Assignment #4

Formatting:

If you're coding in a Jupyter Notebook, just submit your ipynb file with all the outputs. If you're using a py file, include all the outputs in a pdf along with your Python code.

Problem Statement:

You are provided with a dataset containing information about customer behavior on an e-commerce website. The dataset includes features such as customer ID, age, gender, annual income, and spending score. Your task is to perform an in-depth clustering analysis on the dataset and provide insights into customer segmentation based on their behavior.

Dataset:

Utilize the "Mall Customers" dataset.

Tasks:

1- Load the dataset into a Pandas DataFrame and perform exploratory data analysis (EDA) to understand the features and their distributions.

2- Preprocess the data if necessary, handling missing values and encoding categorical variables. Perform feature scaling if required to ensure that all features have the same scale. (10)

3- Apply three different clustering algorithms such as K-means, Agglomerative Hierarchical Clustering, and DBSCAN on the preprocessed data. (60)

4- Evaluate the performance of each clustering algorithm using appropriate metrics such as silhouette score, Davies-Bouldin index, or any other relevant metric. (30)

Grading Criteria:

Total: 100

1- 0
2- 10
3- 60
4- 30