The Titanic

Out[11]:

Out[12]:

Click here to toggle on/off the raw code.

Describing, questioning and analyzing a disaster

Udacity - Data Analyst Nanodegree Program

Aviad Giat - September 2017

In this document (titles are clickable):

Out[19]:

```
{'height': 768, 'scroll': True, 'width': 1024}
```

A sample of 5 records from the new dataframe (Pandas' dataset)

Out[74]:

	Survived	Class	Gender	Age	Fare	Sex	Survived_y_n
0	0	3	male	79.48	7	0	No
1	1	1	female	11.37	71	1	Yes
2	1	3	female	24.08	7	1	Yes
3	1	1	female	73.14	53	1	Yes
4	0	3	male	77.10	8	0	No

Description of the variables in this analysis

Survived	Passenger's survival as an integer (0 = No, 1 = Yes)
Survived_y_n	Passenger's survival as a string (yes, No)
Class	Passenger's Class (1 = 1st; 2 = 2nd; 3 = 3rd)
Gender	Passenger's gender (Male/Female)
Sex	Passenger's gender (0 = Male; 1 = Female)
Age	Passenger's Age
Ages	Passenger's Age group by decades (0-10 = 01, 10-20 = 10, ... , 70-80 = 70)
Fare	The cost of the ticket in dollars

Top
Setup

Discovering and describing the data

Top

```
Dataframe summary, types, NaNs and statistics

Rows and columns
================
There are 8 Columns and 891 Rows in this dataset


The types of data for this dataframe
====================================
Survived          int64
Class             int64
Gender           object
Age             float64
Fare              int32
Sex               int64
Survived_y_n     object
Ages           category
dtype: object


More details about the data frame
=================================
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 8 columns):
Survived        891 non-null int64
Class           891 non-null int64
Gender          891 non-null object
Age             891 non-null float64
Fare            891 non-null int32
Sex             891 non-null int64
Survived_y_n    891 non-null object
Ages            891 non-null category
dtypes: category(1), float64(1), int32(1), int64(3), object(2)
memory usage: 46.6+ KB
None


Statistical summary of the numeric data
=======================================
       Survived   Class     Age    Fare
count    891.00  891.00  891.00  891.00
mean       0.38    2.31   41.05   31.79
std        0.49    0.84   23.25   49.70
min        0.00    1.00    0.01    0.00
```

25%	0.00	2.00	21.14	7.00
50%	0.00	3.00	41.53	14.00
75%	1.00	3.00	61.16	31.00
max	1.00	3.00	79.85	512.00

Questions about the data

Top

1. Were there more people who perished or survived? What was the percent of survivors from all passengers?

2. Who had the best chances to survive?

*** Males VS Females (Gender)**

*** Age by decades (Ages)**

*** Ticket's cost (Fare)**

*** Class - The ticket class (First, Second, Third)**

Analyzing the data and answering the above questions

Top

The dependent variable is the number of survivors. I will analyze 4 independent variables against it. The variables are:

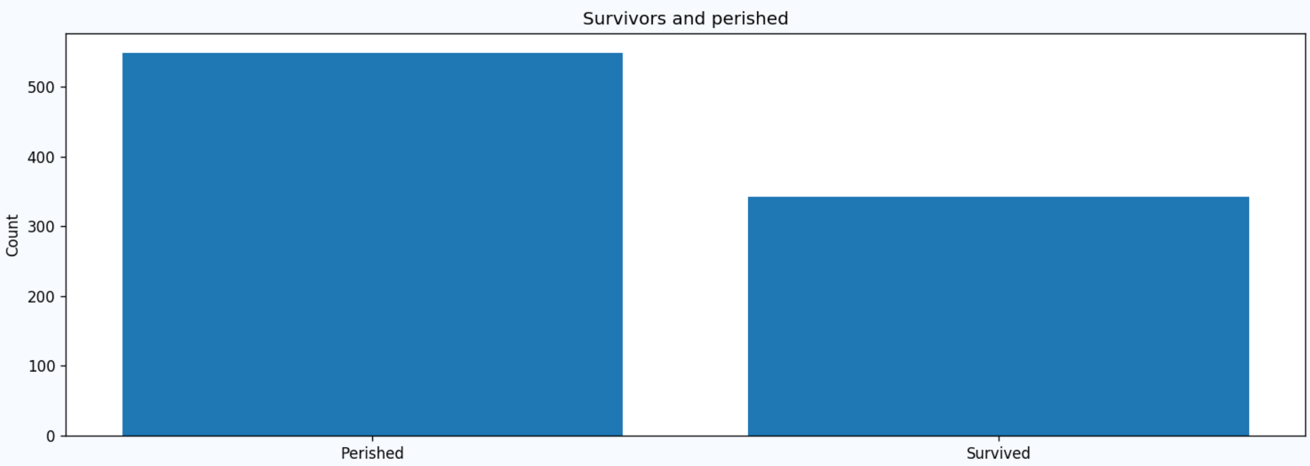
against it. The variables are:
Class | Age | Gender | Fare.

