✅ Task 10: Python EDA — Summary + Outlier Detection

Tools:

- Primary: Google Colab
- Libraries: pandas, numpy, matplotlib
- Alternative: Kaggle Notebook / Jupyter

Dataset:

- House Prices
- Students Performance
- Credit Card Fraud (small sample)

Hints / Mini Guide:

1. Load dataset and check .shape, .info(), .head().
2. Generate descriptive statistics using .describe().
3. Compute missing value % for each column.
4. Plot distributions (histogram + boxplot).
5. Detect outliers using IQR method.
6. Create outlier flag column.
7. Handle outliers (remove/cap) with reason.
8. Create correlation matrix and interpret top correlations.
9. Export cleaned dataset.

Deliverables:

- task10_eda.ipynb
- cleaned_dataset.csv
- eda_findings.txt

Final Outcome:
✅ Practical EDA skill + outlier handling.

Interview Questions Related To Above Task:

- Why EDA is important?
- Explain IQR outlier detection.
- What is skewness?
- Correlation vs causation?
- Why not remove all outliers always?

📌 **Task Submission Guidelines**

- ⏰ **Time Window:**

Youcan complete the task anytime between 10:00 AM to 10:00 PM on the given day. Submission link closes at 10:00 PM

- 🔍 **Self-Research Allowed:**

Youare free to explore, Google, or refer to tutorials to understand concepts and complete the task effectively.

- 🛠️ **Debug Yourself:**

Tryto resolve all errors by yourself. This helps you learn problem-solving and ensures you don't face the same issues in future tasks.

- 💸 **No Paid Tools:**

Ifthe task involves any paid software/tools, do not purchase anything. Just learn the process or find free alternatives.

- 📁 **GitHub Submission:**

Create a new GitHub repository for each task.
Add everything you used for the task — code, datasets, screenshots (if any), and a short README.md explaining what you did.

📤 **Submit Here:**

After completing the task, paste your GitHub repo link and submit it using the link below:

- 👉 [Submission Link]