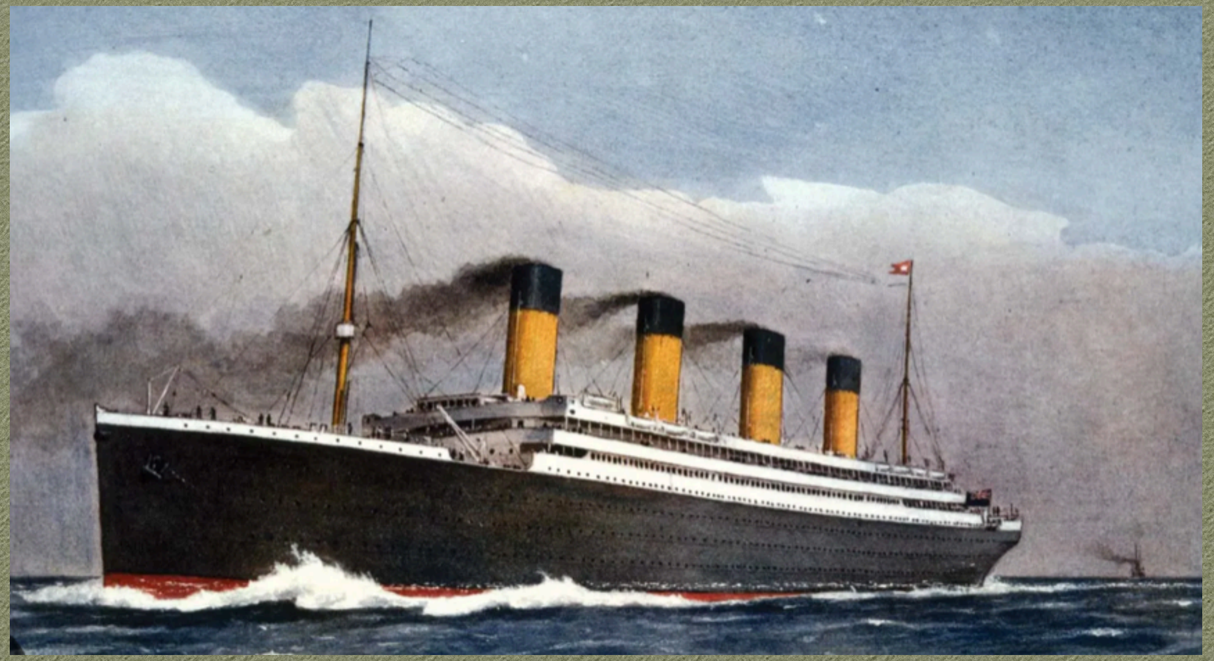
Titanic children survival rate analysis



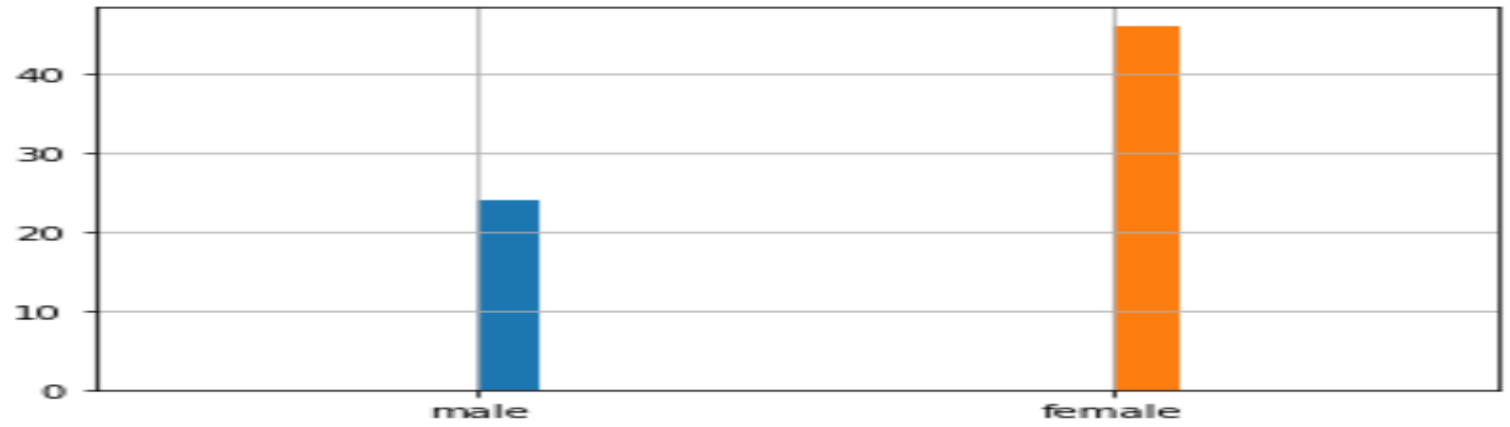
Giwon Lee 40160453

SITUATION

- Given four factors (age, sex, embarkation port, and size of the family), analyzing survival rate of children.

