

Assignment 3.2

Summary of Titanic Dataset

Name: Mohammad Nazim Rehman

Student Id: 100415501

Hypotheses 1: From analysing the Titanic dataset, we can conclude that the survival rate was associated with the class of passengers. With the bar graph at Page 3 and the analysis, we realize the rate of survival was highest in first class, which is possible as it was the most expensive one and would have more safety than the other two classes.

Hypotheses 2: The survival rate is also associated with gender as we have seen from the trends at Page 4 and various visualisations and information that we have done in the analysis as it comes out that Women had a higher survival rate in all age groups and all classes as well as overall. People may usually tend to save Women and Kids in such disastrous situations first. This can also be verified with the correlation matrix as it is very high for survival and sex.

Hypotheses 3: The survival rate is less associated with age as we can see from the bar graph at Page 4 the survival rate for the kids aged 0-12 years old was not very high as compared to other age groups.

Summary:

Variables: 12, Rows: 891

Size of dataset: 10,692

Outliers:

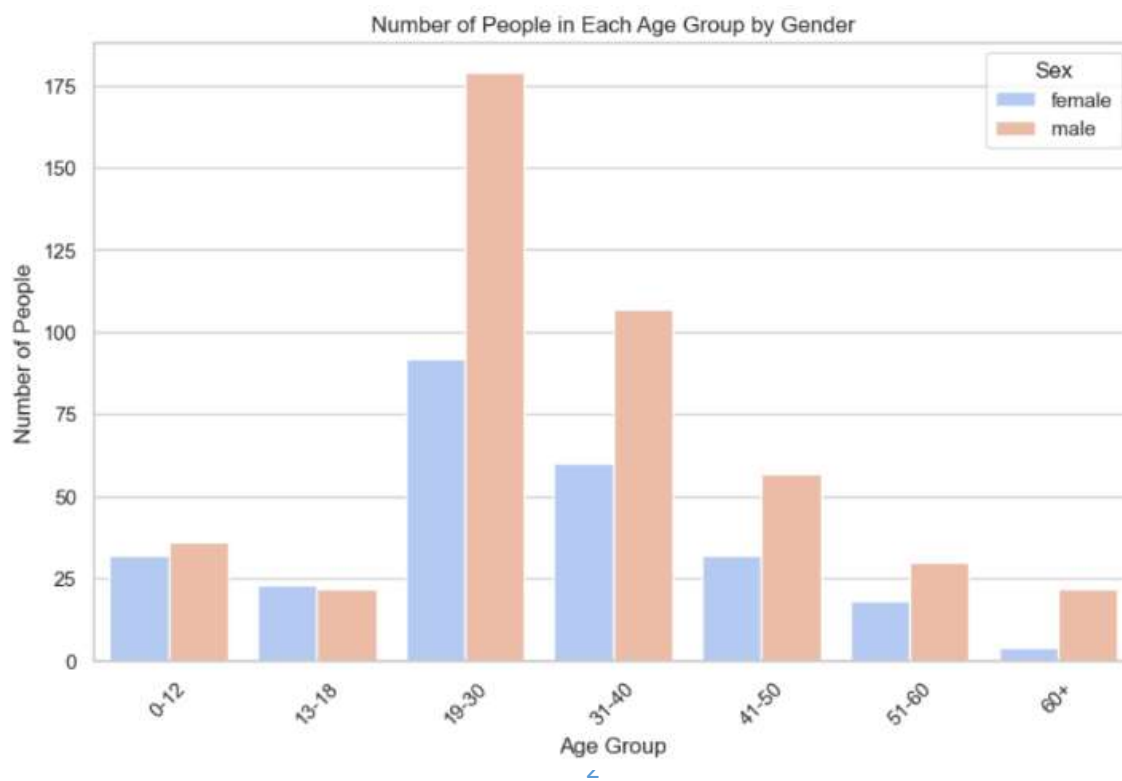
Upon analyzing the features, the **Fare** variable appears to exhibit potential outliers. This is likely due to the distribution of fare prices, where certain ticket prices may seem unusually high. For example, for the **SibSp** feature (siblings and spouses aboard), having a family with eight members might be plausible, especially given the context of 1912. The same reasoning applies to the **Parch** feature (parents and children), where larger families could have been travelling together.

