas al usuario u basadas en su comportamiento en el conjunto de entrenamiento. Es decir, es el conjunto de ítems que el sistema le recomienda al usuario.
  - **T(u):** Es el conjunto de ítems que el usuario realmente ha valorado o seleccionado en el conjunto de prueba. Este conjunto representa lo que el usuario realmente considera relevante.
  - **(R(u)∩T(u)):** Indica la cantidad de ítems relevantes recomendados correctamente.
2. **Valor Promedio por Orden (AOV):** Este indicador mide el valor promedio de las transacciones realizadas en la plataforma. Es útil para identificar el impacto económico directo de las recomendaciones.

Fórmula:

$$AOV = \frac{\text{Ingresos Totales}}{\text{Número Total de Órdenes}} \quad (2)$$

Donde:

- Ingresos Totales: Es la suma total de dinero generado por todas las transacciones realizadas en un período de tiempo específico. Esto incluye todas las ventas procesadas, sin excluir devoluciones o descuentos.
  - Número Total de Órdenes: Representa la cantidad total de transacciones realizadas en la plataforma en el mismo período. Cada compra realizada por un cliente se considera una orden única, independientemente de la cantidad de productos adquiridos en esa transacción.
3. Índice de Rentabilidad del Modelo (IRM): Analiza la relación entre el incremento en ingresos generado por el modelo y los costos de su implementación.

Fórmula:

$$IRM = \left( \frac{\text{Incremento en ingresos atribuible al modelo}}{\text{Costo Total de Implementación}} \right) \times 100 \quad (3)$$

Donde:

- Incremento en ingresos atribuible al modelo: Es la cantidad de dinero que se generó exclusivamente gracias al uso del modelo de recomendación. Se calcula comparando los ingresos antes y después de la implementación del modelo, descontando cualquier aumento que no esté directamente relacionado con el sistema.
- Costo total de implementación: Incluye todos los gastos necesarios para desarrollar, implementar y mantener el modelo. Esto puede abarcar costos de desarrollo, licencias de software, infraestructura computacional, capacitación del personal y tiempo dedicado por el equipo.
- Multiplicado por 100: Convierte el índice en un porcentaje, facilitando su interpretación como una proporción de rentabilidad.

## 6.6 CRONOGRAMA DE IMPLEMENTACIÓN Y PRESUPUESTO.

En la presente sección se detalla el cronograma de actividades a realizar para implementar la propuesta, así como los costos de esta misma.

### 6.6.1 ACTIVIDADES DE IMPLEMENTACIÓN DE PROPUESTA

En la tabla 11 se especifica las diferentes actividades para la implementación de la propuesta, describiendo los tiempos esperados para la misma y a quien estaría a cargo esta actividad.

**Tabla 13 Actividades implementación de propuesta**

Actividad	Inicio	Final	Días	Responsable
Definición de Requisitos y Socialización del Proyecto	10/01/2025	10/01/2025	1	Analista de Datos
Preparación y Configuración del Entorno de Datos	13/01/2025	14/01/2025	2	Ingeniero de Datos
Extracción, Transformación y Carga (ETL) de Datos Históricos	15/01/2025	27/01/2025	9	Ingeniero de Datos
Análisis Exploratorio de Datos y Definición de Features	28/01/2025	06/02/2025	8	Analista de Datos
Ingeniería de Features y Preprocesamiento	07/02/2025	12/02/2025	4	Ingeniero de Datos
Desarrollo y Entrenamiento de Modelos de Recomendación	13/02/2025	26/02/2025	10	Científico de Datos
Evaluación y Validación de Modelos	27/02/2025	05/03/2025	5	Científico de Datos
Implementación de la API de Recomendaciones	06/03/2025	07/03/2025	2	Ingeniero de Datos
Pruebas de la API e Integración	10/03/2025	11/03/2025	2	Ingeniero de Datos
Documentación y Capacitación Interna	12/03/2025	13/03/2025	2	Científico de Datos
Duración del proyecto	10/01/2025	13/03/2025	44	

Fuente: Elaboración propia

El cronograma presentado detalla las actividades necesarias para la implementación del sistema de recomendaciones en LynxPay, asignando responsabilidades a cada rol (Ingeniero de Datos, Analista de Datos y Científico de Datos). Además, se añade un diagrama de Gantt, en las que se muestra los pasos a realizar, así como sus fechas para realizar la implementación.

