Task 3: Exploratory Data Analysis (EDA)

Tools:

- Python (Matplotlib, Seaborn, Pandas)
- Alternatives: Tableau Public

Dataset:

- "Netflix Movies and TV Shows"
- "Iris Dataset"

Hints / Mini Guide:

1. Plot distribution of numerical features using histograms.
2. Analyze categorical features using count plots.
3. Use box plots to identify outliers.
4. Plot correlation heatmap to understand feature relationships.
5. Write insights for each visualization.
6. Identify features important for prediction.
7. Summarize findings in bullet points.

Deliverables:

- EDA Notebook
- Visual insights report

Final Outcome:

Intern understands data patterns and feature behavior.

Interview Questions Related To Above Task:

- What is EDA?
- What is correlation?
- Why detect outliers?
- What is multicollinearity?
- Why visualization is important?

📌 **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]