

# EDA Report - Titanic Dataset

## 1. Basic Information

Rows: 891

Columns: 12

Features:

- Numerical: Age, SibSp, Parch, Fare, PassengerId, Pclass
- Categorical: Sex, Embarked, Cabin, Ticket, Name

## 2. Missing Values

- Age: 177 missing values
- Cabin: 687 missing values- Embarked: 2 missing values

Action Planned:

- Fill 'Embarked' with mode.
- Fill 'Age' with median.
- Drop 'Cabin' due to heavy missing data.

## Analysis

- Age: Mostly between 20 to 40 years, right-skewed.
- Females had a higher survival rate than males.
- 1st class passengers had higher survival chances than 3rd class.
- Younger passengers had better survival rates.
- Passengers who paid higher fares had better survival rates.

### **3. Correlation Insights**

- Pclass is negatively correlated with survival (lower class, less chance).
- Fare is positively correlated with survival.

### **4. Key Insights & Patterns**

- Women and children had survival priority.
- Higher ticket classes ensured better chances of survival.