ASSIGNMENT 3
SUMMARY

I.   Data Observation

The count of values for each variable is 891. Here are the details on each variable.

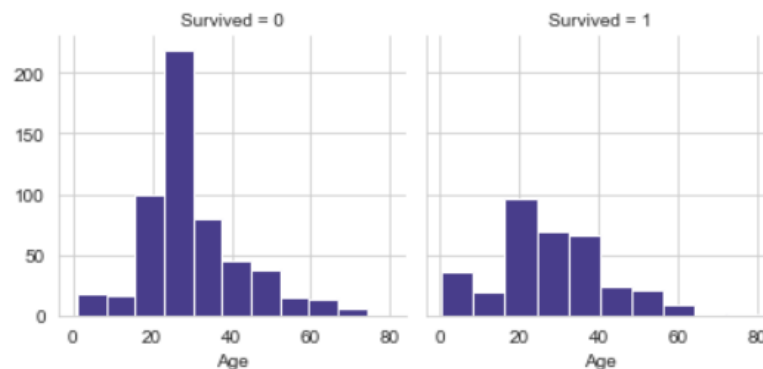|	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	male	Q	S
count	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.066409	0.523008	0.381594	32.204208	0.647587	0.086420	0.722783
std	257.353842	0.486592	0.836071	13.244532	1.102743	0.806057	49.693429	0.477990	0.281141	0.447876
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	22.000000	0.000000	0.000000	7.910400	0.000000	0.000000	0.000000
50%	446.000000	0.000000	3.000000	26.000000	0.000000	0.000000	14.454200	1.000000	0.000000	1.000000
75%	668.500000	1.000000	3.000000	37.000000	1.000000	0.000000	31.000000	1.000000	0.000000	1.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200	1.000000	1.000000	1.000000

CLEANING AND REPLACING NULL VALUES

As we examine the variables, we observed null values. There are 177 null values under the Age Variables, 687 on the Cabin and 2 on the Embarked as shown in the table and graph below. The null values were replaced by mean value based on age and passenger class. If Passenger Class is 1, replace it with value of 37. If Passenger Class is 2, replace it with value of 29. If Passenger Class is 3, replace it with value of 24. Cabin column was dropped as there are 687 null values.
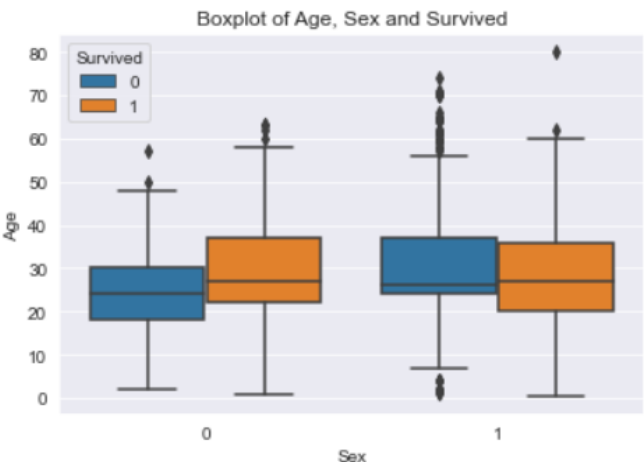
II.   Association with Survival Rate

A. Age and Survival Rate

Based on the graph below, it is more likely that the age that can survive is within the bracket of age 20 to 40.
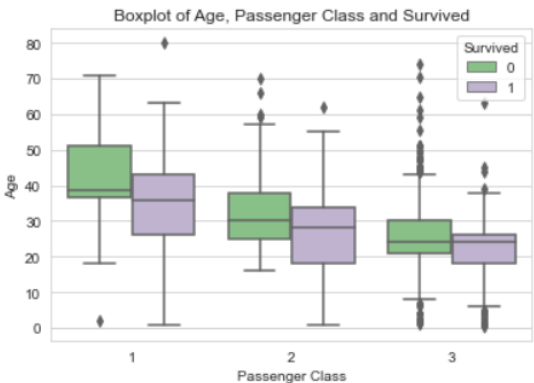
The graph also shows that it is more likely that the age that can survive is within the bracket of age 20 to 40 regardless of gender.
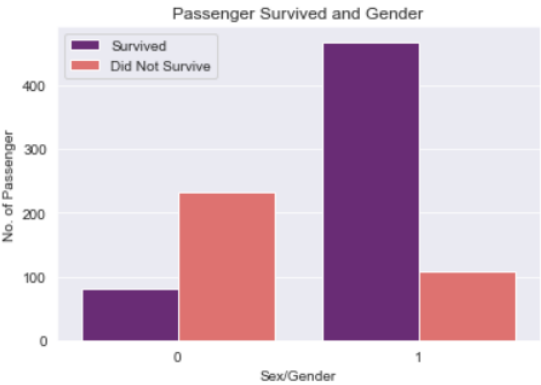

