Titanic EDA

1. Objective

To explore the Titanic dataset and uncover patterns in passenger demographics and survival rates .

2. Data Preprocessing

- Identified and handled missing values:
  - ❖ Age: 177 missing values
  - ❖ Cabin: 687 missing values
  - ❖ Embarked: 2 missing values

3. Data Types

- The dataset includes a mix of:

  Numerical columns: Age, Fare, SibSp, Parch

  Categorical columns: Sex, Pclass, Embarked, Cabin

  Target column: Survived

4. Steps:

- **Data Cleaning**: Handling missing values (Age, Embarked), data type corrections
- **Univariate Analysis**: Distribution of age, class, gender, fare
- **Bivariate Analysis**: Survival rate by gender, class, family size, embarkation point
- **Visualization**: Bar charts, histograms, heatmaps, Pair plot, Boxplot
- **Correlation Analysis**: To identify influential features:
  - ➢ **Weak correlations** observed between:
    - ❖ Age and Fare
    - ❖ SibSp and Parch
    - ❖ Fare and Parch
  - ➢ **Stronger correlations** found between:
    - ❖ Pclass and Fare
    - ❖ Pclass and Age

5. Insights

- Survival chances were heavily influenced by gender, class, and age
- Females had a much higher survival rate compared to males.
- Children (age < 10) had better survival odds.
- Most passengers embarked from Southampton.
- Larger families (more than 3 people) had lower survival rates.