**Customer Segmentation Using Clustering Report**

**Name:** KHANSA AHMAD QURESHI

**Date:** 16th April 2025

**Dataset Used:** Mall\_Customers.csv from Kaggle

**Objective**

The goal of this project was to segment customers based on their demographic and behavioral characteristics, such as age, income, and spending score. The segmentation was done using clustering algorithms, specifically **K-Means clustering**. The output of this task was a model that group’s customers into distinct clusters, helping businesses understand different customer profiles and their needs.

# Steps Taken:

## 1. Exploratory Data Analysis (EDA):

The first step involved loading the dataset (Mall\_Customers.csv) and performing basic cleaning and exploration. We reviewed the columns, checked for missing values, and analyzed the distribution of customer features such as age, income, and spending score.  
We used visualization techniques such as histograms and scatter plots to understand the relationships between features. This helped us identify potential clusters in the data.

## 2. Clustering Algorithm:

**K-Means Clustering:** We applied the K-Means clustering algorithm to segment the customers. The number of clusters was chosen based on the elbow method by plotting the Within-Cluster Sum of Squares (WCSS) against the number of clusters. The optimal number of clusters was found to be 5, i.e. n\_clusters = 5.  
After running the K-Means algorithm, we assigned the cluster labels to each customer and examined the features of each cluster, such as the average age, average income, and average spending score.

## 3. Cluster Interpretation:

We analyzed the average values of the features within each cluster to interpret customer behaviors. The segments were characterized as:

* **Cluster 0:** Older customers with medium income and moderate spending scores.
* **Cluster 1:** High-income, mid-aged customers with high spending scores.
* **Cluster 2:** Young customers with low income but high spending tendencies.
* **Cluster 3:** Young customers with balanced income and moderate spending.
* **Cluster 4:** Older customers with high income but low spending.

These segments can be valuable for businesses in targeting specific customer groups with tailored marketing strategies.

## 4. Visualization:

We visualized the customer segments using scatter plots to help better understand the relationships between age, income, and spending score. These visualizations provided an intuitive way to compare the segments.

# Problems Faced and Solutions:

## 1. Choosing the Right Number of Clusters:

Initially, determining the right number of clusters was challenging. We tried different values for the number of clusters and observed that the elbow method provided a clear indication of where the optimal number of clusters occurred (**5** in our case).

## 2. Handling Data in Streamlit:

There was an issue with calling st.pyplot() without providing a figure, which raised a deprecation warning. The **solution** was to explicitly create a figure object using plt.subplots() and pass it to st.pyplot() to ensure future compatibility and thread safety.

## 3. Interpreting the Clusters:

Interpreting the clusters was not straightforward, as each cluster had different characteristics. However, by looking at the mean values for age, income, and spending score, we were able to make meaningful insights about each customer segment.

# Findings:

* Segment 0: Represents older, middle-income customers who spend moderately. This group may represent retired individuals or those in a stable life phase.
* **Segment 1:** Characterized by high-income, middle-aged customers with high spending power, likely representing professionals or people with premium shopping habits.
* **Segment 2:** This group includes younger customers with low incomes but high spending, which may indicate impulsive or frequent shopping behavior, possibly students or entry-level workers.
* **Segment 3:** This segment consists of young adults with a balanced income and moderate spending habits, likely early-career individuals who are still building their financial independence.
* **Segment 4:** Comprises older customers with high income but low spending tendencies, suggesting a conservative or minimalist approach to spending, possibly retirees or cautious buyers.

# Conclusion:

By using clustering, we successfully segmented customers based on their demographic and behavioral features. These segments provide businesses with insights into their customer base, which can help in targeting customers with appropriate marketing strategies. We also addressed several challenges, such as determining the optimal number of clusters and fixing issues related to Streamlit visualization. The clustering model can be extended and refined by incorporating additional features and using more advanced clustering techniques.