**Logotipo

Descripción generada automáticamente**

**Segmentación de Clientes Colombianos Residentes en el Exterior para la Adquisición de Vivienda en Colombia: Un Enfoque de Machine Learning en Créditos Hipotecarios**

Yuri Patricia Bernal Ospina

Juan David Jimenez Guerra

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor  
Nombres completos, Título académico más alto

Universidad de Antioquia  
Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2024

| **Cita** | (Bernal Ospina & Jimenez Guerra, 2024) |
| --- | --- |
| **Referencia**  **Estilo APA 7 (2020)** | Bernal Ospina, Y. P., Jimenez Guerra, J. D. (2024). Segmentación de Clientes Colombianos Residentes en el Exterior para la Adquisición de Vivienda en Colombia: Un Enfoque de Machine Learning en Créditos Hipotecarios. Universidad de Antioquia, Medellín, Colombia. |

**** 

Especialización en Analítica y Ciencia de Datos, CohorteVII.

Centro de Investigación Ambientales y de Ingeniería (CIA).

|  | Diagrama  Descripción generada automáticamente con confianza media |
| --- | --- |

Centro de Documentación Ingeniería (CENDOI)

**Repositorio Institucional:** http://bibliotecadigital.udea.edu.co

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Diego José Luis Botia Valderrama

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

**Encabezado**

**Dedicatoria**

Texto de dedicatoria centrado.

**Agradecimientos**

Texto de agradecimientos centrado.

**Tabla de contenido**

[Resumen 6](#_3rdcrjn)

[Abstract 7](#_26in1rg)

[1. Descripción del problema 8](#_35nkun2)

[1.1. Problema de negocio 8](#_1ksv4uv)

[1.2. Aproximación desde la analítica de datos 8](#_44sinio)

[1.3. Origen de los datos 8](#_z337ya)

[1.4. Métricas de desempeño 9](#_3j2qqm3)

[2. Objetivos 10](#_4i7ojhp)

[2.1. Objetivo general](#_2xcytpi) [10](#_zb42l4yyd77h)

[2.2. Objetivos específicos 10](#_3whwml4)

[3. Datos 10](#_2bn6wsx)

[3.1. Datos originales 10](#_qsh70q)

[3.2. Datasets 13](#_3as4poj)

[3.3. Analítica descriptiva 15](#_1pxezwc)

[Referencias 21](#_1jlao46)

**Siglas, acrónimos y abreviaturas**

**APA** American Psychological Association

**Cms.** Centímetros

**ERIC** Education Resources Information Center

**Esp.** Especialista

**MP** Magistrado Ponente

**MSc** Magister Scientiae

**Párr.** Párrafo

**PhD** Philosophiae Doctor

**PBQ-SF** Personality Belief Questionnaire Short Form

**PostDoc** PostDoctor

**UdeA** Universidad de Antioquia

# Resumen

El resumen permite identificar la esencia del escrito, es un abstract. Realiza una descripción general de tu proyecto: qué se persigue, qué datos se tiene, qué estrategia se siguió para las iteraciones, que obstáculos hubo, qué resultados se obtuvieron, etc. La longitud es mínimo 150 y máximo 250 palabras.

Incluye al final de dicha página la dirección de los repositorios GitHub.

*Palabras clave*: palabra 1, palabra 2, palabra 3, palabra 4.

# Abstract

El abstract es el mismo resumen pero en idioma inglés. Conserva la misma extensión o aproximación, es decir, mínimo 150 y máximo 250 palabras.

*Keywords***:** …..

# Descripción del problema

## Problema de negocio

La adquisición de un crédito de vivienda es un proceso de larga duración, con ciclos de compra que suelen extenderse por años y trámites que a menudo resultan complejos. Para destacar en este entorno competitivo y dinámico, las empresas dedicadas a otorgar créditos de este tipo deben adoptar un enfoque innovador y orientado al cliente, dedicándose activamente a comprender sus necesidades.

La empresa Viventa LLC, especializada en otorgar créditos para la adquisición de vivienda a ciudadanos colombianos residentes en el exterior, busca segmentar su base de clientes para comprender mejor sus necesidades, comportamientos y características. Mediante esta segmentación, la empresa pretende diseñar estrategias comerciales y de marketing más efectivas, identificando los canales de comunicación adecuados y creando mensajes personalizados para cada segmento detectado, mejorando así los procesos internos de la compañía. La falta de comprensión de sus clientes, por ejemplo, ha llevado a Viventa LLC a invertir grandes sumas de dinero en campañas publicitarias que no han generado el retorno esperado.

