ANÁLISIS DE DATOS MEDIANTE HERRAMIENTAS TECNOLÓGICAS PARA CAMPAÑAS DIGITALES DE UNA COMPAÑÍA INMOBILIARIA

Mireles Jiménez Gilberto [1], Escalante Euán Jesús Francisco[1], Chan Pavón Miriam Victoria[1], Santos Flores Jorge[1] [1] Universidad Autónoma de Yucatán , www.fiq@correo.uady.mx*

Área de participación: **Ingeniería Administrativa**

Resumen

En esta investigación se implementó una selección de herramientas estadísticas avanzadas con el fin de identificar los perfiles de clientes potenciales de una compañía inmobiliaria. Se empleó el enfoque CRISP-DM junto con modelos de regresión y técnicas de agrupamiento para modelar y predecir el comportamiento de los clientes, considerando un conjunto de variables de interés. Los resultados obtenidos revelan evidencia significativa respecto a las contribuciones iniciales en el proceso de compra-venta. Esto permite establecer estrategias para personalizar campañas publicitarias y mejorar la toma de decisiones.

Palabras clave: *Análisis de Datos, Perfiles de Clientes, CRISP-DM, Segmentación de Mercado.*

Abstract

This research implemented a selection of advanced statistical tools to identify potential customer profiles for a real estate company. The CRISP-DM approach was employed along with regression models and clustering techniques to model and predict customer behavior, based on a set of relevant variables. The results obtained reveal significant evidence regarding initial contributions in the buying-selling process. This allows for the establishment of strategies to personalize advertising campaigns and enhance decision-making.

Key words: *Data Analysis, Customer Profiling, CRISP-DM, Market Segmentation.*

Introducción

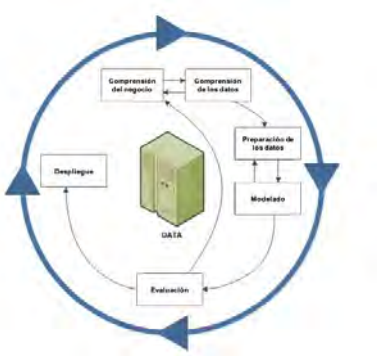
En la era actual, caracterizada por la digitalización y la acumulación masiva de datos, las empresas inmobiliarias enfrentan el desafío de adaptarse a las dinámicas cambiantes del mercado y a las expectativas de los consumidores. El análisis de datos se ha convertido en una herramienta indispensable, permitiendo una mejor comprensión de las necesidades y comportamientos de los clientes y optimizando las estrategias de marketing digital. Estudios recientes demuestran que la adopción de Big Data Analytics impacta de manera positiva en el rendimiento empresarial y en el conocimiento del comportamiento del consumidor, facilitando una toma de decisiones estratégica y bien informada, especialmente en pequeñas y medianas empresas [1]. Este estudio se centra en la aplicación de técnicas avanzadas de análisis de datos para identificar perfiles de clientes potenciales en una compañía inmobiliaria, empleando el modelo CRISP-DM (Cross-Industry Standard Process for Data Mining). Este modelo, según la documentación de IBM, ofrece un ciclo estructurado de minería de datos, favoreciendo la extracción de conocimiento a partir de grandes volúmenes de datos. A través de esta metodología, se busca mejorar la efectividad de las campañas de marketing y contribuir al desarrollo de estrategias comerciales más eficientes y personalizadas, promoviendo la satisfacción y fidelidad del cliente.

Metodología

El análisis sigue la metodología CRISP-DM, ampliamente reconocida como estándar en proyectos de minería de datos, la cual se compone de seis fases críticas: entendimiento del negocio, entendimiento de los datos,

preparación de los datos, modelado, evaluación y despliegue. Este enfoque, según la documentación de IBM, proporciona un marco estructurado y eficaz para la minería de datos [2]. Además, investigaciones como la de Montalvo García han demostrado que CRISP-DM puede adaptarse a diferentes sectores, incluyendo pequeñas y medianas empresas, a través de variantes como CRISP-DM/SMEs, que ofrecen simplicidad, flexibilidad y bajo costo de implementación, lo que valida aún más su aplicabilidad en el sector inmobiliario [3]. Las herramientas tecnológicas avanzadas, como el software R, se emplearon para el análisis y procesamiento de datos estadísticos. Las técnicas de modelado predictivo y clustering permitieron la identificación de patrones y segmentos de clientes, mientras que las visualizaciones gráficas facilitaron la interpretación de los resultados, ayudando en la comprensión de tendencias y relaciones complejas entre las variables.

