

TITANIC DATASET - EXPLORATORY DATA ANALYSIS (EDA) REPORT

1. DATASET INFORMATION

- Various columns: PassengerId, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked
 - Includes both numerical and categorical data
 - Some missing values in columns like Age, Cabin, and Embarked
-

2. DESCRIPTIVE STATISTICS

- Age: Mean ~30, Std Dev ~14, Min = 0.42, Max = 80
 - Fare: Right-skewed distribution, with outliers
 - Pclass: Mostly Class 3 passengers
-

3. MISSING VALUES SUMMARY

- **Age:** Missing values
 - **Cabin:** Majority missing
 - **Embarked:** Few missing
-

4. VALUE COUNTS FOR CATEGORICAL FEATURES

- **Sex:** ~65% male, 35% female
 - **Embarked:** Majority from port 'S'
 - **Pclass:** Most in 3rd class
-

5. VISUALIZATIONS

(Include these plots in your report)

- **Histogram:** Age distribution
 - **Boxplot:** Age by Pclass
 - **Scatterplot:** Fare vs Age by Sex
 - **Heatmap:** Correlation among numerical features
-

6. KEY INSIGHTS

- Older passengers are more common in higher classes.
- Fare is highly variable; many low-fare passengers, few high-paying outliers.
- Age and Fare are weakly correlated.
- Males dominated the passenger list.
- Pclass and Fare are negatively correlated.