Project 3

Framing Questions:

This is a very interesting data set. As mentioned in the description in Kaggle, even though the main reason for the loss of lives was a lack of sufficient number of lifeboats, there seem to be other factors involved as well. Data analysis can be helpful in determining the factors that may have influenced the survival chances of passengers.

Some of the questions that I intend to explore in this dataset are as follows:

(1) Is passenger class a factor?
(2) Is sex of the passenger a factor?
(3) Is age group a factor?

Data Wrangling:

A command was run to find out the missing values. Following results were obtained:

```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
Count            0
dtype: int64
```

For my analysis, since I am working with Class, Sex and Age only the missing values in age is a concern. Therefore, I will remove the missing values during the age analysis.

titanic_df_edited=titanic_df.dropna()

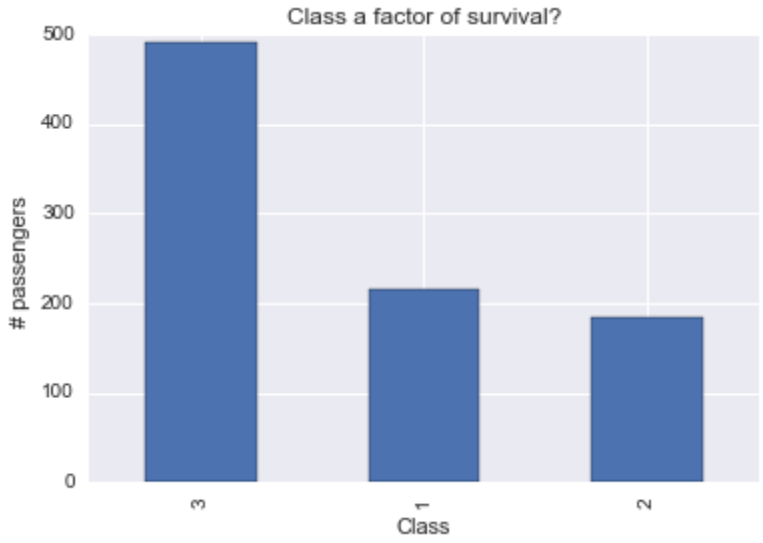
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Cabin          0
```

```
Embarked        0
Count           0
```

Question 1: To determine whether passenger class is a factor or not?

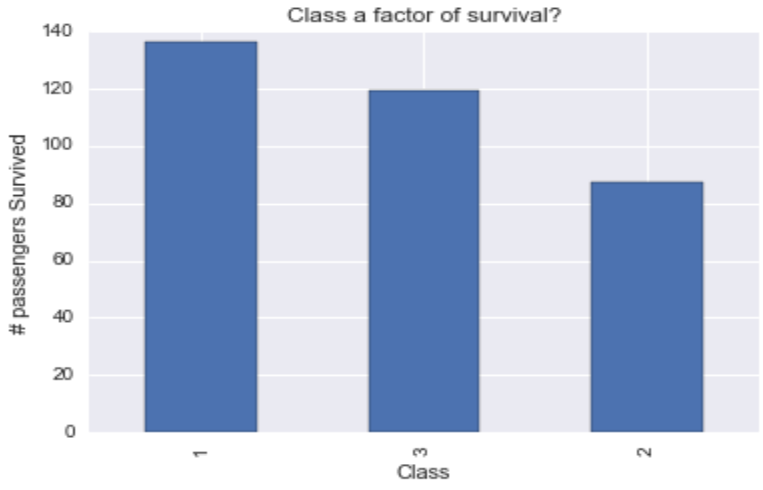
First I wrote a code to sort the passengers by class. Following bar graph shows the distribution:

```
3    491
1    216
2    184
```



After this, I grouped the data by passenger class using the groupby function in pandas. Then the sum function was called to find the total number of passengers that survived in each class. Following results were obtained:

