

Titanic Dataset Analysis Report

Survival Count

Observation: More passengers did not survive (0) than survived (1).

Gender Distribution

Observation: There are more male passengers than female passengers.

Passenger Class Distribution

Observation: Most passengers belong to 3rd class, fewer in 1st and 2nd class.

Age Distribution

Observation: Most passengers are between ages 20 and 40.

Survival by Gender

Observation: Females had a higher survival rate than males.

Survival by Passenger Class

Observation: Passengers in 1st class had higher survival rates compared to 2nd and 3rd classes.

Age vs Survival

Observation: Survivors tend to be younger on average, with fewer elderly survivors.

Survival by Title

Observation: Titles like 'Mrs' and 'Miss' have higher survival counts compared to 'Mr'.

Titanic Dataset Analysis Report

Survival by Deck

Observation: Passengers on decks B, C, and n show higher survival rates, but many decks have unknown data.

Survival by Port of Embarkation

Observation: Passengers who embarked from port 'S' had better survival chances.

Correlation Heatmap

Observation: 'Fare' and 'Pclass' are negatively correlated; 'Fare' shows slight positive correlation with 'Survived'.

Pairplot

Observation: Survivors tend to pay higher fares and belong more often to 1st class.

Age vs Fare Scatterplot

Observation: Survivors tend to cluster at higher Fare values across ages.

Final Summary of Findings

Summary of Findings:

- Survival was strongly influenced by gender, with females having a much higher survival rate.
- Passenger class mattered; first-class passengers had better chances of survival than those in lower classes.
- Age also played a role; younger passengers were more likely to survive.
- Titles extracted from names reflected social status, affecting survival chances (e.g., 'Mrs' and 'Miss' had higher survival).

Titanic Dataset Analysis Report

- Deck and embarkation port had some impact, but data missingness limits strong conclusions.
- Fare paid correlated with survival, indicating wealth or class advantage.
- The overall dataset has some missing values that were addressed via median or mode imputation.