TASK 3

TITANIC DATASET

REPORT OF MY PROJECT

ABOUT DATASET

The Titanic dataset contains information on 891 passengers from the RMS Titanic disaster. It includes demographic details (age, gender, class), travel details (ticket, fare, cabin, embarkation), and survival status. This dataset is widely used for data analysis, visualization based on passenger attributes.

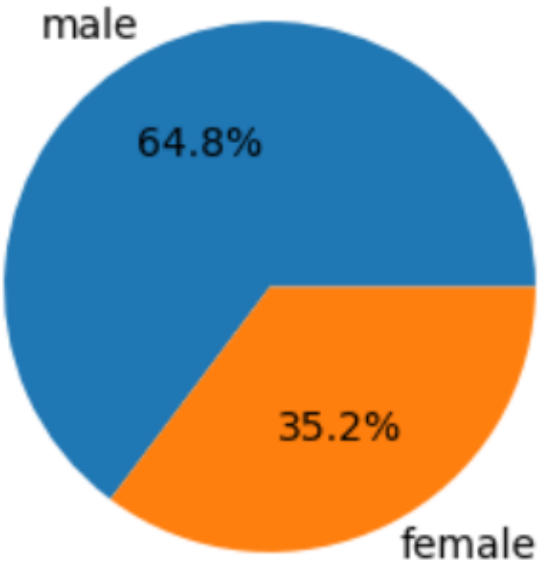
- Importing necessary libraries like pandas, Numpy, matplotlib, plotly, seaborn and future warnings.
- Loading dataset by using pandas library.
- Checking the information, describe, columns name, null values, value counts.

- Drop the column named Cabin that consist more then 75% of null values.

- Filling the null value of column age by using mean because it contains numerical values and column embarked by using mode because it contains categorical values.