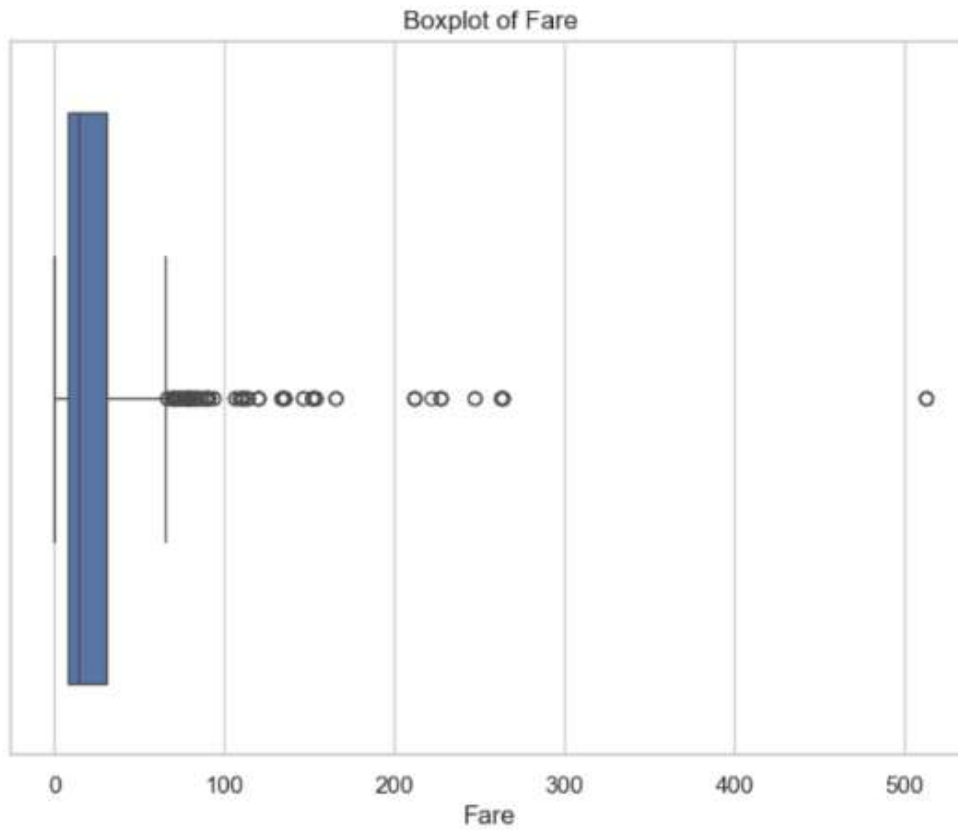
The same goes for age as the data depicts that there were passengers as young as 0 years old and as old as 80 years and after careful consideration and fact check we have found out the actual range for age which was of 0 years – 74 years.

