

Titanic Dataset Analysis - Summary and Observations

Summary of Findings

Summary of Findings:

1. Survival Rate by Class:

- 1st Class: Highest survival rate (~63%).
- 2nd Class: Moderate survival rate (~47%).
- 3rd Class: Lowest survival rate (~24%).

2. Survival and Gender:

- Female passengers had a much higher survival rate than males.
- Most survivors were women and children.

3. Age:

- The average age of survivors is slightly lower than non-survivors.
- Younger passengers were more likely to survive.

4. Fare:

- Passengers who paid higher fares were more likely to survive.
- Most passengers paid low fares; very few paid extremely high fares.

5. Correlation Heatmap:

- Strong positive correlation between survival and being female.
- Fare and Pclass also show moderate correlation with survival.

6. Logistic Regression Confusion Matrix:

- The model was able to predict survival fairly well with selected features.
- Indicates basic features like Pclass, Age, Sex, and Fare hold predictive power.

Visual Observations

Visual Observations:

1. Heatmap of Missing Values:

- Significant missing data in 'Cabin' column.
- Few missing values in 'Age', filled using median.

2. Pairplot:

- Clear relationship between Pclass, Sex, and Survival.
- Females and 1st Class passengers clustered among survivors.

3. Boxplot of Age vs Survival:

- Survivors have slightly lower median age compared to non-survivors.

4. Histogram of Fare:

- Skewed distribution; most passengers paid less than 100.

5. Scatterplot (Age vs Fare):

- Survivors cluster in regions with higher fare and lower to middle age.

6. Correlation Heatmap:

- 'Sex' (female), 'Fare', and lower 'Pclass' are positively associated with survival.