

Titanic Dataset – Exploratory Data Analysis Report

Prepared by: Samidha Koli

Summary of Findings:

- The dataset contains information on Titanic passengers such as age, fare, passenger class, and survival status.
- Several columns (such as Age and Cabin) contain missing values, and approximately 38% of passengers survived.
- The age distribution is skewed toward younger passengers, with most passengers between 20 and 40 years old.
- Boxplot (Age vs Survived) shows survivors have a slightly lower median age, but survival occurs across all ages.
- Scatterplot (Age vs Fare) shows no clear linear relationship between age and fare.
- Pairplot confirms higher fares and lower passenger class (1st and 2nd) are associated with higher survival rates.
- Heatmap shows Fare is positively correlated with Survived, while Pclass is negatively correlated with Survived.

Conclusion:

Passengers who paid higher fares and were in higher classes (1st/2nd) had a much higher chance of survival, while age alone did not significantly influence the outcome.