DATA ANALYTICS AND VISULIZATION

LAB ASSIGNMENT WEEK-1:

HTNO:2403A52332

NAME:BOLLAM RISHIKA

OBJECTIVE: An academic coordinator maintains student marks in Excel. You will first analyze the data using Excel, and then clean and transform it using Pandas in Python.

Tasks in Excel:

- Calculate total and average marks for each student.
- Highlight students scoring below 35 using conditional formatting.

Tasks in Pandas:

- Load the Excel file using pd.read_excel().
- Remove students with missing roll numbers or names.
- Fill missing marks using subject-wise mean.

CODE:

```
from google.colab import files
uploaded = files.upload()
import pandas as pd
df = pd.read_excel("student_marks.xlsx")  # Make sure the name matches exactly
df_cleaned = df.dropna(subset=["ROLLNO", "NAME"])
subjects = ["MATHS", "SCIENCE", "ENGLISH"]
for subject in subjects:
    mean_val = df_cleaned[subject].mean()
    df_cleaned[subject] = df_cleaned[subject].fillna(mean_val)
    print(df_cleaned)
```

```
Choose files  student_marks.xlsx
• student_marks.xlsx(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 9884 bytes, last modified: 05/09/2025 - 100% done
Saving student_marks.xlsx to student_marks (3).xlsx
   ROLLNO     NAME  MATHS  SCIENCE  ENGLISH    SUM        AVG
0     1.0   GANESH   99.0       90     95.0  285.0  94.666667
1     2.0  RISHIKA   98.0       89     80.0  267.0  89.000000
2     3.0  RISHIKA   97.0       88     85.0  270.0  90.000000
3     4.0     ABHI   95.0       84     76.0  255.0  85.000000
4     5.0      RAM   99.0       92     97.0  288.0  96.000000
5     6.0  KRISHNA   98.0       93     95.0  286.0  95.333333
6     7.0    RADHA   98.0       93     95.0  286.0  95.333333
7     8.0    VIJAY   92.0       66     86.0  244.0  81.333333
8     9.0     RAVI   80.0       78     85.0  243.0  81.000000
9    10.0     AJAY   78.0       88     79.0  245.0  81.666667
```

REFLECTION:

You used **Pandas** to automate data cleaning, which is a key skill in data science and academic analytics.

- You learned how to handle **missing data**, a common real-world issue, by filling gaps with meaningful values (subject-wise averages).