Exploratory Data Analysis Report – Titanic Dataset

Objective

The purpose of this analysis was to perform exploratory data analysis (EDA) on the Titanic dataset to identify key patterns, trends, and relationships affecting passenger survival. The analysis used Python libraries such as Pandas, Matplotlib, and Seaborn for statistical and visual exploration.

Dataset Overview:

The dataset consists of 891 passenger records with features such as age, gender, class, fare, and survival status. Initial inspection using .info() and .describe() revealed missing values in the Age, Cabin, and Embarked columns. The average passenger age was around 30 years, and the average fare was approximately $32.

Key Findings:

- Survival Distribution: About 38% of passengers survived, while 62% did not, showing a strong class and gender-based disparity.
- Gender and Survival: Females had a much higher survival rate than males (around 74% of women survived compared to 19% of men). This aligns with the "women and children first" policy during evacuation.
- Passenger Class and Survival: Survival rate was highest among First-Class passengers (63%) and lowest among Third-Class passengers (24%). This indicates that socioeconomic status significantly influenced survival chances.
- Age Patterns: Younger passengers, particularly children, had a higher likelihood of survival compared to older passengers.
- Fare and Survival Relationship: Passengers who paid higher fares tended to survive more often, suggesting a link between wealth and access to safety measures.

Correlations:

Pclass and Fare showed a strong negative correlation (wealthier passengers traveled in higher classes).

Survived was positively correlated with Fare and negatively correlated with Pclass.

Age had a weak correlation with survival.

Missing Data: Columns Age and Cabin contain a large number of missing values, requiring imputation or removal before predictive modeling.

Conclusion:

The analysis of the Titanic dataset reveals that gender, passenger class, and fare amount were the most influential factors affecting survival. Women and children from higher classes had a significantly higher chance of survival, while most third-class male passengers did not survive. These findings highlight the clear impact of socioeconomic status and gender during the disaster.

Overall, this EDA demonstrates how statistical summaries and visualizations can uncover meaningful insights and patterns in real-world data.