I will also try to answer at least one of the above questions with a statistical test.

Let's start with the general number of survivors and victims:

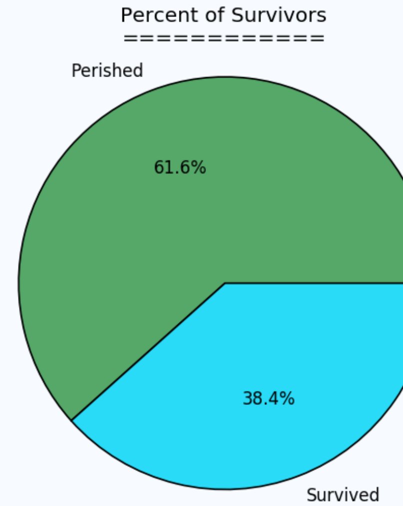
Number of survivors

```
Number of survivors and perished
==================================
Survived:
No     549
Yes    342
Name: Survived_y_n, dtype: int64
```



* As we can see from the above bar plot, more passengers died (61%) on the Titanic than survived (38%). For every 10 people who survived 16 perished.

Survivors and perished as percent



As we can see from the above pie chart, more passengers died (61%) on the Titanic than survived (38%). For every 10 people who survived 16 perished.

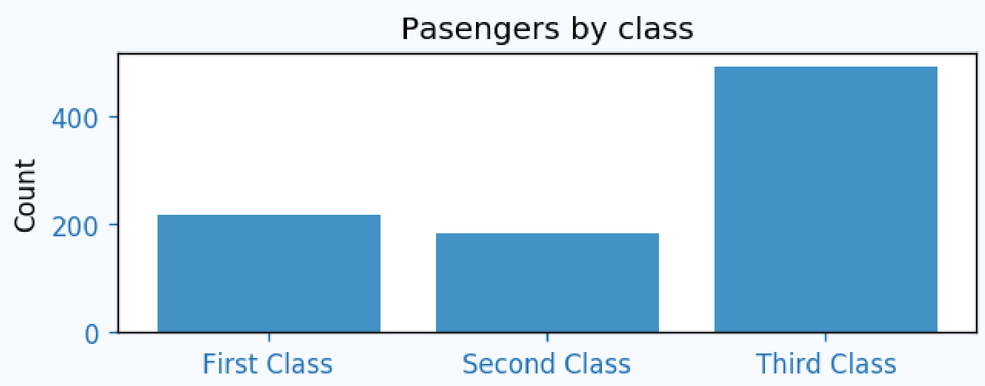
[Top](#)
[Analysing](#)

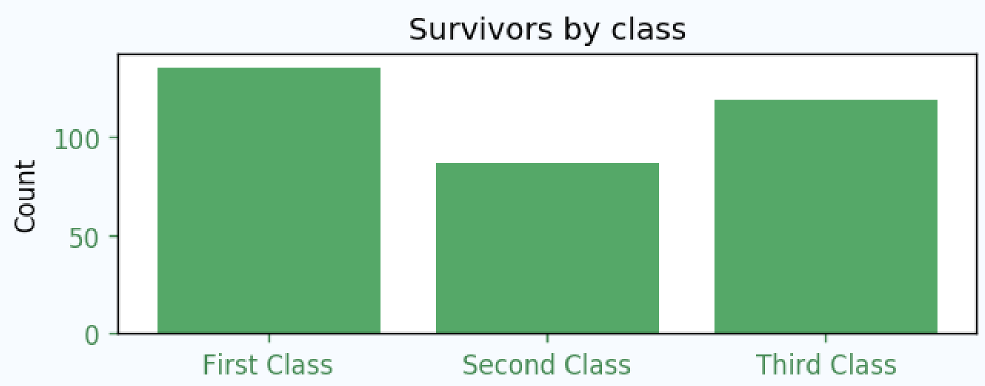
Now, lets break down the numbers and percentages following the 4 variables (Class | Age | Gender | Fare)

