Name: Harshal Gavali

Roll No: 282017

PRN No: 22310324

**Assignment 5: Customer Segmentation Using Clustering**  
  
**Objective**  
  
The goal of this assignment is to identify profitable customer segments in a mall dataset using unsupervised learning techniques. Clustering algorithms are employed to group customers based on their spending behavior. The main focus is to analyze the Spending Score and other customer attributes to derive useful business insights.  
  
**Dataset Description**  
  
The dataset 'Mall\_Customers.csv' includes the following attributes:  
  
- Customer ID: Unique identifier for each customer.  
- Gender: Male or Female.  
- Age: Customer's age.  
- Annual Income (k$): Annual income in thousands.  
- Spending Score (1-100): Score assigned by the mall based on customer behavior and spending.  
  
**Steps Performed**  
  
1. Data Preprocessing:  
 - The dataset is loaded using pandas.  
 - Basic data cleaning, including checking for null values and understanding the data distribution, is performed.  
 - Exploratory Data Analysis (EDA) is conducted using matplotlib and seaborn to visualize distributions and relationships.  
  
2. Feature Selection:  
 - The features selected for clustering include Annual Income and Spending Score, as these are good indicators of purchasing behavior.  
  
3. Clustering Algorithms Applied:  
 - K-Means Clustering:  
 \* The Elbow Method is used to determine the optimal number of clusters (k).  
 \* Customers are grouped based on spending habits and income levels.  
 \* Clusters are visualized using a scatter plot to show separation.  
  
 - Hierarchical Clustering:  
 \* A dendrogram is plotted to visualize the hierarchy and decide the optimal number of clusters.  
 \* Agglomerative Clustering is used to form clusters from bottom-up merging.  
  
4. Evaluation:  
 - Clustering results are visualized to interpret the segments.  
 - Though clustering is unsupervised and doesn't use accuracy metrics directly, visual interpretation and business logic are used to validate the usefulness of clusters.  
  
5. Cross-Validation:  
 - Since unsupervised models don't use labeled data, traditional cross-validation doesn't apply directly.  
 - However, internal validation techniques like silhouette scores or consistency of clusters can be used.  
  
**Conclusion**  
  
By applying K-Means and Hierarchical Clustering on the mall customer dataset, we can identify distinct customer groups. This helps the mall management make strategic decisions such as personalized marketing, loyalty programs, and product placements targeting specific customer groups.