CPSC 4800 – Assignment 3 – Part 2

EDA Report

Hypothesis 1: Determine if the survival rate is associated to the class of passenger.

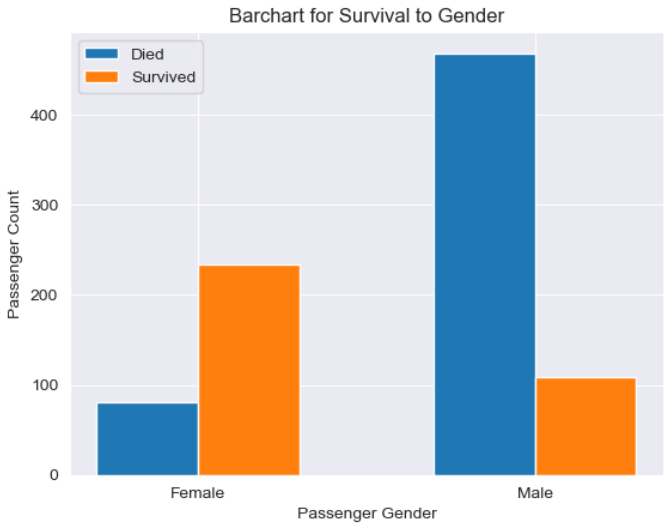
A side-by-side bar graph has been created to determine whether the survival rate of passengers is associated with the class of the passenger. Here, as the group of bars do not look alike, they are not independent from each other.

```
Chi-squared value: 102.88898875696056
Degree of freedom: 2
P-value: 4.549251711298793e-23
```

After this a chi-squared analysis has been run to verify our results. Here, we look at the chi-square statistic and compare it to the decision point value corresponding to a degree of freedom of 2 which is 5.99. Since the test statistic (102.89) is more than the DP, we have enough evidence to conclude that the variables (survival rate and class) are not significantly independent to each other among all subjects in the population.

Based on this analysis we can conclude that survival rate is associated to the passenger class.

Hypothesis 2: Determine if the survival rate is associated to the gender.



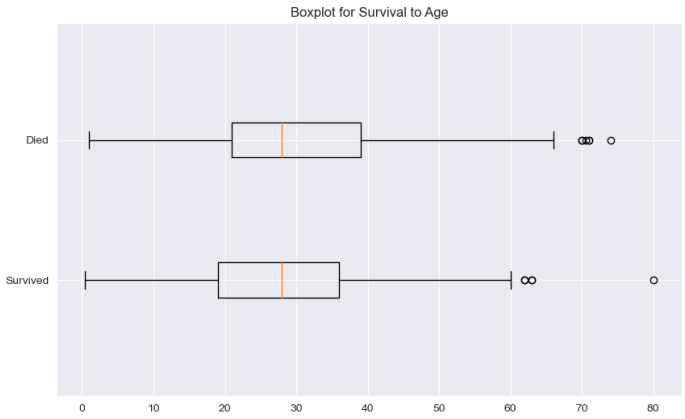
A side-by-side bar graph has been created to determine whether the survival rate of passengers is associated with the gender of the passenger. Here, as the group of bars do not look alike, they are not independent from each other.

```
Chi-squared value: 260.71702016732104
Degree of freedom: 1
P-value: 1.1973570627755645e-58
```

After this a chi-squared analysis has been run to verify our results. Here, we look at the chi-square statistic and compare it to the decision point value corresponding to a degree of freedom of 1 which is 3.84. Since the test statistic (260.72) is more than the DP, we have enough evidence to conclude that the variables (survival rate and gender) are not significantly independent to each other among all subjects in the population.

Based on this analysis we can conclude that survival rate is also associated with the gender of the passenger.

Hypothesis 3: Determine the survival rate is associated with age.



The median age is the same for those who did not survive compared to those who did. There seems to be more variation in age for the passengers that did not survive than for passengers who had survived. Both groups have the presence of outliers but the passengers who did not survive have more outliers in the data set.

	Age
count	290.000000
mean	28.343690
std	14.950952
min	0.420000
25%	19.000000
50%	28.000000
75%	36.000000
max	80.000000

On the left-hand side, we can see some statistics of passengers who had survived. On the right-hand side we can see the statistics of passengers who had not survived. From these statistics we can observe that age was not a factor associated with the survival rate of passengers, as all the statistics are very close together, with minor deviance.

	Age
count	424.000000
mean	30.626179
std	14.172110
min	1.000000
25%	21.000000
50%	28.000000
75%	39.000000
max	74.000000