## Aproximación desde la analítica de datos

Para abordar este problema de segmentación, se desarrollará un modelo de clasificación no supervisado mediante un algoritmo de clustering, con el objetivo de agrupar en segmentos específicos y homogéneos la base de clientes de Viventa LLC, basándose en criterios esenciales para el negocio, tales como edad, género, nivel de ingresos, localización geográfica, intereses, entre otros. Al concluir este proceso de segmentación, la empresa podría estar en condiciones de adaptar y personalizar sus servicios y estrategias comerciales y de marketing para satisfacer de forma más precisa las necesidades particulares de cada grupo identificado. Para este caso, se utilizará la técnica de clustering K-means, que se ajusta a las necesidades y características específicas del negocio.

## Origen de los datos

Los datos utilizados para este proyecto provienen de la base de clientes potenciales y efectivos de Viventa LLC, la cual se encuentra administrada y almacenada en el CRM de Zoho Corporation. La información abarca el historial de clientes que han ingresado a la compañía y contiene detalles como edad, país de residencia, score de crédito, ocupación, ciudad de compra, género, entre otros. Se dispone de datos de clientes desde enero de 2015 hasta abril de 2024.

El departamento comercial es responsable de recolectar dicha información. Este proceso comienza con un contacto inicial del cliente potencial, donde se proporciona toda la información y presentación de los servicios ofrecidos por la empresa. Posteriormente, se envía un enlace con una encuesta que el cliente debe completar para continuar con el proceso. Además, el equipo comercial registra información sobre las actividades realizadas con el usuario, como llamadas y correos electrónicos. Finalmente, el equipo de marketing recopila datos relacionados con el costo de conversión o el costo por clic, especialmente si el cliente proviene de una campaña publicitaria.

## Métricas de desempeño

Dado que se trabajará con datos no etiquetados, se utilizarán medidas que evalúen los resultados de la agrupación sin depender de etiquetas previas. Dentro de estas técnicas de validación interna se incluyen la cohesión dentro de cada cluster y la separación entre los distintos clusters. Además, se llevará a cabo un análisis de las agrupaciones utilizando técnicas como el Análisis de Componentes Principales (PCA) para visualizar la distribución de los datos. Para evaluar la calidad y eficacia de los clusters generados, se utilizarán las siguientes métricas: Coeficiente de Silueta, Índice Calinski Harabasz y Método de Elbow (Codo).

Para medir la efectividad de las campañas de marketing y la satisfacción de los clientes, Viventa LLC utiliza varias métricas clave. En primer lugar, el Costo por Adquisición (CPA) que calcula el costo promedio para adquirir un nuevo cliente, mientras que la Tasa de Conversión evalúa el porcentaje de personas a las que se les desembolsa el crédito hipotecario. Además, el Costo por Clic (CPC) analiza el costo promedio de cada clic en los anuncios, y el Índice de Satisfacción del Cliente (CSAT) mide el nivel de satisfacción de los clientes con los servicios ofrecidos. Estas métricas ayudan a la empresa a entender mejor el impacto de las campañas y a optimizar sus estrategias para alcanzar mejores resultados.

Uno de los propósitos de la segmentación de clientes mediante el modelo de clustering es optimizar las métricas utilizadas por el negocio para evaluar sus procesos de marketing y la satisfacción de sus clientes. Por ejemplo, un valor cercano a 1 en el coeficiente de Silueta indicará una mejor separación entre los clusters. Con este tipo de agrupaciones clave, la empresa espera alcanzar un índice de satisfacción del cliente del 70% o más. Por otro lado, métricas como el costo por adquisición y el costo por clic dependerán del presupuesto asignado a las campañas de marketing. Sin embargo, con grupos de clientes bien definidos, se espera que la rentabilidad que cada cliente aporta a la empresa sea mayor.

# Objetivos

## Objetivo general

## Desarrollar un modelo de segmentación de clientes basado en el uso de técnicas de aprendizaje automático no supervisado, utilizando el registro de clientes de Viventa LLC desde enero de 2015 hasta abril de 2024, para el posterior desarrollo de estrategias comerciales ajustadas a cada segmento fruto de la clasificación automática de los clientes nuevos de la empresa.

## Objetivos específicos

* Realizar un análisis inicial de datos y su preparación para el desarrollo del modelo.
* Determinar el número de los clusters a utilizar para la aplicación del modelo.
* Aplicar y evaluar el modelo de clusterización K-Means adecuado para la clusterización de clientes nuevos y validar sus resultados utilizando técnicas de validación.

# Datos

## Datos originales

Los datos provienen de una descarga en formato CSV de la tabla de clientes de Viventa LLC, la cual se encuentra alojada en Zoho CRM, plataforma usada para la recolección, manipulación y administración de todo el proceso del cliente en la compañía. Este archivo cuenta con 186 columnas y 28,302 filas de registros de clientes desde enero de 2015 hasta abril de 2024, los cuales tienen un peso de 32,792 kb.

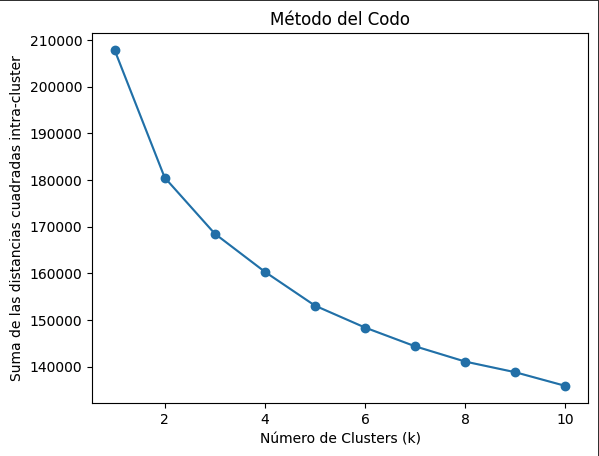
El dataset contiene los siguientes campos:

| Id : Identificador único del usuario |
| --- |
| Website : Campo vacío |
| Candidato Owner Name : Empleado responsable de la gestión del candidato |
| Candidato Source : Fuente por la cual ingresó el cliente al CRM. |
| Candidato Status : Estado de la gestión del ejecutivo con el cliente. |
| Created Time : Fecha en la cual se creó el registro del cliente. |
| Is Converted? : Hace referencia a si el cliente diligenció o no sus datos. |
| Tasks Involved : Hace referencia a sí el cliente tiene relacionada una tarea por él asesor de la empresa. |
| Events Involved : Hace referencia a sí el cliente tiene relacionado un evento por él asesor de la empresa. |
| Calls Involved : Hace referencia a sí el cliente tiene relacionada una llamada por él asesor de la empresa. |
| Activities Involved : Hace referencia a sí el cliente tiene relacionada una actividad por él asesor de la empresa. |
| Modified Time : Fecha de la última actualización en la información del cliente. |
| Converted Account : Identificador de las cuentas con las que se crean anuncios publicitarios |
| Converted Contact : Identificador o llave primaria con el módulo contact |
| Created By : ID del usuario creador del cliente en el sistema. |
| State : Fase en la que se encuentra el caso, informado por el asesor del cliente. |
| ¿En qué ciudad quiere comprar? Ciudad en la que el cliente informa que desea comprar su inmueble. |
| Tipo de documento : Tipo de identificación legal del cliente |
| ¿Por qué medio se enteró de nosotros? : Medio por el cual el cliente informa que conoció la compañía |
| País de residencia : País en el que vive el cliente |
| Fecha Bancolombia : Fecha en la cual el caso es enviado a Bancolombia, si aplica. |
| Mi casa con giros : Campaña en la cual se hacía un descuento según el aliado de crédito para el desembolso de los recursos. |
| Estado de Gestión : Campo donde se reflejaba la gestión del cliente en algunas ocasiones, actualmente está desactualizado. |
| Banco : Aliado de crédito por el cual se realizará el proceso de aprobación y desembolso. |
| Proyecto : Identificador del proyecto en el que el cliente desea comprar. |
| Tipo de Inmueble : Tipo de inmueble que el cliente desea adquirir según el uso/tamaño que el cliente informe que desea adquirir. |
| Torre/Etapa : Etapa o torre dentro del proyecto inmobiliario. |
| Unidad : Número de apartamento o unidad dentro del proyecto inmobiliario. |
| Proyecto no afiliado / Dirección Usado : Información alternativa del inmueble si el proyecto no se encuentra afiliado en las bases internas de la compañía. |
| Valor del Crédito a solicitar : Valor del crédito hipotecario que requiere el cliente. |
| Valor del Inmueble : Valor del inmueble que el cliente desea comprar |
| Última actividad : Fecha de última validación del cliente. |
| Detalle actividades : Campo en el que se van detallando los avances de actividades del cliente. |
| Nombre del Aliado : Campo en el que se diligencia "Viventa" se queda vacío, según el caso continúe. |
| Fecha Informe Bancolombia : Fecha en la cual Bancolombia informa que el cliente es referido suyo. |
| Estado Davivienda : Informa el estado de gestión del cliente, si aplica. |
| Campaign : Campaña publicitaria por la cual ingreso el cliente a la compañía |
| Acquisition : clasificación interna por la cual ingreso el cliente a la compañía |
| Motivo : Motivo por el que se detiene el caso del cliente. |
| Departamento : Clasificación del área comercial responsable del cliente. |
| Region : Clasificación del área geográfica donde se encuentra el cliente. |
| Converted Deal : Identificador o llave primaria con el módulo Deal. |
| Converted Time : Fecha en la cual el cliente es convertido, es decir, llena una encuesta con su información. |
| Ad : Nombre del anuncio publicitario por el cual ingresó el cliente a la empresa. |
| AdGroup Name : Nombre del grupo de anuncios publicitarios por el cual ingresó el cliente a la empresa. |
| Ad Campaign Name : Nombre de la campaña de anuncios publicitarios por el cual ingresó el cliente a la empresa. |
| Ad Click Date : fecha en la cual el cliente ingresó al anuncio publicitario por el cual ingresó a la empresa. |
| Ad Network : Aliado que distribuye el anuncio publicitario por el cual ingresó el cliente a la empresa. |
| Click Type : Forma de ingreso o dispositivo con el que anuncio publicitario por el cual ingresó el cliente a la empresa. |
| Cost per Click : Informa el costo por cada click del anuncio por el que ingresó el cliente. |
| Cost per Conversión : Informa el costo por cada conversión del anuncio por el que ingresó el cliente. |
| Necesita Inmueble : Informa si el cliente desea ser contactado por el departamento inmobiliario para ver ofertas. |
| ¿El cliente tiene propiedad seleccionada? Informa si el cliente ya tiene un inmueble seleccionado o no. |
| Reactivado : Informa si el cliente tuvo un proceso continuo o si en algún momento se reactivó el proceso. |
| Citado a Evento : Informa si en alguna ocasión el cliente fue citado a un evento comercial de la empresa. |
| SI asistió : Informa si el cliente asistió o no al evento en el que se le citó. |
| City : Ciudad donde el cliente informa que vive. |
| Convertido : Informa si el cliente fue convertido o no, este campo es usado en un módulo posterior llamado "Oportunidades". |
| Edad : Edad informada por el cliente. |
| Estado : Estado en el que el cliente informa que vive. |
| Fecha cambio de estado : Fecha en la cual se cambió el estado de la solicitud del cliente. |
| Fecha nacimiento : Fecha de nacimiento informada por el cliente. |
| Last Activity Time : Fecha de la última actividad realizada para el cliente. |
| Necesita Crédito : Campo en el que el cliente indica si necesita crédito o no |
| Sexo : Genero informado por el cliente. |
| Vive en Colombia : Campo en el que el cliente informa si vive o no en Colombia. |
| Ya tiene propiedad : Campo en el cliente informa si posee propiedades inmobiliarias a su nombre o no. |
| Tiempo de conversión de leads en días : Diferencia entre la fecha de creación del cliente en el CRM y su fecha de conversión. |
| Respuesta consulta centrales : Resultado de la consulta del cliente en centrales de riesgo de crédito. |
| Ocupación Principal : Principal fuente de ingresos del cliente. |