1. Child Passengers (Second code block)

```
The total number of children who survived from the incident is:  70  people.
The total number of children who died from the incident is:  69  people.
Number of Male survivor:  24
Number of Feale survivor:  46
```



2. Histogram of # of survived children

- Anyone who is 18 or less are considered children.
- Throughout the analysis, Male and Female ratio is always measured due to significant differences in the data.

Problem- the problems that I found in the dataset

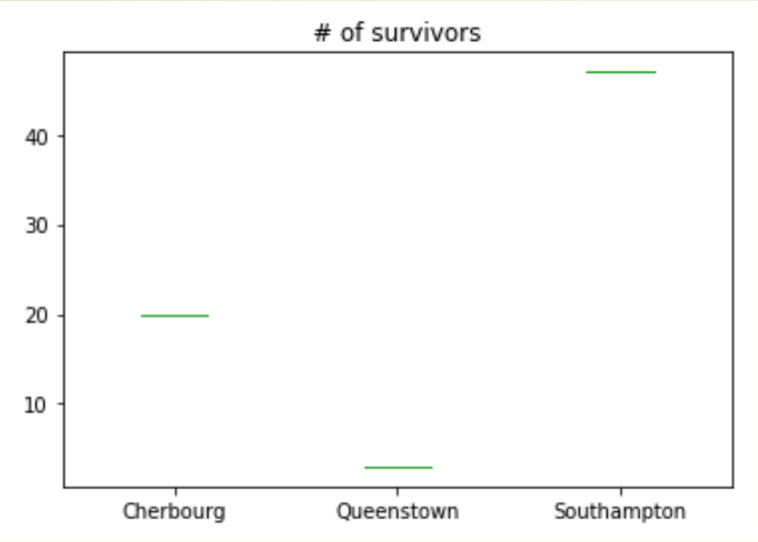
- The dataset I obtained from Kaggle has train and test CSV files. I concatenate two datasets to analyze the survival rate of children.

- It seems there are some missing values from the data set. Some members have no age or has a floating value of the age.

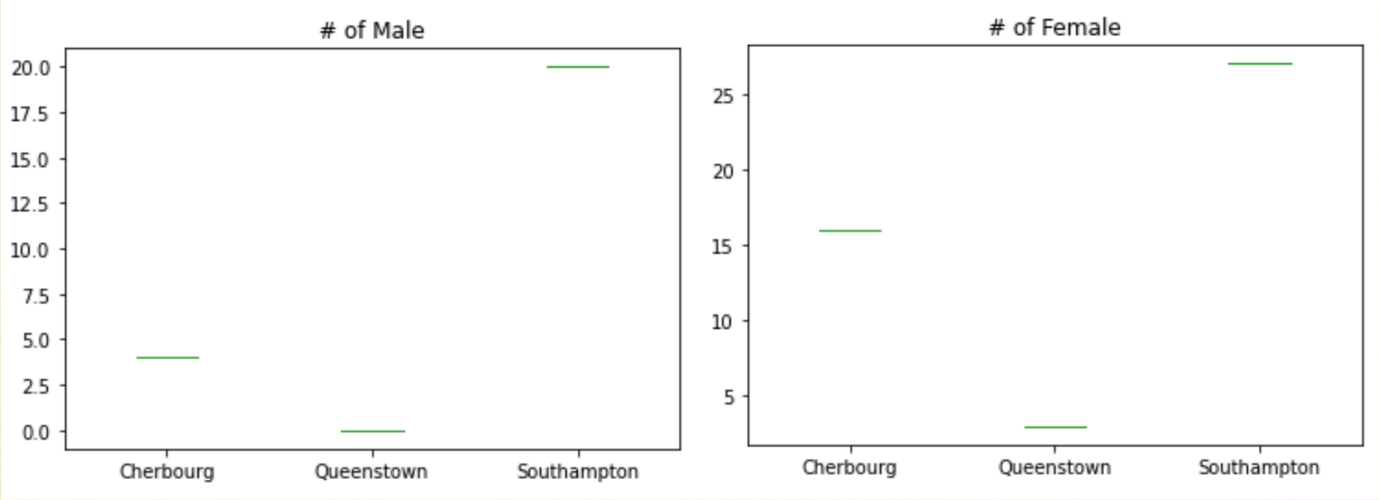
- Below are box plots the embarked locations who are children.

- Number of children survived show in the plots.

- There is significantly different number of passengers embarked at the each of the location (see figure 3.)

