Titanic Dataset - EDA Findings Report

1. Dataset Overview
• Total rows: 891
• Columns: 12
• Dataset contains demographic, ticket, and survival data of Titanic passengers.

2. Missing Values
• Age: 177 missing values
• Cabin: 687 missing values (highly incomplete)
• Embarked: 2 missing values
→ Cabin column is not reliable for analysis without major cleaning.

3. Key Numerical Insights
• Average Age: 29.7 years
• Fare ranges widely (0 to 512). Shows right-skewed distribution.
• Majority of passengers are 3rd class (Pclass = 3).

4. Survival Insights
• Survival Rate: approx. 38% (Survived = 1).
• Females had significantly higher survival rate compared to males.
• 1st-class passengers survived the most, followed by 2nd, then 3rd class.

5. Key Patterns & Relationships
• Sex vs Survival: Strong relationship (Females survived more).
• Class vs Survival: Higher class → higher survival.
• Fare vs Survival: Higher fare passengers had better chances.
• Age shows no strong correlation with survival.

6. Data Quality Notes
• Cabin column is too incomplete; usually dropped in modeling.
• Age column requires imputation (median is commonly used).

7. Summary
The Titanic dataset reveals clear social and economic disparities affecting survival. Passenger sex, class, and fare strongly influenced survival outcomes. Cabin data is mostly missing, and age has moderate missing values but can be imputed. Overall, these patterns provide a solid foundation for predictive modeling or further statistical analysis.