Titanic Dataset Analysis Report

1 Introduction

The Titanic dataset provides valuable insights into the passengers aboard the ill-fated ship, including their demographics, ticket information, and survival outcomes. By analyzing the data, we can explore patterns and relationships between various features to understand the factors that influenced survival rates. This report presents a structured exploratory data analysis (EDA), focusing on univariate, bivariate, and multivariate insights.

1.1 Dataset Features

The dataset consists of the following key features:

- **Survived**: Indicates whether a passenger survived (1) or not (0).
- **Pclass**: Passenger class (1st, 2nd, or 3rd class).
- **Name**: Full name of the passenger.
- **Sex**: Gender (Male/Female).
- **Age**: Age of the passenger in years.
- **SibSp**: Number of siblings/spouses aboard the Titanic.
- **Parch**: Number of parents/children aboard the Titanic.
- **Ticket**: Ticket number.
- **Fare**: Fare paid for the ticket.
- **Cabin**: Cabin number (if assigned).
- **Embarked**: Port where the passenger boarded the ship (C = Cherbourg, Q = Queenstown, S = Southampton).

2 Exploratory Data Analysis (EDA)

2.1 Univariate Analysis

2.1.1 Numerical

- The **age distribution** follows an approximately normal pattern, with most passengers aged between 20 and 40 years and a peak around 30 years.
- **Fare values** are highly right-skewed, indicating most passengers paid lower fares, with a few paying significantly higher prices.
- **SibSp and Parch distributions** show that most passengers traveled alone, with very few having multiple family members on board.

2.1.2 Categorical

- **Survival rate** is very low, with only about 38% of passengers surviving.
- **Males significantly outnumber females**, indicating gender disparity.
- **Third-class passengers form the majority**, followed by first and second class.
- **Most passengers embarked from Southampton**, while Cherbourg and Queenstown had fewer passengers.

2.2 Bivariate Analysis

2.2.1 Numerical vs. Numerical

- **Age and Fare** show no strong correlation, though higher fares were mostly associated with older first-class passengers.
- **SibSp and Parch** are positively related, as passengers with siblings/spouses often had parents/children aboard as well.

2.2.2 Categorical vs. Categorical

- **Female passengers had a significantly higher survival rate than males**, confirming the "women and children first" policy.
- **First-class passengers had the highest survival rate**, while third-class passengers had the lowest.
- **Passengers from Cherbourg had higher survival rates**, likely due to a greater proportion of first-class travelers.

2.2.3 Numerical vs. Categorical

- **Higher fares correlated with better survival chances**, reinforcing the socio-economic influence on survival.
- **Younger passengers, particularly children, had better survival rates**, while older passengers had lower chances.

2.3 Multivariate Analysis

- **Gender and class together played a critical role in survival.** First-class women had the highest survival rates, while third-class men had the lowest.
- **Fare had a strong impact on survival across different classes.** Higher fares were associated with better survival chances, particularly for first-class passengers.
- **Embarkation port showed variations in survival rates.** Passengers from Cherbourg, who were more likely to be in first class, had better survival chances than those from Southampton, where more third-class passengers boarded.
- **Passengers traveling with family members had higher survival rates than those traveling alone.** This suggests that family groups may have had better access to lifeboats or received assistance during the evacuation.

3 Key Insights

- **Gender, class, and fare were the strongest predictors of survival, with women and first-class passengers having the highest survival rates.**
- **Third-class passengers, particularly men, faced the lowest chances of survival, reinforcing the impact of socio-economic status.**
- **Passengers who paid higher fares had a better chance of survival, indicating that wealthier individuals had greater access to lifeboats.**
- **Passengers who traveled with family members had better survival rates compared to those who traveled alone, suggesting a support system may have improved survival odds.**
- **Embarkation location influenced survival, with Cherbourg passengers having better survival rates than those from Southampton, due to a higher proportion of first-class travelers.**

4 Conclusion

This analysis highlights the key factors influencing survival on the Titanic. **Gender, class, and fare played the most significant roles**, with women and first-class passengers having the highest survival chances. The findings align with historical records, showing that socio-economic status and emergency evacuation policies had a crucial impact on survival outcomes.