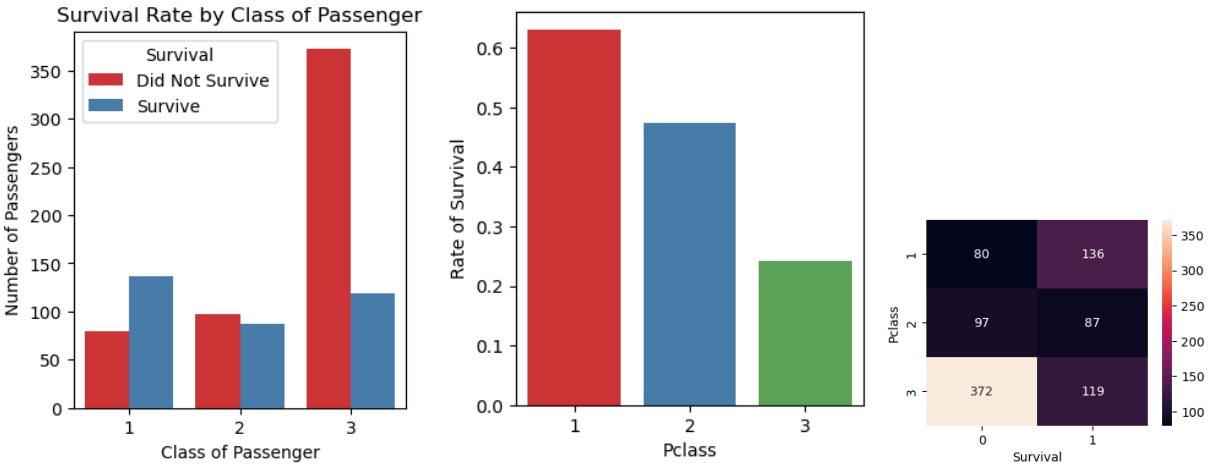
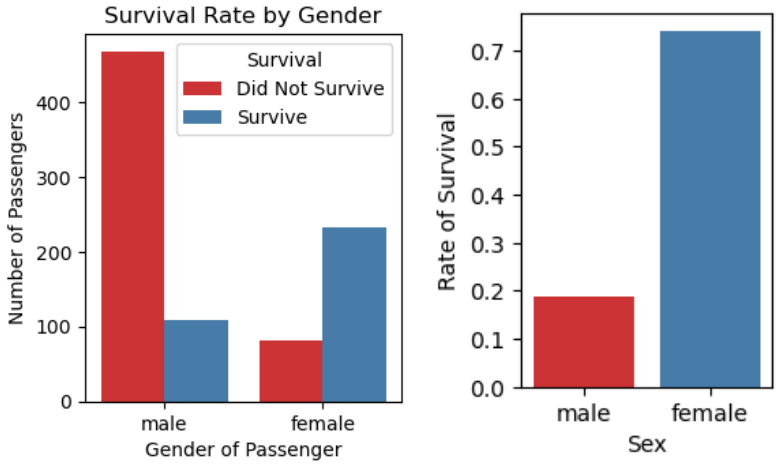
Summary of Titanic Dataset EDA

1. **Survival rate is associated with the class of passenger**



- The left-hand chart displays the count of passengers who survived and those who did not, categorized by their respective classes. It is evident that the largest number of passengers were from third class, and unfortunately, a significant portion of them did not survive the accident.
- As indicated by the left bar chart, passengers in first class have the highest likelihood of survival, while those in third class have the lowest. The survival rate for third-class passengers is notably lower, at less than half of the survival rate for first-class passengers.
- p= '0.000000000000000000000455'. To check the association or correlation between the two categorical variables, we perform chi-square analysis. After performing chi2 analysis, we find that p_value is smaller than 0.05 ($p<0.05$). Therefore, there is a significant association between survival rate and class of passenger.

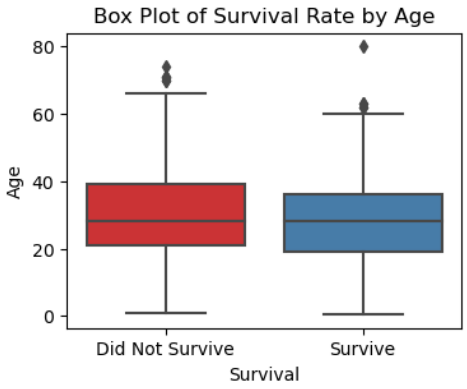
2. **Survival rate is associated with the gender:**

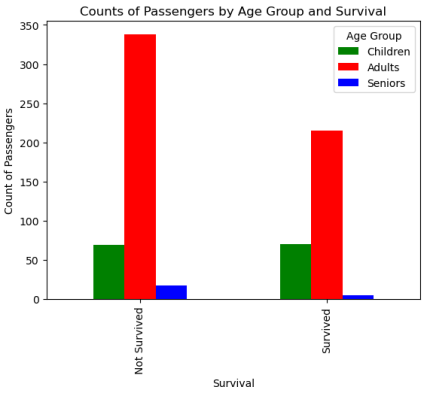


- Observing the right bar plot above, it becomes apparent that the survival rate among females significantly surpasses that of males.
- The bar plot on the left offers a direct comparison of survival and mortality rates for both males and females. It becomes evident that a higher number of males did not survive when contrasted with their chances of survival.
- p= '0.000000000000000000000000000000000000000000000000000000000120'. To check the association or correlation between the two categorical variables, we perform chi-square analysis. After performing chi2 analysis, we find that p_value is smaller than

0.05 ($p<0.05$). Therefore, there is a significant association between gender and survival rate.

3. **Survival rate is associated to the age:**





- In the bar plots shown above, the x-axis is segmented into three age categories: Children (0-18), Adults (18-60), and Seniors (60-100).
- From the information presented in the bar graphs and box plots, it is evident that among these age groups, children exhibit the highest rate of survival.
- p= 0.0106620642852553. To check the association or correlation between the two categorical variables, we perform chi-square analysis. After performing chi2 analysis, we find that p_value is smaller than 0.05 ($p<0.05$). Therefore, there is a significant association between age and survival rate.