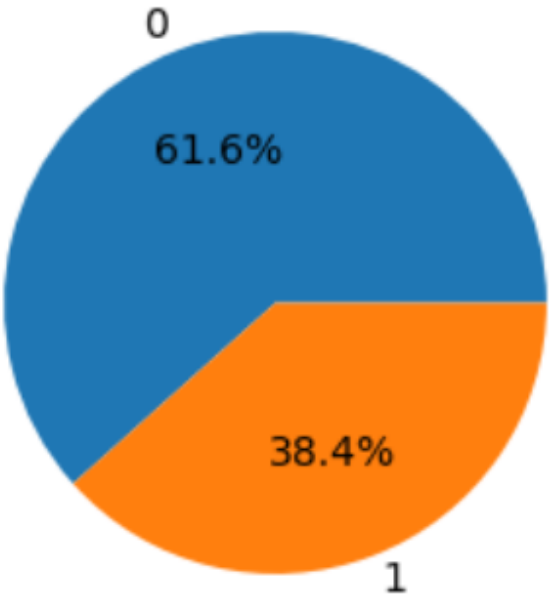
Visualization

1. Gender count



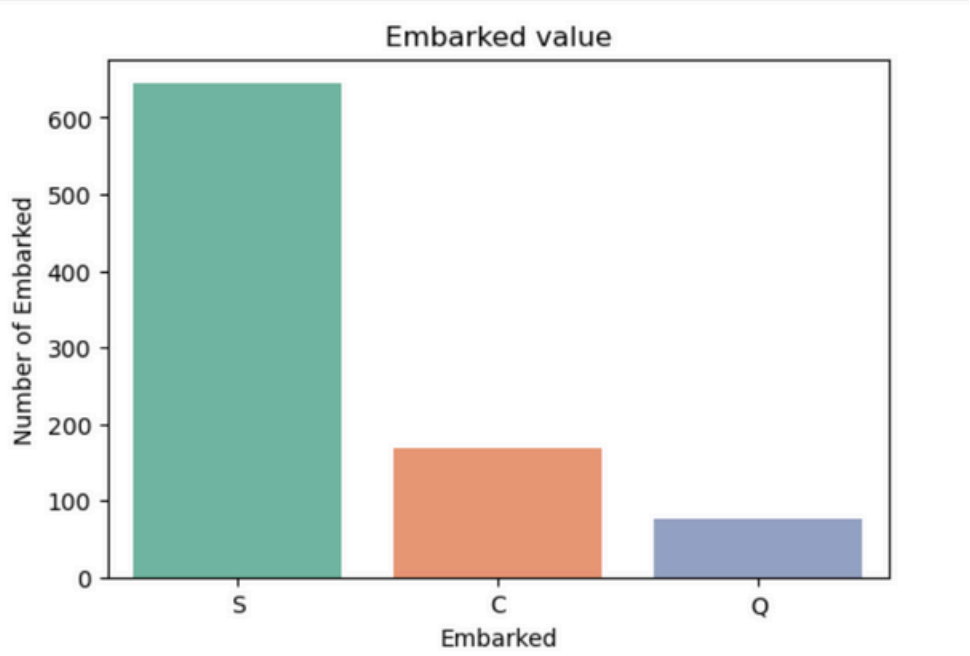
This pie chart shows the gender distribution of Titanic passengers. Males comprised 64.8% of the dataset, while females accounted for 35.2%. The chart visually highlights the male majority among passengers, an important factor in survival analysis, as gender was strongly associated with survival chances in the Titanic disaster.

2. Survival count



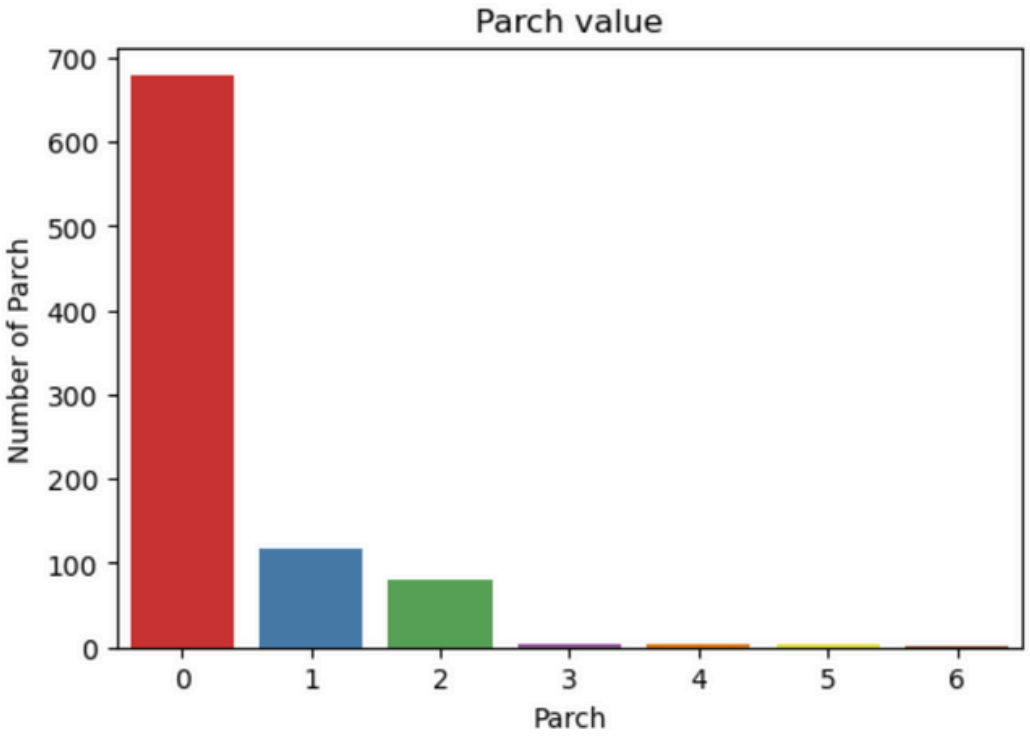
This pie chart represents Titanic passenger survival status. A total of 61.6% (labelled as “0”) did not survive, while 38.4% (labelled as “1”) survived the disaster. The chart clearly shows a higher proportion of non-survivors, reflecting the tragedy’s significant human loss and the limited lifeboat availability during the sinking.

