Summary of Findings: Titanic Dataset (EDA)

1. **Missing Values**:
    - Significant missing data in Age and Cabin.
    - Embarked had a few missing values but can be imputed with the most frequent value.
2. **Survival Rate**:
    - Overall survival rate was approximately **38%**.
    - Survival was **not evenly distributed** across features like gender, class, and age.
3. **Gender and Survival**:
    - **Females had a much higher survival rate** (~74%) than males (~19%).
    - Gender was a strong predictor of survival.
4. **Passenger Class**:
    - **1st class passengers had the highest survival rate**, while 3rd class had the lowest.
    - Wealth and social status likely influenced survival chances.
5. **Age Distribution**:
    - Most passengers were between 20–40 years old.
    - Children (under 10) had **moderately high survival rates**, especially in 1st and 2nd class.
6. **Fare**:
    - Passengers who paid higher fares had higher chances of survival.
    - Some extreme outliers in fare were observed (very high fare values in 1st class).
7. **Embarked Port**:
    - Most passengers boarded from **Southampton (S)**.
    - Passengers from **Cherbourg (C)** had the highest survival rate.
8. **Correlation Matrix Insights**:

- Fare and Pclass were moderately correlated.
- Sex and Pclass showed a relationship with Survived, indicating their importance in predictive modeling.

9. **Outliers**:
    - Detected outliers in Fare and Age using boxplots.
    - Consider capping or transformation if using for ML.
10. **Multivariate Analysis**:
    - Using pairplots and heatmaps, **clusters** of survival could be seen in certain combinations of Fare, Pclass, and Sex.