

Trabajo Formativo Semana 1

Nombre del estudiante: Enzo Rodrigo Silva Soto

Nombre del módulo: TALLER DE ANALÍTICA EN LOS NEGOCIOS

Nombre del docente: Beatriz Torrellas Torres

Semana: Semana 1

Fecha: 19 de Agosto del 2025

Tabla de contenidos

[Introducción 3](#_Toc119486256)

[Desarrollo 4](#_Toc119486257)

[Conclusión 17](#_Toc119486258)

# Introducción

Hoy en día, el análisis de datos se ha convertido en una herramienta fundamental para las organizaciones, ya que permite transformar grandes volúmenes de información en conocimiento útil para la toma de decisiones. A través de técnicas de segmentación como el análisis de clústeres, es posible identificar patrones de comportamiento y diferenciar grupos de clientes según características comunes. Esto resulta especialmente valioso en áreas como el marketing, donde conocer a los distintos perfiles de consumidores ayuda a diseñar estrategias más efectivas, optimizar recursos y mejorar la relación con los clientes.

# Desarrollo

1. Importación de Librerías

Texto

El contenido generado por IA puede ser incorrecto.

1. Se procede a cargar los datos desde el csv, se define que, si los datos no se encuentran en formato csv, se buscara un archivo Excel.

Texto

El contenido generado por IA puede ser incorrecto.

1. Se realiza una exploración inicial de los datos, donde podemos identificar que se trata de un dataset que contiene 5 columnas, no posee valores nulos ni filas duplicadas, junto a ello se verificaron las estadísticas básicas de las columnas numéricas.

Texto

El contenido generado por IA puede ser incorrecto.

Texto

El contenido generado por IA puede ser incorrecto.

1. Se realiza una limpieza de datos, normalizando en primer lugar las minúsculas y los espacios, se validan que existan las tablas dentro del dataframe, si bien en el paso anterior se verifico con el resultado que no se contaban con valores nulos ni duplicados, para eventos de esta actividad se realiza la eliminación de duplicados, y la eliminación de nulos.

Texto

El contenido generado por IA puede ser incorrecto.

Texto

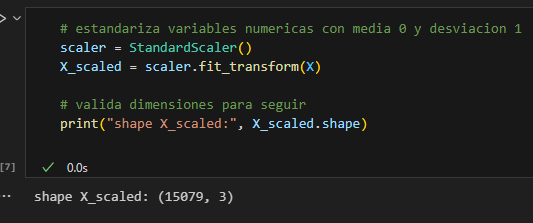
El contenido generado por IA puede ser incorrecto.

1. Se procede a realizar el chequeo de outliers, para detectar datos que extremos.

Texto

El contenido generado por IA puede ser incorrecto.

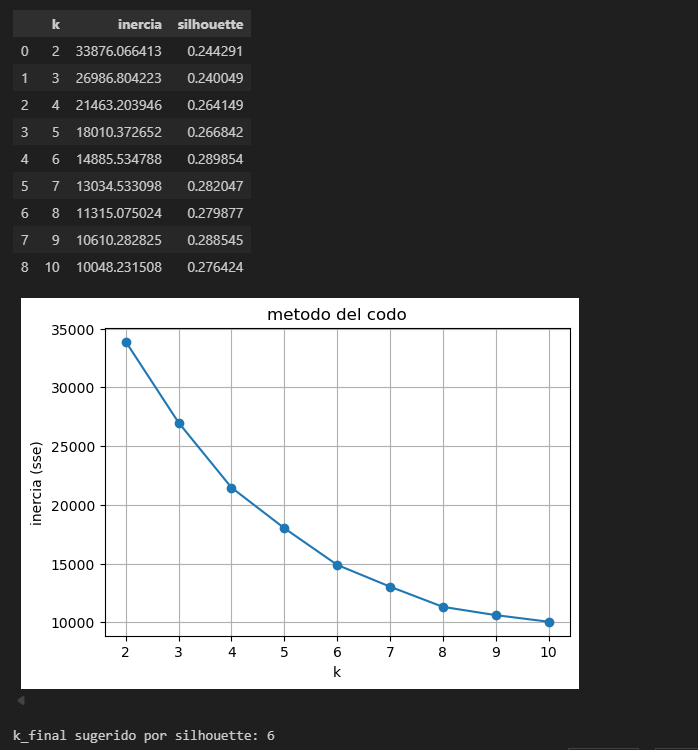
1. Se escalaron las variables numéricas para que todas quedaran en la misma medida. De esta forma ninguna variable influye más que otra en el análisis. El resultado muestra que se trabajara con 15.079 registros y 3 columnas (edad, ingreso anual y puntaje de gasto).



1. Se probo el modelo K-Means con distintos valores de k entre 2 y 10. Para cada k se calculó la inercia (SSE) y el índice silhouette. La grafica del método del codo mostro como la inercia disminuye al aumentar k, y el índice silhouette indico que el valor más adecuado era k = 6. Con esto se definió el número de clusters a utilizar en el análisis.

Texto

El contenido generado por IA puede ser incorrecto.



1. Se entreno el modelo K-Means con k = 6. Como resultado se obtuvo una inercia final de 14.885,53 y un índice silhouette de 0,2899, lo que confirma una separación aceptable entre los grupos. Además, se observó que los tamaños de los clusters quedaron bastante equilibrados, variando entre 2.400 y 2.600 clientes por grupo.

Texto

El contenido generado por IA puede ser incorrecto.

1. Se transformaron los centroides de los clusters a la escala original de las variables (edad, ingreso anual y puntaje de gasto). Esto permite interpretar de forma directa las características promedio de cada grupo de clientes.

