

Titanic Dataset - EDA Report

1. Dataset Overview

The Titanic dataset contains information about passengers including:

- Survival status (0 = No, 1 = Yes)
- Passenger class (1st, 2nd, 3rd)
- Name, Sex, Age
- Number of siblings/spouses aboard
- Number of parents/children aboard
- Ticket, Fare, Cabin, and Embarkation Town.

2. Statistical Summary

The dataset has a mix of categorical and numerical variables.

- Mean Age: ~29.7 years
- Mean Fare: ~32.2
- Survived rate: Around 38%
- Missing Values: Present in 'age', 'embarked', 'deck', 'embark_town'.

3. Visual Findings

- Females had a much higher survival rate compared to males.
- First-class passengers had a better chance of survival compared to second and third classes.
- Younger passengers, particularly children, were more likely to survive.
- Strong correlations observed between 'fare', 'pclass', and 'survived'.

4. Key Trends

- Survival was influenced by sex, class, and age.
- Higher fare was positively associated with survival probability.
- Third-class passengers had the lowest survival rates.

5. Conclusion

Titanic Dataset - EDA Report

Gender, socioeconomic status, and age played critical roles in survival probability during the Titanic disaster. Future modeling and predictive analytics can leverage these features for accurate survival predictions.