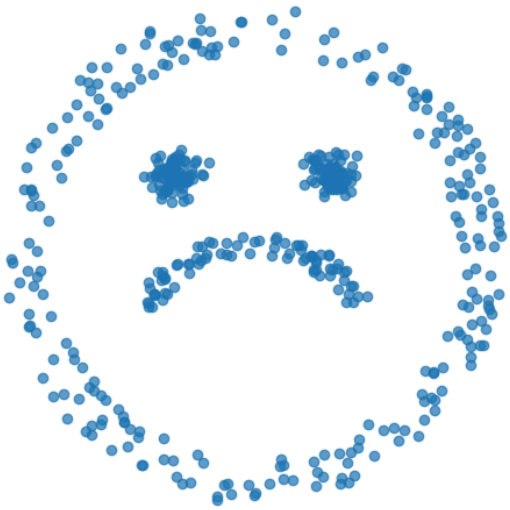
Task 0: Clustering the Smiley Dataset

You are given **smiley_dataset.csv**, which contains points that, when visualized, form a smiley face. Your task is to apply **both DBSCAN and K-Means clustering** on the dataset.

To visualize the dataset, use the provided code

```
csv_path = "smiley_dataset.csv"

smile = pd.read_csv(csv_path)

plt.figure(figsize=(6,6))

plt.scatter(smile['X'], smile['Y'], alpha=0.7)

plt.title('Smiley Face Dataset')

plt.xlabel('X')

plt.ylabel('Y')

plt.axis('equal')

plt.show()
```



From the image, we can observe **four clusters**—one for each **eye**, one for the **smile**, and one for the **outer set**.

Clustering Parameters:

- For **K-Means**, use **k = 4** (since we expect four clusters).
- For **DBSCAN**, use **eps = 0.3** and **min_samples = 5**.
- Experiment by **changing these parameters** in DBSCAN to understand their effect.

Key Concepts:

- `eps` **(epsilon)** defines the maximum **radius** around a point to consider neighbors for forming a cluster.
- `min_samples` specifies the **minimum number of points** required to form a dense region (core point).

Questions to Answer:

1. How do **K-Means and DBSCAN** handle the dataset differently?
2. Why might **DBSCAN** perform better on non-linear clusters?

Final Task:

- Write a post on **LinkedIn** explaining the differences between **K-Means and DBSCAN**, with your insights from this task.
- Tag and mention **GDGOC_Fast-Peshawar** page in your post.

Task 1: Optimize Sweet Item Placement for Eid Sales

For the occasion of Eid, the store owner wants to optimize the placement of sweet items.The goal is to use the FP-Growth algorithm to identify frequently bought sweet items so the store owner can strategically arrange them together for increased sales during Eid.

Steps to Follow:

1. Use the provided list of transactions representing customer purchases.
2. Apply the FP-Growth algorithm to find frequent itemsets.
3. Analyze the results to determine which sweet items are often bought together.
4. Explain how these insights can help the store owner optimize product placement for Eid sales.

List of Transactions (Eid Sweet Purchases)

Use the following list of transactions in your code:

```
transactions = [
      ["gulab jamun", "barfi", "jalebi"],
      ["gulab jamun", "laddu", "halwa", "kheer"],
      ["barfi", "jalebi", "laddu"],
      ["gulab jamun", "barfi", "laddu", "jalebi"],
      ["halwa", "kheer", "soan papdi"],
      ["gulab jamun", "barfi", "jalebi", "rasmalai"],
      ["barfi", "jalebi", "soan papdi", "peda"],
      ["laddu", "kheer", "barfi"],
      ["gulab jamun", "barfi", "jalebi", "rasmalai"],
      ["halwa", "soan papdi", "kheer"]
]
```

Questions to Answer:

1. What are the top frequent sweet itemsets found using FP-Growth?

2. How can the store owner arrange sweet items together to increase Eid sales?

3. Experiment by changing the min_support value and observe how it affects the frequent itemsets.

Task 2: Write a Medium Post on Evaluation Metrics.

Write a detailed Medium post explaining precision, accuracy, recall, and F1-score.
The post should cover:

1. Definitions & Formulas: Clearly define each metric with mathematical formulas.

2. Where to Use Each: Explain scenarios where accuracy, precision, recall, or F1-score should be prioritized.

3. Problem Statement & Trade-offs: Create a real-world problem where focusing on one metric over another could lead to incorrect conclusions.

4. Real-World Examples: Provide practical use cases to illustrate why certain metrics matter more in different cases.

Deliverables:

1. LinkedIn Post Link
2. Medium Post Link
3. Jupyter notebook