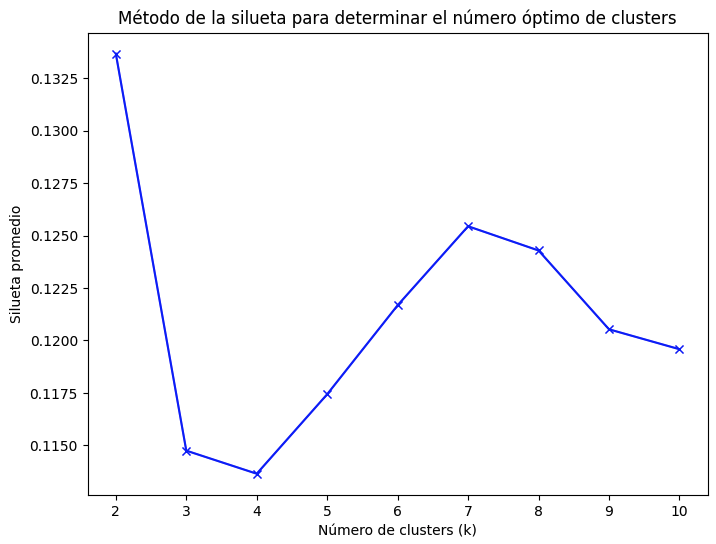
## Datasets

El dataset principal se construyó a partir de los datos originales, siguiendo los siguientes pasos:

* **Limpieza y preprocesamiento de datos:** Se realizó una exhaustiva limpieza y preprocesamiento de los datos, que incluyó la eliminación de valores atípicos y duplicados, el manejo de valores faltantes, y la estandarización de variables categóricas.
* **Normalización de datos:** para aplicar k-means, se normalizaron los datos numéricos utilizando la función ‘normalize’, garantizando que todas las características contribuyeran de manera equitativa al cálculo de distancias.
* **One-hot encoding para variables categóricas:** Se aplicó one-hot encoding para codificar las variables categóricas, asegurando su adecuado tratamiento en el análisis.
* **Determinación del número óptimo de clusters:** Se llevó a cabo la selección del número óptimo de clusters utilizando técnicas como el método del codo y el coeficiente de silueta.



A través del método del codo se observa una gran disminución en la suma de las distancias cuadradas cuando se incrementa el número de clusters de 1 a 2, y luego a 3. En este caso, el número óptimo de clusters para los datos podría ser 3, según la interpretación del gráfico después de este punto, agregar más clusters no proporciona una mejora significativa.

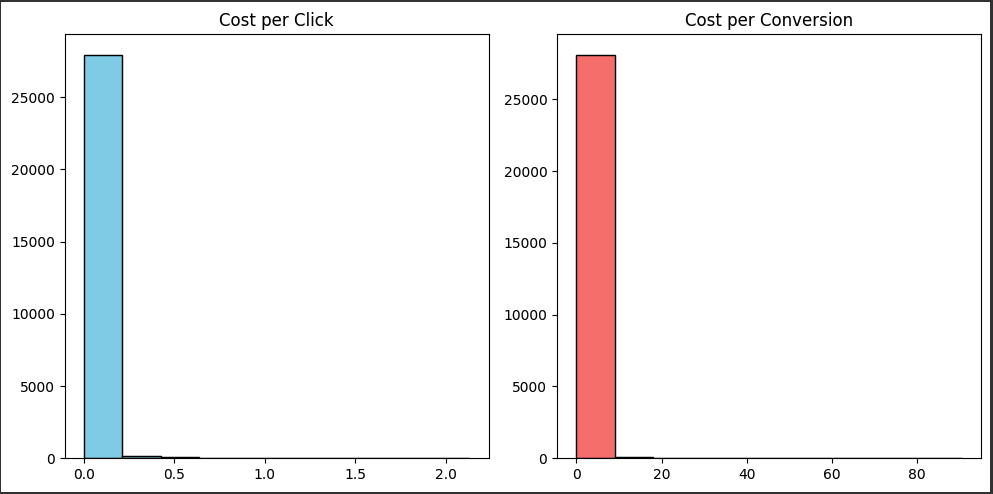


Por otro lado,basado en el coeficiente de silueta, el número óptimo de clusters podría ser K = 2, ya que este valor maximiza el coeficiente de silueta promedio.

