Exploratory Data Analysis Report - Titanic Dataset

This report presents the findings from an exploratory data analysis (EDA) conducted on the Titanic dataset. The primary objective was to extract patterns, trends, and anomalies using visualization and statistics.

1. Data Overview:

- The dataset contains passenger information including survival status, class, gender, age, fare, and embarkation port.

- Missing values were handled in 'Age' and 'Embarked'; 'Cabin' was dropped due to excessive missingness.

2. Univariate Analysis:

- Majority of passengers were in 3rd class.

- Most passengers were males.

- Age distribution is right-skewed with a median around 28.

- Fare distribution is heavily skewed due to some very high values.

3. Bivariate Analysis:

- Females had a higher survival rate than males.

- 1st class passengers had the highest survival rate.

- Survival was less likely in 3rd class.

4. Multivariate Insights:

- Survival is positively correlated with passenger class and fare.

- Pairplots and heatmaps highlight relationships among age, fare, class, and survival.

5. Summary:

- Gender and class are strong predictors of survival.
- Data is slightly skewed in numerical columns.
- Visual analysis helps identify useful features for modeling.

Conclusion:

EDA helps in understanding the dataset's structure and supports feature selection for machine learning models.