Summary for Numerical Variables

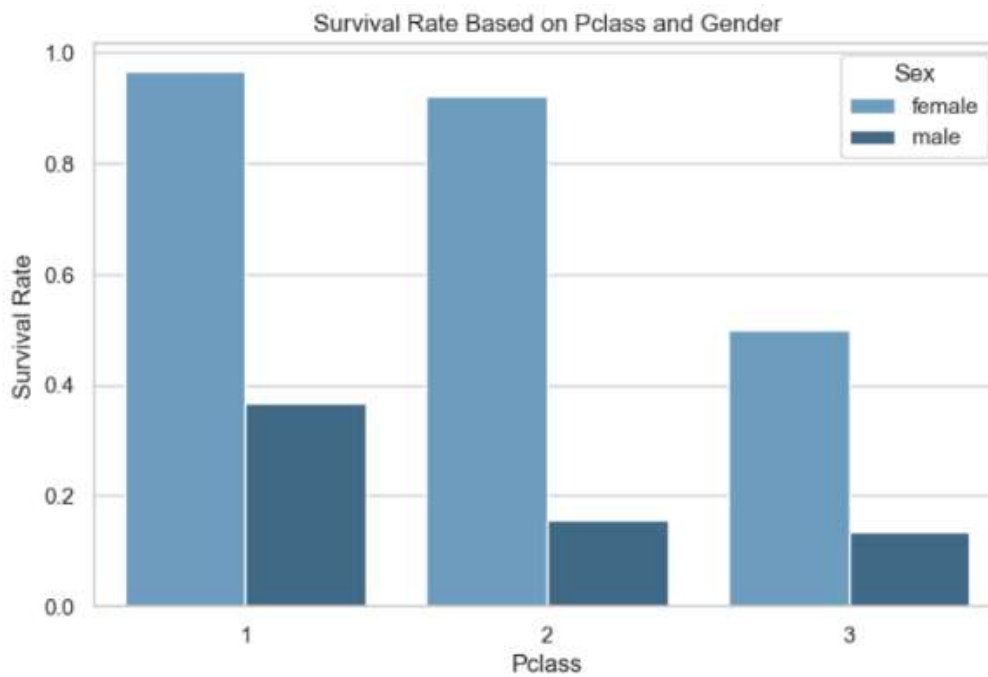
	PassengerId	Age	SibSp	Parch	Fare
count	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	29.699118	0.523008	0.381594	32.204208
std	257.353842	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	38.000000	1.000000	0.000000	31.000000
max	891.000000	80.000000	8.000000	6.000000	512.329200

```
Sex
male    577
female  314
Name: count, dtype: int64
Survived
0      549
1      342
Name: count, dtype: int64
Pclass
3      491
1      216
2      184
Name: count, dtype: int64
Embarked
S      644
C      168
Q       77
Name: count, dtype: int64
```

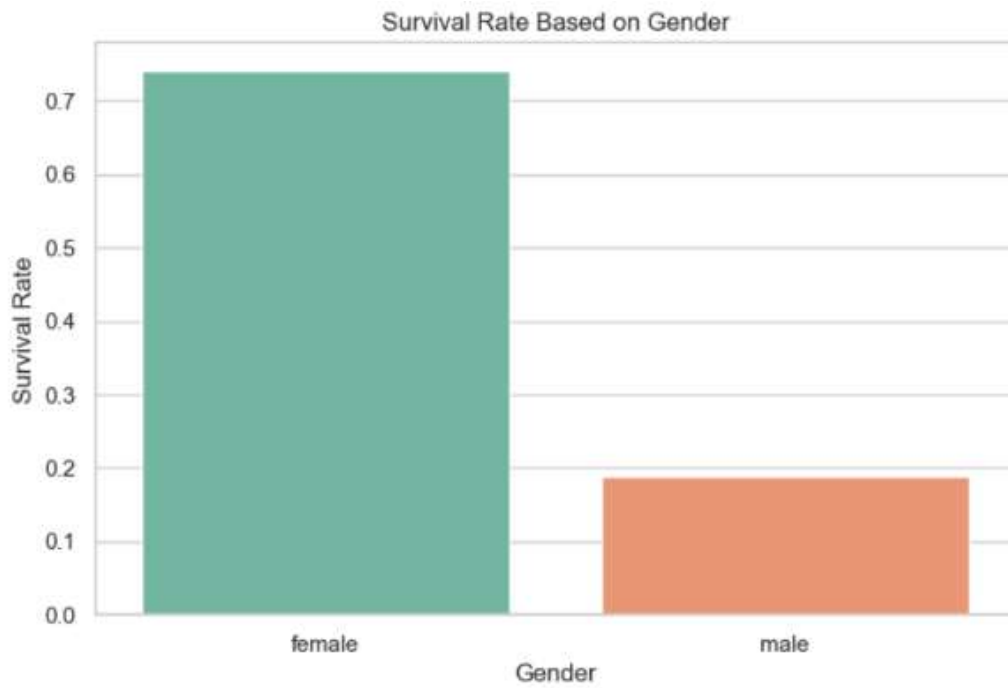




Hypotheses 1:



Hypotheses 2:



Hypotheses 3:

