Exploratory Data Analysis (EDA) Report

Introduction

The dataset was analyzed to identify patterns, relationships, and potential predictors of survival. The focus was on handling missing values, univariate exploration, and bivariate relationships across features.

Missing Values

- Missing values were identified and handled accordingly.
- Age values were imputed using the median.
- Other categorical missing values (if any) were filled or excluded depending on context.

Univariate Analysis

Univariate analysis was carried out for individual features:
- **Age:** Broad distribution, missing values imputed.
- **Fare:** Positively skewed with outliers.
- **Class (Pclass):** Majority in 3rd class.
- **Sex:** Slightly more males.
- **Family (SibSp, Parch):** Most traveled alone or in small groups.

Bivariate Analysis

- **Sex vs Survival:** Females show higher survival.
- **Class vs Survival:** Higher class had better survival chances.
- **Fare vs Survival:** Higher fares linked to better survival.
- **Age vs Survival:** Subtle effect; children and elderly showed different patterns.
- **Family Size:** Small families had slightly better outcomes.

Observations

- Gender and Class are the strongest indicators of survival.
- Fare provides weaker predictive power.
- Age has limited influence except at extremes.
- Outliers in fare expected (premium tickets).

Conclusion

The analysis reinforces well-documented insights about the Titanic tragedy:
- Women and children, especially from higher classes, had better survival.
- Economic status influenced outcomes.
- Age and family played smaller roles but added context.

This EDA provides a cleaned dataset for predictive modeling.