CPSC 4800

Assignment 3 Part-2

Introduction

The sinking of the Titanic is one of the most infamous shipwrecks in history.

On April 15,1912, the RMS Titanic sank while on her first voyage after colliding with an iceberg. The Titanic was considered an engineering marvel and widely regarded as unsinkable. This complacency might have been one of the reasons why there were not enough lifeboats on the ship. As a result of this there were a lot of casualties.

The titanic.csv dataset contains passenger information along with the information that whether they survived or died in the accident. In this report I have analyzed 3 of these factors and tried to see their correlation with the survival rate. The 3 factors analyzed in this report are:

1 - Pclass

2 - Sex

3 - Age

Below is the summary statistics for the numerical/continuous variables in the given Titanic dataset:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

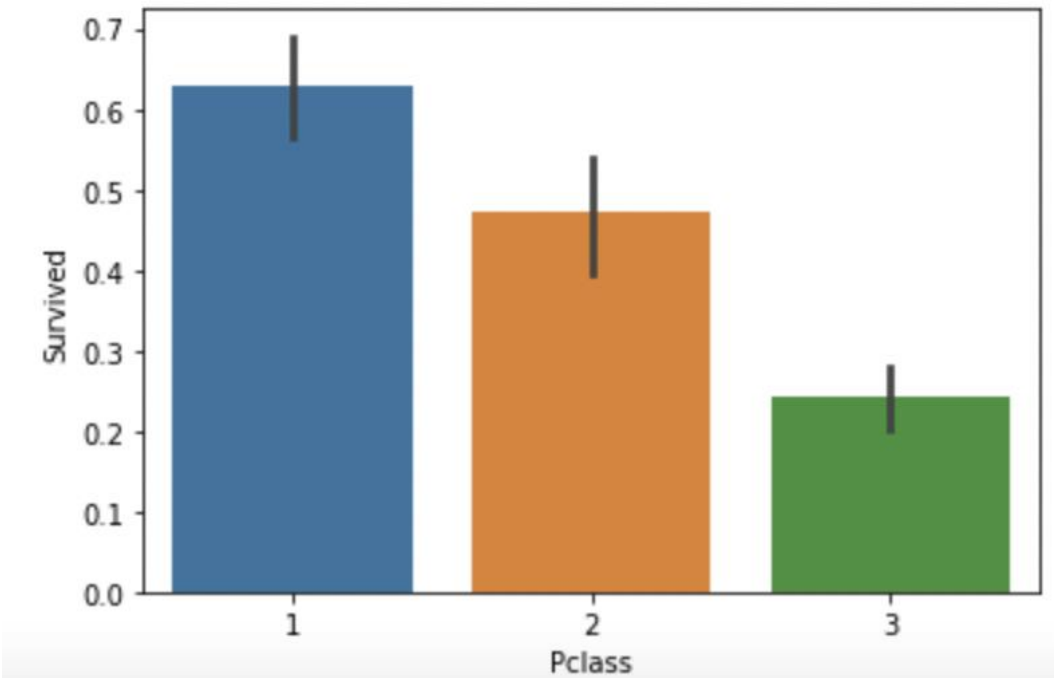
We can see there are 891 records/observations in the dataset. We notice that the mean/average age of the passengers was 29.6 years. Also, of the 891 passengers,

age data is available for 714 passengers. The class data is available for all the 891 passengers. There are 3 classes of passengers: 1,2 and 3 with 1 being the highest class.

1 - Is survival rate associated with the class of passenger?

Below is a table and Bar Chart with the percentage of passengers in each class who survived:

Pclass	Survived
1	0.629630
2	0.472826
3	0.242363



We can clearly see that passengers in first class had the maximum survival rate of 62.9 percent. Passengers in second class had a survival rate of 47.28% whereas those in third class had a survival rate of only 24.23%.

