Titanic Data Analysis (EDA - Train Set)

Objective:

The goal of this project is to perform exploratory data analysis (EDA) on the Titanic train dataset to uncover insights into passenger survival rates and the factors influencing survival. The analysis aims to explore key features like age, gender, passenger class, and fare to identify patterns and trends.

Dataset:

- **Source**: Kaggle Titanic Dataset

- **Dataset Type**: Train Set (Used for analysis)

Observations from the Analysis:

1. **Missing Values**: The dataset has missing values in the 'Age' and 'Cabin' columns. 'Age' was filled with the median, while 'Cabin' was dropped due to excessive missing data.

2. **Survival Rate**: 38% of passengers survived.

   ○ Higher survival rate among women compared to men.

   ○ 1st class passengers had a significantly higher survival rate than 2nd and 3rd class passengers.

3. **Passenger Classes**: Most passengers were in 3rd class, followed by 2nd and 1st class.

   ○ 1st class passengers had the highest survival rate.

4. **Age Distribution**: The age distribution is right-skewed, with more young passengers than older ones.

   ○ Survival rate slightly higher for younger passengers.

5. **Gender**: Women had a higher survival rate (~74%) compared to men (~18%).

6. **Fare**: Fare distribution is skewed, with a few passengers paying significantly more. Higher fares are correlated with 1st class and higher survival rates.

Insights for Business/Action:

- **Policy Adjustments**: Policies could have been made to improve survival rates for lower class passengers.

- **Gender Disparity**: The higher survival rate for women and children suggests prioritization during the evacuation.