Texto

El contenido generado por IA puede ser incorrecto.

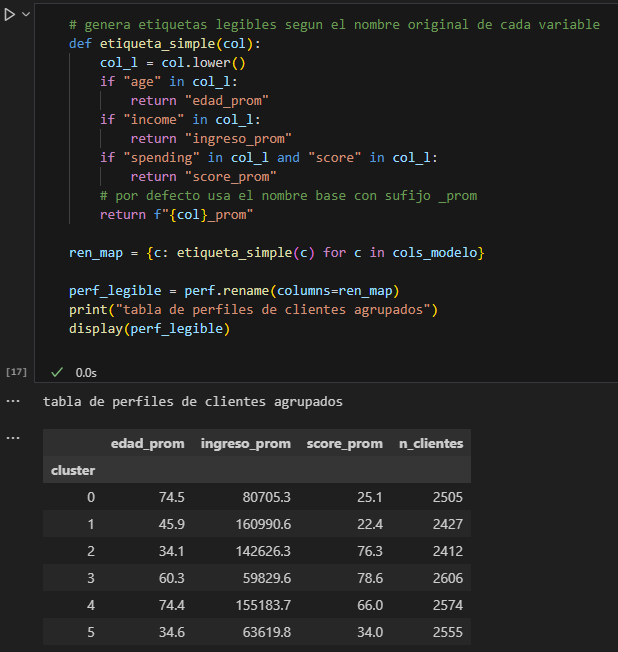
1. Se calcularon los valores promedio y medianos de edad, ingreso anual y puntaje de gasto para cada cluster, junto con la cantidad de clientes por grupo. Estos resultados permiten describir de manera clara el perfil de cada segmento.

Texto

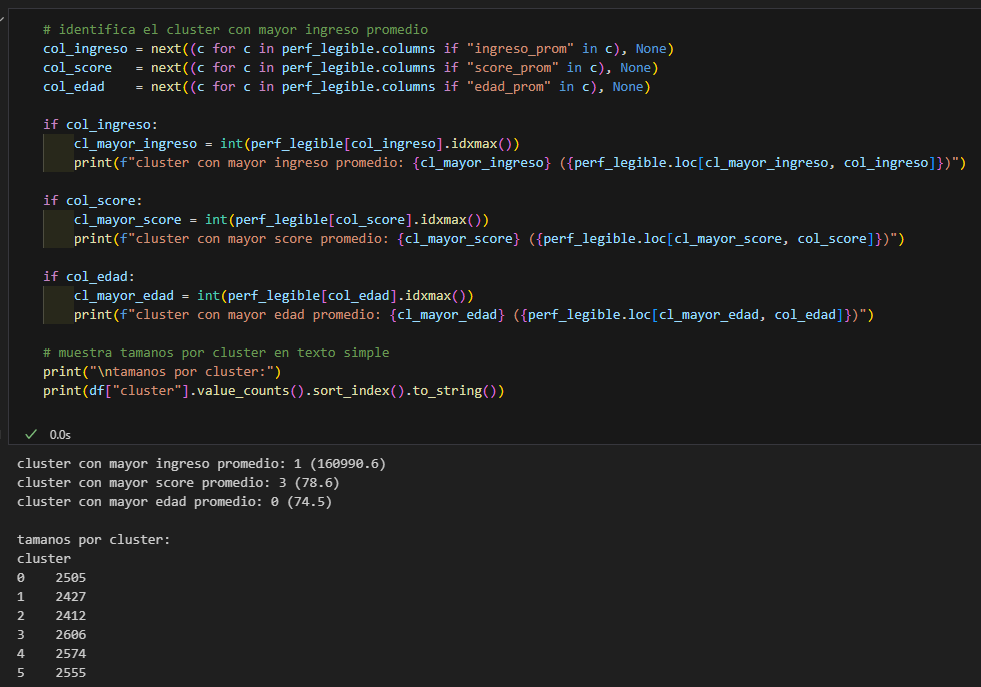
El contenido generado por IA puede ser incorrecto.



1. Se genero una tabla resumen con los perfiles de clientes agrupados. En ella se muestran las medias de edad, ingreso anual y puntaje de gasto para cada cluster, junto con la cantidad de clientes en cada grupo. Esta tabla facilita la interpretación y presentación de los resultados.



1. Se identificaron los clusters con los valores promedio más altos en cada variable: el cluster 1 con el mayor ingreso promedio, el cluster 3 con el mayor puntaje de gasto, y el cluster 0 con la mayor edad promedio. Además, se mostraron los tamaños de cada cluster, lo que permite dimensionar la cantidad de clientes que compone cada grupo.



**Análisis de Resultados**

* El cluster 1 tiene el ingreso promedio más alto, pero su puntaje de gasto es bajo. Esto muestra que son personas con alto poder adquisitivo, pero que no gastan tanto.
* El cluster 3 es el que más gasta en promedio, aunque no tiene los ingresos más altos. Representa clientes que consumen bastante dentro de su nivel.
* El cluster 0 agrupa a las personas de mayor edad, con ingresos y gasto más moderados.

# 

# Conclusión

La actividad permitió aplicar técnicas de segmentación de clientes utilizando el algoritmo K-Means, logrando identificar grupos con características distintas en cuanto a edad, ingresos y nivel de gasto. El ejercicio demostró la utilidad del análisis de datos para comprender mejor a los clientes y apoyar la toma de decisiones en un contexto real. Además, resultó relevante retomar y reutilizar algoritmos y códigos ya aprendidos en el ramo de Minería de Datos, donde también se trabajó con K-Means, ya que esto facilitó el desarrollo del trabajo actual y reforzó los conocimientos adquiridos previamente.