✅ **Task 7: Logistic Regression – Titanic Survival Prediction**

Dataset:
- Primary: Kaggle Titanic Dataset
- Alternative: Seaborn Titanic – load_dataset('titanic')

Tools:
- Python, Pandas
- Scikit-learn
- Matplotlib / Seaborn
- Alternatives: Kaggle Notebook, RapidMiner (free)

Hints / Mini Guide:

1. Load Titanic dataset and review important features like Age, Sex, Fare, and Survived.
2. Handle missing values in Age and Embarked using median/mode strategies.
3. Remove unnecessary columns like PassengerId or Name if they do not help prediction.
4. Encode categorical features using OneHotEncoding to convert them into numeric form.
5. Apply StandardScaler on numeric columns like Fare and Age for stable training.
6. Split dataset into train-test sets ensuring class distribution stays similar.
7. Train Logistic Regression model and generate predictions on test data.
8. Evaluate using accuracy, precision, recall, F1-score, and confusion matrix.
9. Plot ROC curve and calculate AUC score for more reliable evaluation.

Deliverables:

- Notebook
- Confusion matrix image
- ROC curve + AUC score

Final Outcome:
- Intern learns binary classification and proper performance evaluation.

Interview Questions Related To Above Task:

- Why Logistic Regression is used for classification?
- What is precision and recall?
- What is ROC curve?
- What does AUC mean?
- What is confusion matrix?

📌 **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]

⭐⭐⭐⭐⭐