

**Evaluación Final: Proyecto Caso de estudio**

Caso de estudio: Aplicar técnicas de analítica y visualización de datos, para aplicar un algoritmo de aprendizaje automático no supervisado, que se utiliza para descubrir patrones en un conjunto de datos.

**Alumnos:**

Francisco Javier Solis Bamaca.

Paola Jadziry Culebro Ovando.

**Docente:** Dr. Christian Mauricio Castillo Estrada

**Grado y grupo:** 8° D

**Fecha de entrega:** 24 de mayo del 2024

**Materia:** Big Data

**Introducción**

En el contexto académico y universitario, donde la aplicación práctica de conocimientos teóricos es esencial para el desarrollo profesional, nuestro proyecto, titulado "Técnicas de Analítica y Visualización de Datos para Aplicar un Algoritmo de Aprendizaje Automático", surge como una oportunidad única para integrar y expandir nuestros estudios en análisis de datos y machine learning. Este proyecto tiene como objetivo principal explorar y aplicar técnicas avanzadas de análisis de datos y visualización utilizando Python, para identificar patrones de comportamiento de los usuarios, con el fin de mejorar la toma de decisiones estratégicas dentro de cualquier organización.

A través de este proyecto, no solo estamos adquiriendo habilidades técnicas esenciales en Python, sino que también estamos desarrollando una comprensión profunda de cómo los datos pueden ser utilizados para obtener insights valiosos que pueden guiar las decisiones empresariales.

**Descripción del proyecto**

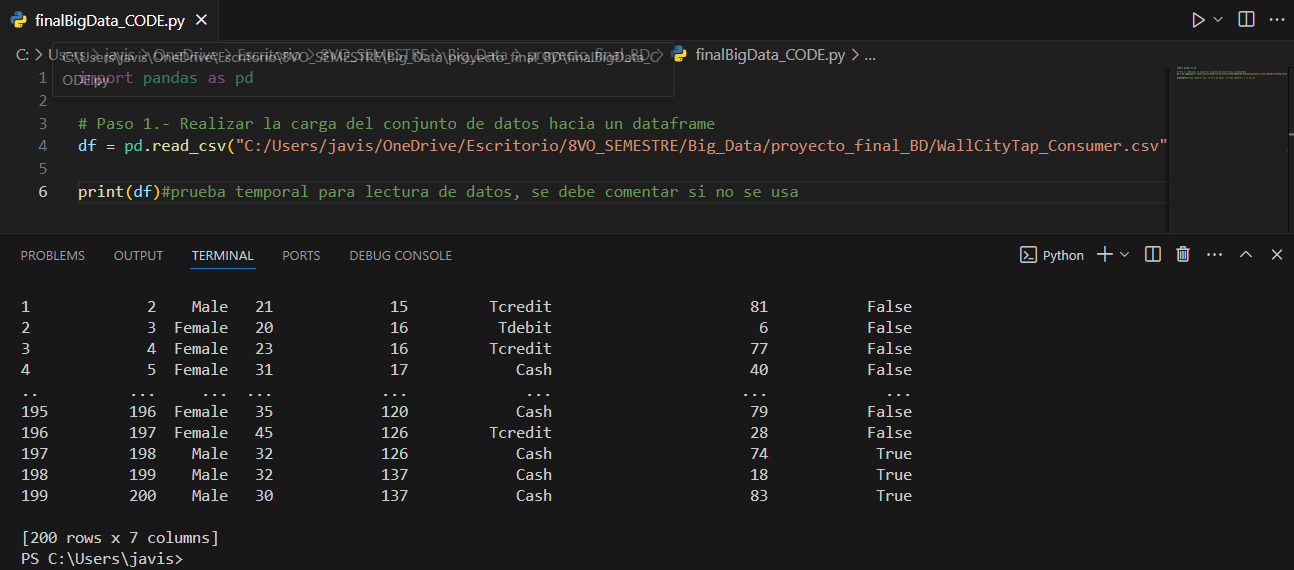
Propósito: El propósito del presente proyecto final de la UC: BigData, es aplicar el uso de algoritmo de aprendizaje no supervisado denominado K-means; con la finalidad de analizar un conjunto de datos de consumidores frecuentes relacionados con la empresa WallCityTap S.A. Con la finalidad de dar respuesta a una serie de cuestionamientos e identificar patrones de comportamiento que coadyuven al proceso de toma de decisiones de los directivos; para ello, se realizará el desarrollo de un Dashboard de indicadores utilizando técnicas de visualización de datos y el framework Shiny for R or Python.

