

Task 1: Data Cleaning & Preprocessing

- **Objective:** Learn how to clean and prepare raw data for ML.
- **Tools:** Python, Pandas, NumPy, Matplotlib/Seaborn

Hints/Mini Guide:

1. Import the dataset and explore basic info (nulls, data types).
2. Handle missing values using mean/median/imputation.
3. Convert categorical features into numerical using encoding.
4. Normalize/standardize the numerical features.
5. Visualize outliers using boxplots and remove them.

Dataset: You can use any dataset relevant to the task, e.g., Titanic Dataset

link to download: [click here to download dataset](#)

What You'll Learn: Data cleaning, handling nulls, encoding, feature scaling.

Interview Questions:

1. What are the different types of missing data?
2. How do you handle categorical variables?
3. What is the difference between normalization and standardization?
4. How do you detect outliers?
5. Why is preprocessing important in ML?
6. What is one-hot encoding vs label encoding?
7. How do you handle data imbalance?
8. Can preprocessing affect model accuracy?

Submit Here:

After completing the task, paste your GitHub repo link and submit it using the link below:

-  [\[Submission Link\]](#).

📌 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:

- 👉 [\[Submission Link\]](#).

Best
of
Luck