3. Embarked value



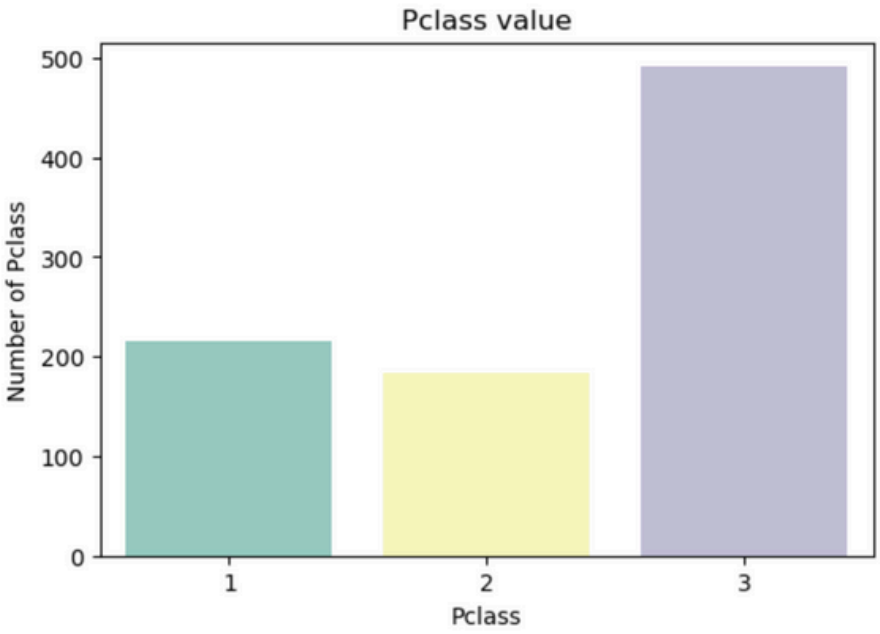
This bar chart displays the embarkation points of Titanic passengers. Most passengers, over 600, boarded at port S (Southampton), followed by around 170 from port C (Cherbourg) and about 80 from port Q (Queenstown). The dominance of Southampton passengers reflects its role as the ship's primary departure port.

4. Parch value



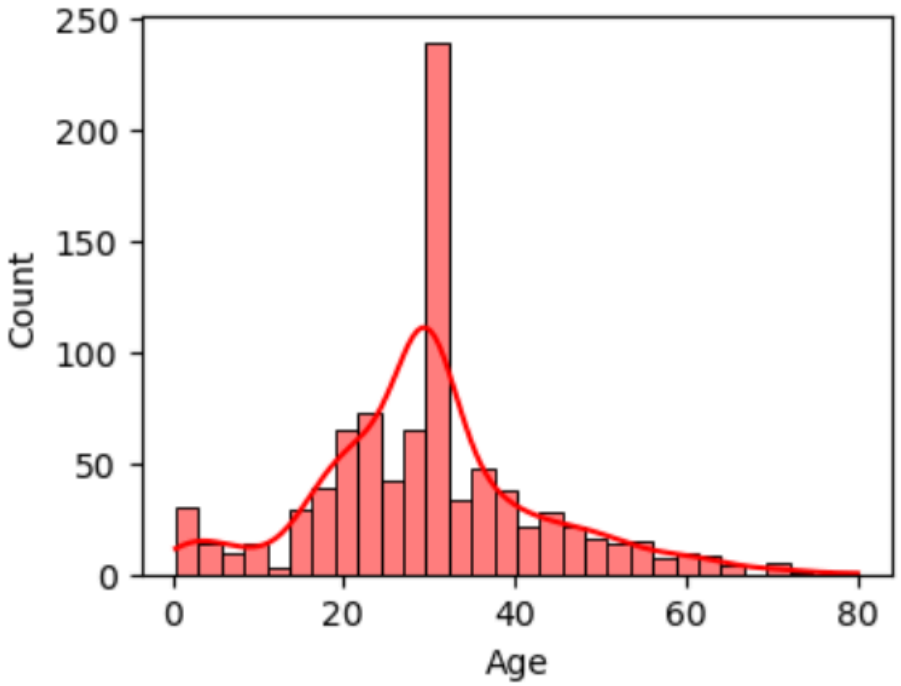
This bar chart shows the distribution of Titanic passengers by “Parch” (parents/children aboard). Most passengers (~680) had no parents or children, while smaller groups had 1 or 2. Very few had between 3 and 6. This indicates that most passengers traveled without close family dependents.

