ASSIGNMENT 3

I. Data Observation

A. VARIABLES

Based on the data, there 12 variables such as the following:

1. Passenger ID – ID number of Passenger
2. Survived – Whether the passenger survived or not
3. PClass – Passenger Class - Categorical
4. Name – Name of passenger
5. Sex – Gender of passenger
6. Age – Age of Passenger
7. Sibsp - Siblings and Spouse Aboard
8. Parch -Parents and Children Aboard
9. Ticket – Ticket Number
10. Fare – Passenger Fare
11. Cabin – Cabin of Passenger
12. Embarked – Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

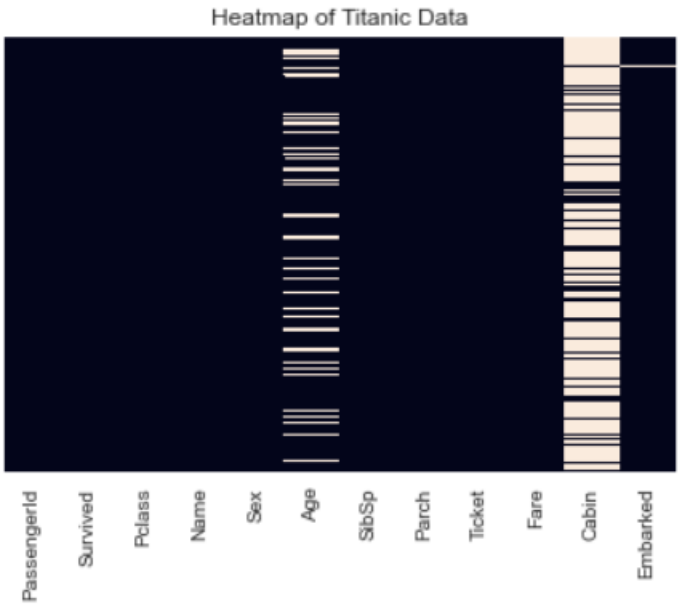
Also, the count of each variable is 891. Here are the details on each variable.

