

Summary of findings

Title - Exploratory data analysis on Public dataset

Dataset used - Titanic Dataset from Kaggle

This analysis explored the Titanic dataset to understand the factors linked to passenger survival. Before examining patterns, the data was cleaned by filling missing values in the Age and Embarked columns and removing the Cabin column due to excessive missing entries. Duplicate rows were also eliminated to ensure accuracy.

A simple look at individual features showed that most passengers were young adults, and there were more males than females on board. When survival patterns were compared, clear differences appeared. Female passengers had a much higher chance of surviving than males. Younger passengers also showed slightly better survival outcomes compared to older individuals. The passenger class played an important role as well: people traveling in the first class had noticeably higher survival rates than those in lower classes. Fare values supported this trend, as passengers who paid higher fares were more likely to survive.

The correlation heatmap confirmed these relationships, showing positive links between fare and survival and a negative link between passenger class and survival. Overall, the analysis suggests that gender, class, and economic status were major factors influencing who survived the Titanic disaster.