Clustering Analysis

## Understanding and Implementing K-Means, Hierarchical, and DBSCAN Algorithms

**Objective:**

The objective of this assignment is to introduce to various clustering algorithms, including K-Means, hierarchical, and DBSCAN, and provide hands-on experience in applying these techniques to a real-world dataset.

**Datasets :**

**Data Preprocessing:**

1. Preprocess the dataset to handle missing values, remove outliers, and scale the features if necessary.
2. Perform exploratory data analysis (EDA) to gain insights into the distribution of data and identify potential clusters.
3. Use multiple visualizations to understand the hidden patterns in the dataset

**Implementing Clustering Algorithms:**

* Implement the K-Means, hierarchical, and DBSCAN algorithms using a programming language such as Python with libraries like scikit-learn or MATLAB.
* Apply each clustering algorithm to the pre-processed dataset to identify clusters within the data.
* Experiment with different parameter settings for hierarchical clustering (e.g., linkage criteria), K-means (Elbow curve for different K values) and DBSCAN (e.g., epsilon, minPts) and evaluate the clustering results.

**Cluster Analysis and Interpretation:**

* Analyse the clusters generated by each clustering algorithm and interpret the characteristics of each cluster. Write you insights in few comments.

**Visualization:**

Visualize the clustering results using scatter plots or other suitable visualization techniques.

Plot the clusters with different colours to visualize the separation of data points belonging to different clusters.

**Evaluation and Performance Metrics:**

Evaluate the quality of clustering using internal evaluation metrics such as silhouette score for K-Means and DBSCAN.

Answer: <https://colab.research.google.com/drive/1_qQ5hBqtZgrkLbxS_znzjXweL4MmBrru?usp=sharing>