Para el éxito del proyecto, es crucial la colaboración entre el equipo de implementación y los equipos relevantes de LynxPay. Esta colaboración asegurará la correcta integración del sistema de recomendaciones en las operaciones y estrategias de la empresa.

Después de la implementación técnica, la capacitación al personal y la puesta en marcha

del sistema, las tareas principales pasarán a ser el mantenimiento del modelo y el seguimiento de los resultados. Se estima que estas actividades se realizarán dos veces al año, comenzando en 2025. Este mantenimiento regular asegurará la precisión y la relevancia continua de las recomendaciones a lo largo del tiempo.

El cronograma actualiza las fechas de inicio al 10 de enero de 2025, con una duración total



estimada del proyecto de 44 días, finalizando aproximadamente el 13 de marzo de 2025. Esto proporciona a LynxPay un marco temporal claro para la implementación del sistema de recomendaciones y la posterior planificación de sus estrategias basadas en los resultados de este.

**Figura 34 Diagrama de Gantt de actividades de implementación de propuesta**

Fuente: Elaboración propia

### 6.6.2 COSTOS DE LA IMPLEMENTACIÓN DE LA PROPUESTA

Esta sección presenta un desglose de los costos incurridos durante la implementación del proyecto. Los cálculos se basan en los salarios promedio vigentes en Honduras, lo cual permite obtener una estimación realista de la inversión necesaria. El costo unitario representa el salario diario correspondiente al perfil profesional responsable de cada tarea, mientras que el costo total de cada actividad se obtiene multiplicando este costo unitario por la duración en días de esta. Para mayor detalle sobre las referencias salariales y los costos de equipo considerados en la propuesta, se remite al lector a los anexos del 2 al 6.

**Tabla 14 Costos de Propuesta**

Actividad	Días	Rol	Costo Unitario (Lps)	Costo Total (Lps)	Referencia
Definición de Requisitos y Socialización	1	-	-	-	
Preparación y Configuración del Entorno de Datos	2	-	-	-	

<b>ETL de Datos Históricos</b>	9	Ingeniero de Datos	L1,000.00	L9,000.00	<a href="https://www.glassdoor.com/Sueldos/san-pedro-sula-ingeniero-de-datos-sueldo-SRCH_IL.0,14_IC2362995_KO15,33.htm">https://www.glassdoor.com/Sueldos/san-pedro-sula-ingeniero-de-datos-sueldo-SRCH_IL.0,14_IC2362995_KO15,33.htm</a>
--------------------------------	---	--------------------	-----------	-----------	---

Fuente: Elaboración propia

### Continuación de la Tabla 10. Costos de Propuesta

Actividad	Días	Rol	Costo Unitario (Lps)	Costo Total (Lps)	Referencia
<b>Análisis Exploratorio de Datos y Definición de Features</b>	8	Analista de Datos	L668.49	L5,347.92	<a href="https://hn.computrabajo.com/salarios/analista">https://hn.computrabajo.com/salarios/analista</a>
<b>Ingeniería de características y Preprocesamiento</b>	4	Científico de Datos	L1,291.00	L5,164.00	<a href="#">Salario de Científico de Datos</a>
<b>Desarrollo y Entrenamiento de Modelos de Recomendación</b>	10	Científico de Datos	L1,291.00	L12,910.00	
<b>Evaluación y Validación de Modelos</b>	5	Científico de Datos	L1,291.00	L6,455.00	
<b>Implementación de la API de Recomendaciones</b>	2	Ingeniero de Software	L800.00	L1,600.00	<a href="#">Salario de Ingeniero de Software, Honduras.</a>
<b>Pruebas de la API e Integración</b>	2	Ingeniero de Software	L800.00	L1,600.00	
<b>Documentación y Capacitación Interna</b>	2	Ingeniero de Software	L800.00	L1,600.00	
<b>Total, Lempiras</b>	44		L7,941.49	L43,676.92	
<b>Total, dólares</b>	44		\$313.89	\$1,726.36	

