**Project Title: Customer Segmentation Using K-Means and Hierarchical Clustering**

**1. Introduction**

This project aims to segment customers using K-Means and Hierarchical clustering based on their annual income and spending score. The goal is to identify distinct customer groups that can be targeted with tailored marketing strategies and to compare the effectiveness of different clustering methods.

**2. Project Overview**

**2.1 Purpose**

The purpose of this project is to use K-Means and Hierarchical clustering to segment customers into meaningful groups. These segments can help businesses understand their customer base better and design targeted marketing campaigns.

**2.2 Scope**

The scope of the project includes:

* Data collection and preprocessing
* Application of K-Means clustering algorithm
* Determination of the optimal number of clusters using the Elbow Method and Silhouette Score
* Cluster profiling and analysis
* Application of Hierarchical clustering
* Comparison of K-Means and Hierarchical clustering
* Visualization of clusters using PCA

**3. Deliverables**

**3.1 Data Collection and Preprocessing**

* Collection of the Mall Customers dataset.
* Preprocessing of the dataset, including handling missing values and normalizing features.
* Documentation of the data preprocessing steps and code.

**3.2 K-Means Clustering**

* Implementation of the K-Means clustering algorithm using the annual income and spending score features.
* Determination of the optimal number of clusters using the Elbow Method and Silhouette Score.
* Training the K-Means model and assigning cluster labels to customers.
* Detailed documentation of the clustering process, including code and parameters used.

**3.3 Optimal Number of Clusters**

* Use the Elbow Method to determine the range of potential optimal clusters.
* Calculate Silhouette Scores for different numbers of clusters.
* Discuss the criteria for selecting the optimal number of clusters.

**3.4 Cluster Profiling and Insights**

* Analysis of the characteristics of each cluster.
* Provide insights into the customer segments based on their spending behavior and income levels.
* Documentation of the cluster profiling and insights.

**3.5 Hierarchical Clustering**

* Implementation of Hierarchical clustering using the same features.
* Comparison of the clusters formed with those obtained from K-Means.
* Discussion of the differences between the clustering methods.
* Documentation of the hierarchical clustering process, including code and parameters used.

**3.6 Visualizing Clusters with PCA**

* Application of PCA to the Mall Customers dataset to reduce its dimensionality.
* Visualization of the clusters from both K-Means and Hierarchical clustering in the PCA-reduced space.
* Documentation of the PCA process and visualizations.

**4. Dataset**

The dataset used for this project is the [Mall Customers Dataset](https://www.kaggle.com/datasets/shwetabh123/mall-customers) . This dataset contains various features of customers, including their annual income and spending score, which can be used for clustering.

**5. Project Partner**

* **Zaryab Ahmad Khan**
  + Role: Data collection, preprocessing, implementation of K-Means and Hierarchical clustering, determination of the optimal number of clusters, cluster profiling, PCA, visualization, analysis, and documentation.

**6. References**

* Mall Customers Dataset. Retrieved from [Kaggle.](https://www.kaggle.com/datasets/shwetabh123/mall-customers)