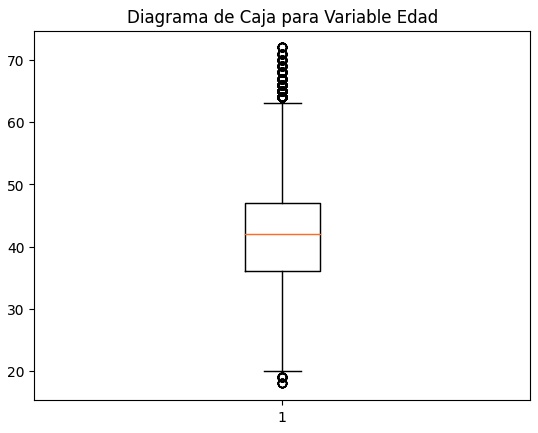
## Analítica descriptiva

Se realizó un análisis preliminar del conjunto de datos, compuesto por 28,302 filas y 186 columnas. El dataset incluye 29 columnas con valores float64, 4 columnas con valores int64 y 153 columnas con datos de tipo object. Con el fin de garantizar la calidad de los datos, se tomaron varias medidas:

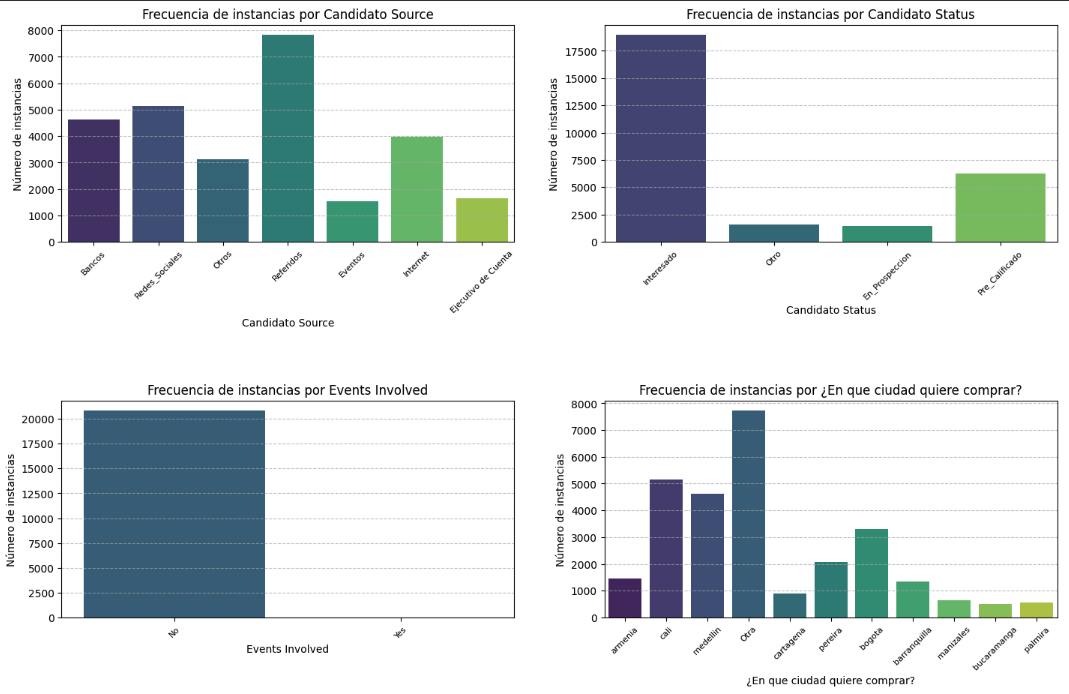
* Eliminar las columnas que contenían más del 50% de datos nulos, con especial atención a aquellas con más del 95% de valores faltantes. También se excluyeron columnas de ID 's, aquellas con un único valor en todos los registros y las relacionadas con fechas, ya que no aportan información relevante para el análisis de clustering.
* Se realizó un análisis gráfico de las variables numéricas, evidenciándose una alta correlación entre algunas de ellas. Por ejemplo, un histograma mostró que las variables “Cost per Click” y “Cost per Conversión” contenían información redundante. Por lo tanto, se decidió conservar sólo una de ellas y eliminar la otra.



* Se representó la distribución de algunas variables numéricas utilizando diagramas de bigotes (boxplots) con el fin de identificar valores atípicos. Posteriormente, se ajustaron los outliers conforme a las indicaciones del negocio.

