

Exploratory Data Analysis on the Titanic Dataset

Objective:

The goal of this analysis is to uncover insights, patterns, and trends from the Titanic dataset using statistical summaries and data visualizations. We aim to identify factors influencing survival rates on the Titanic.

Dataset Overview:

Total Records: 891

Columns: 12

Missing Values:

- Age: 177 missing
- Cabin: 687 missing (dropped due to sparsity)
- Embarked: 2 missing (can be filled or dropped)

No Duplicate Rows

Initial Observations:

The dataset contains demographic and travel information such as age, sex, fare, passenger class, etc.

Survival rate is encoded in the Survived column (0 = No, 1 = Yes).

Statistical Summary:

Feature	Mean	Std Dev	Min	25%	50%	75%	Max
---------	------	---------	-----	-----	-----	-----	-----

Age	29.70	14.53	0.42	20.13	28.00	38.00	80.00
-----	-------	-------	------	-------	-------	-------	-------

Exploratory Data Analysis on the Titanic Dataset

Fare 32.20 49.69 0.00 7.91 14.45 31.00 512.33

SibSp 0.52 1.10 0 0 0 1 8

Parch 0.38 0.81 0 0 0 0 6

Visual Analysis & Findings:

- Gender Distribution:

Males were more numerous than females on board.

- Age Distribution:

Most passengers were between 20 and 40 years old.

- Fare Distribution:

Fare values are right-skewed with some extreme outliers.

- Survival Rate by Gender:

Females had significantly higher survival rates than males.

- Age vs. Survival:

Survivors tend to be slightly younger on average.

- Correlation Heatmap:

Survival is moderately correlated with Fare and Pclass.

SibSp and Parch show minor correlation with survival.

Exploratory Data Analysis on the Titanic Dataset

- Pairplot Analysis:

Shows clusters in age-fare-pclass-survived space.

Data Cleaning Steps:

- Filled missing Age values with the median age.
- Dropped Cabin column due to excessive missing values.

Summary of Key Findings:

- Pclass: Most passengers were in 3rd class, who had lower survival rates.
- Sex: Females had a significantly higher survival rate.
- Age: Younger passengers had higher chances of survival.
- Fare: Higher fare-paying passengers (likely 1st class) had better survival chances.
- Correlation: Strongest correlations with survival were Sex, Pclass, and Fare.

Conclusion:

The analysis shows that socio-economic status (Pclass and Fare) and gender played a major role in survival rates on the Titanic. This aligns with the historical narrative that 'women and children first' was followed during evacuation.