Minería de datos (Modelo CRISP-DM)



(IBM, 2021)

Preparación

Los datos fueron limpiados y preprocesados para garantizar su calidad y precisión, un paso crucial según Espinosa-Zúñiga (2020) en la aplicación de la metodología CRISP-DM para la segmentación geográfica de datos. Este proceso incluyó la eliminación de registros duplicados o incompletos y la normalización de las variables para facilitar la comparación y análisis, además de la transformación de datos categóricos en formatos numéricos adecuados para los análisis posteriores [4].

Desarrollo

La recolección de datos se realizó a través de fuentes internas de la empresa, incluyendo bases de datos de ventas y marketing, y registros de interacciones con clientes. Los datos recolectados abarcaron variables como enganches, mensualidades, tiempo de inversión, conocimiento de Yucatán y propósitos de compra de propiedad. La preparación de los datos implicó la limpieza de datos incompletos o incorrectos, la transformación de variables para su análisis y la normalización de datos para facilitar las comparaciones y el análisis. Se aseguró la calidad y la relevancia de los datos mediante la verificación de su precisión y la actualización periódica de los mismos.

El análisis de correlación se utilizó para identificar y evaluar las relaciones entre las variables financieras significativas, como el enganche, la mensualidad y el tiempo de inversión, tal como sugiere Cheng y Chen (2009) en su estudio sobre la segmentación del valor del cliente. Mediante el uso de gráficos de correlación y cálculos de coeficientes de correlación de Pearson, se determinó la fuerza y dirección de estas relaciones, proporcionando una base sólida para el modelado predictivo posterior. Los resultados del análisis mostraron correlaciones positivas entre el enganche y la mensualidad, sugiriendo que mayores enganches están asociados con mayores mensualidades, lo cual es consistente con las expectativas del mercado inmobiliario [5].

Justificación de la Selección de Variables

Para garantizar la relevancia y la significancia de las variables incluidas en el modelo de regresión utilizado en este análisis, se implementaron métodos estadísticos rigurosos. El principal método utilizado fue el Análisis de Varianza (ANOVA), que permitió evaluar la contribución de cada variable al modelo en relación con la variable dependiente 'Mensualidad'.

