Task 4: Feature Encoding & Scaling

Tools:

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

Dataset:

- "Adult Income Dataset"

Hints / Mini Guide:

1. Identify categorical and numerical features.
2. Apply Label Encoding where order exists.
3. Apply One-Hot Encoding where order does not exist.
4. Scale numerical features using StandardScaler.
5. Compare model readiness before and after scaling.
6. Explain impact of scaling on ML algorithms.
7. Save processed dataset.

Deliverables:

- Preprocessed dataset
- Notebook

Final Outcome:

Intern understands feature engineering basics.

Interview Questions Related To Above Task:

- Label vs One-Hot encoding?
- Why scaling is required?
- What is normalization?
- Which algorithms need scaling?
- What is feature engineering?

📌 **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]