Dataset Overview

- The dataset contains **150 students** with **10 columns**, including **Marks, Attendance (%), Status (Pass/Fail), Age, Subject, and City**.
- Some **missing values** are present in **Marks (10), Email ID (10), Attendance (%) (11), Status (11), Age (5), Subject (2), and City (3)**.

Exploratory Data Analysis (EDA) Report

1. Summary Statistics

- **Marks:**
  - **Mean:** 71.49, **Median:** 72, **Std Dev:** 20.41
  - **Min:** 31, **Max:** 120
  - **Skewness:** -0.03 (approximately normal distribution)
  - **Variance:** 416.74 (high variability in marks)
- **Attendance (%):**
  - **Mean:** 77.96, **Median:** 80, **Std Dev:** 16.36
  - **Min:** 30, **Max:** 120 (outlier detected)
- **Age:**
  - **Mean:** 21.39, **Min:** 18, **Max:** 25

2. Categorical Data Distribution

- **Status Distribution:**
  - **Pass:** 94 students
  - **Fail:** 45 students
  - More students have passed than failed.
- **Most Common First Names:**
  - **Top 3 Names:** David (9), James (7), Olivia (6)

3. Correlation Analysis

- **Marks vs Attendance (%) Correlation: 0.36** (moderate positive relationship)
  - Higher attendance is associated with higher marks.
- **Marks vs Age Correlation: -0.07** (negligible relationship)
- **Attendance vs Age Correlation: -0.11** (slight negative relationship)

4. Status-Based Comparisons

- **Pass Group:**
  - **Avg Marks:** 82.15
  - **Avg Attendance:** 81.51%
- **Fail Group:**
  - **Avg Marks:** 49.34
  - **Avg Attendance:** 69.95%
  - Students who failed had significantly lower marks and attendance.

Key Takeaways

1. **Attendance and Marks are positively correlated**—students with better attendance tend to score higher.
2. **Failing students have lower average attendance and marks**, suggesting attendance may impact performance.
3. **Some outliers exist in the dataset**, especially in marks and attendance, which may need further cleaning.
4. **Age has little impact on marks or attendance.**