El Diccionario del conjunto de datos, se describe a continuación:

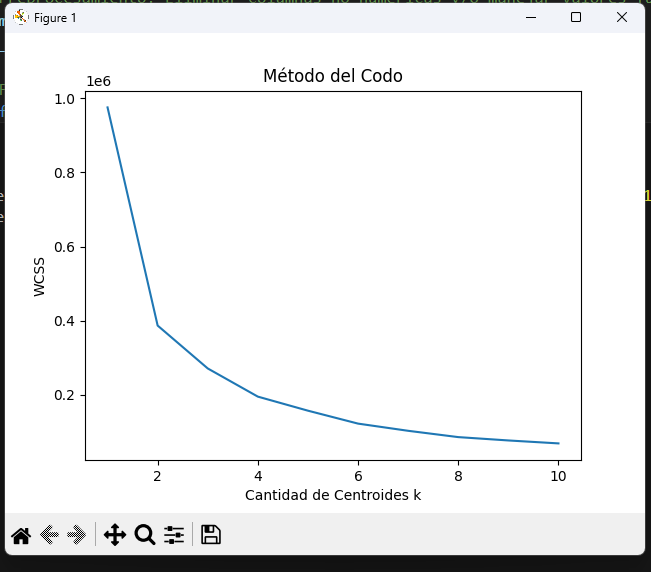
* **CustomerID:** un identificador único para cada cliente.
* **Gender:** El género del cliente.
* **Age:** La edad del cliente.
* **Annual\_Income:** El ingreso anual del cliente (en miles de dólares).
* **Spending\_Score (1–100):** Puntuación asignada al cliente en función de sus hábitos de gasto. La puntuación varía de 1 a 100, y una puntuación más alta indica un cliente que gasta más.
* **Payment\_Methods:** El tipo de medio de pago que más utiliza el consumidor.
* **Online\_Purchases:** Indicador de compras por mes mediante uso del E-commerce.