5. Pclass value



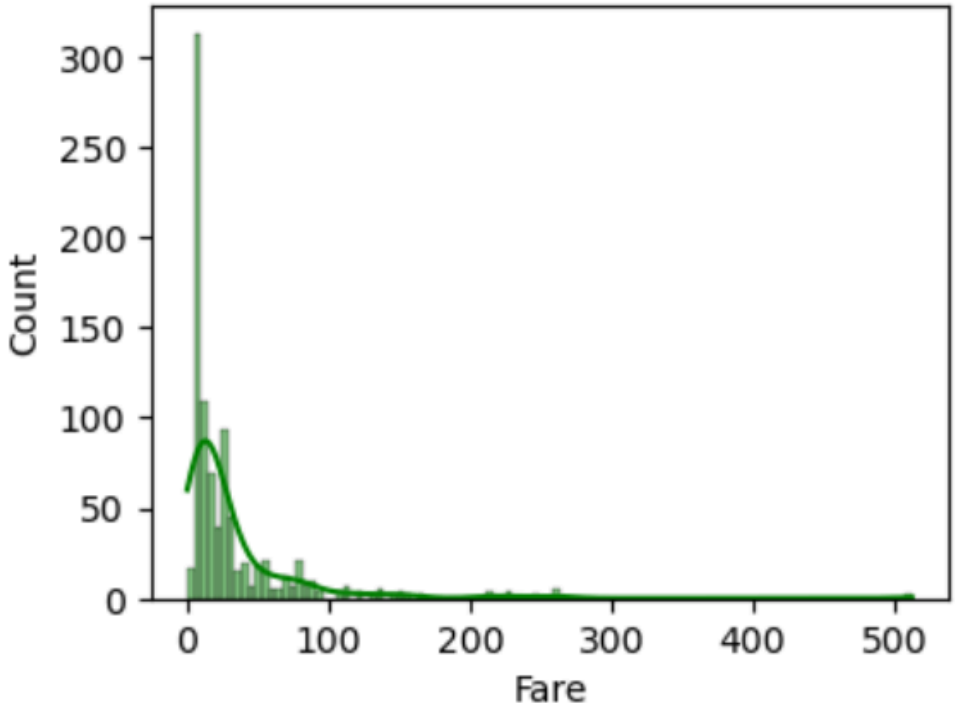
This bar chart shows Titanic passenger distribution by class (Pclass). Third class dominates with nearly 500 passengers, followed by first class (~210) and second class (~180). This indicates a larger proportion of lower-class travelers, reflecting economic demographics of passengers on board.