## 6.7 CONCORANCIA DE LOS SEGMENTOS DE LA TESIS CON LA PROPUESTA

**Tabla 15 Matriz de Concordancia**

Capítulo I			Capitulo II	Capítulo III			Capítulo V	Capítulo VI	
Título Investigación	Objetivo General	Objetivos Específicos	Teorías /Metodologías de sustento	Variables	Poblaciones	Técnicas	Conclusiones	Nombre de la propuesta	Objetivos propuesta
Propuesta de Implementación de Técnicas de Clusterización para la Recomendación Personalizada de Productos para una Plataforma de E-Commerce en Honduras	Validar la viabilidad y efectividad de las técnicas de clusterización como herramienta para la segmentación de clientes en la plataforma LynxPay, con el propósito de generar recomendaciones de productos personalizadas que optimicen la experiencia del usuario, aumenten las tasas de conversión y mejoren la competitividad de las tiendas en línea en Honduras.	Recopilar y analizar de forma exhaustiva los datos disponibles en la plataforma LynxPay, tales como transacciones realizadas y productos añadidos al carrito, con el objetivo de identificar las principales variables que influyen en el comportamiento de compra de los usuarios.	Teoría de la Segmentación	Segmentación Transaccional			A pesar de las limitaciones en los datos disponibles, este estudio demostró el valor de la segmentación de clientes y la personalización de recomendaciones en el contexto de LynxPay	Plan de implementación de sistema de recomendación híbrido basado en modelos de clusterización para plataforma LynxPay	Implementar en la plataforma LynxPay un sistema de recomendación híbrido que retorne un listado de productos a partir de datos transaccionales, demográficos, geográficos y de comportamiento de los clientes de cada inquilino en LynxPay

**Continuación de la Tabla 13. Matriz de Concordancia**

Capítulo I			Capítulo II	Capítulo III			Capítulo V	Capítulo VI	
Título Investigación	Objetivo General	Objetivos Específicos	Teorías /Metodologías de sustento	Variables	Poblaciones	Técnicas	Conclusiones	Nombre de la propuesta	Objetivos propuesta
		Aplicar diferentes algoritmos de clusterización, como K-MEANS y DBSCAN, con el propósito de segmentar la base de clientes en grupos homogéneos basados en variables clave de comportamiento, como la frecuencia de compra, el tipo de productos adquiridos y las preferencias de navegación.	Teoría de la Segmentación	Segmentación Transaccional			El sistema de recomendación demostró una mejora significativa en la precisión de las recomendaciones. Al utilizar métricas como Precision@10, se verificó que los algoritmos empleados, especialmente K-Means, pudo identificar patrones de comportamiento en los usuarios y segmentarlos adecuadamente. Esto permitió que las recomendaciones ofrecidas fueran más relevantes y alineadas con los intereses y necesidades individuales de los clientes.		Visualizar los resultados de la investigación y sistema de recomendación híbrido, como la precisión de sus recomendaciones.

**Continuación de la Tabla 13. Matriz de Concordancia**

Capítulo I			Capítulo II	Capítulo III			Capítulo V	Capítulo VI	
Título Investigación	Objetivo General	Objetivos Específicos	Teorías /Metodologías de sustento	Variables	Poblaciones	Técnicas	Conclusiones	Nombre de la propuesta	Objetivos propuesta
		Desarrollar y validar modelos de recomendación personalizados para los segmentos identificados, analizando el impacto de estos modelos en las métricas clave de rendimiento, tales como la precisión de las recomendaciones.		Segmentación Geográfica			La precisión alcanzada confirma que el uso de modelos híbridos y técnicas avanzadas de análisis de datos no solo incrementa la efectividad de las sugerencias, sino que también mejora la experiencia del usuario en la plataforma. Estos resultados validan la viabilidad de emplear algoritmos de aprendizaje automático en contextos de comercio electrónico con datos limitados, siempre que se seleccionen adecuadamente las métricas y los modelos.		Establecer si existe infraestructura de IT para ejecutar el sistema de recomendación.

