

1. I analyzed the Titanic data.
2. Questions I posed:

Calculate how many people survived and how many didn't

Calculate Survival vs. Age / Pclass / SibSp / Parch / Fare

Calculate what percentage different categorical data (Female / Male) (Pclass 1,2,3) (SibSp 0,1,2,3,4,5,8) (Parch 0,1,2,3,4,5,6) survived

Calculate how different age group survived

Scatter plot of 'Pclass' and 'Age'

Scatter plot of 'SibSp' and 'Parch'

3. The dataframe I used throughout the project is divided by the `groupby()` function with 'Survived' factor.

I did `count()` to count the survived and non-survived people, and `mean()` to compare the mean difference of other factors like age, Pclass, and Fare.

Mean comparison and categorical comparison were repetitive, so I created a function for each.

Then I made a conditional division of data, using `titanic_df[titanic_df['Age'] < age]`

Then I could calculate how different age group survived.

I used `matplotlib scatter()` function to plot.

4. The data wrangling I did was to divide the age groups and compare survival rate of the two groups

5. Summary statistics and plots communicating your final results

The statistics are round values, first element = not survived / second element = survived.

Just mean calculation showed lower age (30.6, 28.3), higher Pclass (2.5, 2.0), less siblings/spouses (0.55, 0.47), more parents/children (0.33, 0.46), more fare (22.1, 48.4) led to better survival.

When I divided each group with `groupby()`, the results weren't exactly the same.

Sex looks like the most significant factor. 74% female survived while only 19% male survived.

For Pclass, definitely, the higher class survived better than the lower (1: 63%, 2: 47%, 3: 24%).

For Siblings/Spouses, lower wasn't always better. Having no SibSp (0: 34%) didn't survive better than having one. (1: 54%)

For Parents/Children, same result. (0: 34%, 1: 55%)

Younger people survived better. (Under 10: 61%, Over 60: 27%)

From the first plot, Pclass 1 has mostly adults survived. Pclass 2 and 3 have mostly babies/teens survived.

From the second plot, having some family members were more likely to survive, but not too many or none.

A list of Web sites, books, forums, blog posts, github repositories, etc. that you referred to:

<https://www.kaggle.com/c/titanic/data>

http://matplotlib.org/api/pyplot_api.html

<http://stackoverflow.com/questions/13747860/plot-specific-columns-and-rows-from-a-file-with-pylab>

<http://stackoverflow.com/questions/31609600/jupyter-ipython-notebook-not-plotting>

<http://stackoverflow.com/questions/19410042/how-to-make-ipython-notebook-matplotlib-plot-inline>

<http://stackoverflow.com/questions/12018992/print-combining-strings-and-numbers>

<http://stackoverflow.com/questions/18106975/compare-float-and-float64-in-python>

<http://stackoverflow.com/questions/14847457/how-do-i-find-the-length-or-dimensions-size-of-a-numpy-matrix-in-python>

<http://stackoverflow.com/questions/16970982/find-unique-rows-in-numpy-array>

<http://stackoverflow.com/questions/16476924/how-to-iterate-over-rows-in-a-dataframe>