# Exploratory Data Analysis Report – Titanic Dataset

## 1.Introduction

The sinking of the RMS Titanic in 1912 remains one of the most studied maritime disasters due to the significant loss of life and the social patterns reflected in survival outcomes. The passenger list represents a diverse group differing in gender, age, class, and travel arrangements, making it a valuable dataset for understanding how these factors affected survival.

This report performs an exploratory data analysis (EDA) on a subset of the Titanic dataset containing 418 passengers and 12 features. The aim is not prediction, but clear interpretation of trends and relationships within the data. The analysis uses Python with Pandas, Matplotlib, and Seaborn to explore:

- Data structure and completeness
- Distribution of key variables
- Survival differences across demographic and class categories
- Correlations between numerical features

The results provide insight into how social status, economic factors, and demographics shaped survival chances during the event.

## 2.Problem Statement

The goal of this analysis is to examine the Titanic passenger dataset to identify the factors that influenced survival during the disaster. The dataset contains demographic and travel-related information for passengers, but survival outcomes vary widely across individuals.

The core problem is:

Which passenger characteristics are most strongly associated with surviving the Titanic sinking?

To answer this, the analysis focuses on statistical exploration and visual examination of key features such as sex, passenger class, age, fare, and family relationships. The findings should help explain survival patterns and highlight which attributes had meaningful impact.

## 3.Objective of the Study

This analysis has three main objectives:

1. Understand the dataset structure, quality, and missing information
2. Identify patterns and relationships linked to survival
3. Summarize the key factors that influenced passenger outcomes

The focus is on extracting meaningful, interpretable insights rather than building a predictive model.

## 4.Scope of Analysis

The study includes:

- Data profiling using .info() and .describe()
- Visual exploration through histograms, boxplots, bar charts, and heatmaps
- Survival analysis across sex, class, fare, and embarkation point
- Correlation analysis of numerical variables

The analysis does **not** include:

- Imputation of missing values
- Feature engineering
- Machine learning modeling

These areas are noted for future work.

## 5.Dataset Overview

The dataset includes:

- 418 rows
- 12 columns

Key features:

- Passenger demographics: Sex, Age
- Travel information: Pclass, Ticket, Fare, Embarked
- Family size: SibSp, Parch
- Target variable: Survived (0 = No, 1 = Yes)

Most columns are complete. Major missing values appear in:

- Cabin (almost entirely missing)
- Age (partial missing)
- Fare (one missing)

## 6. Statistical Summary

Numerical features show:

- Wide age range, mostly adults
- Fare values heavily skewed with a few high-ticket passengers
- Most passengers traveled alone or with small family groups

The overall survival rate in this subset is approximately **36%**.

## 7. Univariate Analysis

Histograms and summary statistics show:

- Age distribution centered around adulthood, with fewer children and elderly passengers
- Fares concentrated in the lower range, indicating many lower-cost ticket holders
- Most passengers belonged to **3rd class**, followed by 1st and 2nd

## 8. Bivariate Analysis

### Survival by Sex

The most striking result:

- **All female passengers survived**
- **No male passengers survived**

Sex appears to be the strongest survival indicator in this dataset.

### Survival by Class

Survival rates by class:

- 1st Class: ~47%
- 2nd Class: ~32%
- 3rd Class: ~33%

Higher-class passengers show a clear survival advantage.

### Survival by Fare

Passengers who paid higher fares were more likely to survive, which aligns with the class findings.

## 9. Correlation Analysis

A correlation heatmap highlights:

- Positive relationship between Survival and Fare
- Negative relationship between Survival and Pclass
- Strong negative correlation between Pclass and Fare, confirming price differences across classes

Most other numerical features show weak correlations.

## 10. Key Insights

The analysis indicates:

- Sex is the dominant factor in survival outcomes
- Passenger class and fare have meaningful influence
- Family-related features (SibSp, Parch) show minimal impact
- Missing Cabin values limit its usefulness

## 11. Conclusion

The dataset reveals clear survival patterns:

- Female and higher-class passengers had a greater chance of survival
- Economic status and ticket price appear linked to survival probability

These findings highlight social and structural factors affecting survival during the Titanic disaster.