**Continuación de la Tabla 13. Matriz de Concordancia**

Capítulo I			Capitulo II	Capítulo III			Capítulo V	Capítulo VI	
Título Investigación	Objetivo General	Objetivos Específicos	Teorías /Metodologías de sustento	Variables	Poblaciones	Técnicas	Conclusiones	Nombre de la propuesta	Objetivos propuesta
		Desarrollar y validar modelos de recomendación personalizados para los segmentos identificados, analizando el impacto de estos modelos en las métricas clave de rendimiento, tales como la precisión de las recomendaciones.					La viabilidad económica de esta iniciativa se ve respaldada por la infraestructura tecnológica existente de la empresa y el potencial de generar ingresos adicionales. Los resultados obtenidos sugieren que la inversión en esta área no solo es rentable a corto plazo, sino que también posiciona a LynxPay como una plataforma más competitiva y atractiva para sus emprendedores.		Capacitar al recurso humano desarrollando la plataforma, que permita el uso del sistema de recomendación.

**Continuación de la Tabla 13. Matriz de Concordancia**

Capítulo I			Capitulo II	Capítulo III			Capítulo V	Capítulo VI	
Título Investigación	Objetivo General	Objetivos Específicos	Teorías /Metodologías de sustento	Variables	Poblaciones	Técnicas	Conclusiones	Nombre de la propuesta	Objetivos propuesta
		Proponer estrategias de mejora basadas en los resultados obtenidos durante la validación de los modelos, sugiriendo posibles optimizaciones para la implementación futura de los sistemas de recomendación en LynxPay, con miras a incrementar la personalización y mejorar la experiencia del usuario					Se han generado propuestas estratégicas para enriquecer los modelos de recomendación de LynxPay con el objetivo de ofrecer una experiencia de usuario más personalizada y relevante. La incorporación de datos demográficos, geográficos y de navegación más detallados permitirá una segmentación más precisa y una adaptación de las recomendaciones a las preferencias individuales y al contexto de cada usuario.		Integración del sistema de recomendación en la plataforma.

## BIBLIOGRAFÍA

- Aurélien, G. (2023). *Hands-On Machine Learning with Sckit-Learn, Keras and TensorFlow*. O'Reilly.
- Badrul Sarwar, G. K. (2021). Item-Based Collaborative Filtering Recommendation. *GroupLens Research Group/Army HPC Research Center*.
- Barilliance. (October de 2024). *Casos de Estudio*. Obtenido de Barilliance: <https://www.barilliance.com/case-studies/>
- Beregovskaya Irina, K. M. (2021). Review of Clustering-Based Recommender Systems. 22.
- Berrios Lagos, D. M., Osorio Molina, E. D., & Medina Casco, M. A. (2022). Diagnóstico exploratorio del comercio electrónico en Tegucigalpa durante la pandemia. *Tesis Doctoral. Universidad Tecnológica Centroamericana UNITEC*.
- Chaffey, D. (2016). En D. Chaffey, *Digital Marketing: Strategy, Implementation and practice*. Pearson Education.
- Dbeaver. (22 de Noviembre de 2024). *Dbeaver*. Obtenido de Documentación: <https://dbeaver.com/docs/dbeaver/#about-dbeaver>
- Digital Commerce 360. (01 de July de 2024). *Latin America Database*. Obtenido de Latin America Database: <https://www.digitalcommerce360.com/product/latin-america-database/>
- Duy-Nghia Nguyen, V.-H. N.-S. (2024). A personalized product recommendation model in e-commerce based on retrieval strategy. *Journal of Open Innovation: Technology, Market, and Complexity*.
- Espinoza Freire, E. E. (2019). LAS VARIABLES Y SU OPERACIONALIZACIÓN EN LA INVESTIGACIÓN EDUCATIVA. SEGUNDA PARTE. *REVISTA CONRADO | Revista pedagógica de la Universidad de Cienfuegos*, 171-180.
- Flores Bautista, P., Sánchez Ayala, J., Jimenez DeLucio, J., & Rojo Cisneros, S. (2023). Comprendiendo la fidelización de clientes: elementos clave, estadísticas y clasificaciones. *XIKUA Boletín Científico de la Escuela Superior de Tlahuelilpan*.
- Golding, T. (2024). *Building Multi-Tenant SaaS Architectures*. California: O'Reilly.
- Haefner, N., Wincent, J., Parida, V., & Gassmann, O. (2021). Artificial intelligence and innovation management: A review, framework, and research agenda. *Technological Forecasting & Social Change*.
- Han, J., Pei, J., & Kamber, M. (2012). *Data Mining: Concepts and Techniques* (Third Edition ed.). Morgan Kaufmann. doi:<https://doi.org/10.1016/C2009-0-61819-5>
- Hennig, C. M. (2015). CLUSTER ANALYSIS: ADVANCED CONCEPTS. En C. M. Hennig, *Handbook of Cluster Analysis* (pág. 258). CRC Press.
- Hernández Sampieri, R., Collado, C. F., & Baptista Lucio, M. (2014). *Metodología de la investigación* (6ª ed.). (6ª edición. ed.). México D.F., México: McGraw-Hill.
- Hunter, D. J. (22 de Noviembre de 2024). *Matplotlib*. Obtenido de Historia: <https://matplotlib.org/stable/project/history.html>
- Koren, Y. B. (2009). Matrix factorization techniques for recommender systems. *Computer*.
- Laudon, K., & Traver, C. (2020). E-commerce 2020-2021. En K. C. Laudon, & C. G. Traver, *E-commerce 2020-2021* (pág. 45). Pearson.
- Linden G, S. B. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, 76-80.
- LynxLabs. (22 de Noviembre de 2024). *Página de inicio*. Obtenido de LinkedIn: <https://www.linkedin.com/company/lynx-labs/about/>
- Ortiz, P. A. (s.f.). *Método por Tres Valores Mejorado un método estocástico para la estimación temprana de la duración de los proyectos*.
- Orús, A. (23 de Abril de 2024). *Statista*. Obtenido de <https://es.statista.com/temas/9072/comercio->

