

# Data Cleaning & Preprocessing Report

## 1. Objective

Learn how to clean and prepare raw data for machine learning using Python, Pandas, NumPy, and Seaborn.

## 2. Dataset Used

Titanic Dataset (commonly used for classification problems in ML).

## 3. Steps Performed

1. Imported the dataset and explored data types, nulls, and descriptive statistics.

2. Handled missing values:

- Filled 'Age' with median.
- Filled 'Embarked' with mode.
- Dropped 'Cabin' due to excessive missing values.

3. Encoded categorical features:

- Converted 'Sex' using Label Encoding.
- Applied One-Hot Encoding on 'Embarked'.

4. Standardized numerical columns ('Age' and 'Fare') using StandardScaler.

5. Detected and removed outliers using Z-score method.

6. Visualized cleaned data:

- Boxplots for 'Age' and 'Fare'.
- Histograms to show distributions.
- Correlation heatmap using numeric columns only.