Task 1: Understanding Dataset & Data Types

Tools:

- Python (Pandas, NumPy)
- Jupyter Notebook / Google Colab
- Alternatives: VS Code + Python, Kaggle Notebooks

Dataset:

- "Titanic Dataset"
- "Students Performance Dataset"

Hints / Mini Guide:

1. Load the dataset using Pandas and display first and last few records to understand the structure of rows and columns.
2. Identify numerical, categorical, ordinal, and binary features manually by inspecting column names and values.
3. Use df.info() and df.describe() to understand data types, null values, and statistical summaries.
4. Check unique values in categorical columns to understand data distribution.
5. Identify target variable and input features for ML suitability.
6. Analyze dataset size and discuss whether it is suitable for machine learning.
7. Write clear observations about data quality issues like missing values or imbalance.

Deliverables:

- Jupyter Notebook
- Dataset analysis report (1 page)

Final Outcome:

Intern understands data structure, types, and ML readiness of datasets.

Interview Questions Related To Above Task:

- Difference between numerical and categorical data?
- What is a target variable?
- Why is data understanding important before modeling?
- What is data imbalance?
- What does df.describe() show?

📌 **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]