

Task 5: Exploratory Data Analysis (EDA)

- **Objective:** Extract insights using visual and statistical exploration.
- **Tools:** Python (Pandas, Matplotlib, Seaborn)
- **Deliverables:** Jupyter Notebook + PDF report of findings
- **Hints/Mini Guide:**
 - a. Use `.describe()`, `.info()`, `.value_counts()`
 - b. Use `sns.pairplot()`, `sns.heatmap()` for visualization
 - c. Identify relationships and trends
 - d. Plot histograms, boxplots, scatterplots
 - e. Write observations for each visual
 - f. Provide summary of findings

Dataset: TITANIC DATASET([CLICK FOR LINK](#)) / OR ANY RELEVANT DATA SET OF UR OWN.

Outcome: Gain skill in finding patterns, trends, and anomalies.

Interview Questions:

1. What is EDA and why is it important?
2. Which plots do you use to check correlation?
3. How do you handle skewed data?
4. How to detect multicollinearity?
5. What are univariate, bivariate, and multivariate analyses?
6. Difference between heatmap and pairplot?
7. How do you summarize your insights?

📌 Task Submission Guidelines

- 🕒 **Time Window:**

You can 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:**

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

- 🔧 **Debug Yourself:**

Try to 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:**

If the 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