El resultado del análisis ANOVA reveló que la variable 'Enganche' es estadísticamente significativa con un valor F de 69.3953 y un p-valor menor a 0.001, indicando una fuerte influencia en la predicción de la mensualidad (ver Tabla 1). Las variables 'Tiempo_inversion' y 'Tiempo_Construir_Vivir', aunque no alcanzaron significancia estadística en este modelo, fueron consideradas en el análisis debido a su relevancia conceptual y la presencia de una correlación positiva moderada entre ellas, como se observa en la matriz de correlación, lo que sugiere una interacción posible que podría ser crucial en un contexto más amplio del comportamiento del consumidor.

```
> print(anova_result)
Analysis of Variance Table

Response: Mensualidad
                      Df     Sum Sq    Mean Sq F value    Pr(>F)
Enganche               1 1.7734e+10 1.7734e+10 69.3953 4.756e-16 ***
Tiempo_inversion       1 8.7294e+07 8.7294e+07  0.3416    0.5591
Tiempo_Construir_Vivir 1 1.5244e+08 1.5244e+08  0.5965    0.4402
Residuals            652 1.6662e+11 2.5555e+08
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Tabla 1. Tabla de Análisis de Varianza

La matriz de correlación proporciona una visualización clara de las relaciones entre las variables analizadas. Este análisis destaca la asociación significativa entre 'Enganche' y 'Mensualidad', junto con otras interacciones potenciales que podrían influir en decisiones de inversión y preferencias de los clientes.

La combinación de estos enfoques cuantitativos robustece el proceso de selección de variables, asegurando que las decisiones basadas en estos datos reflejan con precisión las dinámicas del mercado y son apoyadas por una metodología estadísticamente validada.

1.1 MODELO DE REGRESIÓN LINEAL MÚLTIPLE

Se desarrolló un modelo de regresión lineal múltiple con el objetivo de predecir la mensualidad en función de variables independientes significativas identificadas previamente en el análisis de correlación. El modelo incorpora términos de interacción para evaluar efectos combinados, como los del conocimiento sobre Yucatán y el enganche inicial. La evaluación del modelo incluyó la interpretación de coeficientes, análisis de significancia estadística y diagnóstico de residuos, confirmando la adecuación del modelo. Este modelo reveló cómo distintos factores influyen en la mensualidad, facilitando decisiones informadas en estrategias de precios y marketing.

Detalles del Modelo de Regresión Lineal Múltiple y sus Coeficientes

El modelo de regresión lineal múltiple que se empleó en este análisis busca predecir la variable 'Mensualidad' en función de diversas variables independientes. La ecuación del modelo se representa de la siguiente manera utilizando el editor de ecuaciones:

$$\text{Mensualidad} = \beta_0 + \beta_1(\text{Enganche}) + \beta_2(\text{Tiempo de inversión}) + \beta_3(\text{Tiempo para Construir o Vivir}) + \epsilon$$

donde β_0 es el intercepto, β_1,β_2, y β_3 son los coeficientes de las variables independientes, y ϵ es el término de error.

Este modelo incorpora:

- **Intercepto (β_0)**: Representa la mensualidad base cuando todas las variables independientes son cero.
- **Enganche (β_1)**: Este coeficiente muestra un impacto directamente proporcional sobre la mensualidad, sugiriendo que un mayor enganche se asocia con una mayor mensualidad.
- **Tiempo de Inversión (β_2) y Tiempo para Construir o Vivir (β_3):** Aunque sus coeficientes son menores, proporcionan información valiosa sobre cómo los tiempos relacionados con la inversión afectan las decisiones financieras.