The graph below illustrated that it is estimated that the age that can survive in each passenger class are the following:

    a. Majority is 28 to 42 years old
    b. Majority is 19 to 35 years old
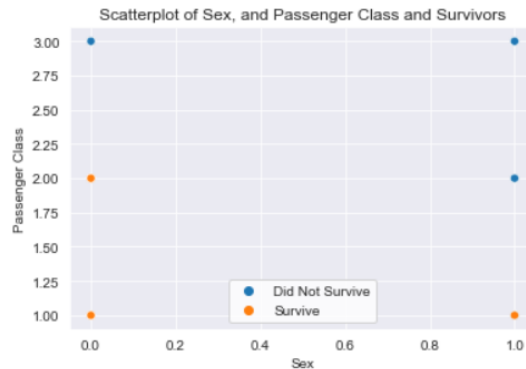   c. Majority is 19 to 26 years old

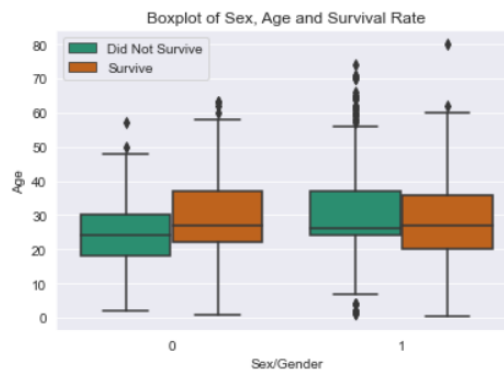## B. Sex and Survival Rate

Based on the graph below, it shows that the Female is likely to survive than Male.

In addition, by considering also the passenger class, more females in Passenger Class 1 survived.
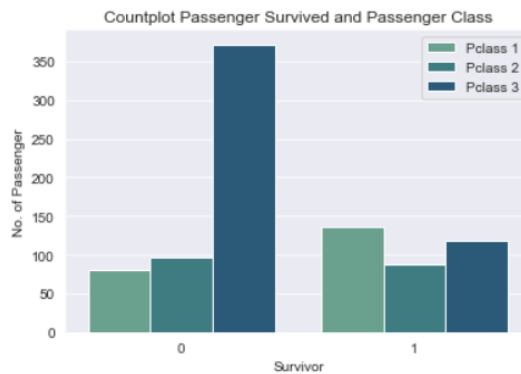
Scatterplot of Sex, and Passenger Class and Survivors

Further, as we inspect the age for those who survived under Female category, it is estimated that the age is 20-38 years old.
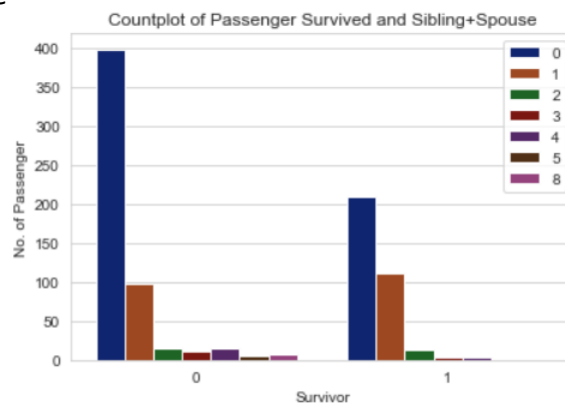

Boxplot of Sex, Age and Survival Rate

C. Passenger Class and Survival Rate

Based on the graph below, it shows that it is more likely that the passengers under passenger class 1 will likely survive, followed by the passengers in class 3.

Countplot Passenger Survived and Passenger Class

D. No. of Siblings and Spouse Aboard and Survival Rate

The graph illustrates that those passengers with 0 and 1 spouse/siblings are are more likely to survive

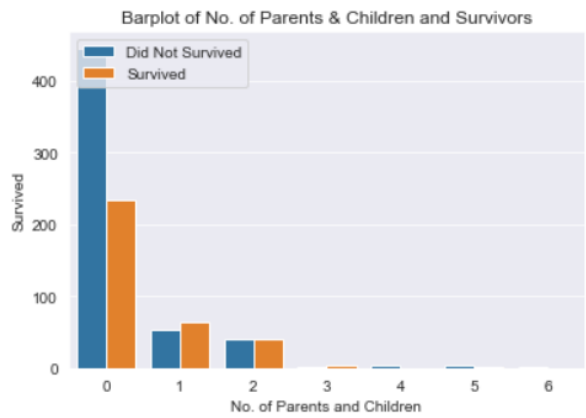Countplot of Passenger Survived and Sibling+Spouse

It is also interesting note that those passengers in passenger Class 1 with 0-3 only of siblings including spouse has high survival rate comparing to other class.

Scatterplot of Siblings/Spouse, and Passenger Class and Survivors

E. No. of Parents and Children Aboard and Survival Rate

The graph shows that those passengers with 0 and 1 parent/s and children will more likely to survive.

Barplot of No. of Parents & Children and Survivors

The graph below depicts that those in passenger Class 1 with 0-2 parents and children aboard may have a more survival rate.

Scatterplot of Parents/Children, and Passenger Class and Survivors