■ Titanic Dataset - Exploratory Data Analysis (EDA) Report

This report presents the Exploratory Data Analysis (EDA) of the Titanic dataset. The objective of this analysis is to uncover meaningful insights regarding passenger demographics, travel classes, fares, and survival rates during the Titanic disaster. Various statistical and visualization techniques were applied to better understand the dataset and detect patterns.

■ Dataset Overview

- Survived: Survival (0 = No, 1 = Yes) - Pclass: Ticket class (1 = 1st, 2 = 2nd, 3 = 3rd) - Sex: Gender of passenger - Age: Age in years - SibSp: # of siblings / spouses aboard - Parch: # of parents / children aboard - Fare: Passenger fare - Embarked: Port of Embarkation (C = Cherbourg, Q = Queenstown, S = Southampton)

■■ EDA Steps Performed

1. Data Cleaning: - Handled missing values in Age, Embarked, and Cabin columns - Converted data types where necessary - Removed duplicate entries 2. Univariate Analysis: - Age distribution (histograms, KDE plots) - Passenger class distribution - Gender distribution - Fare distribution 3. Bivariate/Multivariate Analysis: - Survival rate by gender - Survival rate by passenger class - Survival rate by embarkation port - Survival relation with age and fare - Combined analysis (Gender + Class) 4. Correlation Analysis: - Heatmap of numeric features (Age, Fare, Pclass, Survived, etc.) 5. Outlier & Skewness Analysis: - Detected outliers using boxplots - Observed fare skewness and considered log-transformations

■ Key Insights

- Overall survival rate was approximately 38%. - Females had a much higher survival rate compared to males. - First-class passengers had significantly higher survival chances than second- and third-class passengers. - Children (under 12 years) had better survival rates compared to adults and seniors. - Passengers who paid higher fares had better survival chances, highlighting socio-economic influence. - Passengers embarking from Cherbourg showed better survival compared to Southampton and Queenstown. - Family size influenced survival, with small families (1–3 members) faring better than those alone or in large families.

■ Conclusion

The EDA of the Titanic dataset highlights how demographic and socio-economic factors influenced survival during the disaster. Gender, passenger class, age, and fare were key determinants of survival. This analysis demonstrates the importance of exploratory data analysis in uncovering hidden patterns and guiding further predictive modeling tasks.