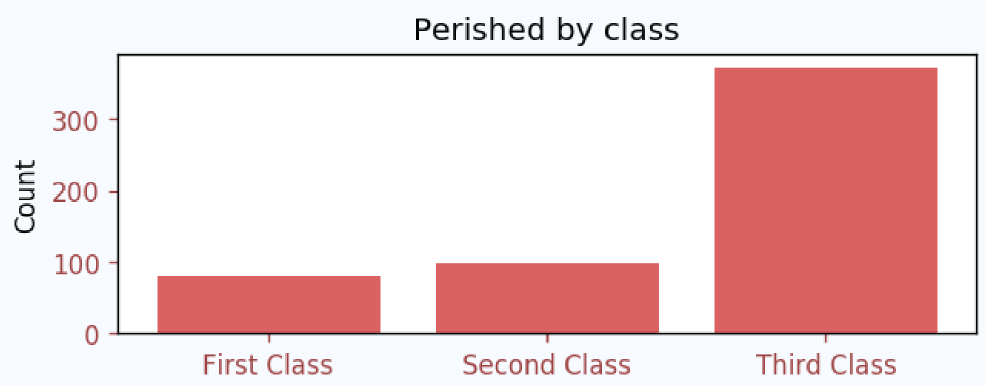
Survivors by Class

[Top](#)

Which class members' survival rate was the highest?







```
Number of passengers in each class
===================================

1    216
2    184
3    491
Name: Class, dtype: int64
```

```


Number of survivors from each class
====================================

```

Out[35]:

Survived_y_n	No	Yes	All
Class			
1	80	136	216
2	97	87	184
3	372	119	491
All	549	342	891

```


Percent of survivors from within each class
============================================

```

Out[36]:

Survived_y_n	No	Yes
Class		
1	37	63
2	53	47
3	76	24

First class passengers had the highest survival rate (63%), while passengers from class 3 had less than 24% chances to survive. Passengers from class 2 had almost 50% chances to survive. Was it a chance, and could the survival odds flip between the classes in a similar disaster like that? Let's try to answer this exact question with a statistical test. Since this is a nominal type of data I will use the Chi-Square test:

```
Chi-Square Test - Number of survivors by Class
===============================================


Null Hypothesis:
Ho: There is no statistically significant difference between any of the classes' survival rate.
Ha: There is a statistically significant difference in the survival rate between any of the 3 cla
sses of passengers.
The survival rate should be 33.3% for the 3 classes.
```

```
Contingency table of the classes' survival
===================================================
Survived    0    1  All
Class
1          80  136  216
2          97   87  184
3         372  119  491
All       549  342  891


The Chi-Square (Goodness of fit), Probability, Degrees of oreedom, and the Expected frequen
cies
Critical Value for the chi square = 5.991
===================================================================
==========================

The Chi-Square distribution's Critical Value is 5.991
```

- We can see here that the chi-square statistic is 102.888 with 2 degrees of freedom, which is a lot more than the Chi Square Critical Value of 5.991.
  The total number of (observed) survivors is 342.
  With a critical value of 5.991, the probability (p-value) is smaller than 0.0001, which is considered as extremely statistically significant at $p < 0.05$.
- It could be interesting to check other disasters across different categories and compare the results with this test, to see whether First class ride improves one's survival rate.
- On the Titanic, Class did affect one's chances of survival.

Top
Class

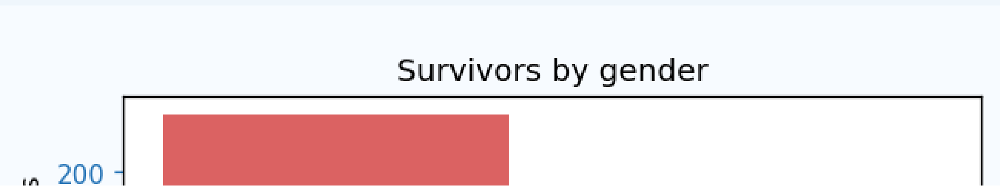
Survivors by Gender

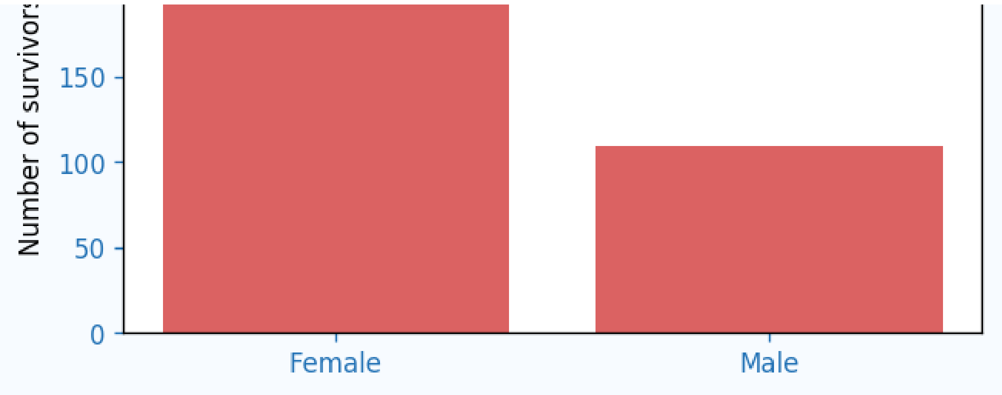
Top

Which gender had better survival rates?