- electronico-en-el-mundo/#topicOverview
- Otzen, T., & Manterola, C. (2017). Técnicas de Muestreo sobre una Población a Estudio. *Int. J. Morphol.*, 227-232.
- Paz, J. (2022). Protección de Datos y Privacidad de la Información en Honduras. *Consortium Legal*, <https://consortiumlegal.com/2022/09/06/proteccion-de-datos-y-privacidad-de-la-informacion-en-honduras/>.
- Philip, K., Gary, & Armstrong. (2017). *FUNDAMENTOS DE MARKETING* (17th ed. ed., Vol. 17th ed.). Pearson.
- Python Org. (22 de Noviembre 1 de 2024). *Python*. Obtenido de Preguntas generales: <https://docs.python.org/3/faq/general.html>
- Ricci, F. R. (2011). Introduction to Recommender Systems Handbook. En Ricci, F. R. L., & B. Shapira, *Introduction to Recommender Systems Handbook* (pág. 289). Springer.
- Ricci, F., Rokach, L., & Shapira, B. (2016). Recommender Systems Handbook. *Springer Science+Business Media*.
- Roland T. Rust, P. C. (2005). Optimizing the Marketing Interventions Mix in Intermediate-Term CRM. *Marketing Science*.
- Salah, M., Hamada, M., Taj Eddin, I., & Kasem, E. (2023). Customer profiling, segmentation, and sales prediction using AI in direct marketing. *Journal of Retailing and Consumer Services*.
- Schafer, J., Konstan, J. A., & Riedl, J. (2001). E-Commerce Recommendation Applications. Data Mining and Knowledge Discovery. *GroupLens Research Project*.
- Scikit Learn. (2 de Noviembre de 2024). *Scikit Learn*. Obtenido de Scikit Learn Home Page: <https://scikit-learn.org/stable/>
- Scikit Learn. (22 de Noviembre de 2024). *Scikit Learn*. Obtenido de Preguntas frecuentes: <https://scikit-learn.org/stable/faq.html>
- Serrano, S. (8 de June de 2023). *Personalized Product Recommendations Tactics for Profits*. Obtenido de Barilliance.
- Smaili, M. Y., & Hachimi, H. (2023). New RFM-D classification model for improving customer analysis and response prediction. *Ain Shams Engineering Journal*.
- Snyder, K., & Aditham, K. (28 de Marzo de 2024). *forbes*. Obtenido de [www.forbes.com](http://www.forbes.com): <https://www.forbes.com/advisor/business/ecommerce-statistics/>
- Tatham, S. (22 de Noviembre de 2024). *Preguntas frecuentes*. Obtenido de PuTTY: <https://www.chiark.greenend.org.uk/~sgtatham/putty/faq.html#faq-meaning>
- Valencia Arias, A., Uribe Bedoya, H., Gonzalez Ruiz, J. D., Sanchez Santos, G., Chapoñan Ramírez, E., & Martínez, E. (2024). Artificial intelligence and recommender systems in e-commerce. Trends and research agenda. *Intelligent Systems with Applications*.
- Visual Studio Code. (22 de Noviembre de 2024). *Visual Studio Code*. Obtenido de Preguntas frecuentes: <https://code.visualstudio.com/docs/supporting/faq>
- Wedel, M., & Kamakura, W. A. (2000). *Market Segmentation Conceptual and Methodological Foundations*. Springer Science+Business Media.
- Werner Reinartz, V. K. (2000). On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing. *Journal of Marketing*, 17–35.
- Yonghong Tian, B. Z. (2019). College Library Personalized Recommendation System Based on Hybrid Recommendation Algorithm. *ScienceDirect*.
- Zelaya, O. (2020). Honduras: Comercio electrónico y firma electrónica. *Central Law*.
- Zengyuan Wu, L. J. (2022). Research on Segmenting E-Commerce Customer through. *Wiley*, 10.
- Zhou, Z. (2004). E-commerce & Information Technology in Hospitality & Tourism. En Z. Zhou, *E-commerce & Information Technology in Hospitality & Tourism* (pág. 118). Thomson/Delmar Learning.

