

# Titanic Data Cleaning & Preprocessing

This repository contains a complete Python script for cleaning and preprocessing the Titanic dataset for Machine Learning tasks using Pandas, NumPy, Matplotlib, and Scikit-learn.

## **Objective:**

Prepare raw Titanic data for training ML models by:

- Handling missing values
- Encoding categorical features
- Normalizing numerical columns
- Detecting and removing outliers

## **Tools & Libraries:**

- Python
- Pandas
- NumPy
- Matplotlib
- Scikit-learn

## **Files:**

- Task1.py : Full preprocessing script
- titanic.csv : Titanic dataset file
- README.pdf : Project documentation

## **How to Run:**

1. Make sure titanic.csv is in the same folder as the script.
2. Run the script:

Task1.py

## **Notes**

The dataset used can be downloaded from:

<https://www.kaggle.com/competitions/titanic/data>

## **Output:**

The script outputs a cleaned DataFrame ready for ML model training.