Let's start with the percent of the survivors and the number of survivors from each gender and compare them with the number of passengers on board after leaving the last port of embarkation.

Number of survivors





Survivors by gender in numbers

```
Number of survivors by Gender
==============================

Survived    1
Gender
female    233
male      109
```

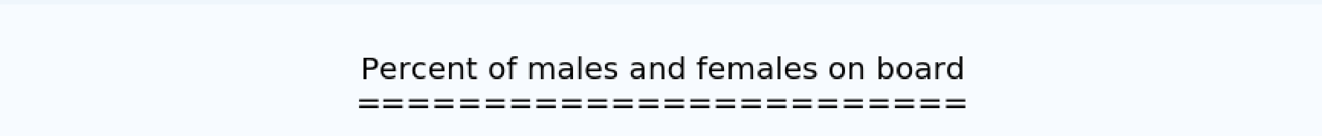
Survivors by gender in percent

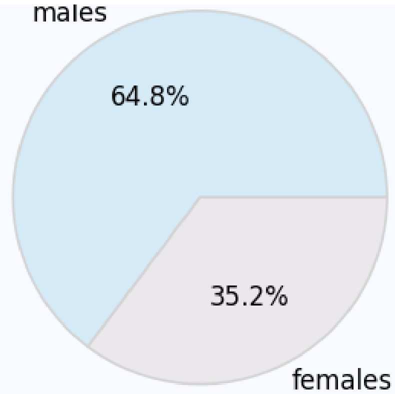
```
Survived      1
Gender
female    68.13
male      31.87
```

Number of males and females on board

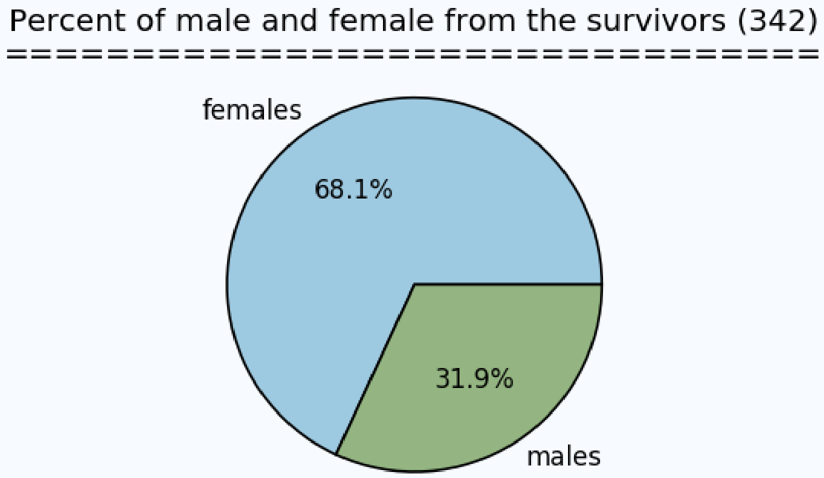
```
male      577
female    314
Name: Gender, dtype: int64


There were 263 more males than females on board
```

Passengers by Gender



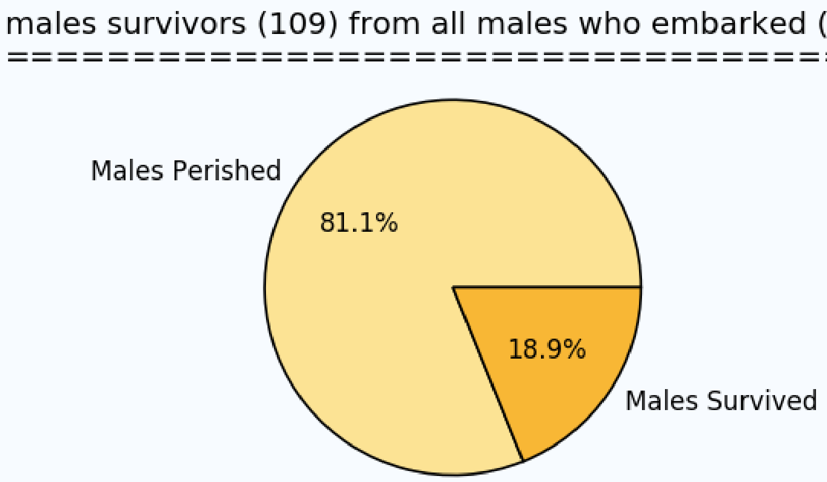


Survivors by gender



We can see that more females than males survived. But is it significant difference or a statistical error? Before answering this question with a statistical test, let's look at the survival's numbers and percentages of both genders and within each gender:

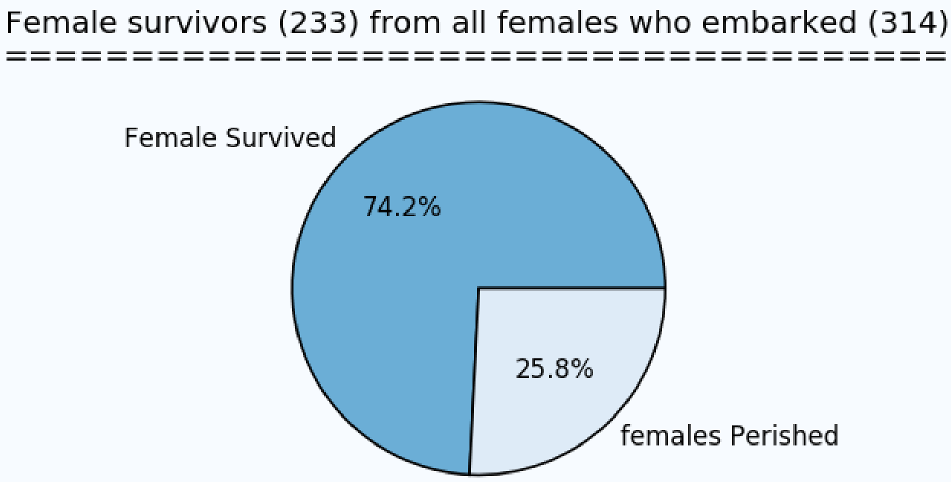
Male survival



Number of Male survivors
==========================

Survived_y_n	No	Yes	All
Gender			
male	468	109	577
All	468	109	577

Female survival



Number of Female survivors
============================

Survived_y_n	No	Yes	All
Gender			
female	81	233	314
All	81	233	314

Critical Value for the chi square = 3.841

- 74% of the females who embarked on the first and last trip of the Titanic survived, compared to only 19% of the males.
  This shows that females had 4 times better chance to survive on this cruise.
  When the Titanic left the last harbor, there were 577 males on the ship (out of 891 passengers), almost twice the number than females (314). Yet, 68% of the total survivors were females (233).
  Now, let's check with a statistical test if the difference between the two genders' survival rate is significantly different and is not due to chance.
  I will use the chi-Square test here as well:

Chi-Square Test - Number of survivors by Gender
===============================

Null Hypothesis:

Ho: There is no statistically significant difference between males and females' survival rate.
Ha: There is a statistically significant difference between males and females' survival rate.

```
Contingency table of males and females survival
==============================================
Survived    0    1  All
Gender
female     81  233  314
male      468  109  577
All       549  342  891



The Chi-Square (Goodness of fit), Probability, Degrees of freedom, and the Expected frequen
cies
============================================================
=======================
(260.71702016732104, 1.1973570627755645e-58, 1, array([[ 193.47474747,  120.5252525
3],
       [ 355.52525253,  221.47474747]]))
```

The Chi square result is 260 with 1 degree of freedom.
The Chi Square critical value for 95% and 1 degree of freedom is 3.841. The one-tailed P value is less than 0.0001.
The association between males, females, and survival is considered to be extremely statistically significant.
In other words, females did not survive in such a great proportion by chance. There had to be a cultural code of behavior that said, females first.

Top
Gender

Gender Survival by Class

Drag and drop the 'Survived', 'Gender' and Class to the left column.

You can change the view to a bar chart and other visualizations under the drop down menu at the left

```
Count of survivors by 'Gender', 'Class'
==========================================
Gender  female  male
Class
1           91    45
2           70    17
3           72    47
```

```
Survivors by Class and Gender
```

Out[49]:

Gender	female	male	sum	percent male	percent female

Class Gender	female	male	sum	percent male	percent female
Class					
1	91	45	136	33	67
2	70	17	87	20	80
3	72	47	119	39	61

Looking at the above plot and table we can see that men from first class perished 3 times more than women. The second class had the worse ratio with 5 men perished for every woman and in the third class man perished in ration of 2.5 men to 1 woman. Women in second class had the best survival rate of 80%, compare to only 20% men from the same class.

Survivors by Age

Top

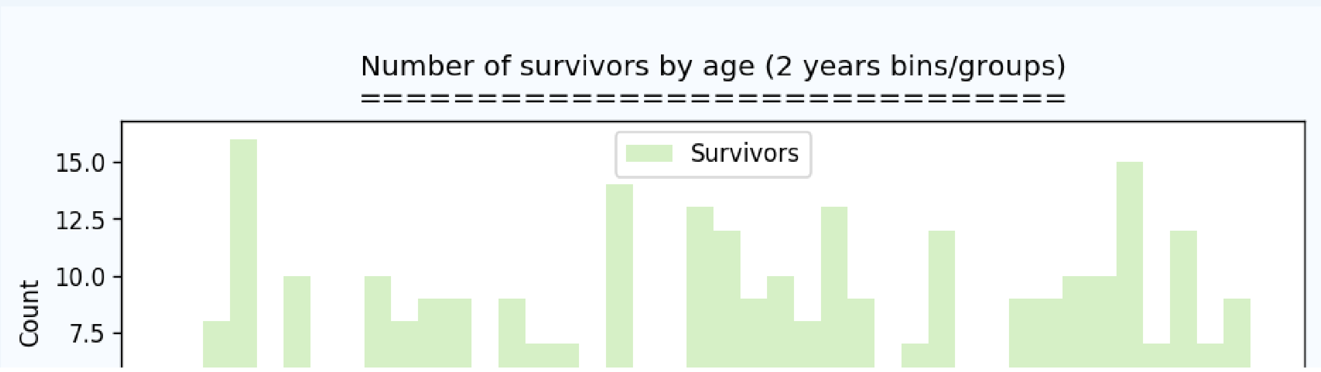
Passengers in which group age had the best chances of survival?

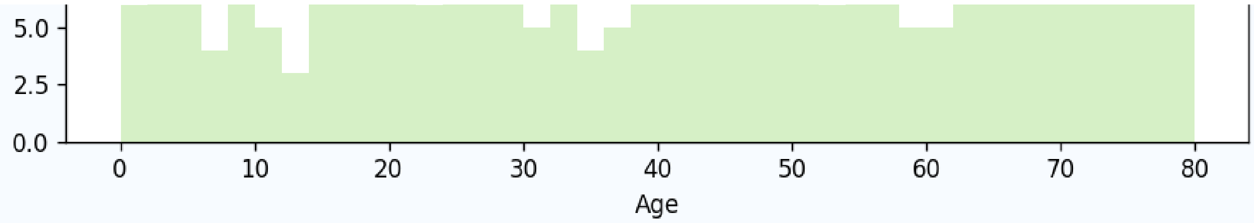
Age variable basic statistics

Out[50]:

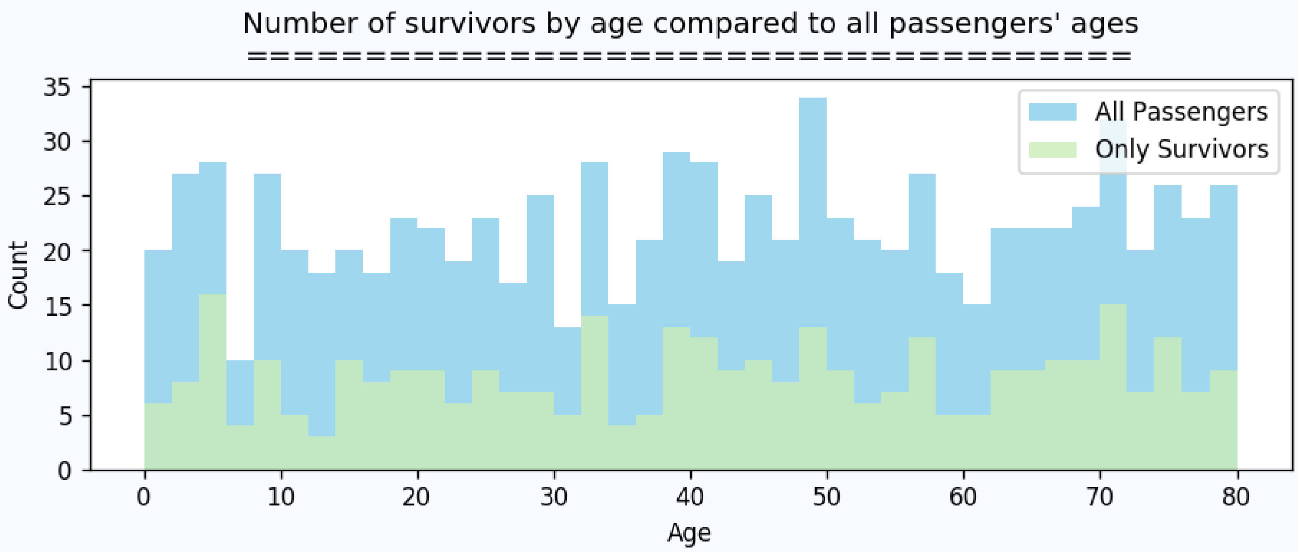
```
count    891.00
mean      40.55
std      23.24
min       0.00
25%      21.00
50%      41.00
75%      61.00
max      79.00
Name: Age, dtype: float64
```