**Desarrollo:**

**Paso 1.- Realizar la carga del conjunto de datos hacia un dataframe:**

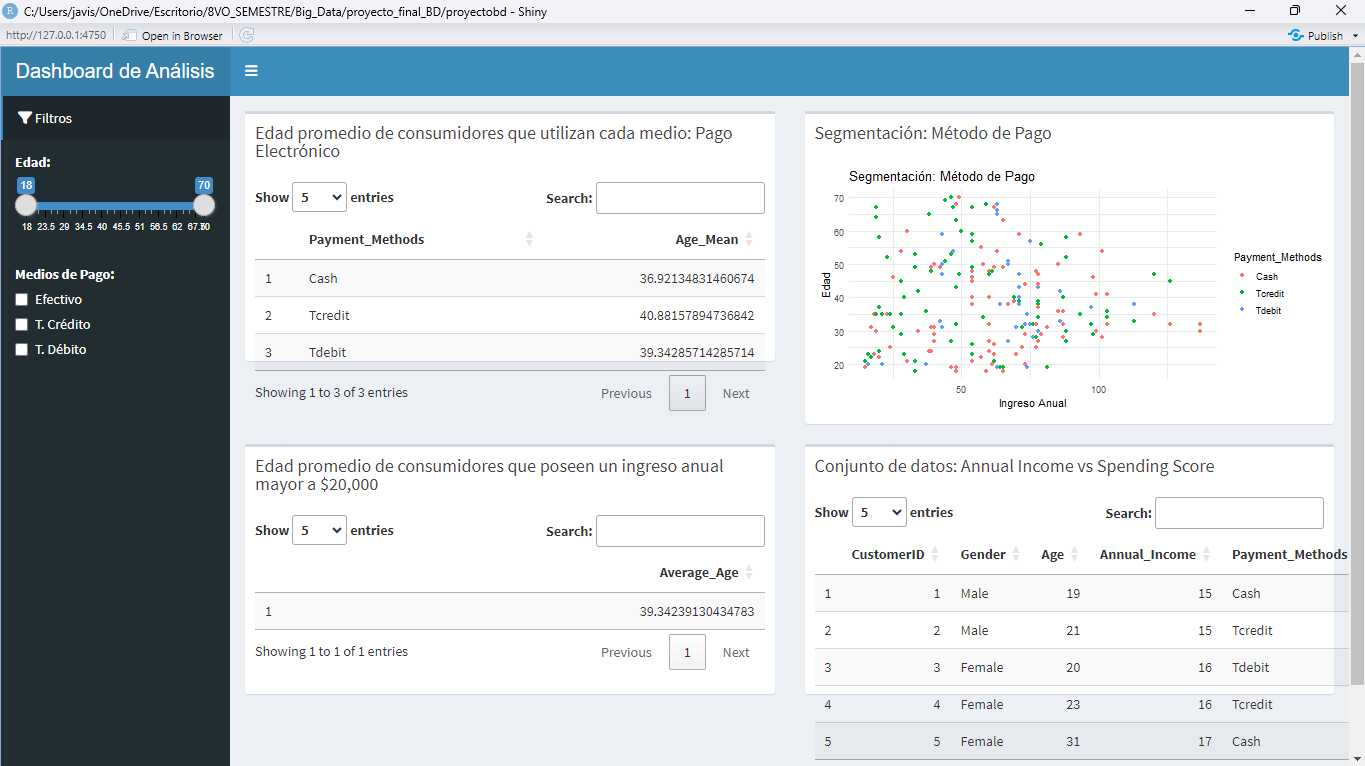
****

**Paso 2.- Identificar el Valor K (Centroides) utilizando el método del codo:**

****

**Paso 3.- Mockup del Dashboard de Visualización de Datos (Shiny for R or Python)**

**\*Importante:** de aquí en adelante nos decidimos a cambiarnos a R usando Rstudio para usar shiny para la creación de nuestro Dashboard



**Interpretación de Dashboard para toma de decisiones**

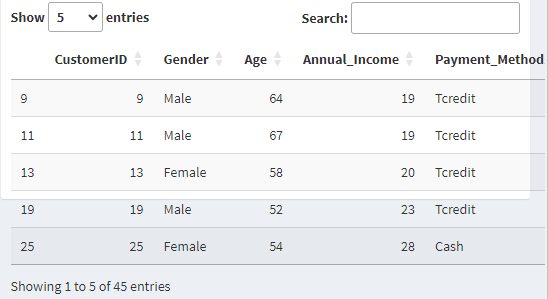
Los directivos desear dar respuesta a los siguientes cuestionamientos, con la finalidad de formular una próxima campaña de nuevos productos y ofertas para sus consumidores frecuentes.

**Cuestionamientos de los Directivos de WallCityTap S.A:**

**1. ¿Es viable financieramente lanzar una promoción del 2 X 1 + meses a interés**

**para los consumidores con una edad mayor a 50 años?**

No es tan viable tomando en cuenta el número de compradores, usando el filtro de edad de personas mayores a 50 años



Es posible que eso se aplique a personas de menor edad, las cuales gastan más

Podemos observar en la imagen, el numero de entradas de compras de las personas de 50 años o más que es de 45 entradas.

**2.- ¿Cuál es el medio de pago preferido de nuestros consumidores?**

Para dar solución a esta pregunta debemos considerar usar el filtro de método de pago y ver el numero de entradas de cada uno:

Para pago en efectivo: 

Para pago con tarjeta de crédito: 

Para pago con tarjeta de débito: 

Podemos observar que el método ganador es el efectivo con 89 entradas.

Es importante recalcar que debemos usar los filtros de manera correcta, tomando en cuenta todas las edades:

**3.- Identificar y describir los consumidores potenciales, para generar mayores**

**ventas y obtener mayores ingresos.**

Para identificar y describir los consumidores potenciales que podrían generar mayores ventas e ingresos, es crucial analizar varios factores presentes en la información proporcionada. Estos factores incluyen el género, la edad, el ingreso anual, los métodos de pago preferidos, el puntaje de gasto y si realizan compras en línea. A continuación, se presenta un análisis basado en estos criterios:

Edad

* Jóvenes (19-24 años): Este grupo tiene un puntaje de gasto promedio más alto, especialmente aquellos que utilizan métodos de pago digitales. Es probable que sean consumidores activos y receptivos a nuevas ofertas y promociones.
* Adultos jóvenes (25-34 años): Muestra una diversidad en sus métodos de pago, con un equilibrio entre efectivo y digitales. Su puntaje de gasto también es notablemente alto, indicando una alta capacidad de compra.
* Adultos medios (35-49 años): Este grupo tiene un puntaje de gasto moderado, con una preferencia por métodos de pago digitales. Son posibles clientes estables con un historial de compras regulares.
* Adultos mayores (50+ años): Aunque su puntaje de gasto puede ser más bajo, este grupo representa una oportunidad significativa debido a su estabilidad financiera y disposición a realizar compras importantes.

Ingresos Anuales

Ingresos bajos a medianos (15-19k): Este segmento muestra una variación en sus métodos de pago, con algunos prefiriendo efectivo y otros métodos digitales. Su puntaje de gasto es moderado, lo que indica una capacidad de compra limitada.

Ingresos altos (20k+): Los consumidores con ingresos más altos tienen un puntaje de gasto significativamente más alto, lo que sugiere una mayor capacidad de compra y una disposición a gastar en productos premium o servicios.