Titanic Dataset Analysis Summary

Major Cleaning and Feature Engineering Steps

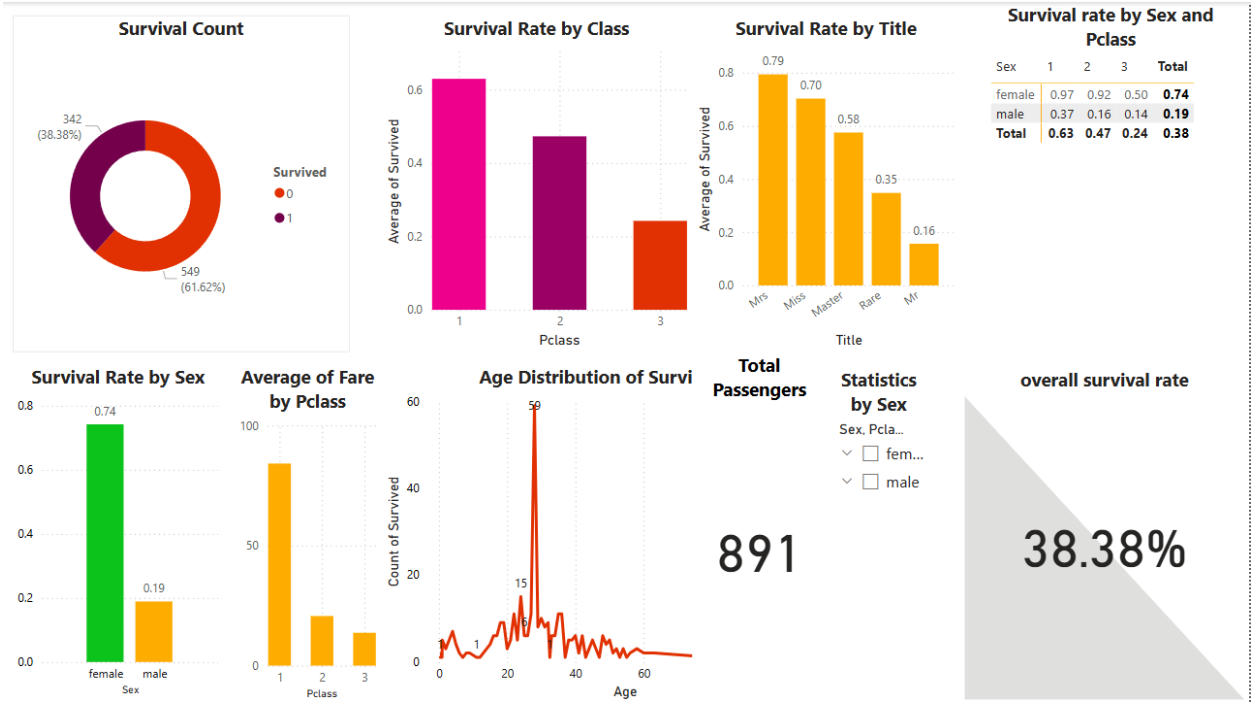
1. **Missing Data Handling: -**
   -Filled missing 'Age' values using median per 'Title'.
   -Dropped or imputed missing 'Embarked' values.
   – Dropped 'Cabin' due to excessive nulls.
2. **Feature Engineering:**
   - Extracted 'Title' from 'Name' (e.g., Mr., Miss, etc.).
   - Binned 'Age' for distribution analysis.
   - Encoded categorical features ('Sex', 'Embarked', 'Title') for modeling.
3. **Column Selection:**

- Retained key predictors: 'Passenger class', 'Sex', 'Age', 'Fare', and 'Title'.