We will start with a simple histogram of the distribution of survivors by age:

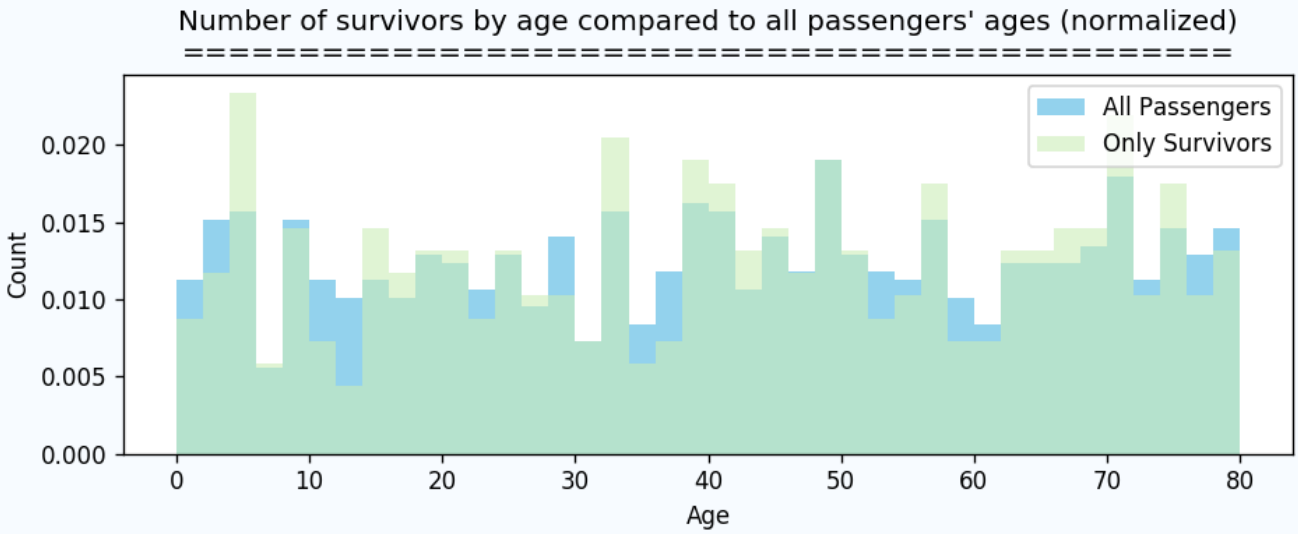




Now let's compare the survivors' ages and the entire population (all passengers, survivors and perished)



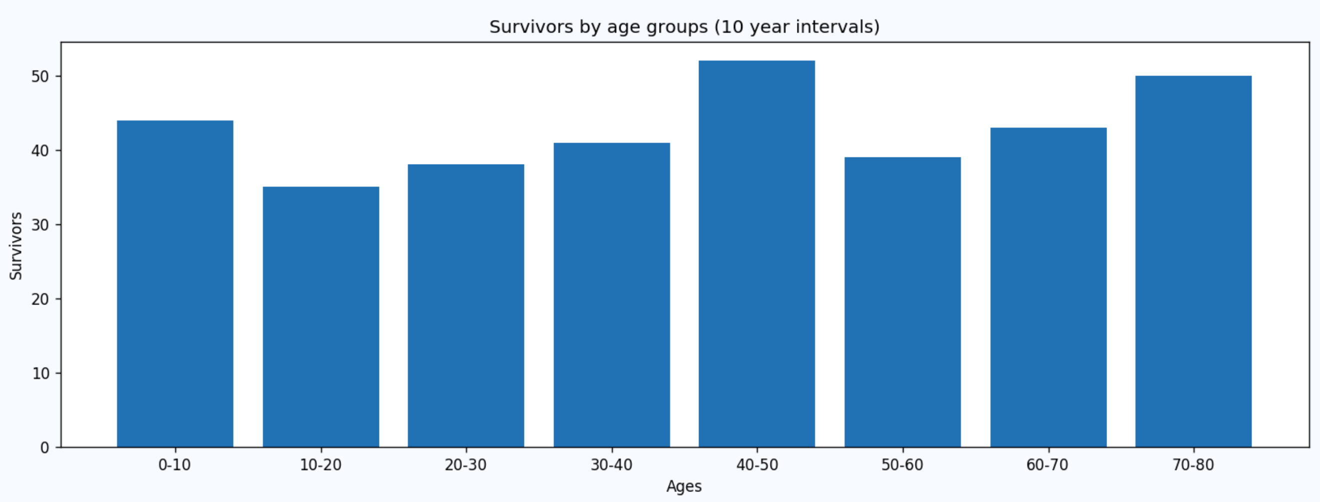
Adjusting both Survivors and All Passengers' values to the same scale for comparison