Para asegurar la validez del modelo, se realizó un Análisis de Varianza (ANOVA) que confirmó que el Enganche es una variable predictor significativa, mientras que las otras variables requieren una revisión más detallada o la recolección de más datos para establecer efectos claros.

1.2 ANÁLISIS DE CLUSTERING

Utilizando técnicas de clustering, específicamente el método K-means, se segmentaron los clientes en grupos distintos basados en sus características financieras y preferencias de inversión. La elección del método K-means se fundamenta en su eficacia reconocida para identificar agrupaciones naturales dentro de grandes conjuntos de datos, según se discute en la obra de Kaufman y Rousseeuw [6]. Se determinó el número óptimo de clusters a través del método del codo y se analizaron los centroides de cada cluster para entender las características comunes dentro de cada grupo. Este análisis ayudó a identificar segmentos de clientes con necesidades y comportamientos similares, lo que permitirá a la empresa dirigir sus esfuerzos de marketing de manera más efectiva y personalizada hacia cada segmento.

Conexión entre el Modelo de Regresión y el Análisis de Clustering

El análisis de clustering se utilizó para identificar grupos homogéneos de clientes basándose en características similares, incluyendo las variables utilizadas en el modelo de regresión. Esta segmentación permitió entender mejor cómo las características de los clientes influyen en la mensualidad y ofreció una validación práctica de los perfiles de clientes identificados a través del modelo de regresión.

- Cada cluster identificado posee características únicas que se alinean con patrones específicos en las predicciones del modelo de regresión. Por ejemplo, clusters con altos valores de enganche generalmente mostraron mayores mensualidades, corroborando la relación establecida por el modelo de regresión.
- La integración de estos métodos proporciona una comprensión más rica y matizada de cómo diferentes segmentos de clientes interactúan con las variables financieras, lo que permite una planificación y targeting más efectivo en campañas y productos.

Este enfoque metodológico dual, combinando regresión y clustering, no solo refuerza la robustez de nuestros hallazgos, sino que también mejora la capacidad de la empresa para implementar estrategias diferenciadas basadas en datos concretos y análisis estadístico riguroso.

Resultados y discusión

El análisis de los datos reveló hallazgos clave sobre el comportamiento de los clientes y las dinámicas del mercado inmobiliario. En el análisis de correlación, se confirmó una correlación positiva entre el enganche y la mensualidad, sugiriendo que los clientes dispuestos a realizar un enganche elevado también tienden a aceptar mensualidades más altas, lo cual puede estar relacionado con un interés en propiedades de mayor valor. Asimismo, el modelo

de regresión lineal múltiple proporcionó una herramienta predictiva robusta, identificando el enganche y el conocimiento sobre Yucatán como factores influyentes en las decisiones de compra. La segmentación de clientes, lograda mediante el análisis de clustering, permitió identificar cuatro perfiles diferenciados de clientes, cada uno con características y necesidades financieras particulares, lo que facilita la creación de ofertas personalizadas y comunicaciones dirigidas.

Diversas visualizaciones gráficas facilitaron la interpretación de los resultados. Entre estas, los mapas de calor de correlación resaltaron las interrelaciones significativas entre las variables, permitiendo una evaluación detallada de las dinámicas entre el enganche, la mensualidad y otras variables demográficas y financieras. Los gráficos de dispersión con líneas de regresión y los diagramas de caja también contribuyeron a clarificar la relación entre variables como el enganche y la mensualidad, proporcionando una representación visual clara de la distribución y las tendencias.