	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

B. NULL VALUES

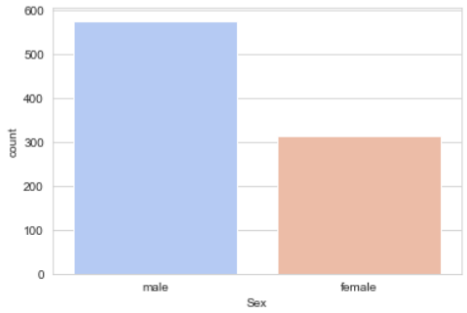
As we examine the variables, we observed null values. There are 177 null values under the Age Variables, 687 on the Cabin and 2 on the Embarked as shown in the table and graph below.

```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

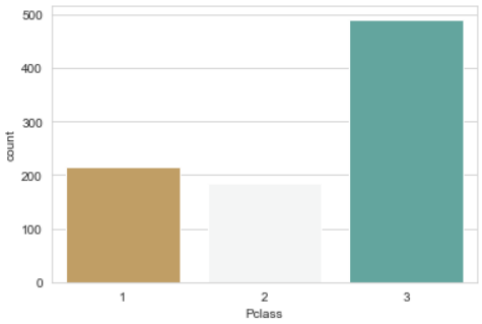


C. DATA OBSERVATION IN VARIABLES

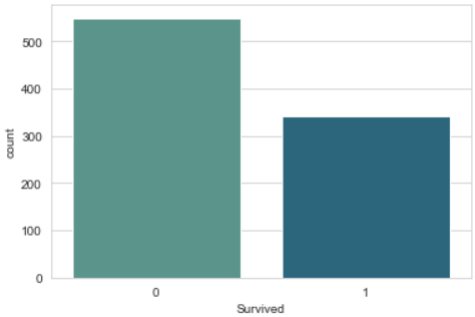
2. **Sex** – The categories under sex is Male and Female.



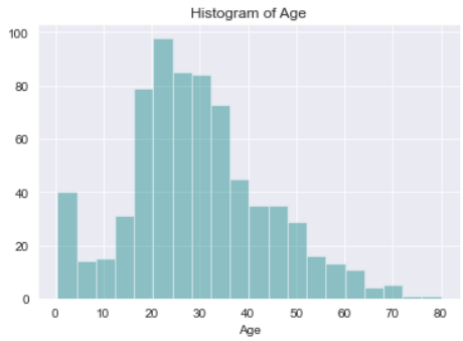
1. **PClass** – The categories under PClass is 1, 2 and 3.



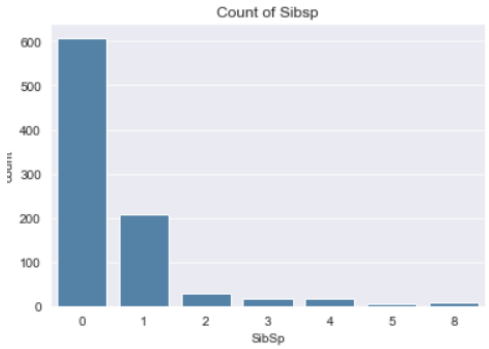
4. **PClass** – The categories under PClass is 1, 2 and 3.



3. **Age –** Below are the histogram of Age



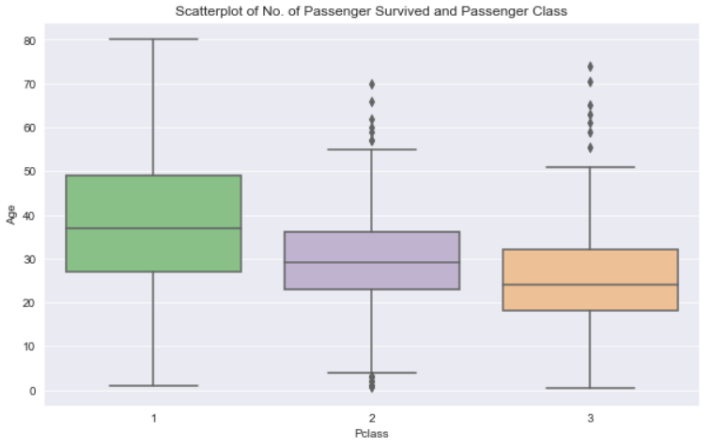
5. **SibSp –** Siblings and Spouse Aboard have minimum number of 0 and maximum number of 8



6. **Fare –** The minimum number of Fare is 0 and maximum number is 512.

II. Data Cleaning

