I) Analysis of the data set
Here are the columns in the dataset.

- `PassengerId` -- A numerical id assigned to each passenger.
- `Survived` -- Whether the passenger survived (1), or didn't (0). We'll be making predictions for this column.
- `Pclass` -- The class the passenger was in -- first class (1), second class (2), or third class (3).
- `Name` -- the name of the passenger.
- `Sex` -- The gender of the passenger -- male or female.
- `Age` -- The age of the passenger. Fractional.
- `SibSp` -- The number of siblings and spouses the passenger had on board.
- `Parch` -- The number of parents and children the passenger had on board.
- `Ticket` -- The ticket number of the passenger.
- `Fare` -- How much the passenger paid for the ticker.
- `Cabin` -- Which cabin the passenger was in.
- `Embarked` -- Where the passenger boarded the Titanic.

Using these, What questions can be solved?
1) What percent of males and females survived?
2) What percent of females were from 1st and 2nd class who survived
3) Average age of females who survived
4) people who have more siblings vs people who survived
5) people who have more parch vs people who survived
6) the combination of gender, age and Pclass which resulted in highest rates of survival and the least rates of survival

III) Data Wrangling:

We observed that there are 714 records, for age out of total records,
One simple way of dealing with this is to take the median of the these values and replace it for NULL records.

describe() does not tell about the columns 'Name', Sex, Embarked, Ticket
One way to deal with this is Convert Sex, i.e male = 0, female =1, so that we can answer the questions 1,2,3,6

Solutions:
1) What percent of males and females survived?

Solution:
male_survived = 109
, total_male = 577
, female_survived = 233
, total_female = 314
, total_survived = 342

Male_survivors = float(male_survived)/total_male = 0.1889
Female_survivors = float(female_survived)/total_female = 0.742
Compute Chi*2 Test

2) What percent of survived females were from 1st and 2nd class

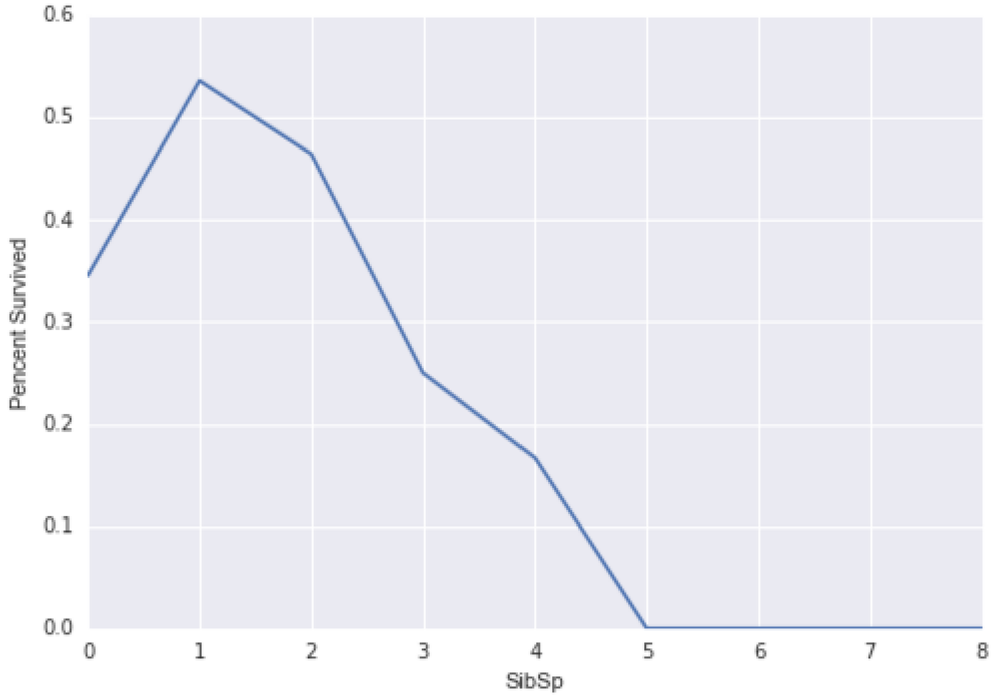
Pclass_1_percent_females_survived = 0.669117647059
Pclass_2_percent_females_survived = 0.804597701149
Pclass_3_percent_females_survived = 0.605042016807

The only significant result from this is Pclass_2_percent_females_survived.

