CPSC 4800 – Fall 2021
Assignment 3

Name: Emilio Sagre
Student Number: 100371610

Introduction

The dataset is about the details of the passengers who board the RMS Titanic ship operated by the White Star Line. The ship sank in the North Atlantic Ocean on 15th of April 1912, according to Wikipedia. The data consist of details of who survived and did not survive during the accident with the following details:

1. Survival - Survival (0 = No; 1 = Yes)
2. Pclass - Passenger Class (1 = 1st class; 2 = 2nd class; 3 = 3rd class)
3. Name - Name
4. Sex - Gender
5. Age - Age
6. Sibsp - Number of Siblings/Spouses Aboard
7. Parch - Number of Parents/Children Aboard
8. Ticket - Ticket Number
9. Fare - Passenger Fare
10. Cabin - Cabin Code
11. Embarked - Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)

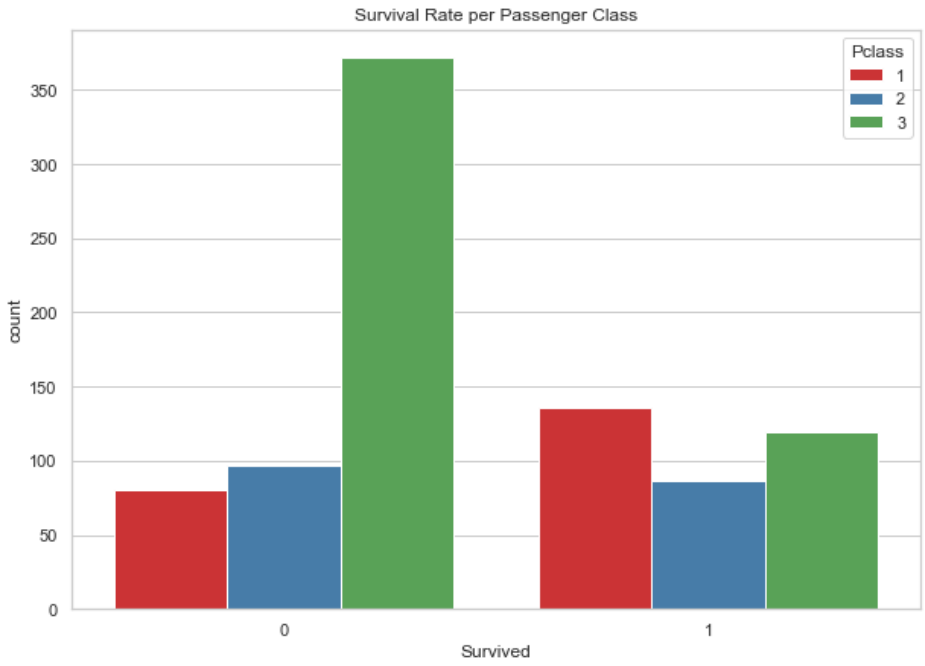
The dataset has 891 rows.

Analysis

1. To determine if the survival rate associated to the class of passenger grouped dataset based survival column in comparison to its passenger class.

Determine if the survival rate is associated to the class of passenger

In [169]:
```
# CHECK SURVIVAL RATE PER PASSENGER CLASS
sns.set_style('whitegrid')
sns.countplot(x='Survived', hue='Pclass', data=titanic, palette='Set1').set(Title = 'Survival Rate per Passenger Class')
```



In [156]:
```
titanic[['Pclass', 'Survived']].groupby(['Pclass'], as_index = False).mean().sort_values('Survived', ascending = False)
```

Out[156]:

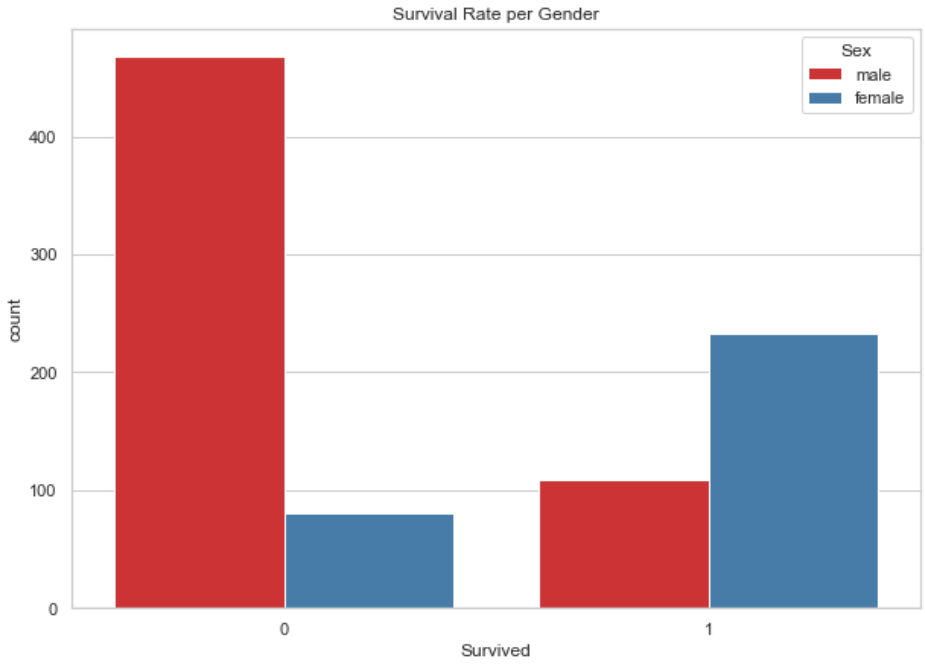
	Pclass	Survived
0	1	0.629630
1	2	0.472826
2	3	0.242363

Observation:
Based on the data, 62.9% from class 1 had survived (highest among the class) compared to the lowest class at 24.2% only.

2. To determine if the survival rate associated to the gender, I grouped the data survivors based on their sex.

Determine if the survival rate is associated to the gender

In [168]:
```
# CHECK SURVIVAL RATE PER GENDER
sns.set_style('whitegrid')
sns.countplot(x='Survived', hue='Sex', data=titanic, palette ='Set1').set(Title = 'Survival Rate per Gender')
```



In [154]:
```
titanic[['Sex', 'Survived']].groupby(['Sex'], as_index = False).mean().sort_values('Survived', ascending = False)
```

Out[154]:

	Sex	Survived
0	female	0.742038
1	male	0.188908

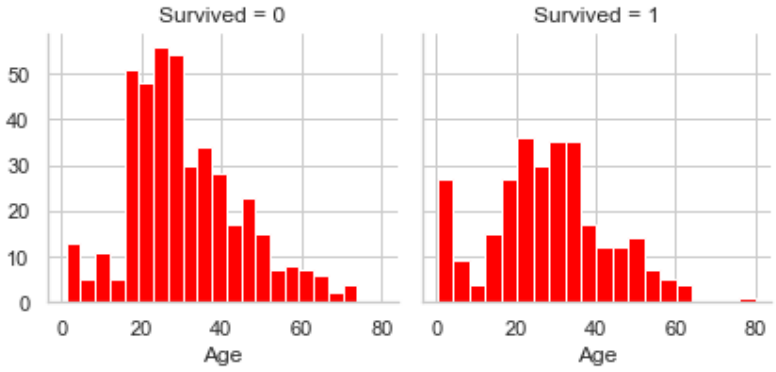
Observation:
Grouping the data by gender, 74.2% of the female has survived compared to male at 18.8% only.

3. To determine the survival rate associated to the age, I created side-by-side histogram based to the number of survivors. Then I created a boxplot to check the age mean and interquartile range of the survivors.

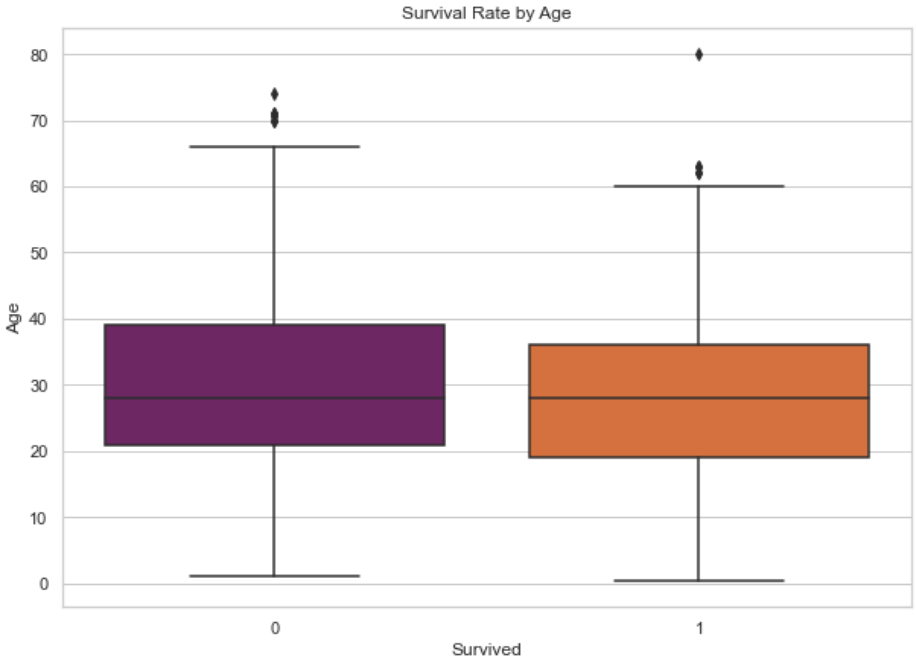
Determine the survival rate is associated to the age

