

TASK-2 EXPLORATORY DATA ANALYSIS (EDA)

Data Loading

The dataset was loaded using `pd.read_csv()` and basic structure checked using `head()`, `info()`, and `describe()`.

Summary Statistics

Basic descriptive statistics like:

- Mean, Median, Std, Min, Max
- Count of missing values were generated for numeric features.

Data Visualization

Multiple plots were created to understand the data better:

- Histograms for distribution of numerical features
- Boxplots to detect outliers
- Pairplots for observing feature relationships
- Correlation Matrix Heatmap for checking feature correlations

Pattern Detection

Key observations were made such as:

- Missing values in 'Age' and 'Cabin'
- Outliers in 'Fare' and 'Age'
- Strong negative correlation between Pclass and Survived
- Higher survival rates among females and higher-class passengers
- Passengers in 1st class had significantly higher survival chances.
- Females had a much higher survival rate than males.
- The 'Cabin' column had many missing values.
- 'Fare' and 'Age' features contained outliers that need handling during preprocessing.