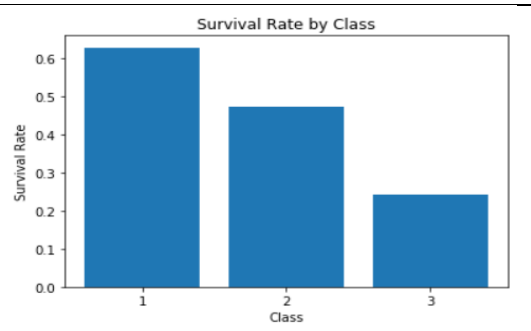


Hypothesis 1: the survival rate is associated to the class of passenger

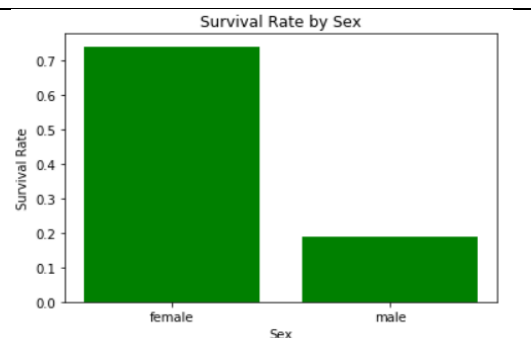
P class	Survivals	Total number of passengers	Survival Rate
1	136	216	0.63
2	87	184	0.47
3	119	491	0.24



From the table and bar graph, we can conclude that the more expensive tickets passengers bought, the more survival rate passengers had. Therefore, hypothesis 1 can be concluded to be true.

Hypothesis 2: the survival rate is associated to the gender

Gender	Survivals	Total number of passengers	Survival Rate
Female	233	314	0.74
male	109	577	0.19

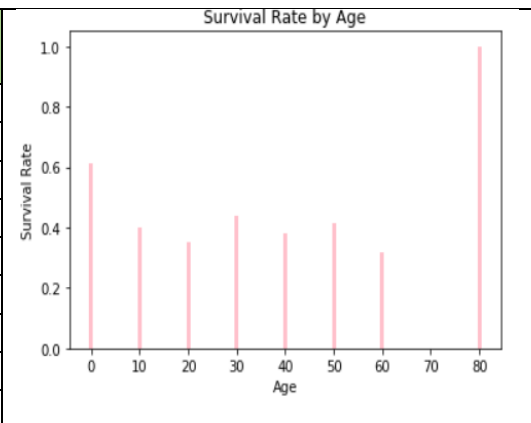


Female passengers have about 4 times more chances to survive from the accident according to the table and graph. Hence, hypothesis 2 can be considered to be true.

Hypothesis 3: the survival rate is associated to the age

Method1: Using 'Age' as a categorial variable

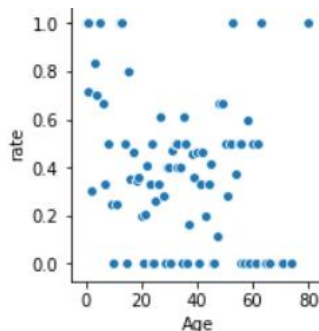
Age Class	Survivals	Total number of passengers	Survival Rate
0	38	62	0.61
10	41	102	0.40
20	77	220	0.35
30	73	167	0.44
40	34	89	0.38
50	20	48	0.42
60	6	19	0.32
70	0	6	0
80	1	1	1



The result shows that the passengers whose age is falling onto 10 or 80 in age class have more chances to survive. However, I cannot assure that the younger or the older passengers are, the more chances they had because passengers in their 30 survived more than passengers in their 20 and passengers in their 40. Thus, it is necessary to check correlation coefficient

Method2: Using 'Age' as a numerical variable

The correlation coefficient between 'Age' and 'rate', which means survival rate here, is about -0.37. It indicates that age and survival rate have negative and weak relation to each other as we cannot dictate clear linear form in scatter plot.



	Age	Survived	total	rate
Age	1.000000	-0.248333	-0.198554	-0.367963
Survived	-0.248333	1.000000	0.906307	0.301918
total	-0.198554	0.906307	1.000000	0.042893
rate	-0.367963	0.301918	0.042893	1.000000