

## Objective:

To extract insights from the Titanic dataset using visual and statistical exploration.

## Tools Used:

Python, Pandas, Matplotlib, Seaborn

### 1. Dataset Overview

Dataset: Titanic Dataset (Kaggle)

Total Rows: 891

Key Columns: Survived, Pclass, Sex, Age, SibSp, Parch, Fare, Embarked

#### Missing Values:

Age: ~19% missing

Cabin: >75% missing

Embarked: 2 missing values

### 2. Statistical Summary

Mean Age: ~29.7 years

Fare ranged from 0 to 512

#### Survival Rate:

Survived (1): ~38%

Did Not Survive (0): ~62%

### 3. Key Observations from Visualizations

#### i) Pairplot:

Females and younger passengers show a higher survival cluster.

Passengers who paid higher fares were more likely to survive.

#### ii) Heatmap (Correlation Matrix):

Sex (female) and Fare showed positive correlation with Survived.

Pclass had a moderate negative correlation with Survived (lower class, lower survival chance).

#### iii) Histogram (Age Distribution):

Most passengers were aged between 20 and 40 years.

Distribution skewed slightly towards younger ages.

iv) Boxplot (Age vs Survived):

Survivors had a lower median age than non-survivors.

More variation in age among those who died.

v) Scatterplot (Age vs Fare):

Survivors cluster around lower age and higher fare.

Passengers with higher fare often belonged to 1st class, influencing survival.

#### 4. Summary of Findings

Factor	Impact on Survival
Sex	Females had a significantly higher survival rate
Pclass	1st class passengers had better chances of survival
Age	Younger passengers were more likely to survive
Fare	Higher fare correlated with survival
Family	Moderate effect from SibSp and Parch (being alone reduced chances slightly)

#### 5. Conclusion

The survival on Titanic was not random; it was influenced by:

Social status (Pclass)

Gender

Age

Ticket price (Fare)