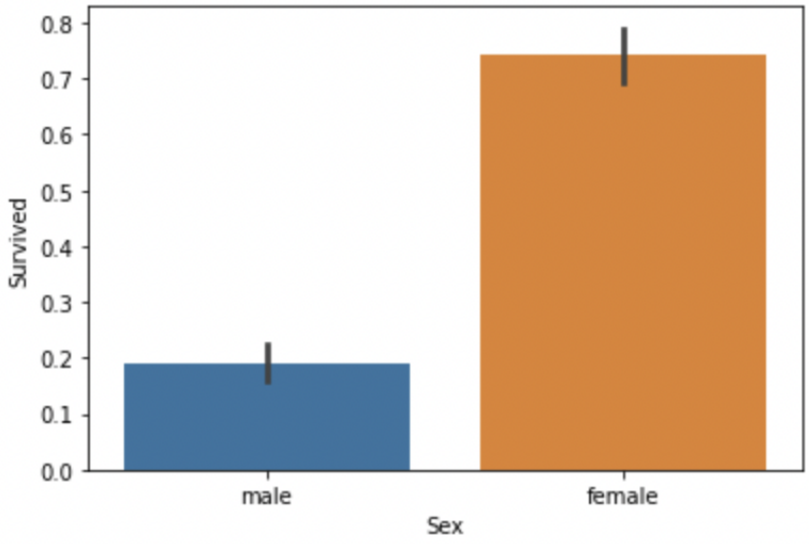
2 - Is the survival rate associated with the gender of passenger?

In the below table we can see the survival rate percentage based on class and gender:

		Survived
Sex	**Pclass**	
female	**1**	96.81
	2	92.11
	3	50.00
male	**1**	36.89
	2	15.74
	3	13.54

In Hypothesis 1 test we noticed that passengers in 1st class had a higher survival rate and here we can see that gender is also related to the survival rate. Females in First class had the maximum chances of survival at 96% whereas males in 3rd class had the least survival rate at 13.54%.

In general females had a survival rate of 74.20% compared to males who had a survival rate of 18.89%. The Bar Chart below depicts the survival rates based on Sex/Gender. We can see that Females had a higher survival rate.



3 - Is the survival rate associated to the age?

For analyzing the relationship between age and survival rate in this report I have converted the age data to integers and for each age group calculated the survival percentage.

So, suppose there are 5 passengers aged 15 of whom 4 survived the accident. In this case the survival rate for age 15 would be [4/5=.8] 80%. Similarly, the survival rate percentage for all ages is calculated and plotted against age in a scatter plot. Below are the top 5 ages where the survival rate percentages were highest:

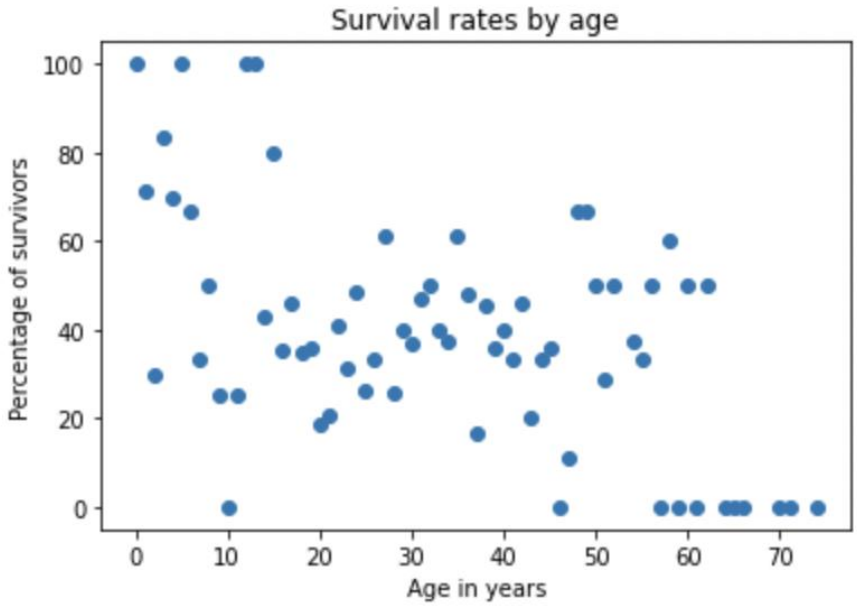
```
df_survival_age.sort_values(by=['Percentage'],ascending=False).head(5)
```

	Survived	Deaths	Total	Percentage	Age
0	7	0	7	100.0	0
12	1	0	1	100.0	12
63	2	0	2	100.0	63
53	1	0	1	100.0	53
13	2	0	2	100.0	13

Checking for the correlation between age and survival percentage we see that the correlation between the variables is –0.392 i.e., the variables are negatively correlated which means that as one variable increases the other decreases. In the below scatter plot for Age vs Survival Percentage, we notice that passengers of lower age had the highest survival percentages.

Correlation between Age and Survival Percentage

	Age
Survived	-0.392468



Conclusion

In this report we can see that passengers in First Class had the maximum survival rate. Similarly, the survival rates for female and younger passengers were better than male and older passengers.

While there was some element of luck involved in surviving; some groups of people were more likely to survive than others.