6. Age distribution



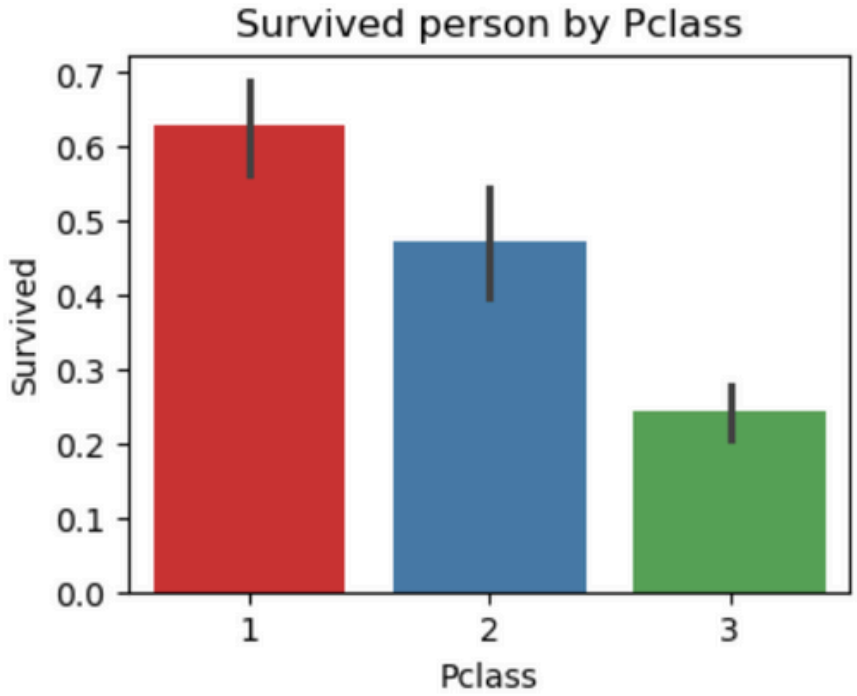
This histogram with KDE shows Titanic passengers' age distribution. Most passengers were aged between 20 and 40, with a notable spike around 30. Fewer children and elderly passengers are seen, and the distribution skews slightly right, indicating a small proportion of older individuals on board.

7. Fare distribution



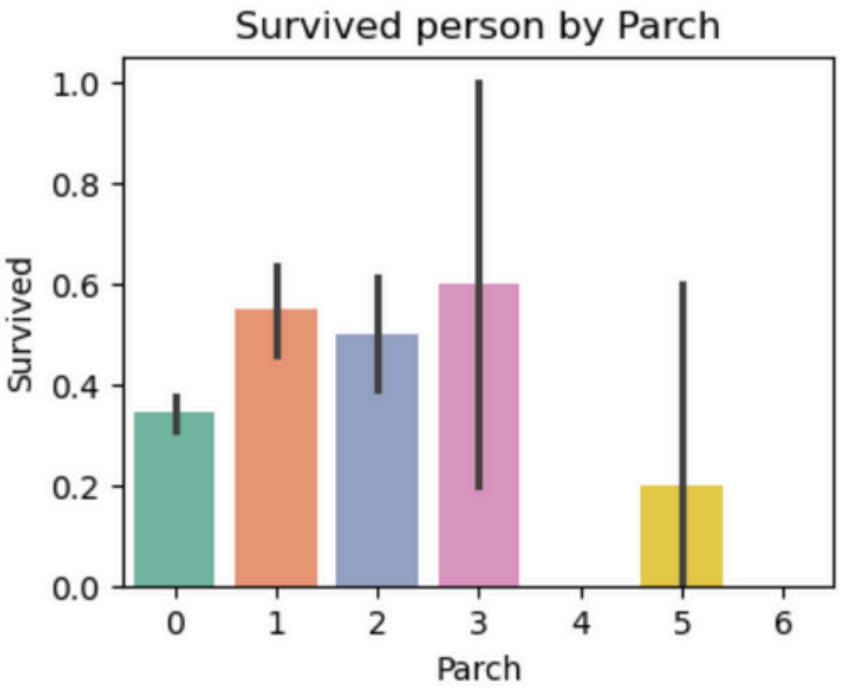
The histogram with KDE shows Titanic passengers' fare distribution, heavily skewed right. Most fares were below 50, with a sharp peak near very low values. A small number of passengers paid significantly higher fares, reaching over 500, indicating strong fare inequality among ticket classes.

