# Customer Segmentation Using Clustering (K-Means)

## 1. Introduction

Customer segmentation is essential for businesses to understand consumer behavior and target marketing efforts effectively. This project applies **K-Means** on the **Mall Customers Dataset** to identify distinct customer groups based on **Annual Income and Spending Score**.

## 2. Approach Used

### Data Preprocessing & Exploration

- **Loaded the dataset** and checked for missing values.
- **Converted categorical variables**, like Gender, into numerical form.
- **Plotted distributions** of Age, Income, and Spending Score for insights.

### Choosing the Optimal Number of Clusters (Elbow Method & Dendrograms)

- Used the **Elbow Method** by plotting the **Within-Cluster-Sum-of-Squares (WCSS)** to find the optimal K value.

### Applying Clustering Algorithms

1. **K-Means Clustering**
    - Implemented K-Means with **K=5 (from Elbow Method)**.
    - Used kmeans.fit_predict() to assign customers to clusters.
    - Extracted cluster **centroids** and visualized them.

### Visualizing & Interpreting Clusters

- **Scatter plots** were used to show how customers are grouped.
- **Cluster centers** were plotted to highlight segment patterns.
- **Business insights** were derived from customer spending behavior.

# 3. Challenges Faced

**Choosing the Right K (Number of Clusters)**

- The Elbow Method sometimes gives **ambiguous results**.

**Handling Outliers & Noise**

- Some customers showed unusual spending behavior, which could affect clustering.
- Considered **scaling techniques (Standardization/Normalization)** for improvement.

**Interpreting Cluster Results**

- Some clusters overlapped, making clear segmentation difficult.
- Applied **feature scaling** and **alternative clustering metrics** (e.g., silhouette score) to refine segmentation.

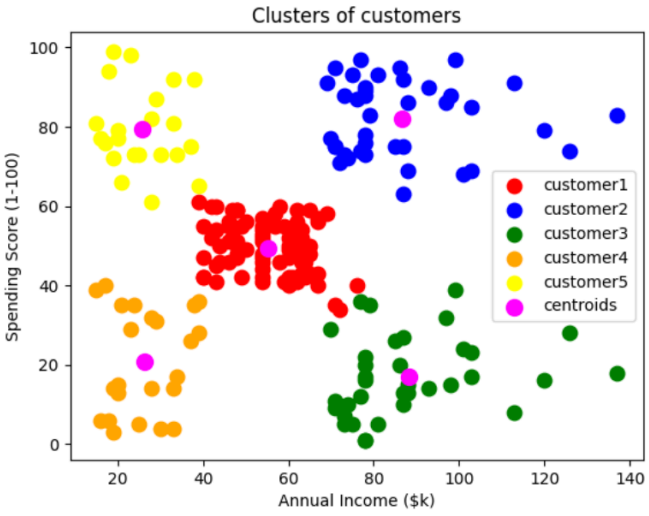
# 4. Model Performance & Improvement

### Model Evaluation (K-Means)

- **K-Means** was **faster and scalable** but required **manual K selection**.

# 5. Conclusion

This project successfully identified **five customer segments** based on spending behavior. The results can help businesses **target customers more effectively**, offer **personalized promotions**, and **optimize marketing strategies**. Further improvements can be made by **exploring advanced clustering techniques** and adding more features. The following clusters were discovered by this customer segmentation:




- **Red Cluster (customer1)**: Customers with **moderate income** and **moderate spending scores**. Likely **average shoppers**.

- **Blue Cluster (customer2)**: Customers with **high income** and **high spending scores**. Likely **premium customers** who spend a lot.

- **Green Cluster (customer3)**: Customers with **high income** but **low spending scores**. Likely **conservative spenders** or **financially cautious individuals**.

- **Orange Cluster (customer4)**: Customers with **low income** and **low spending scores**. Likely **budget-conscious shoppers**.

- **Yellow Cluster (customer5)**: Customers with **low income** but **high spending scores**. Likely **impulsive buyers or young consumers** who prioritize spending.

**Pink Points (Centroids)**

- These are the **cluster centroids**, representing the central point of each segment.
- They help businesses understand the typical spending behavior and income level of each customer group.