Name: Gaurav Gautam
StudentId: 100379780

Summary of Titanic Dataset Analysis

This report summarizes the dataset provided and is also available at Kaggle. Below are the variables and their corresponding data type. As part of exploratory data analysis, validation of datatype, Null Values, Accuracy check is being done. Also during the EDA, visualization through data frame and seaborn is also leveraged to validate the correlation between the variables.

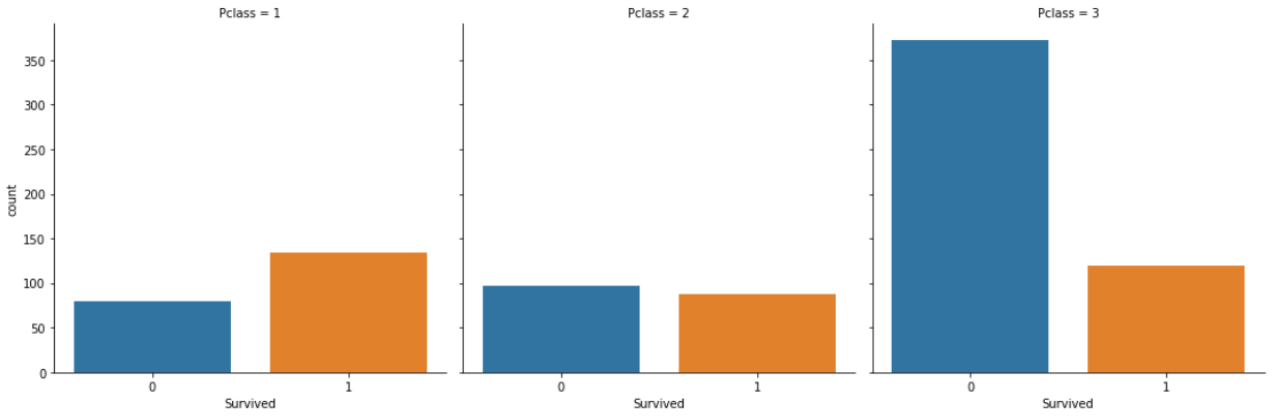
In this dataset, we can have analysis against the "Survived" variable and statistically prove or disapprove that other variables like Gender, P-class, etc will contribute to the higher chances of higher survival rate. In the dataset, the "Survived" variable has two values i.e

0	Not Survived
1	Survived

Following Hypothesis is carried forward:

1) **If the survival rate is associated with the class of passengers (Pclass vs Survival)**:

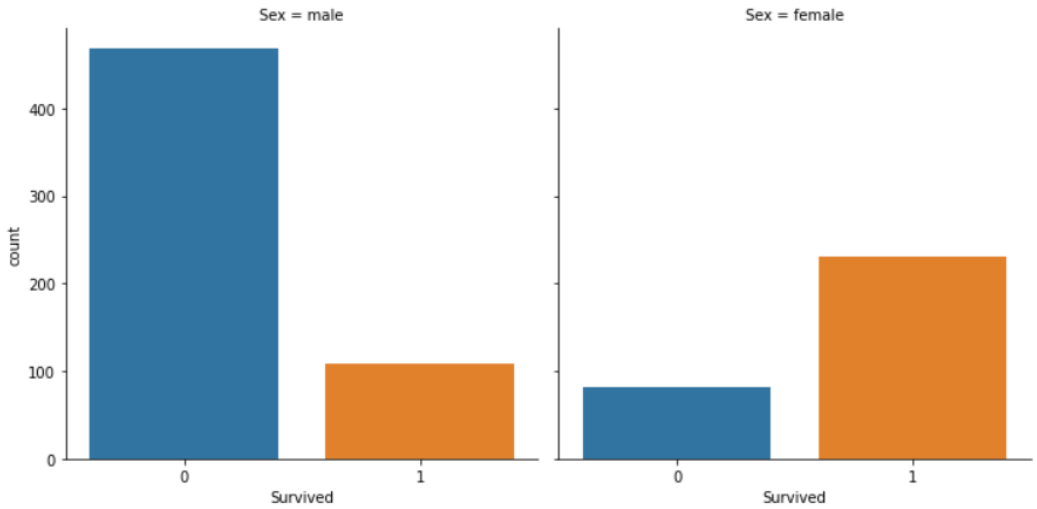
```
In [103]: sns.catplot(x='Survived',col='Pclass',kind='count',data=df);
```



We can see that the survival rate decreases with the Pclass. The Chi-Square test also proves these variables to be Dependent with a p-value greater than 0.05.

2) **If the survival rate is associated with the gender of passengers (Sex vs Survival)**:

```
In [104]: sns.catplot(x='Survived',col='Sex',kind='count',data=df);
```



As we can see in the above graph, the Female survival rate is significantly higher than the Male survival rate. The Chi-Square test also proves these variables to be Dependent with a p-value greater than 0.05.

3) **If the survival rate is associated with the Embarked variable (Embarked vs Survival)**:

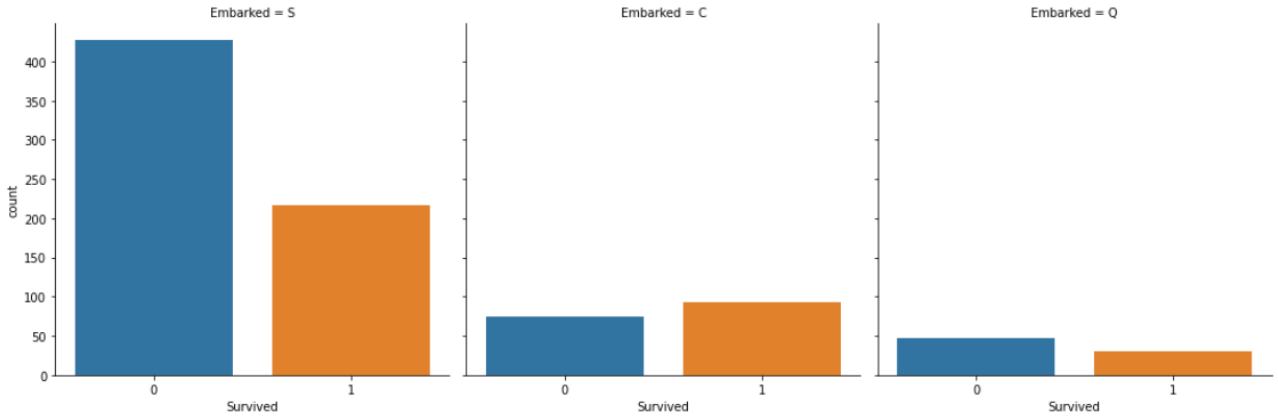
Embarked means from where the passengers mounted from.

S = Southampton port

C = Cherbourg Port

Q = Queenstown Port

```
In [107]: sns.catplot(x='Survived',col='Embarked',kind='count',data=df);
```



It is evident from the above graph, the Survival rate is correlated with the Embarked variable as passengers mounted from Southhamptopn port survival rate is significantly higher. The Chi-Square test also proves these variables to be Dependent with a p-value greater than 0.05.