📘 Instructions

You are provided with a dataset (force2020_data.csv) **with no target labels**.

🧠 Your Tasks

✅ Step 1: Load and Explore the Data

- Load the dataset using Pandas.
- Explore the structure and summary statistics of the dataset.
- Visualize the distribution of features using histograms or pair plots.

✅ Step 2: Preprocess the Data

- Standardize or normalize the dataset using StandardScaler or MinMaxScaler.

✅ Step 3: Apply Clustering Algorithms

Perform the following clustering methods:

🔹 **K-Means Clustering:**

- Run K-Means for a range of K values (e.g., 2 to 10).
- For each K, record the **Within-Cluster Sum of Squares (WCSS)** and **Silhouette Score**.
- Plot the **Elbow Curve** and **Silhouette Scores** to determine the optimal K.
- Visualize the clusters using a scatter plot (use any two features).

🔹 **Agglomerative Hierarchical Clustering:**

- Apply Agglomerative clustering with the optimal number of clusters (as found from silhouette analysis).
- Try different linkage methods (e.g., 'ward', 'complete', 'average') and compare performance using Silhouette Score.
- Plot the clustering result using scatter plot and color the clusters.

🔹 **DBSCAN:**

- Apply DBSCAN clustering.
- Experiment with different eps and min_samples values.
- Evaluate using Silhouette Score (note: it may be undefined if only one cluster is found).
- Visualize clusters with scatter plots.

✅ Step 4: Compare and Evaluate Clustering Algorithms

- Use **Silhouette Score** and **WCSS** to compare clustering performance.
- Summarize your observations in a markdown cell in the notebook:
    - Which algorithm worked best for your data?
    - How many clusters were found?

- Any issues like noise/outliers?

✅ **Step 6: Upload Your Work to GitHub**

1. Save your notebook (e.g., force2020_clustering_analysis.ipynb)
2. Create a GitHub repository (if you don't have one).
3. Upload the notebook file.
4. Copy the GitHub link to the notebook file.
5. Submit the link as your assignment.