**Create a K-means clustering algorithm to group customers of a retail store based on their purchase history.**

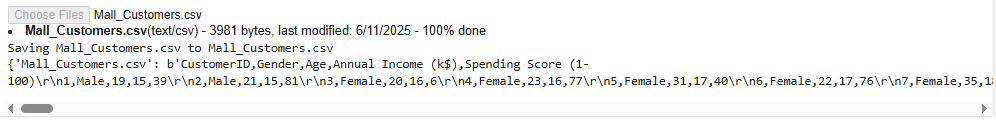
**Dataset :-** [**https://www.kaggle.com/datasets/vjchoudhary7/customer-segmentation-tutorial-in-python**](https://www.kaggle.com/datasets/vjchoudhary7/customer-segmentation-tutorial-in-python)

**Step 1:** Kaggle Setup

# Upload kaggle.json (download it from your Kaggle account)

from google.colab import files

files.upload()



**Step 2:**

!mkdir -p ~/.kaggle

!cp kaggle.json ~/.kaggle/

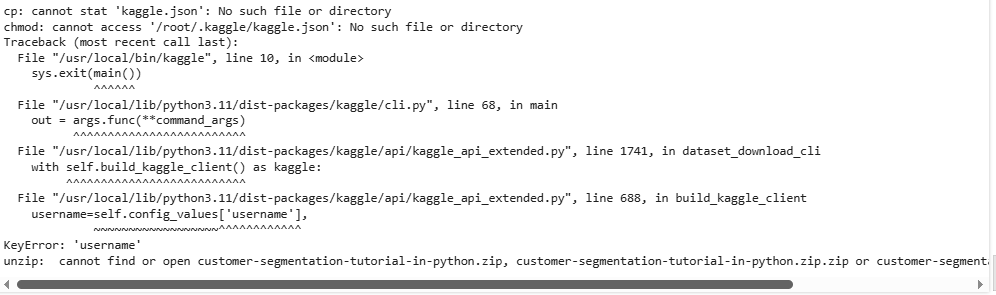
!chmod 600 ~/.kaggle/kaggle.json

# Download the dataset

!kaggle datasets download -d vjchoudhary7/customer-segmentation-tutorial-in-python

# Unzip the dataset

!unzip customer-segmentation-tutorial-in-python.zip



**Step 3:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.cluster import KMeans

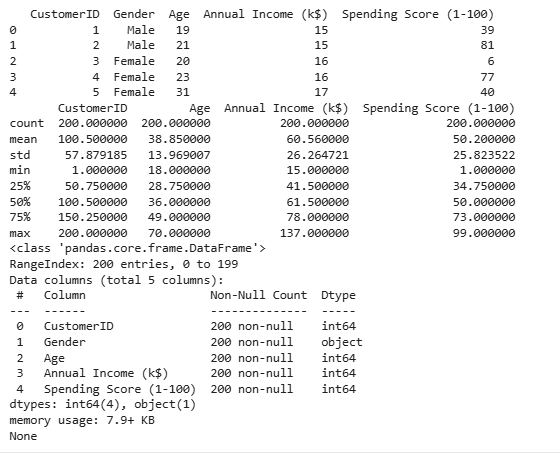
from sklearn.preprocessing import StandardScaler

df = pd.read\_csv('Mall\_Customers.csv')

print(df.head())

print(df.describe())

print(df.info())



**Step 4:**

X = df[['Annual Income (k$)', 'Spending Score (1-100)']]

# Optional: Feature Scaling

scaler = StandardScaler()

X\_scaled = scaler.fit\_transform(X)

wcss = []  # Within-cluster sum of squares

for i in range(1, 11):

    kmeans = KMeans(n\_clusters=i, init='k-means++', random\_state=42)

    kmeans.fit(X\_scaled)

    wcss.append(kmeans.inertia\_)

