

# Summary Report: Analysis of the Titanic Dataset

## 1. Introduction:

In this report, we analyze the Titanic dataset to understand the demographics of passengers aboard the Titanic and explore factors influencing survival. The dataset contains information about passengers including their age, gender, ticket class, fare, and survival status.

## 2. Dataset Overview and Initial Observations

- i. The Titanic dataset consists of the following columns:
- ii. PassengerId: Unique identifier for each passenger
- iii. Survived: Survival status (0 = No, 1 = Yes)
- iv. Pclass: Ticket class (1 = 1st, 2 = 2nd, 3 = 3rd)
- v. Name: Passenger's name
- vi. Sex: Passenger's gender
- vii. Age: Passenger's age
- viii. SibSp: Number of siblings/spouses aboard
- ix. Parch: Number of parents/children aboard
- x. Ticket: Ticket number
- xi. Fare: Fare paid for the ticket
- xii. Cabin: Cabin number
- xiii. Embarked: Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton)

Initial observations reveal missing values in the Age and Cabin columns, which will require cleaning and preprocessing.

## 3. Data Cleaning and Preprocessing

Data cleaning and preprocessing steps include:

➤ Handling Missing Values:

Rows with missing Age values were dropped due to its significance in analysis.

Cabin column was not used in this analysis due to a high number of missing values.

➤ Data Type Conversion:

Categorical variables (Sex, Embarked) were converted to numerical format for analysis.

## 4. Statistical Analysis

Summary statistics of the dataset:

- i. Age: Mean age = 29.70 years, Median age = 28.00 years

- ii. Fare: Mean fare = \$32.20, Median fare = \$14.45
- iii. Survival Rate: 38.38% of passengers survived

Correlation analysis:

- i. Positive correlation between Fare and Survived (0.26)
- ii. Negative correlation between Pclass and Survived (-0.34)

## 5. Data Visualization

Visualizations with interpretations:

- i. Age Distribution:

Most passengers were in their 20s and 30s.

Survival rate was higher among children and young adults.

- ii. Fare vs. Survival:

Passengers who paid higher fares had a higher chance of survival.

- iii. Correlation Matrix:

The heatmap shows the correlation between different numerical features.

## 6. Key Findings and Insights

Based on the analysis:

- Passengers in higher ticket classes (1st class) had a better chance of survival.
- Females had a significantly higher survival rate compared to males.
- Passengers who paid higher fares tended to survive more.
- Age was a factor in survival, with children having a higher chance of survival.

## 7. Conclusion

In conclusion, the analysis of the Titanic dataset provides insights into the factors influencing survival. Females, children, and passengers in higher classes had higher survival rates. The fare paid and age were also significant factors affecting survival. This analysis enhances our understanding of the tragic event and the demographics of passengers aboard the Titanic.