Mapas de calor de correlación: que destacan las fuertes interrelaciones entre las variables consideradas.

Los mapas de calor de correlación, como el que se muestra en el Gráfico 1, son herramientas visuales cruciales para identificar y evaluar las interrelaciones entre diversas variables financieras y demográficas en nuestro estudio. En este gráfico específico, se destaca la matriz de correlación que exhibe las interacciones entre variables como el 'Enganche', la 'Mensualidad', 'Tiempo_inversion', y 'X.Conoces.Yucatán.', entre otras.

El valor de correlación de 0.33 entre 'Tiempo_inversion' y 'Tiempo_Construir_Vivir' sugiere una relación moderadamente positiva, indicando que a medida que el tiempo de inversión aumenta, también lo hace el tiempo estimado para construir o vivir, lo cual puede reflejar la planificación a largo plazo de los inversores en propiedades. Por otro lado, una correlación positiva de 0.17 entre 'Enganche' y 'Mensualidad' demuestra que los clientes que optan por un enganche más alto tienden a aceptar una mayor mensualidad, lo que podría indicar una disposición a invertir en propiedades de mayor valor.

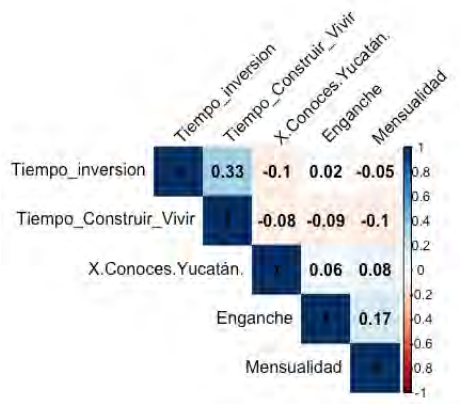


Gráfico 1. Matriz de Correlación que muestra la relación entre 'Enganche' y 'Mensualidad'.

Correlaciones Observadas:

- *Tiempo_inversion y Tiempo_Construir_Vivir*
  Correlación positiva (0.33)
- *Enganche y Mensualidad*
  Correlación positiva (0.17)

El análisis reveló que variables como el enganche y el conocimiento sobre Yucatán influyen significativamente en las decisiones de compra. Esto concuerda con hallazgos previos, como los de Chacaltana Espinoza, quien identifica que ciertos atributos, en particular la sostenibilidad y eficiencia en el uso de recursos, son prioritarios para los clientes y determinan su interés en productos inmobiliarios. Esta relación directa entre atributos específicos y la decisión de compra destaca la importancia de comprender los factores que los clientes valoran más en el mercado inmobiliario [7]. De manera similar, este estudio identifica variables clave que pueden guiar la segmentación y personalización de campañas en el sector inmobiliario.

Boxplot de Variable Mensualidad por variable "¿Conoces Yucatán?"

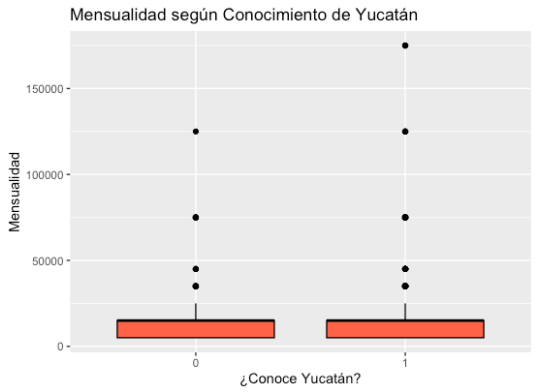


Gráfico 2. Boxplot de Variable Mensualidad por variable "Conoces Yucatán"

El presente gráfico nos ayuda a entender si el conocimiento regional puede ser un factor para considerar en campañas de marketing o estrategias de venta.

Gráficos de dispersión con líneas de regresión: que muestran la relación lineal entre variables como el enganche y la mensualidad.

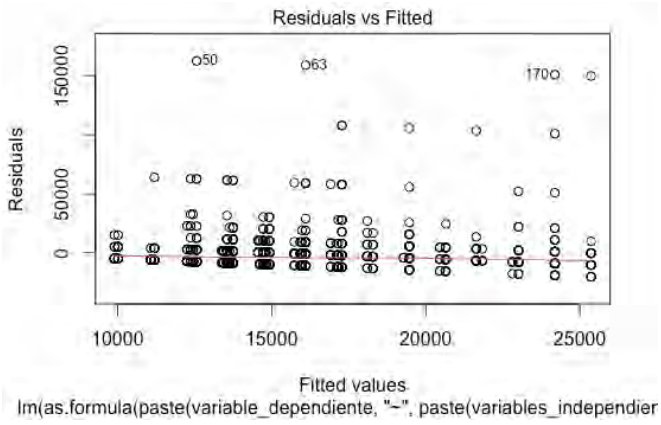


Gráfico 3. Gráfico de Coeficientes de Residuales

En este gráfico, se observa que la mayoría de los residuales están agrupados alrededor de cero y no muestran patrones claros de curvatura o agrupamiento, lo cual es un buen indicativo de que el modelo generalmente se ajusta bien a los datos.

El gráfico de residuos muestra una leve tendencia de mayor dispersión en valores ajustados altos (mayores a 15,000), lo cual podría sugerir que el modelo no ajusta adecuadamente los valores más elevados o que existe heterocedasticidad (varianza no constante de los errores). Este hallazgo resalta la importancia de considerar una transformación de los datos o el uso de un modelo robusto en investigaciones futuras para mejorar la adecuación del modelo.