A. Age



Based on the graph beside, replace the null values:

1. If Passenger Class is 1, replace it with value of 37
2. If Passenger Class is 2, replace it with value of 29
3. If Passenger Class is 3, replace it with value of 24

B. Cabin

Given that Cabin has 687 null values, it is best to drop and do not use this data.

C. Sex

Replace Sex by 1 for Male and 0 for Female

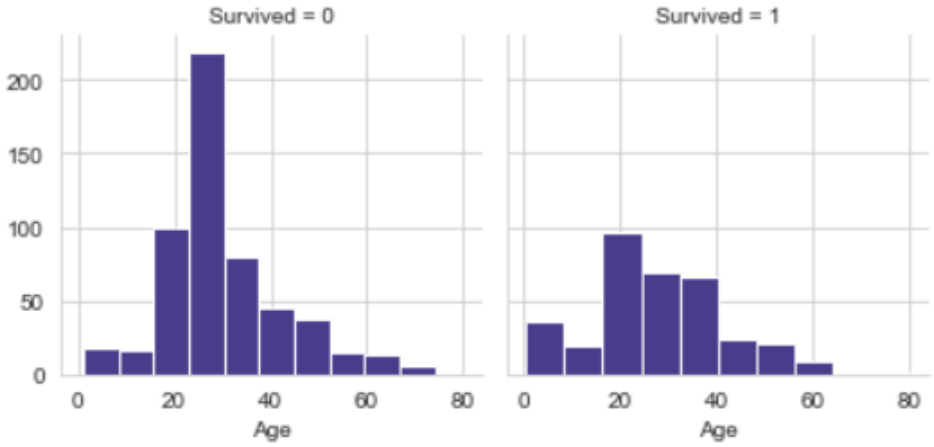
D. Embarked

Separate the column for port of entry of C, Q and S and replace 0 for False and 1 for True.

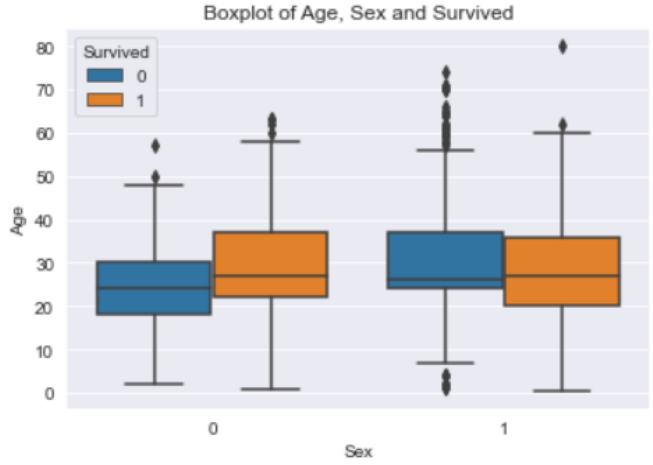
III. Association with the Survival Rate

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.

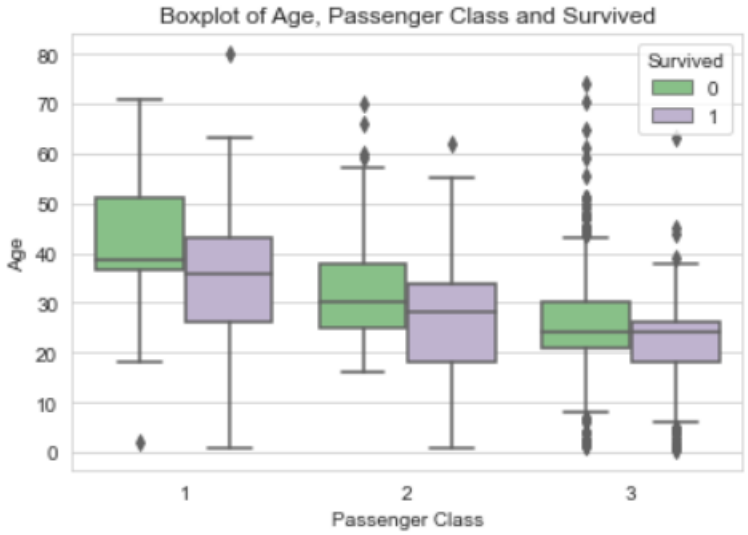


The graph also shows that It is more likely that the age that can survive is within the bracket of age 20 to 40 regardless of gender.



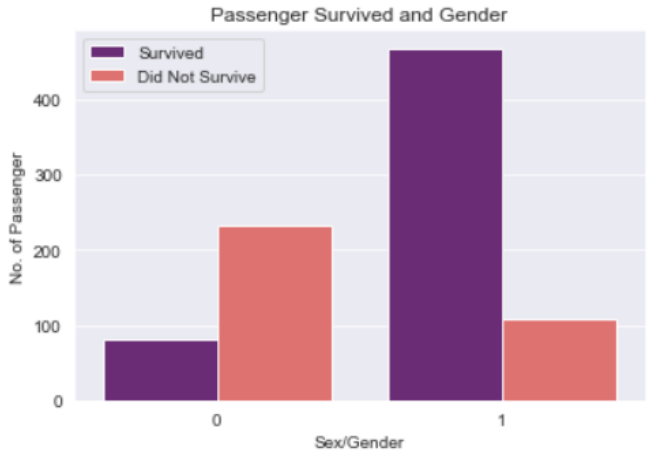
The graph below illustrated that it is estimated that the age that can survive in each passenger class are the following:

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

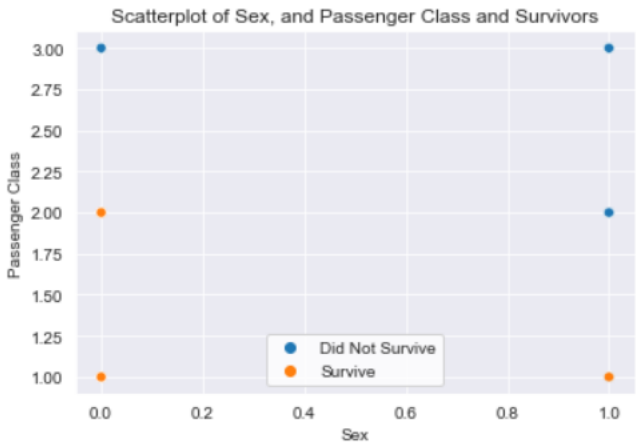


B. Sex and Survival Rate

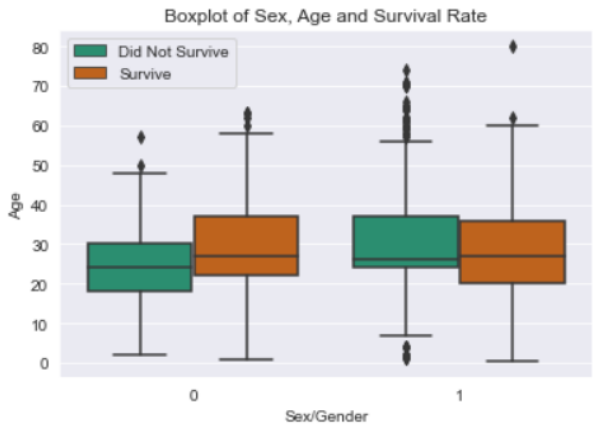
Based on the graph below, it shows that the Female is likely to survive than Male.



In addition, by considering also the passenger class, more females in Passenger Class 1 survived.

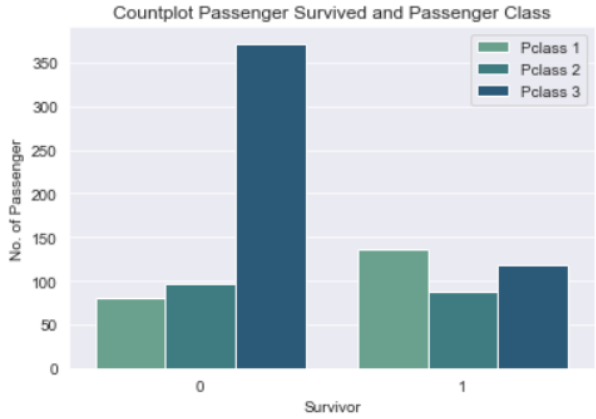


Further, as we inspect the age for those who survived under Female category, it is estimated that the age is 20-38 years old.



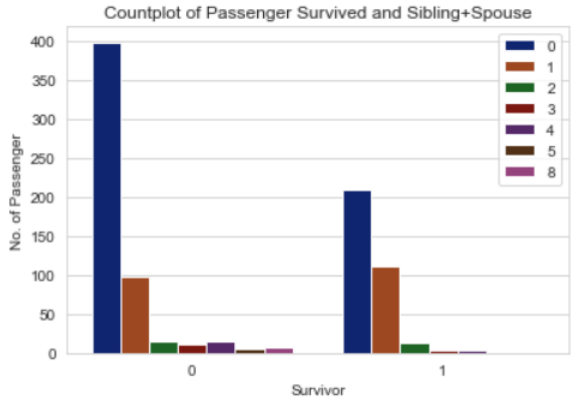
C. Passenger Class and Survival Rate

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.



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

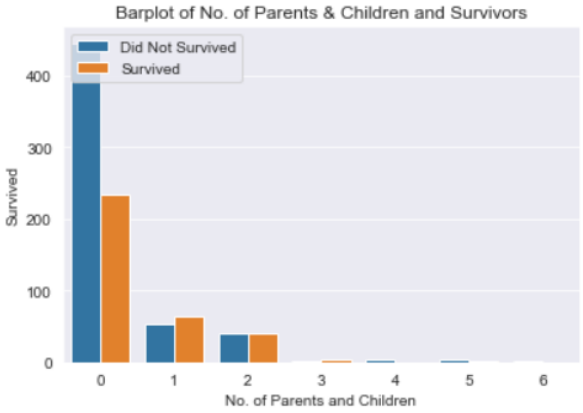
The graph illustrates that those passengers with 0 and 1 spouse/siblings are are more likely to survive.



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.

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.



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