# ANEXOS

## Anexo 1



### CARTA DE AUTORIZACIÓN DE LA EMPRESA O INSTITUCIÓN

Nombre y apellido del Director o Gerente: Jorge Alejandro Caballero Murillo  
Puesto Laboral: CEO  
Empresa o Institución: LYNXLABS SA  
Dirección principal de la Empresa o Institución: 21102-3 Centro Morazán, Tegucigalpa, Francisco Morazán  
Ciudad: Tegucigalpa Departamento: Francisco Morazán Día: 18 Mes: Noviembre Año: 2024

Estimado Señor(a): Jorge Alejandro Caballero Murillo

Reciba un cordial y atento saludo. Por medio de la presente deseamos solicitar su apoyo, dado que somos alumnos de UNITEC y nos encontramos desarrollando el Trabajo de Tesis previo a obtener nuestro título de maestría en Análítica de Negocios.

Hemos seleccionado como tema PROPUESTA DE IMPLEMENTACIÓN DE TÉCNICAS DE CLUSTERIZACIÓN PARA LA RECOMENDACIÓN PERSONALIZADA DE PRODUCTOS PARA UNA PLATAFORMA DE E-COMMERCE EN HONDURAS, por lo que estaríamos muy agradecidos de contar con el apoyo de la empresa que usted representa para poder desarrollar nuestra investigación. En particular, dicha solicitud se circunscribe a peticionar que se nos autorice a realizar: Extracción de datos (mismos que no saldrán de la empresa) con finalidad de análisis.

A la espera de su aprobación, me suscribo de Usted.

Atentamente,

Keneth Mauricio Pérez Brizuela

Firma, nombre y apellidos

No. de cuenta: 12313104

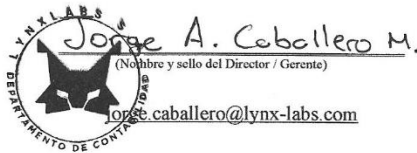
Andy Javier Reyes Erazo

Firma, nombre y apellidos

No. de cuenta: 12313102

Por este medio, LynxLabs SA (empresa / institución),

Autoriza la realización dentro de sus instalaciones o del uso de información de la empresa en el proyecto de investigación de Tesis de Postgrado antes mencionado.

  
(Nombre y sello del Director / Gerente)  
jorge.caballero@lynx-labs.com

Correo electrónico de Director/Gerente

Vo.Bo

## Anexo 2

**GET** /recommendations/{cliente\_id} Obtiene recomendaciones para un cliente específico.

Recibe el ID de un cliente y retorna una lista de recomendaciones personalizadas basadas en el clúster al que pertenece.

