Exploratory Data Analysis (EDA) Report

1. Summary of the Dataset

- The dataset consists of 100 students with information on their **Marks**, **Attendance (%)**, and **Status (Pass/Fail)**.
- There are **missing values** in the "Marks," "Email ID," "Attendance (%)", and "Status" columns.
- The **Marks** range from **33 to 120**, with an **average of 75.85** and a **standard deviation of 20.01**.
- The **Attendance (%)** varies from **30 to 120**, with an **average of 79.43%**.

2. Key Statistical Findings

- **Marks Distribution:**
  - The **skewness** of marks is **-0.11**, indicating a slightly left-skewed distribution.
  - The **variance** is **400.21**, showing significant variation in student performance.
  - Some students have **marks above 100**, which may indicate **data entry errors**.
- **Categorical Variables:**
  - **Pass/Fail Status:**
    - **62 students passed**, while **35 failed**.
  - **First Names:**
    - Some names appear **more than once** (e.g., Xavier, Quinn, Zachary).

3. Correlation Analysis

- There is a **moderate positive correlation (0.67) between Marks and Attendance (%)**.
  - This suggests that students with higher attendance tend to perform better.

4. Grouped Comparisons (Pass vs. Fail)

Status	Average Marks	Average Attendance (%)
Fail	53.77	65.13
Pass	87.60	86.15

- **Students who passed had significantly higher marks and attendance compared to those who failed**.

5. Key Inferences

1. **Higher attendance correlates with better performance**, suggesting that **consistent attendance is crucial for academic success**.
2. **Some marks exceed 100**, which may require **data cleaning**.
3. **Attendance below 65% is associated with failing status**, indicating that **low attendance is a risk factor for failure**.
4. **There is some duplication in student names**, which might suggest **errors in data entry or duplicate records**.