

TITANIC – Exploratory Data Analysis Report

1. Introduction

The Titanic dataset contains passenger details used to understand survival patterns. This report summarises data cleaning, visualisation, and key insights.

2. Dataset Overview

- Rows: 891
- Columns: 12
- Important features: Age, Sex, Pclass, Fare, Embarked, Survived

3. Data Cleaning

- Age missing values filled with *median*
- Embarked filled with *mode*
- Cabin dropped due to 687+ missing values

4. Visual Analysis

✓ Survival Rate

More passengers died than survived.

✓ Gender vs Survival

Females survived significantly more than males.

✓ Pclass vs Survival

1st class > 2nd class > 3rd class survival.

✓ Age Distribution

Most passengers were young adults (20–40).

✓ Heatmap

Strong correlation between **Fare** and **Survival**.

5. Final Insights

1. Women had the highest probability of survival.
2. Higher social class significantly improved survival chances.
3. High ticket fare ↗ survival chances.
4. Children survived more often than elders.
5. Large families had lower survival probability.

6. Conclusion

The Titanic dataset clearly shows that **gender, class, and ticket fare played a major role** in determining survival.