3) Average age of females who survived

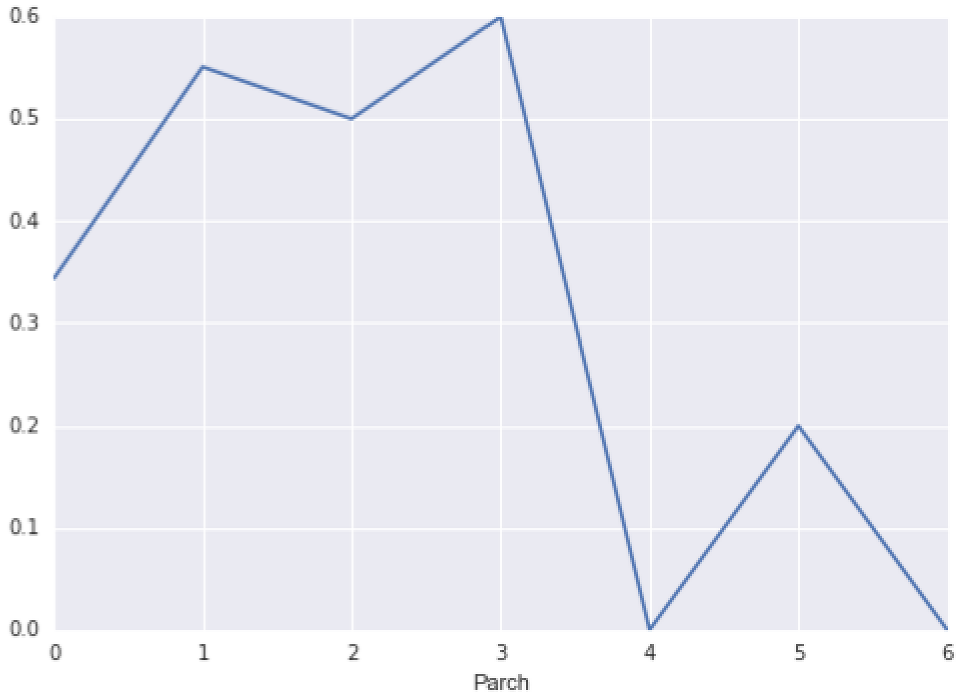
#We have put the median values in the age column so need results might not be true representative
It is observed that the average of Female who survived is 28.7167381974
It is observed that the average of Male who survived is 27.382293578

4) people who have more siblings vs people who survived



We observe that, people who have 1-2 siblings have better chance of survival. Well it is likely as they would be saved by their siblings