7. Survived by Pclass



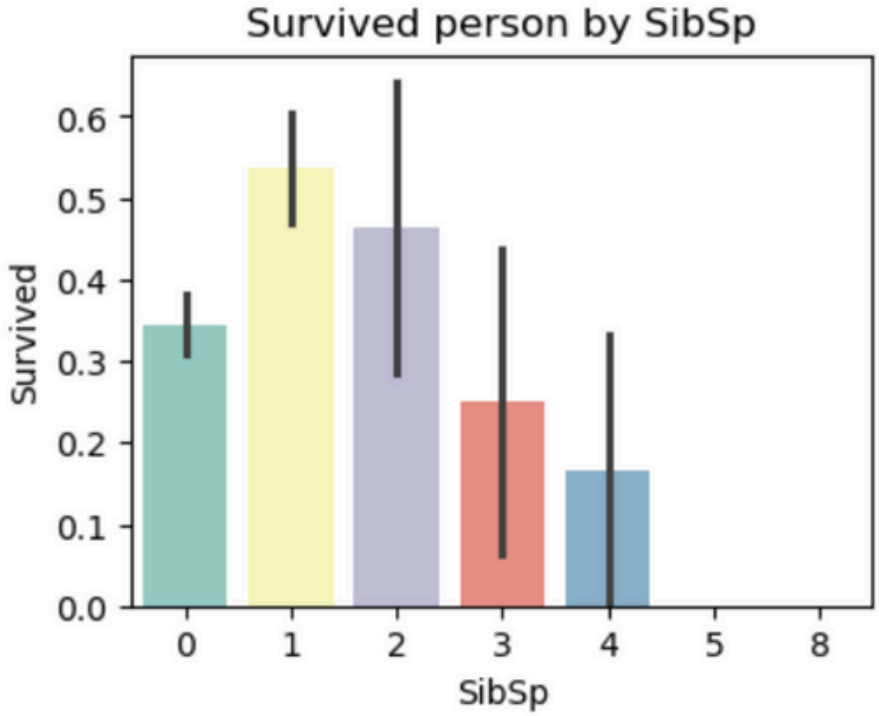
The chart shows Titanic survival rates by passenger class. First-class passengers had the highest survival rate (~63%), followed by second class (~47%), while third class had the lowest (~24%). This suggests a strong link between socio-economic status and survival chances during the disaster.

7. Survived by Parch



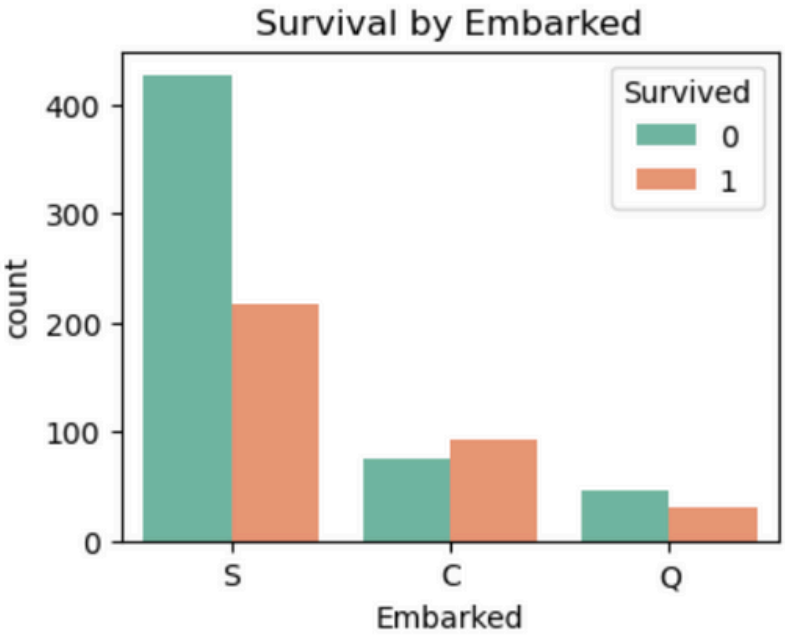
The chart shows Titanic survival rates by “Parch” (parents/children aboard). Survival peaked for passengers with 1–3 family members, especially Parch=3 (~60%). Solo travelers (Parch=0) had lower survival (~35%), while large families (Parch=5 or more) had the lowest (~20%), suggesting moderate family presence increased survival chances.

