# Titanic Dataset Preprocessing

This repository contains the preprocessing work done on the Titanic dataset.

## 📊 Dataset

Datasource : CSV file downloaded

## ✅ Tasks Completed Today (Day 1)

- Loaded the Titanic dataset using pandas.

- \*\*Handled missing values\*\*:

- Filled missing values in the `Age` column using the median.

- Filled missing values in the `Embarked` column using the most frequent value.

- \*\*Dropped the `Cabin` column\*\* due to too many missing values.

- \*\*Label Encoding\*\*:

- Encoded the `Sex` column (male/female → 0/1).

- \*\*One-Hot Encoding\*\*:

- Applied one-hot encoding to the `Embarked` column (now represented as `Embarked\_Q`, `Embarked\_S`).

- \*\*Feature Scaling\*\*:

- Standardized the `Age` and `Fare` columns using `StandardScaler`.

- \*\*Outlier Removal\*\*:

- Removed outliers from `Age` and `Fare` columns using the IQR method.

## 📁 Files

- `titanic\_preprocessing.py`: Python code with all preprocessing steps.

- `titanic.csv`: Dataset used.

- `README.md`: Documentation (this file).

## 🧠 Tools Used

- Python

- Pandas

- Seaborn

- Matplotlib

- sklearn.preprocessing

-StandardScaler

- Jupyter Notebook

---