```
1    136
3    119
2     87
```



This chart does not provide any useful information. Even though the number of passengers that survived from third class are lower than those from 1st and 2nd class, it would be wrong to draw any conclusions from this data alone.

Perhaps a per cent comparison would be better!

% survivors in 1st class= 62.96%
% survivors in 2nd class= 47.28%
% survivors in 3rd class= 24.23%

Conclusion: The above plot shows that there may be a correlation between passenger's class and their survival chances. One may draw a conclusion that upper class passengers were given a priority over lower classes in lifeboats. (We also saw this in the movie!)

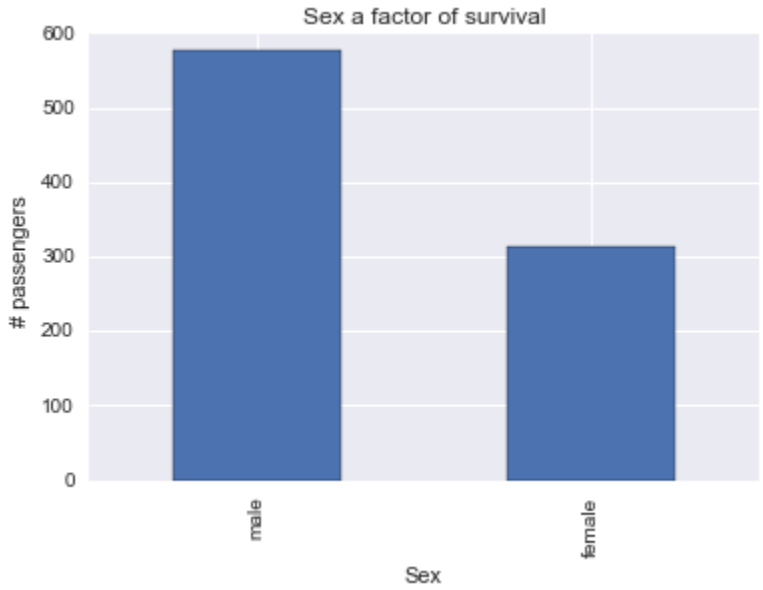
Similar analysis can be performed to answer question 2,
Question 2: To determine whether passenger sex is a factor or not?

For the rest of the analysis similar to question 1, let's skip the discription and look at the graphs.

To make it easier, I added a column in the data spreadsheet which will give the total count.

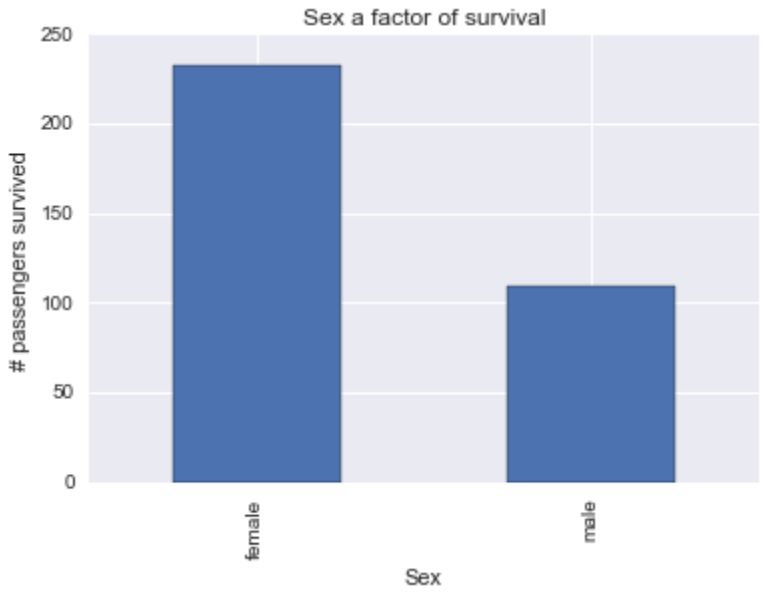
Following results are obtained:

```
male      577
female    314
```



Now let's look at the number of passengers survived:

```
female    233
male      109
```

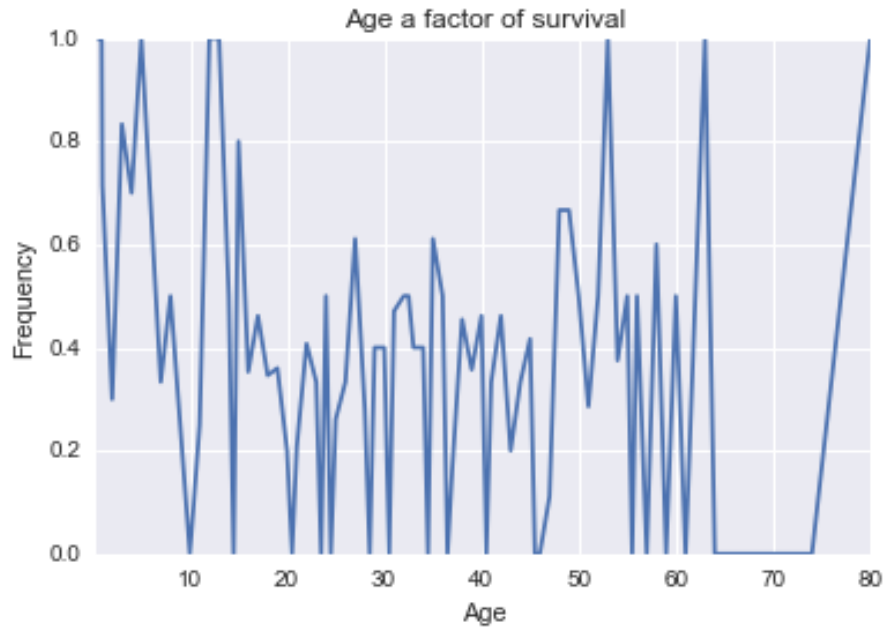


% female survivors = 74.2038%
% male survivors = 18.89%

Conclusion: Based on the above chart it is clear that female survivors ar significantly higher than male survivors. Hence we could say that sex may be a factor of survival rate.

Question 3: Is age group a factor?

Let's try to look at the mean of survival rate:



Conclusion: From the above graph it's clear that the survival rate for children and older people is infact higher!
This could be due to the higher priority given to children and older people for lifeboats!

Limitations of the Analysis:

In the age group analysis, the number of missing values were very hight (177 out of 891 data points) . Hence the conclusions drawn from that analysis is not very accurate as we may be missing important information.

Also, there may be other factors that could be taken into account. Like whether the passenger was active or not at the time the ship hit the iceberg. Also, the exact location of each passenger (like how far their cabin was from the deck) may be something to look into.