7. Survived by SipSp



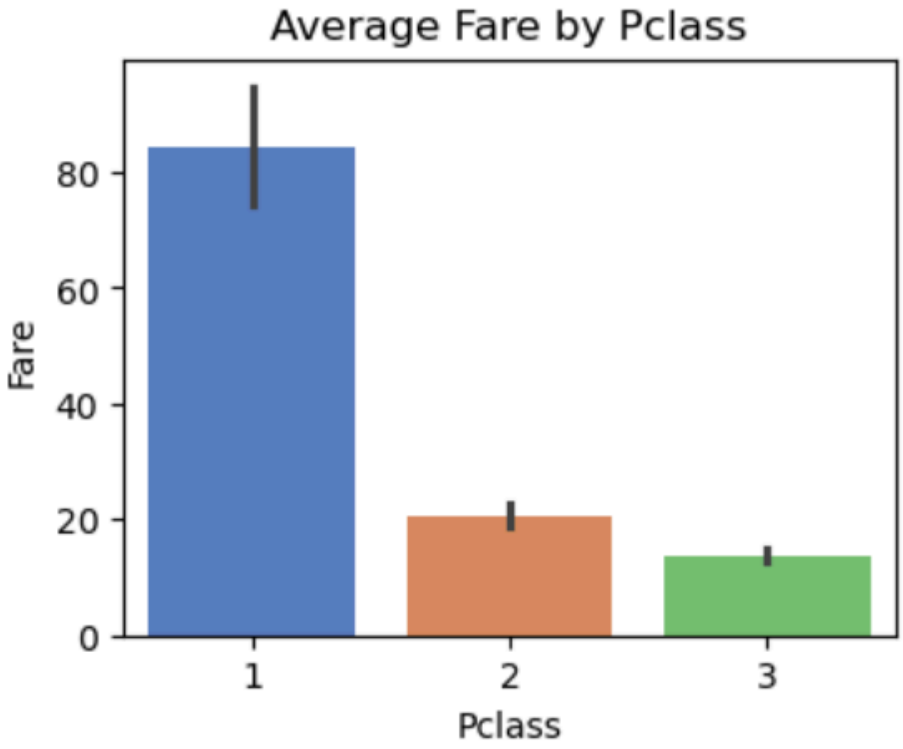
This chart shows Titanic survival rates by SibSp (siblings/spouses aboard). Survival peaked at SibSp=1 (~54%), followed by SibSp=2 (~47%). Passengers alone (SibSp=0) had ~34% survival, while larger groups (SibSp≥3) faced sharply lower chances. This suggests having a small number of companions improved survival likelihood compared to being alone or in large groups.