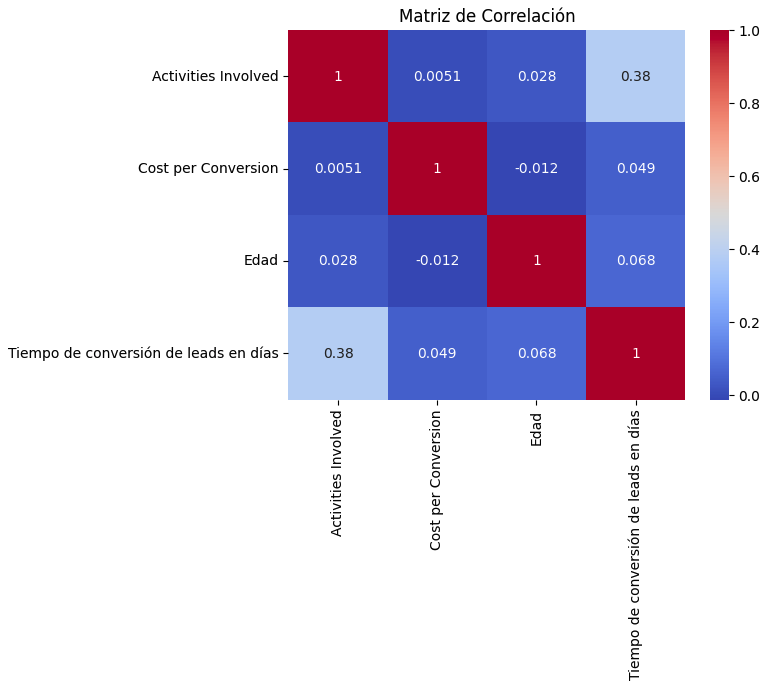


* Se procedió a estandarizar las variables categóricas mediante su agrupación en conjuntos homogéneos, con el propósito de facilitar su procesamiento y comprensión. Luego, se llevó a cabo una visualización del recuento de instancias por variables categóricas, con el objetivo de obtener una visión general de la distribución de las diferentes categorías. Este enfoque permitió identificar desequilibrios en la distribución de los datos.



Al analizar las frecuencias de las variables categóricas a través de su representación gráfica, se observa que en algunas columnas la mayoría de los datos pertenecen a una sola categoría. En respuesta a esto, se inició un proceso de revisión de estas variables con el objetivo de evaluar su relevancia para la inclusión en el modelo.

* Además, se procedió a generar una matriz de correlación para las variables numéricas, con el propósito de analizar las relaciones lineales dentro del conjunto de datos, facilitando la identificación de posibles patrones o tendencias significativas.



Por ejemplo, las correlaciones entre las variables numéricas son bastante bajas, a excepción de la correlación moderada entre “Activities Involved” y “Tiempo de conversión de leads en días”.

* Finalmente, para abordar los datos faltantes, se asignaron valores utilizando la moda para variables categóricas, el promedio para variables numéricas y, en casos excepcionales, se creó una nueva categoría denominada “Unknown”.

