

# Description of the Customer Segmentation Analysis Code

## Purpose:

Analyzes customer spending patterns based on age using K-Means clustering and visualizes results with a scatter plot.

## Libraries Used:

- pandas: For data manipulation and loading the CSV file.
- matplotlib.pyplot: For basic plotting (elbow curve and scatter plots).
- seaborn: For enhanced scatter plot visualization with clusters.
- sklearn.cluster.KMeans: For performing K-Means clustering.

## Step 1: Load the Dataset:

- Reads Mall\_Customers.csv (downloaded from Kaggle) into a DataFrame using `pd.read_csv()`.
- Assumes the file is in the same directory as the script.

## Step 2: Data Verification:

- Checks for missing values with `data.isnull().sum()` to ensure data quality.

## Step 3: Data Selection:

- Extracts Age and Spending Score (1-100) columns into a variable X for analysis.

## Step 4: Elbow Method:

- Loops through 1 to 10 clusters, fits K-Means, and calculates inertia (within-cluster sum of squares).
- Plots the elbow curve to determine the optimal number of clusters (k).
- Helps user visually decide the best k (typically where the curve bends, e.g., 4).

## Step 5: Apply K-Means Clustering:

- Sets `n_clusters=4` (adjustable based on elbow plot) and fits K-Means to X.
- Predicts cluster labels for each customer and adds them to the DataFrame as Cluster.

## Step 6: Visualize Clusters:

- Uses `seaborn.scatterplot` to create a scatter plot of Age vs Spending Score.
- Colors points by cluster, adds a title, labels, and legend for clarity.
- Displays the plot to show customer segments visually.

## Step 7: Cluster Insights:

- Prints the cluster centers (average Age and Spending Score for each cluster).
- Helps interpret which age groups have higher or lower spending scores.

#### Output:

- Two plots:
  - Elbow curve to choose the number of clusters.
  - Scatter plot showing customer segments.
- Text output of cluster centers for analysis.

#### Customization Notes:

- Replace Mall\_Customers.csv with your own dataset if needed.
- Adjust n\_clusters based on the elbow plot or specific requirements.
- Extend to more features (e.g., income) for deeper analysis.

**Vipul Mehra**