# Plot the elbow curve

plt.figure(figsize=(8,5))

plt.plot(range(1, 11), wcss, marker='o')

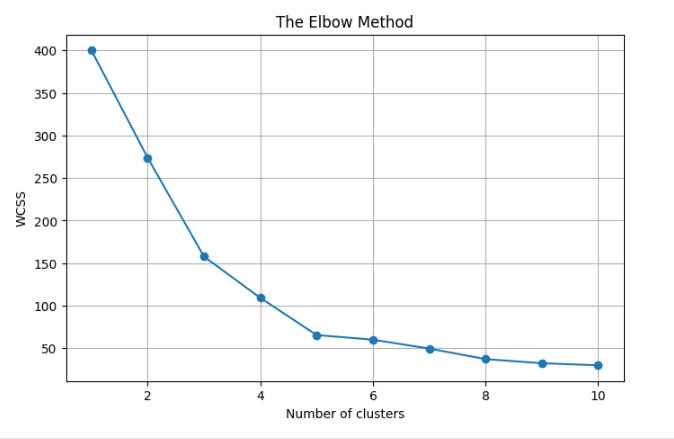
plt.title('The Elbow Method')

plt.xlabel('Number of clusters')

plt.ylabel('WCSS')

plt.grid(True)

plt.show()



**Step 5:**

kmeans = KMeans(n\_clusters=5, init='k-means++', random\_state=42)

y\_kmeans = kmeans.fit\_predict(X\_scaled)

# Add cluster labels to dataframe

df['Cluster'] = y\_kmeans

plt.figure(figsize=(8,6))

sns.scatterplot(x=X\_scaled[:, 0], y=X\_scaled[:, 1], hue=y\_kmeans, palette='Set1')

plt.scatter(kmeans.cluster\_centers\_[:, 0], kmeans.cluster\_centers\_[:, 1],

            s=300, c='yellow', label='Centroids', marker='\*')

plt.title('Customer Segments (K-Means Clustering)')

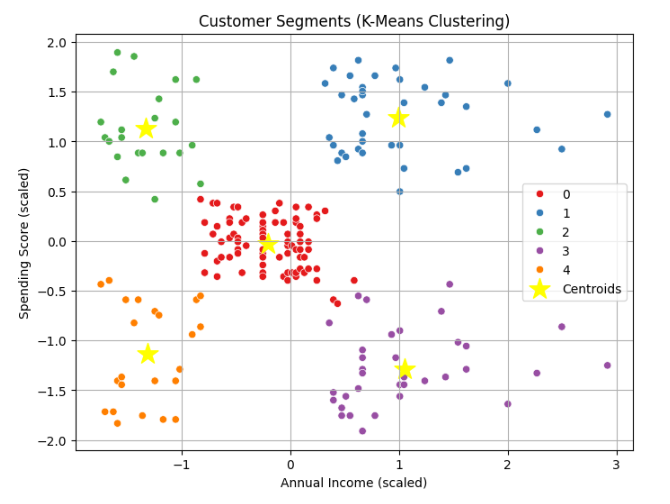
plt.xlabel('Annual Income (scaled)')

plt.ylabel('Spending Score (scaled)')

plt.legend()

plt.grid(True)

plt.show()



**SUMMARY:**

The document outlines a project to implement a K-Means clustering algorithm for customer segmentation using the Aegean Wi-Fi Intrusion Dataset. The process begins by setting up access to the dataset via Kaggle and proceeds with standard Python data analysis tools including pandas, NumPy, Matplotlib, and Seaborn. After loading and exploring the dataset, the relevant features—Annual Income and Spending Score—are scaled using StandardScaler. The optimal number of clusters is determined through the Elbow Method by plotting the Within-Cluster Sum of Squares (WCSS). A K-Means model with 5 clusters is then trained, and the resulting clusters are visualized with a scatter plot, highlighting the centroids. This approach helps group customers based on similar purchasing behavior for targeted marketing.