| import pandas as pd import numpy as np import seaborn as sns import matplotlib.pyplot as plt import warnings from sklearn.datasets import make\_blobs from sklearn.cluster import KMeans from sklearn.metrics import silhouette\_score from google.colab import files  warnings.filterwarnings('ignore') !pip install unidecode  #Cargar los datos archivo = 'https://raw.githubusercontent.com/JuanJimenezGuerra/momentos\_evaluativos/main/momentos\_evaluativos/ME03/Datos\_Grupo03.csv' df= pd.read\_csv(archivo, sep= ",", low\_memory=False)  #Eliminar columnas con más del 50% de registros nulos: columnas\_con\_nulos = df.columns[df.isnull().sum() > df.shape[0]/2] df = df.drop(columnas\_con\_nulos, axis=1)  #Eliminar columnas que contengan nombres, ID's, un único valor o fechas: df.drop(['Id','Candidato Owner Name','Converted Deal','Converted Account','Converted Contact','Created By'], axis='columns', inplace=True) df.drop(['Is Converted?','Nombre del Aliado'], axis='columns', inplace=True) df.drop(['Created Time','Modified Time', 'Ultima actividad', 'Detalle actividades', 'Fecha Informe Bancolombia', 'Converted Time',  'Fecha cambio de estado','Fecha nacimiento','Last Activity Time'], axis='columns', inplace=True)  #Corregir y estandarizar variables categóricas, por ejemplo, columna 'Sexo': sexo\_map = {'F':'F',  'M':'M',  'H':'M'} df['Sexo'] = df['Sexo'].map(sexo\_map)  #Revisar gráficamente las variables numérica, ejm: Variables relacionadas con el costo por click: fig, axs = plt.subplots(1, 2, figsize=(10, 5))  # Histograma para Cost per Click axs[0].hist((df['Cost per Click']), bins=10, color='skyblue', edgecolor='black') axs[0].set\_title('Cost per Click')  # Histograma para Cost per Conversion axs[1].hist((df['Cost per Conversion']), bins=10, color='salmon', edgecolor='black') axs[1].set\_title('Cost per Conversion') plt.tight\_layout() plt.show()  #Realizar un Boxplot para explorar variables numéricas, en este caso la variable edad:  plt.boxplot(df['Edad']) plt.title('Diagrama de Caja para Variable Edad') plt.show()  # Visualización del número de instancias por variables categóricas: num\_variables = len(Varscat) num\_filas = (num\_variables + 1) // 2 num\_columnas = 2 fig, axs = plt.subplots(num\_filas, num\_columnas, figsize=(15, num\_filas \* 5))  for i, variable in enumerate(Varscat):  fila = i // num\_columnas  columna = i % num\_columnas  sns.countplot(data=df, x=variable, palette='viridis', ax=axs[fila, columna])  axs[fila, columna].set\_title(f'Frecuencia de instancias por {variable}', fontsize=12)  axs[fila, columna].set\_xlabel(variable, fontsize=10)  axs[fila, columna].set\_ylabel('Número de instancias', fontsize=10)  axs[fila, columna].tick\_params(axis='x', rotation=45, labelsize=8)  axs[fila, columna].grid(axis='y', linestyle='--', alpha=0.7) plt.tight\_layout(pad=3.0) plt.show()  # Realizar matriz de correlación para variables numéricas: sns.heatmap(df[['Activities Involved', 'Cost per Conversion', 'Edad', 'Tiempo de conversión de leads en días']].corr(), annot=True, cmap='coolwarm') plt.title('Matriz de Correlación') plt.show()  # Imputar valores faltantes:  df['Candidato Source'].fillna(df['Candidato Source'].mode()[0], inplace=True) df['Candidato Status'].fillna(df['Candidato Status'].mode()[0], inplace=True) df['¿Por que medio se entero de nosotros?'].fillna(df['¿Por que medio se entero de nosotros?'].mode()[0], inplace=True) df['País de residencia'].fillna('Unknown', inplace=True) df['Estado Davivienda'].fillna(df['Estado Davivienda'].mode()[0], inplace=True) df['Acquisition'].fillna(df['Acquisition'].mode()[0], inplace=True)  #Escalar variables numéricas: num\_vars = df.select\_dtypes(include = ['float64','int32','int64']).columns.tolist() pd.set\_option('display.float\_format', lambda x: '%.4f' % x) from sklearn.preprocessing import normalize df[num\_vars] = normalize(df[num\_vars]) df[num\_vars].head(2)  # Aplicación de la función de usuario Dummies: one-hot encoding para las variables categóricas:  df = pd.get\_dummies(df, drop\_first=1) df.info() |
| --- |

# 

# Referencias

Portland State University. (n.d.). *Machine Learning - 12 Unsupervised Learning*. Portland State University. https://web.pdx.edu/~gerbing/Books/ML/12-cluster.html

Rodríguez, D. (2023, June 16). *Identificar el número de clústeres con Calinski-Harabasz en k-means e implementación en Python*. Analytics Lane. Retrieved April 25, 2024, from https://www.analyticslane.com/2023/06/16/identificar-el-numero-de-clusteres-con-calinski-harabasz-en-k-means-e-implementacion-en-python/

Moya, M. J. (2020, 01 01). *Aprendizaje no supervisado en el perfilamiento de clientes para profit scoring*. Repositorio Uniandes. Retrieved April 3, 2024, from https://repositorio.uniandes.edu.co/server/api/core/bitstreams/7fb9cbec-6760-478d-9962-c699b61bc1e1/content