# Customer Segmentation Using Clustering

## ****1. Introduction****

Customer segmentation is the process of dividing customers into groups based on common characteristics. In this project, we use clustering techniques to segment customers based on **age, gender, annual income, and spending score**.

## ****2. Approach Used****

We used **K-Means clustering**, a popular unsupervised learning algorithm, to segment customers into five groups. The steps followed were:

**Data Preprocessing**:

* Loaded the dataset
* Converted categorical gender data into numerical values (0 for Male, 1 for Female)
* Selected relevant features for clustering

**Feature Scaling**:

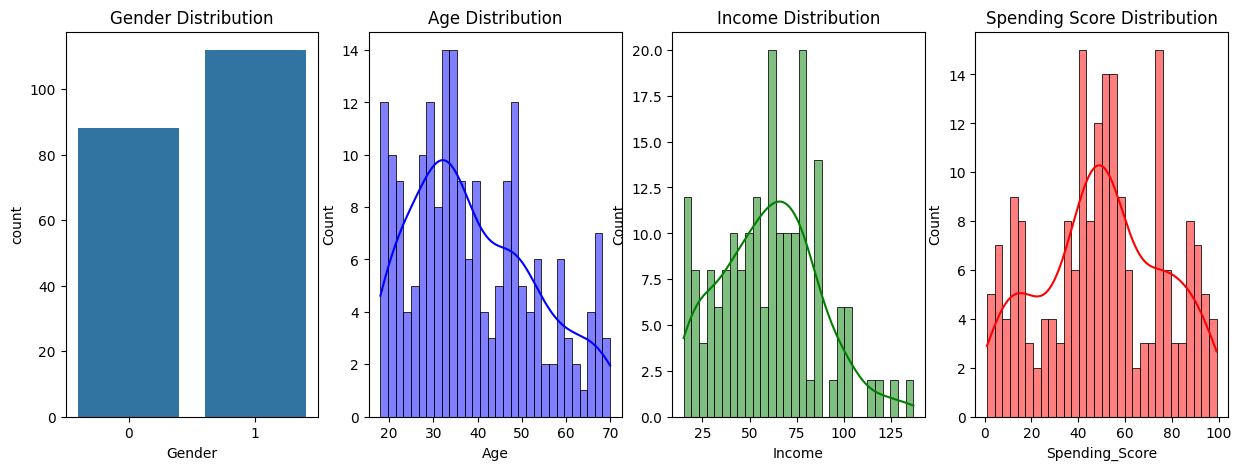
* Standardized numerical features (Age, Annual Income, Spending Score, and Gender) to ensure fair clustering

**Clustering Algorithm**:

* Used K-Means clustering with k=5 to form customer groups
* Determined the optimal number of clusters using the Elbow Method

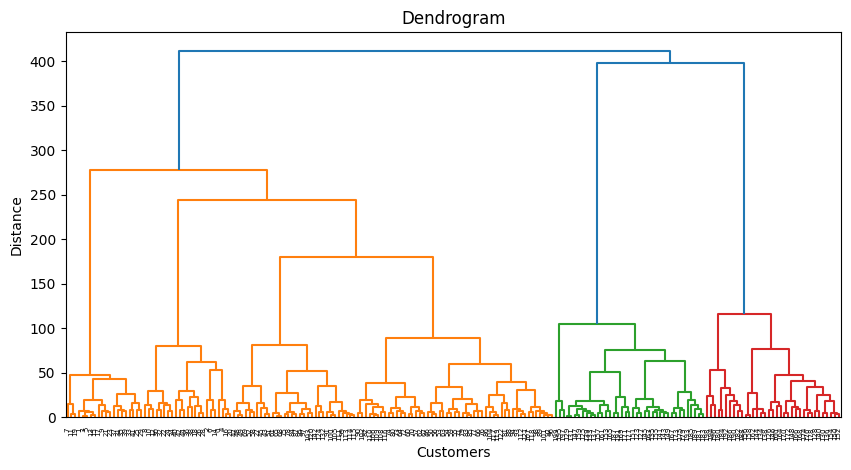
**Visualization**:

* Plotted histograms to show the distribution of Gender, **Age, Income, and Spending Score**



* Used scatter plots to visualize customer segments based on **Age vs Spending Score**, **Income vs Spending Score**, and **Gender Influence**

## 



## 

## ****3. Challenges Faced****

* **Choosing the right number of clusters**: Used the Elbow Method to determine k=5.
* **Feature scaling**: Raw data had different scales, so we standardized it.
* **Gender inclusion**: Initially, clustering was done without gender, but later it was included for better insights.

## ****4. Model Performance & Improvements****

The K-Means model effectively grouped customers into 5 segments.

**Possible Improvements**:

* Using **Hierarchical Clustering** for better interpretability
* Applying **DBSCAN** to detect outliers in data
* Adding **more features** (e.g. shopping frequency) for refined segmentation

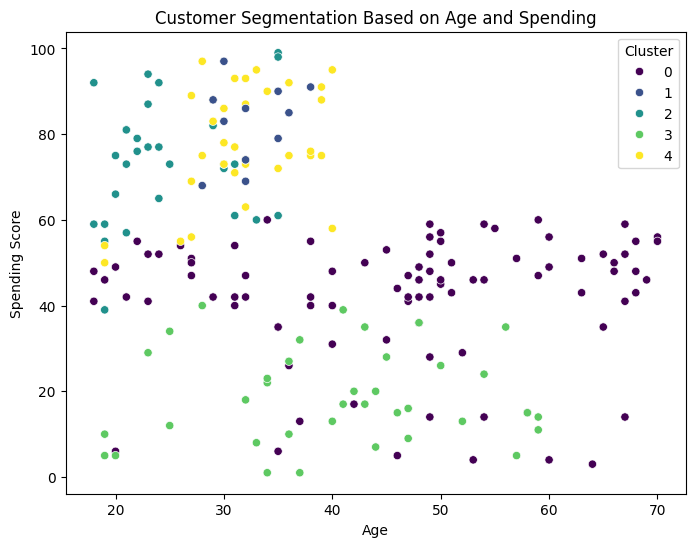
## ****5. Visualizations****

### ****1. Data Distributions****

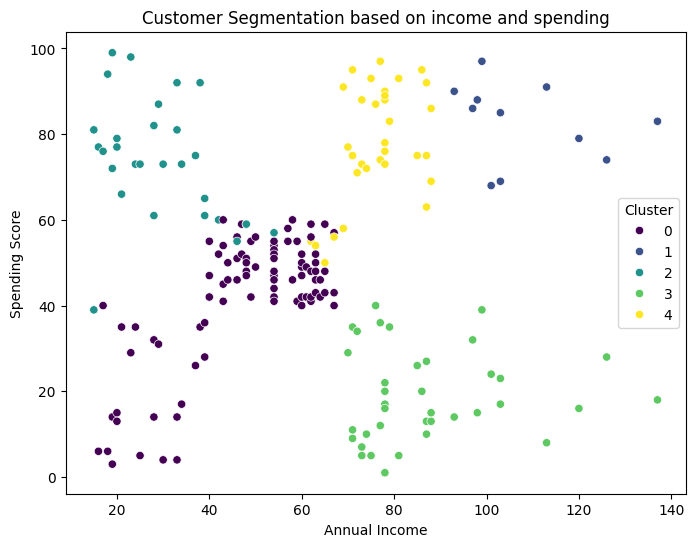
* **Gender, Age, Income, and Spending Score histograms** help understand the data spread.

### ****2. Customer Segments****

* **Age vs Spending Score**: Shows how spending habits vary with age.



* **Annual Income vs Spending Score**: Identifies high-income vs low-income spenders.



* **Gender Influence**: Helps analyze if gender impacts spending behavior.