**Parameters** Try it out

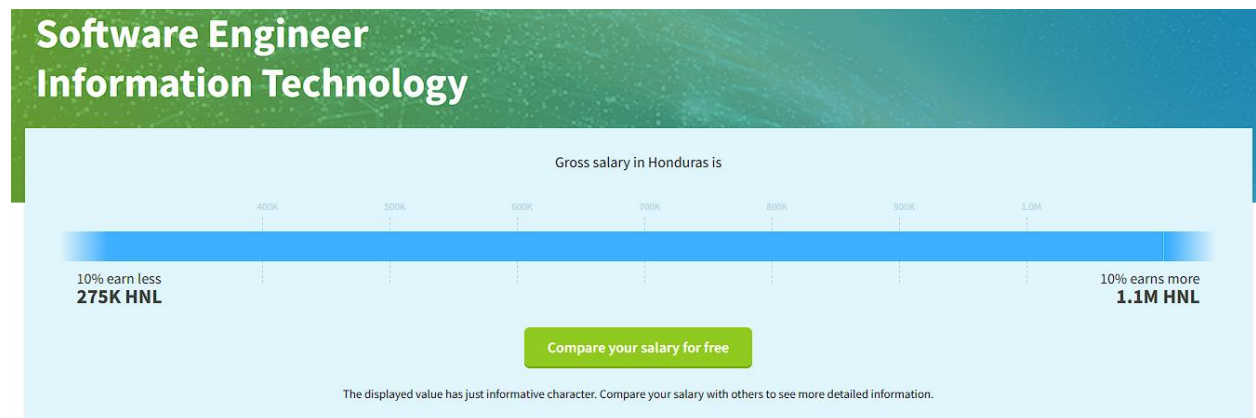
Name	Description
<b>cliente_id</b> * required	ID único del cliente.
integer(\$int64) (path)	<input type="text" value="cliente_id"/>

**Responses**

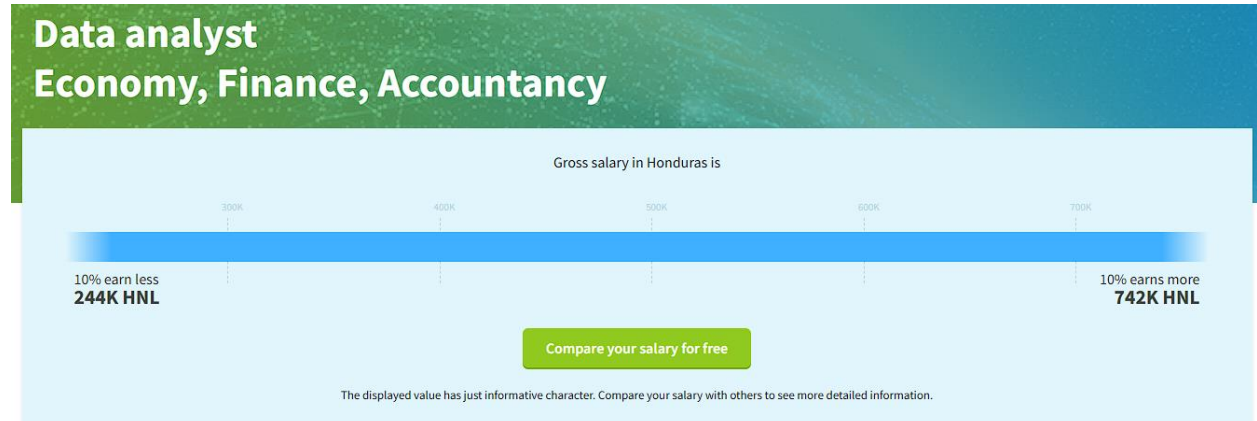
Code	Description	Links
200	Lista de recomendaciones generada exitosamente. <small>Media type</small> <input type="text" value="application/json"/> <small>Controls Accept header</small> <small>Example Value</small>   <small>Schema</small>	No links

```
{
  "cliente_id": 0,
  "cluster_id": 0,
  "recommendations": [
    {
      "item_id": 0,
      "item_name": "string",
      "relevance_score": 0
    }
  ]
}
```

## Anexo 3



Anexo 4



Anexo 5

## Sueldos para Data Scientist en Honduras ⓘ

Confianza baja · 1 sueldos enviados · Actualizado el 10 de jun de 2023

Experiencia

Todos los años de experiencia ▼

Sueldo base

**HNL 24 K - HNL 25 K/mes**

HNL 24 K/mes Sueldo base promedio