Visual Insights

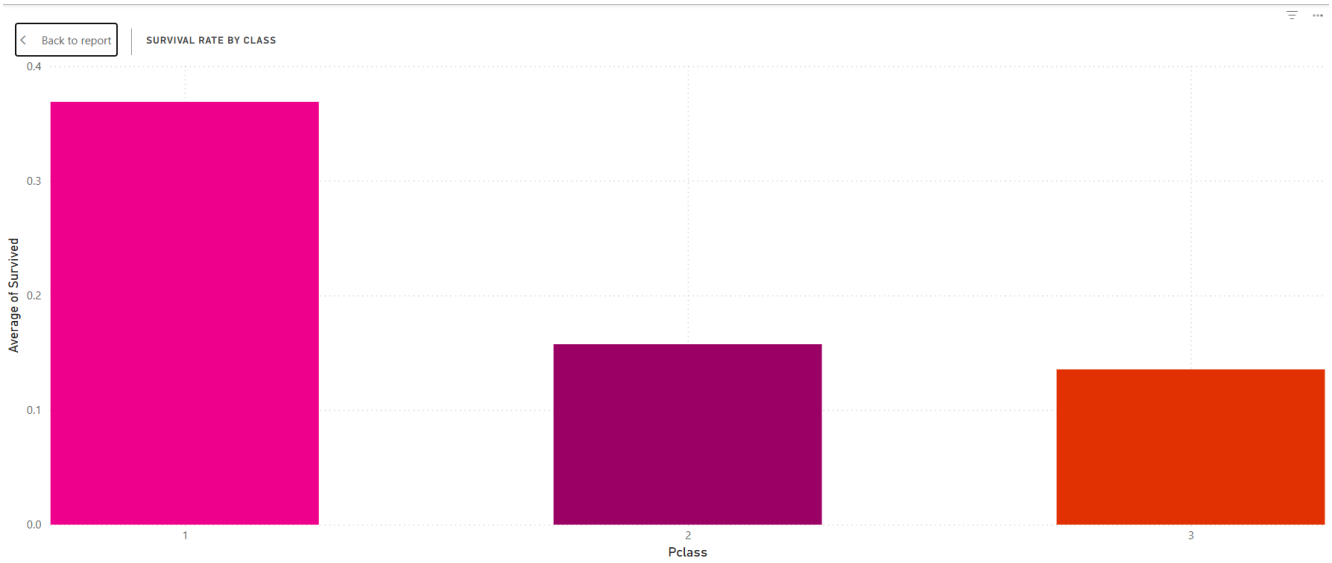
1.Gender Strongly Affects Survival:

- There was on 891 passengers on board ,Females had a survival rate of 74%, while males only 19% and The overall survival rate on the Titanic was 38.38%



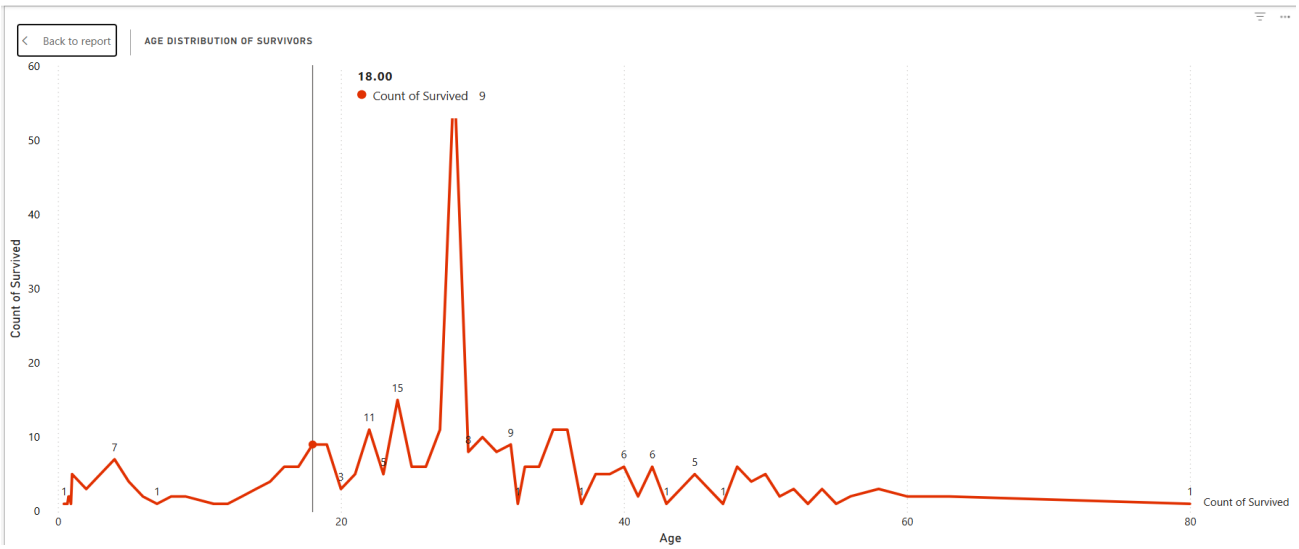
2.Class and Title Matter:

- 1st Class survival rate approximately 63%, compared to 3rd Class approximately 24%. - Titles like 'Mrs.' and 'Miss' had significant higher survival r rates than 'Mr.'.



3. Age Distribution of Survivors:

- Most survivors were young adults, especially females under 40. - Reflects 'women and children first' protocol.



Conclusion & Recommendation

Passengers in 1st class, female, and with titles like Miss or Mrs. were far more likely to survive. For future survival modeling, these features are highly predictive and should be prioritized.

Most Surprising Insight

The Title feature turned out to be more informative than expected. For example, 'Master' (young boys) had better survival odds than adult males in higher classes, showing how social role and perceived vulnerability influenced rescue decisions.