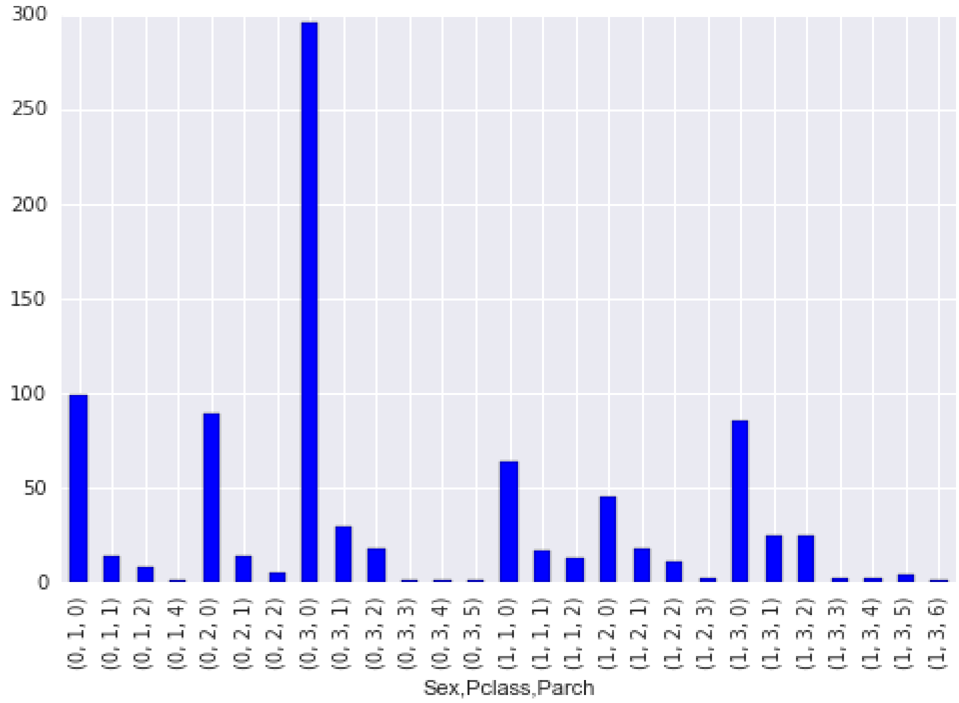
5) people who have more parch vs people who survived

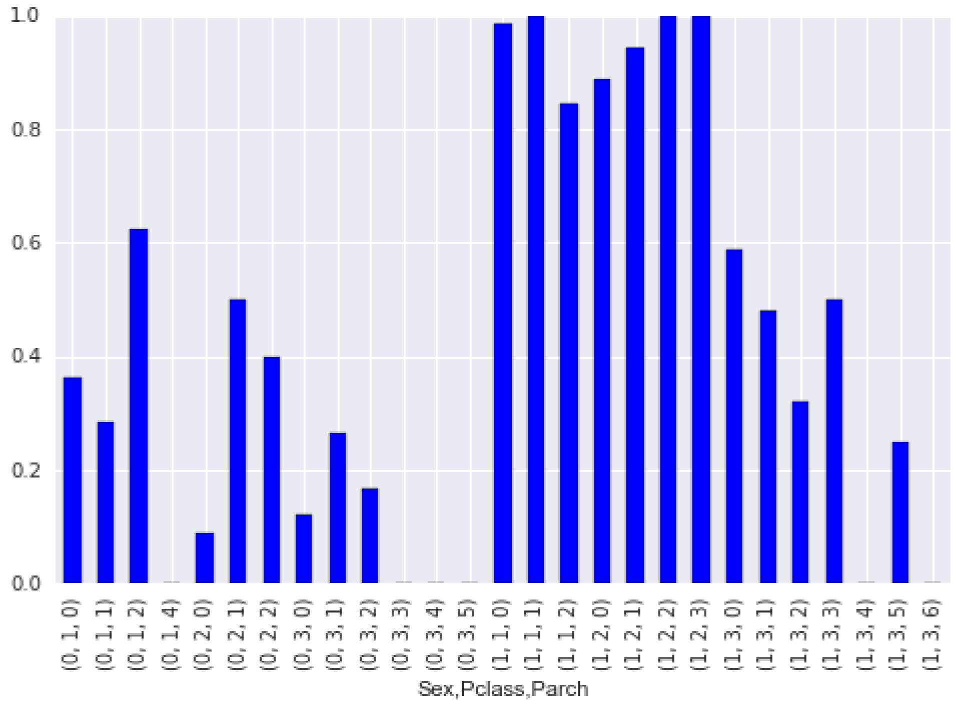


We observe that chances of survival are more if parch is between 1 to 3. It is likely that the ones who survived were saved by their parents or children, but if parch > 4, the survival dropped dramatically, possibly indicating lack of clarity whom to save.

6) the combination of gender, Parch and Pclass which resulted in highest rates of survival and the least rates of survival

Gender: male = 0, female = 1
Parch : Number of parents/ children
Pclass : 1, 2, 3 class cabins





Clearly, we observe that female in 1st and 2nd class have highest rate of survival
Also, people with Parch = 1, or have better chances of survival