**Mini Project – Python Libraries**

**Project - Student Performance and Attendance Analysis**

Analyse student performance and attendance using attached excel file which has 2 tabs in the document.



The goal is to:

1. Calculate total and percentage marks.
2. Derive attendance percentages.
3. Classify students based on performance.
4. Identify students with low attendance and highlight top performers.
5. Provide visual insights into the data.

**Assignment Questions**

**Data Preparation**

1. Load the student marks and attendance data from the provided Excel files into two separate Pandas DataFrames.
2. Merge the two DataFrames on the Name column to create a single DataFrame.
3. Replace attendance values ('Y' or 'N') with numeric values (1 for 'Y' and 0 for 'N').
4. Deal with the missing values.
5. Do the column cleaning like name column is the combination of upper and lower case.

**Data Transformation**

1. Create separate columns for below:
   * Total marks for each student.
   * Percentage marks for each student.
   * Attendance percentage for each student based on attendance columns.
   * Calculate the weighted percentage – Attendance(40%), Mini test01(10%), Mini test02(10%), Live test(20%), Assignment (20%).
   * Use this weighted percentage for further calculations.
2. Classify each student's performance into categories:
   * "Excellent" for percentages ≥ 85.
   * "Good" for percentages between 71 and 84.
   * "Average" for percentages between 50 and 70.
   * "Needs Improvement" for percentages < 50.

**Analysis**

1. Identify students with attendance below 75% but weighted percentage >50%.
2. Highlight the top three students based on percentage marks.
3. Impact of attendance on Tests/Assignment marks.

**Visualization**

1. Create a **bar chart** displaying weighted percentages for top 5 students.
2. Create a **pie chart** showing the distribution of students across the four performance categories.
3. Create **box plots** for each test (Live Test, Mini Test 1, Mini Test 2, Assignment) to visualize the spread and detect potential outliers in scores.
4. Create a **chart** to show the students where attendance is less than 50%.
5. Any other visualization/analysis which you can infer from data for the management.