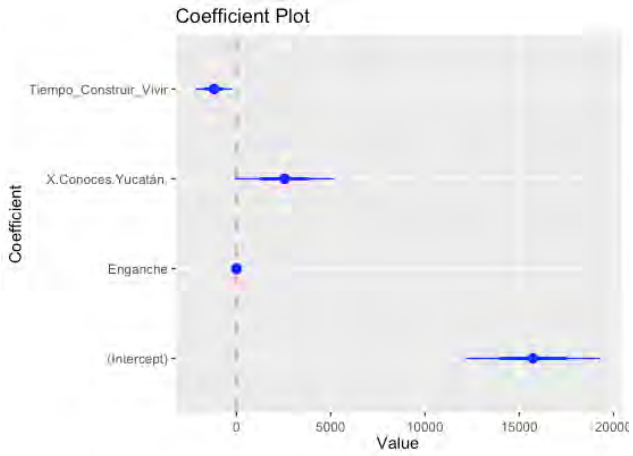


Gráfico 4. Gráfico de Coeficientes del modelo

El "Gráfico de Coeficientes" muestra la magnitud y la incertidumbre asociada a los coeficientes del modelo de regresión lineal utilizado para predecir la mensualidad en función de diferentes variables.

Histogramas y diagramas de caja: que ofrecen una vista detallada de la distribución de las variables y los clusters de clientes.

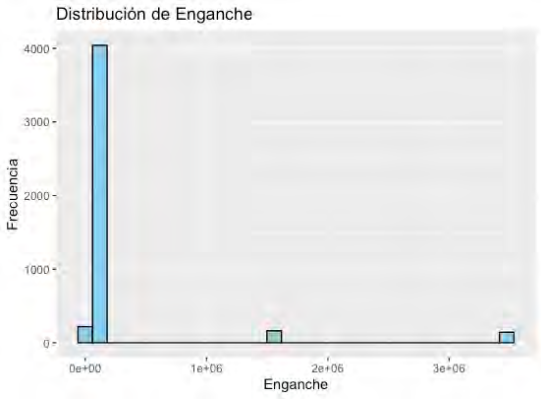


Gráfico 5. Histograma de Distribución de Enganche

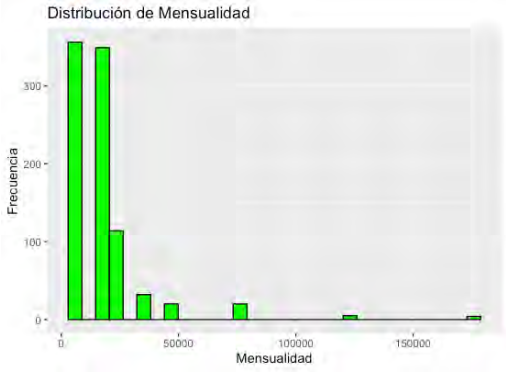


Gráfico 6. Histograma de Distribución de Mensualidad

La segmentación de clientes mediante análisis de clustering permitió identificar cuatro perfiles diferenciados, cada uno con características y necesidades financieras particulares. Estos segmentos facilitan la creación de ofertas personalizadas y comunicaciones dirigidas. La literatura apoya esta metodología; Benslama y Jallouli destacan cómo el clustering, aplicado a datos no estructurados de redes sociales, proporciona una clasificación efectiva que apoya la toma de decisiones en marketing y permite extraer conocimientos útiles de grandes volúmenes de datos [8].

Tabla de descripción de Clusters

Variables	Cluster 1 (Inversores Moderados)	Cluster 2 (Altos Inversores Inmobiliarios)	Cluster 3 (Inversores Premium)	Cluster 4 (Inversores Iniciales)
Miembros	389	161	137	213
Enganche promedio	$75,000	$1,500,000	$3,500,000	$25,000
Mensualidad promedio	$18,136	$17,049	$23,394	$8,000
Tiempo de inversión promedio	4.5 meses	4.7 meses	4.5 meses	4.4 meses
Conocimiento de Yucatán	69% conocen Yucatán	71% conocen Yucatán	74% conocen Yucatán	60% conocen Yucatán
Tiempo estimado para construir o vivir	2.6 años	2.6 años	2.4 años	2.9 años
Características Principales	Este cluster agrupa a clientes con inversiones moderadas y un conocimiento considerable sobre Yucatán, preferentemente para proyectos a mediano plazo.	Este grupo se caracteriza por altos montos de enganche, indicando un segmento de mercado de alto valor con una estabilidad financiera significativa y una rápida rotación de inversión.	Este es el cluster con los enganches y mensualidades más altas, reflejando un segmento de mercado exclusivo interesado en propiedades de lujo y/o inversiones rápidas.	Representa a los clientes con el enganche más bajo y la mensualidad más baja, ideal para aquellos que buscan opciones de inversión accesibles o son nuevos en el mercado inmobiliario.

Tabla 2. Tabla de descripción de clusters

De manera similar, Piedra Paravicino muestra que el clustering permite no solo la creación de perfiles únicos, sino también la personalización de estrategias en el sector inmobiliario, adaptando la comunicación y las ofertas a las necesidades de cada grupo [9]. Este enfoque permite a las empresas inmobiliarias optimizar sus campañas y mejorar la efectividad en la captación de clientes, alineándose con las necesidades particulares de cada segmento identificado en este estudio.

Diagrama de Cajas

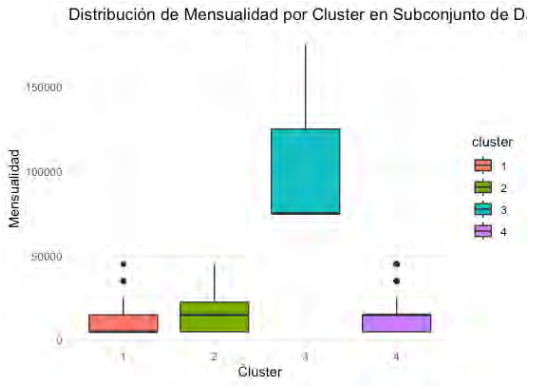


Gráfico 7. Diagrama de cajas de districución de mensualidad por cluster

Según RPubs, esta gráfica muestra cómo la mensualidad varía entre los diferentes clusters identificados en el análisis. Cada caja representa la distribución de la mensualidad dentro de un cluster, proporcionando una visualización clara de la mediana, los cuartiles y los valores atípicos [10].

Estas visualizaciones no solo complementan la discusión textual de los resultados sino que también proporcionan una forma clara y concisa de demostrar las tendencias, patrones y anomalías en los datos.