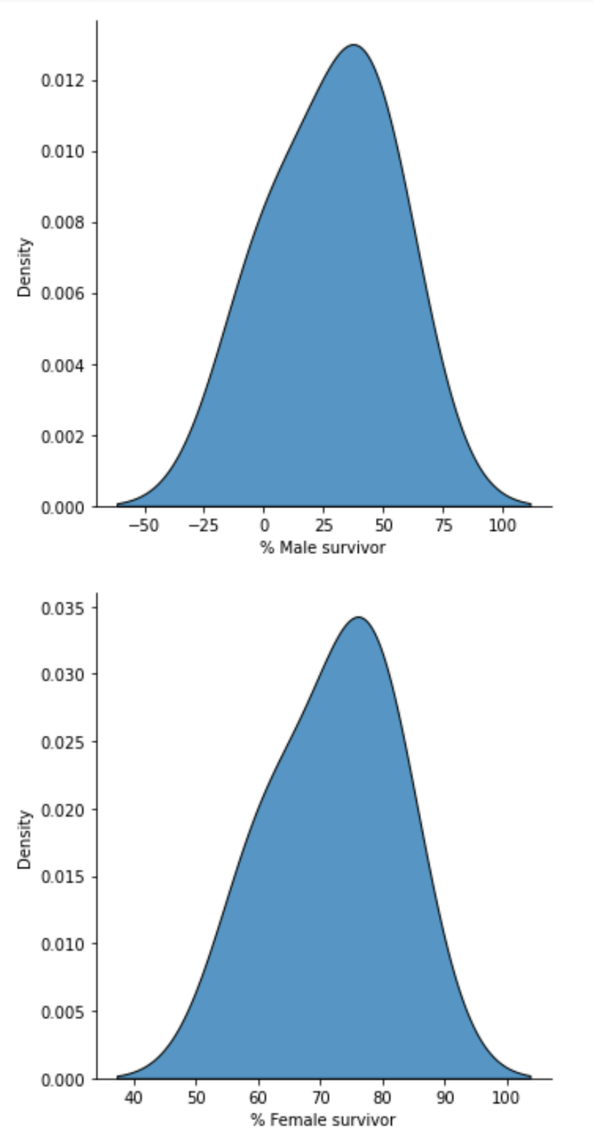


3. # of survivals at different Embarked locations.



4. Difference of male and female survival #

Solution

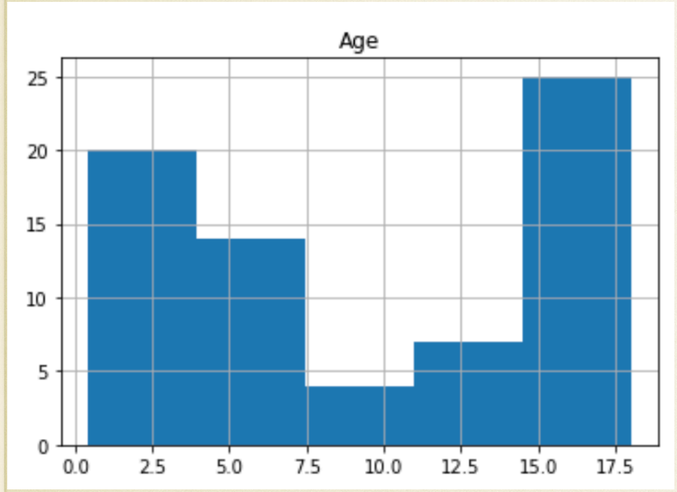


5. Bell Curve Survival Rate Male vs Female

```
numFemale = df.loc[(df['Survived'] == 1) & (df['Age'] < 19) & (df['Sex'] == 'female')]
numMale = df.loc[(df['Survived'] == 1) & (df['Age'] < 19) & (df['Sex'] == 'male')]
```

- Above the code shows how missing value of age is neglected: It is decided that not to implement the age that is missing because even though I replace the missing age value with average age of passengers', they are not considered as a child (avg 28.7).
- Male / Female difference have been measured overall because there is always a significant difference of survival rate between the two.
- Female has higher survival rate than male wherever they embarked at different locations.
- Southampton has more male passengers (child) than female but the percentage of male survivor rate is half of the female survivor rate.
- This is not a solution but this clearly shows the different survival rate by Male vs Female.
- At this point, the embarked location seems to be meaningless.

Approach



6. Age Histogram

- 8 years or less have survived apart from who are 7 or 15 years old.
- Surprisingly, teenager who are between 15 to 18 survived the most.

Survived members with siblings/parents:

	Sex	Survived w family	Survived w/o family	Percentage of survival w/o family
0	Male	22	2	9
1	Female	35	11	31

Dead members with siblings/parents:

	Sex	Dead w family	Dead w/o family	Percentage of Dead w/o family
0	Male	28	23	82
1	Female	18	10	56

7. With or Without family members

- Survival rate is high when you have a family member(s).
- Death rate is high when you do not have a family member(s).

- To approach to analyze:
- First needed to separate the children's dataset from the entire dataset
- Using pandas data frame it is easy to visualize the number of survivals male vs female by 'embarked locations', and 'number of parents/siblings'.
- Using Seaborn and Pandas to create plots and histograms.