Findings and Patterns from Titanic Dataset

1. **Survival Percentage Pie Chart and Count plot:**
    - Observation- The pie chart gives the overall proportion of passengers who survived and who did not, while the count plot provides the raw numbers for each category.
    - Finding- More than 60% of the passengers did not survive. That is 3 out of every 5 people, totaling up to 549 victims, with 342 survivors.
2. **Histograms of Numeric Features:**
    - Observation- These plots show the distribution of individual numeric features namely Survived, PClass, Age, Fare, SibSp and Parch.
    - Finding- Most of the passengers were in the third class, around the age of 30, with tickets costing approximately 50 pounds and without any family members.
3. **Boxplots of Numeric Features (Before and After Outlier Handling):**
    - Observation- Boxplots display the distribution of data, including median, quartiles, and potential outliers.
    - Finding- There are visible outliers present in the data, such as a passenger's fare going beyond 500 pounds. After cleaning, the median of each column is present, for example, 10 being the median fare.
4. **Bar Plots and Count plots of Survived vs. Sex:**
    - Observation- These plots compare the survival rate and counts based on the Sex feature.

- Finding- Clearly, females had a considerably higher survival rate (more than 70%) as compared to males (less than 30%).

5. **Hist plot of Age vs. Survived:**
    - Observation- This plot shows the distribution of Age for both survived and non-survived passengers.
    - Finding- People around the age of 30 were the most vulnerable to the mishap, but also, were the highest number of survivors.
6. **Bar Plots and Count plots of Survived vs. PClass:**
    - Observation- These plots show the survival rate and counts based on the Passenger Class feature.
    - Finding- Passengers in higher classes (1st class) had a much higher survival rate (almost 50%) than those in lower classes (3rd class).
7. **Bar Plots and Count plots of Survived vs. Embarked:**
    - Observation- These plots show the survival rate and counts based on the Port of Embarkation feature.
    - Finding- Even though there is no correlation, Passengers from Q had the highest survival ratio (about 45%), while S had the most victims (more than 300).
8. **Correlation Matrix Heatmap:**
    - Observation- This heatmap displays the correlation coefficients between pairs of numeric features.
    - Finding- Correlation for Fare- PClass is strongly negative and for Fare-SibSp, Fare-Age and Fare-Survived are weakly positive.
9. **Pairplot of Selected Features by Survival:**
    - Observation- This creates scatter plots for all pairs of selected numeric features, with colours based on the survivability of the

passengers. Histograms of individual features are also shown on the diagonal.

- Finding- This provides a visual representation of the relationships between pairs of features and how these relationships differ for survivors and non-survivors. It can help identify clusters or patterns related to survival across multiple dimensions.