The titanic data set which I chose to analysis consist of information on 1309 passengers abord the titanic. The variables in this data set include gender, ticket class, age, survivability and destination The code I wrote provides an exploration on the demographics of the passengers abord and illustrates some trends and anomalies within the data.

**1.1: Data Cleansing**

The data which I received had many missing values, which could affect the visualisation negativity as shown below. I coded a heat- map to represent the missing values and the effectiveness of the following code which I wrote.

Chart

Description automatically generated

Firstly, the embarked column had basically no missing values, which in the grand scheme of things would not affect the data, therefore inserted “Southampton”. Secondly, the deck colour had a significant number of missing values which would be essential useless when visualising trends, therefore I removed the entire variable. Thirdly, I used a random sampling technique to fill in the values in the age variable, by calculating the mean and standard deviation and randomizing ages between the values of, mean – standard deviation and mean + standard deviation. This produced a heat map which looks like:

Chart

Description automatically generated

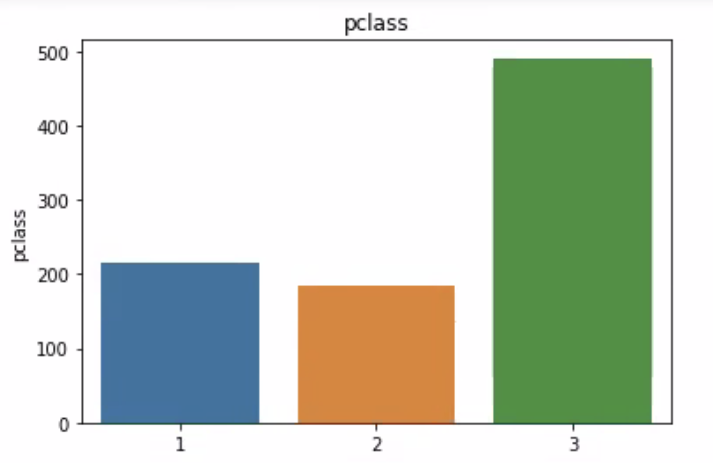
**2.1: Data Visualisation:**

For my demographics I used a wide range of techniques which include, bar plots, cat plots, boxplots, count plots, violin plots and scatterplots. I divided up the variables using a data frame and named them under categorical and numerical data columns. Some questions which I wanted to investigate by using this data set is some of the following.

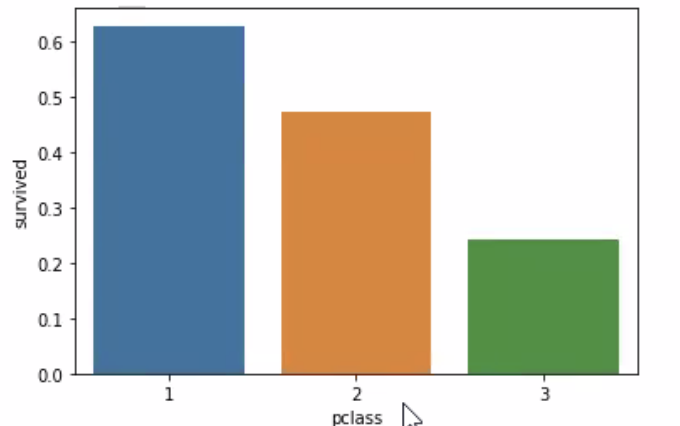
1. Does gender impact your survivability?
2. Does ticket class impact your survivability?
3. Does age impact your survivability?
4. Does the location of embarkation impact your survivability?

**2.2: Ticket class impacting survivability**

From the graph shown below, we can determine that first class, second class and third class represent approximately 24,20,55% of the passengers on board respectfully.



To understand the correlation between the survivability and the ticket class another visualisation is required as shown below:



Although the data uses odd values on the y axis, the data which is shown is approximately accurate. Many studies, which analysed the data stated that there is a correlation between the survivability and ticket class. The correlation was the higher the passengers ticket class was, the higher the chance of survival, which can be depicted in the graph above.

**2.3: Gender impacting survivability**

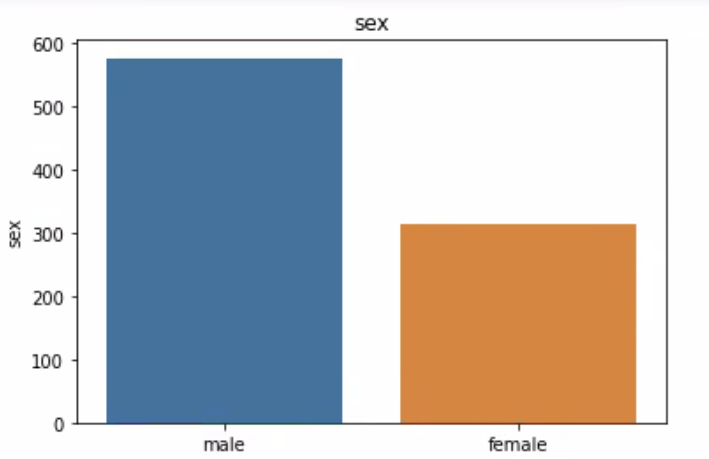


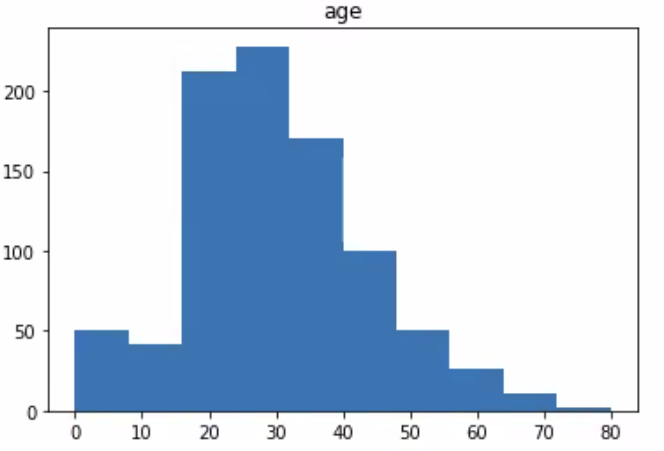
Figure : Gender

Chart, box and whisker chart

Description automatically generated

Figure 2 above shows the correlation between sex and survivability. It shows that women has a higher chance of survival by an enormous amount.

**2.4: Age and Survivability**



**Chart, radar chart

Description automatically generated**

The violin plot above shows that early 20-year-olds were more likely to die compared to everyone below 20 who have a higher chance of survivability. Overall, children had a higher survival rate.

**2.5: Embark Location and Survivability**

**Chart, bar chart

Description automatically generated**

Southampton = S

Queenstown = Q

Cherbourg = C

**Chart, bar chart

Description automatically generated**

Both graphs above shows that passage from Cherbourg, Queenstown and Southampton made up 19 , 9 and 72 % of the headcount respectfully. More typically however, 27,9 and 72% of the survivors were from Cherbourg, Queenstown and Southampton respectfully. After some analysis, it was discovered that passengers from Cherbourg were more likely to purchase higher class tickets, therefore leading to a higher survivability rate.