Trabajo a futuro

Es esencial continuar con la investigación para integrar modelos predictivos más complejos y explorar el impacto de nuevas variables que podrían influir en la dinámica del mercado, como las tendencias económicas globales o los cambios demográficos. El futuro del análisis de datos en el sector inmobiliario promete ser aún más integral, combinando métodos tradicionales con innovaciones tecnológicas como el aprendizaje automático y la inteligencia artificial para refinar aún más las estrategias de negocio y marketing.

Conclusiones

El análisis realizado identifica relaciones significativas entre variables financieras y preferencias de clientes en el sector inmobiliario. Tanto el enganche como la mensualidad se confirmaron como predictores relevantes de interés en propiedades, mientras que el análisis de clustering logró segmentar a los clientes en grupos con diferentes capacidades y preferencias financieras. Estos hallazgos validan el uso de técnicas de minería de datos para una comprensión más profunda del comportamiento del cliente en el mercado inmobiliario.

Con base en los resultados, se recomienda a la compañía inmobiliaria:

- Implementar estrategias de marketing diferenciadas para cada cluster identificado, maximizando la relevancia de las comunicaciones y ofertas.
- Ajustar las condiciones de financiamiento y los precios de las propiedades en función de las capacidades financieras y preferencias de cada segmento, mejorando la tasa de conversión.
- Utilizar los resultados del modelo de regresión para anticipar la demanda de propiedades según las condiciones financieras del mercado, ajustando las estrategias operativas y comerciales en consecuencia.

Referencias

[1] L. M. H. Pérez y J. M. Pérez, “Adopción de Big Data Analitycs en las PyMEs,” Investigación Administrativa, vol. 53, no. 134, pp. 1-21, 2024.

[2] IBM Documentation, "CRISP-DM Help Overview," IBM, 2021. [En línea]. Disponible en: https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview.

[3] J. F. Montalvo García, “CRISP-DM/SMES: una metodología de proyectos de analítica de datos para las PYME,” Revista de Ciencia y Tecnología, 2021.

[4] Espinosa-Zúñiga, J. J. (2020). Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública. Ingeniería, investigación y tecnología, 21.

[5] Cheng, C.-H, y Chen, Y.-S, "Classifying the segmentation of customer value via RFM model and RS theory," Expert Systems with Applications, vol. 36, no. 3, pp. 4176-4184, 2009, doi: 10.1016/j.eswa.2008.04.003.

[6] Kaufman, L., y Rousseeuw, P. J., "Finding Groups in Data: An Introduction to Cluster Analysis," John Wiley & Sons, 2009.

[7] I. G. Chacaltana Espinoza, “Identificación de atributos del producto inmobiliario que influyen en la decisión de compra de vivienda en Lima Metropolitana: aplicación de la herramienta a un caso,” Pontificia Universidad Católica del Perú, 2017. [En línea]. Disponible en: https://hdl.handle.net/20.500.12404/9279

[8] T. Benslama y R. Jallouli, “Clustering of Social Media Data and Marketing Decisions,” en Digital Economy. Emerging Technologies and Business Innovation, ICDEc 2020, Lecture Notes in Business Information Processing, vol. 395, pp. 53–65, dic. 2020.

[9] C. A. Piedra Paravicino, “Segmentación de clientes potenciales del sector inmobiliario en Lima Metropolitana,” Pontificia Universidad Católica del Perú, 2022. [En línea]. Disponible en: https://hdl.handle.net/20.500.12404/11783

[10] R. Delgado, "Introducción a los Modelos de Agrupamiento en R," RPubs, s.f.-b. [En línea]. Disponible en: https://rpubs.com/rdelgado/399475.