In [204]:
```
#  CHECK SURVIVAL RATE PER AGE THROUGH HISTOGRAM
sns.set_style('whitegrid')
sAge = sns.FacetGrid(titanic, col = 'Survived')
sAge.map(plt.hist, 'Age', bins = 20, color = 'red')

```



In [209]:
```
# BOXPLOT FOR SURVIVAL RATE
plt.figure(figsize=(10,7))
sns.boxplot(x= 'Survived', y = 'Age', data=titanic, palette = 'inferno').set(title = 'Survival Rate by Age')
```



In [208]:
```
titanic[['Survived', 'Age']].groupby(['Survived'], as_index = True).mean().sort_values('Age', ascending = False)
```

Out[208]:

Survived	Age
0	30.626179
1	28.343690

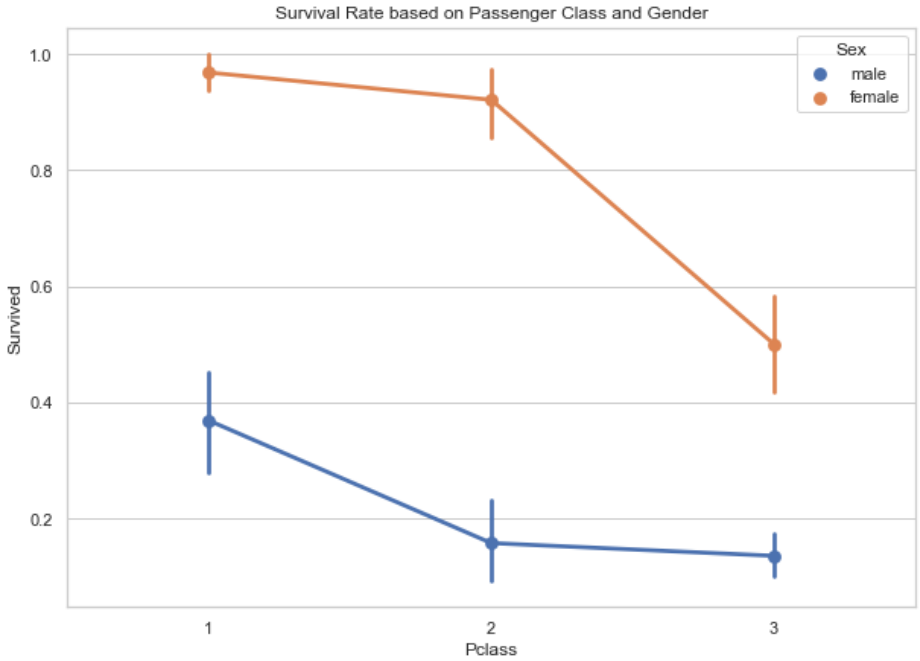
Observation:

Based on the histogram, while majority of the passenger who survived and did not survive are concentrated from ages 20 to 40. However, the passengers with lower age group had a high survival rate. The average age of those who survive is at 28 compared to did not at 31.

4. Survival rate based on Gender and Passenger Class

In [212]:
```
# CHECK IF SURVIVAL RATE IS ASSOCIATED WITH GENDER AND PASSENGER CLASS

plt.figure(figsize = (10,7))
sns.set_style('whitegrid')
sns.pointplot(data = titanic, x = 'Pclass', y = 'Survived', hue = 'Sex').set(title = 'Survival Rate based on Passenger Class
```



In [213]:
```
titanic[['Survived', 'Sex', 'Pclass']].groupby(['Pclass', 'Sex']).mean()
```

Out[213]:

		Survived
Pclass	**Sex**	
1	**female**	0.968085
	male	0.368852
2	**female**	0.921053
	male	0.157407
3	**female**	0.500000
	male	0.135447

Observation:

If we will compare the survival rate based on gender and passenger class, the highest survival rate were the females in class 1 at 96.8% compared to female at the lowest class at only 50%. However, the males at the highest class had a lower survival rate at 36.8%.