7. Survived by Embarked



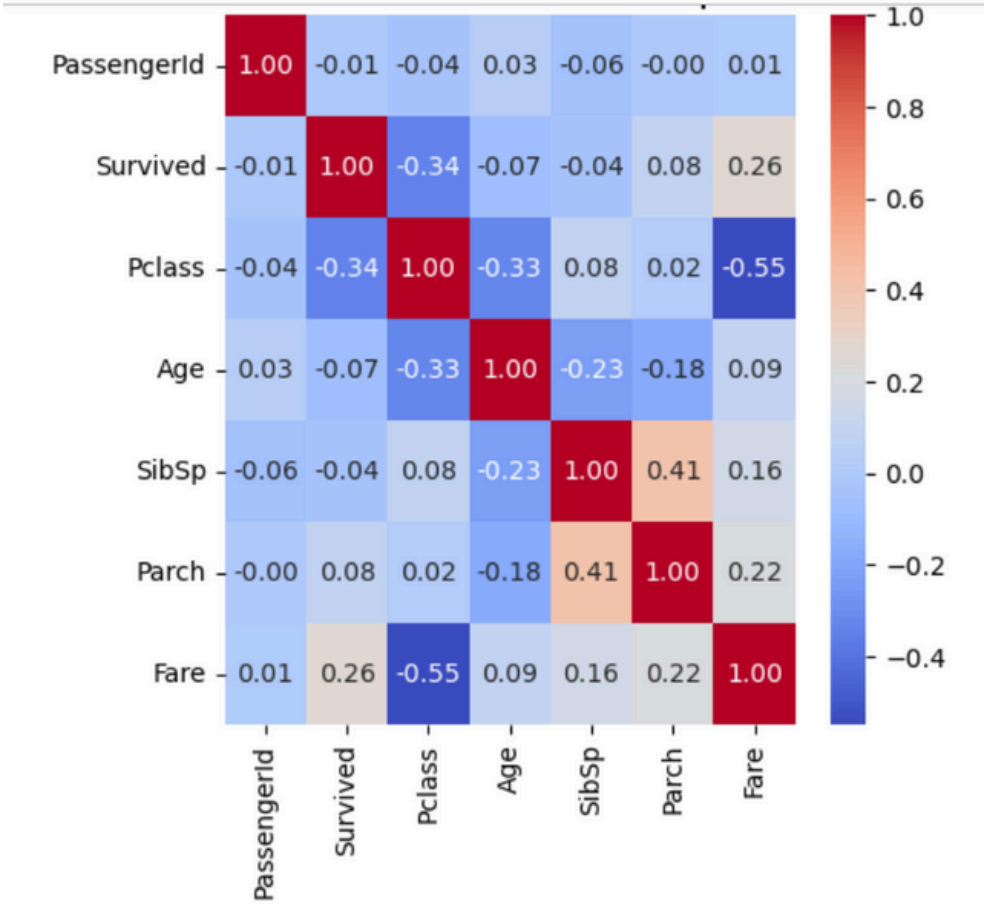
The chart shows Titanic survival by embarkation port. Most passengers boarded at Southampton (S), where more died than survived. Passengers from Cherbourg (C) had a higher survival rate, with survivors outnumbering non-survivors. Queenstown (Q) had the fewest passengers, with more deaths than survivors, indicating embarkation location influenced survival chances.

7. Average Fare by Pclass



The bar chart shows average fares by passenger class on the Titanic. First-class passengers paid the highest fares, averaging above 80, reflecting luxury accommodations. Second-class fares averaged around 20, while third-class fares were lowest, near 13. This clear fare gap highlights the economic differences among passenger classes aboard the ship

7. Correlation Heatmap



The heatmap shows correlations between Titanic dataset features. Fare has a strong negative correlation with Pclass (-0.55), indicating higher fares in higher classes. Survival correlates positively with Fare (0.26) and weakly with Parch (0.08). SibSp and Parch have a notable positive correlation (0.41), suggesting families often traveled with both siblings/spouses and parents/children.

SUMMARY

The Titanic dataset contains variables like PassengerId, Survived, Pclass, Age, SibSp, Parch, and Fare. The correlation heatmap shows Fare negatively correlates with Pclass (-0.55) and moderately with Survived (0.26). SibSp and Parch are moderately related (0.41), indicating family travel patterns. Most other correlations are weak, suggesting limited linear relationships among features, which helps in selecting relevant variables for predictive modeling while avoiding multicollinearity issues

JUPYTER NOTEBOOK LINK