EDA Observations on Titanic Dataset:

1. Survival Count (Countplot of Survived)

- The number of passengers who did not survive (Survived = 0) is significantly higher than those who did (Survived = 1).
- Indicates a class-imbalanced dataset, with more non-survivors (~62%) than survivors (~38%).

2. Age Distribution (Histogram of Age)

- Most passengers were between 20 and 40 years old.
- The age distribution is slightly right-skewed.
- A notable number of children (under 10 years) were also onboard.

3. Fare Distribution (Boxplot of Fare)

- Majority of fares were under $100.
- Several high-value outliers exist (some above $500).
- Median fare lies around $15–$30, suggesting most passengers bought cheaper tickets.

4. Survival by Gender (Countplot: Sex vs Survived)

- A significant majority of females survived, while most males did not.
- Gender had a strong impact on survival outcome.

5. Survival by Passenger Class (Countplot: Pclass vs Survived)

- Passengers in 1st class had the highest survival rate.
- Survival probability decreased with class: 3rd class passengers had the lowest.

6. Age vs Survival (Violin Plot: Survived vs Age)

- Survivors are clustered mostly in the 20–40 age range.
- Children under 10 showed higher survival likelihood.
- Older passengers had varied outcomes.

7. Correlation Heatmap

- Survived is positively correlated with Fare and Sex_female, and negatively correlated with Pclass.
- Weak positive correlation with SibSp and Parch.
- Fare and Pclass are moderately negatively correlated, as expected.

8. Pairplot (Selected Features)

- Survivors generally belonged to higher Fare and lower Pclass.
- Visual separations evident in Fare vs Pclass among survivors.
- Some features show clear clustering based on survival outcome.

9. Survival by Family Size

- Small families (2–4 members) had better survival rates.
- Solo travelers (FamilySize = 1) had lower chances of survival.
- Very large families (7+) were also less likely to survive.

Summary of Findings: Titanic Dataset EDA

The exploratory analysis of the Titanic dataset revealed significant insights into the factors influencing passenger survival.

1. Class and Gender Are Key Predictors of Survival

- First-class passengers had the highest survival rate, highlighting the impact of socio-economic status.
- Female passengers were far more likely to survive than males, reflecting the "women and children first" evacuation protocol.

2. Age Plays a Role

- Children (especially under 10) had higher survival rates.
- Most passengers were between 20 and 40 years old; survivors were often in this range.

3. Fare Is Positively Correlated with Survival

- Passengers who paid higher fares (likely in higher classes) had better survival outcomes.
- This supports the idea that class and ticket price affected rescue priority.

4. Family Size Matters

- Passengers with small families (2–4 members) had a better chance of survival.

- Solo travelers and passengers with large families had lower survival rates, possibly due to lack of support or logistical challenges.

5. Feature Relationships and Patterns

- Moderate correlation between Fare and Pclass (higher class → higher fare).
- Weak but relevant correlations exist between Survived and other features such as Sex, Fare, and Pclass.

6. Data Imbalance and Outliers

- The dataset is imbalanced, with more non-survivors than survivors.
- Fare contains several outliers, indicating a few extremely expensive tickets.

7. No Strong Multicollinearity Observed

- Heatmap shows moderate correlations between features, but no strong multicollinearity issues, making the data suitable for predictive modeling.