- We can see from the above 3 histograms that passengers and survivors distributions have more or less the same shape. From the 3rd (normalized) distribution of both datasets, we can see that there is some symmetry between the 2 distributions. This might suggest that the age groups with most passengers had most of the survivors and groups with less members had less survivors.
  We can also see that there is a wide gap between the number of passengers and the number of survivors (when it is not normalized). This says that there were more people who died than survived across the board of ages.

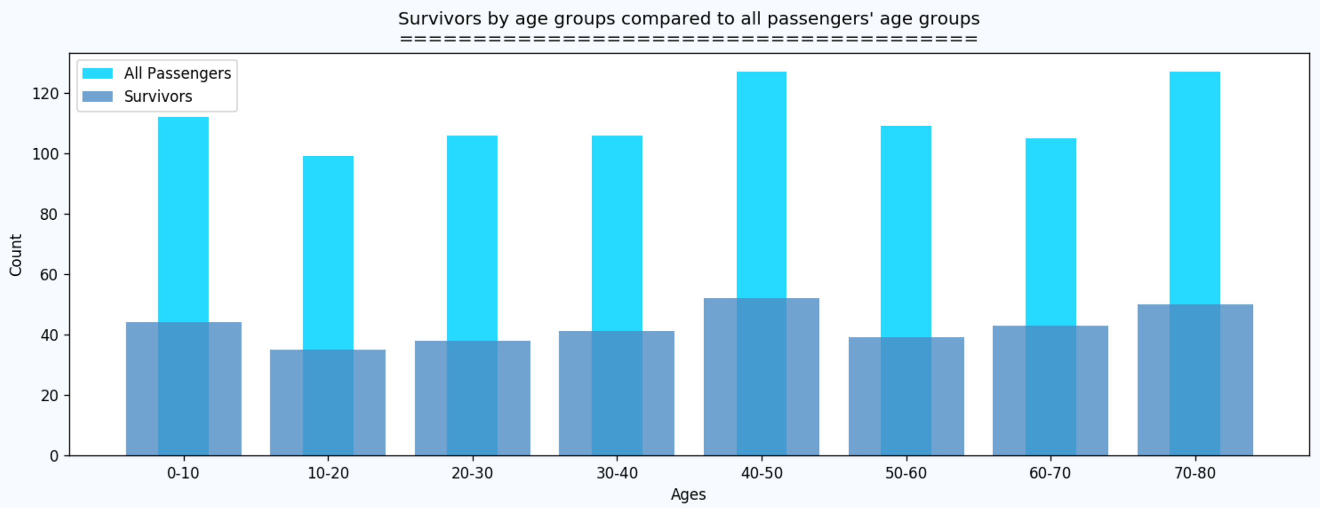
Let’s try to dig in and see if this is really the case or not using the Pandas' function .cut() from the top of this document.
This function creates ranges of ages by decades, up to 80, thus 8 age groups.
First, let's examine how the distribution of this new column looks like:

Survivors by age groups



Out[55]:

```
[0-10, 20-30, 10-20, 40-50, 70-80, 30-40, 50-60, 60-70]
Categories (8, object): [0-10 < 10-20 < 20-30 < 30-40 < 40-50 < 50-60 < 60-70 < 70-80]
```



Survivors by age groups in numbers

Out[57]:

```
Ages
10-20    35
20-30    38
50-60    39
30-40    41
60-70    43
```

```
0-10     44
70-80    50
40-50    52
Name: Survived, dtype: int64
```

- After dividing the Age variable into ranges of 10 years, we can see that there are no exceptional outliers or trends. The distribution seems random. The largest age groups of survivors are the 10s and the 40s. But being the group that had the highest number of survivors does not mean necessarily that the chances were better than other age groups' members.
The groups of 40-50 and 0-10 have the highest number of survivors. But which group members had the best chances of survival within those groups? To find that out I will find the age group's percent of survival from the total number of passengers (both survived and perished) in the specific group.

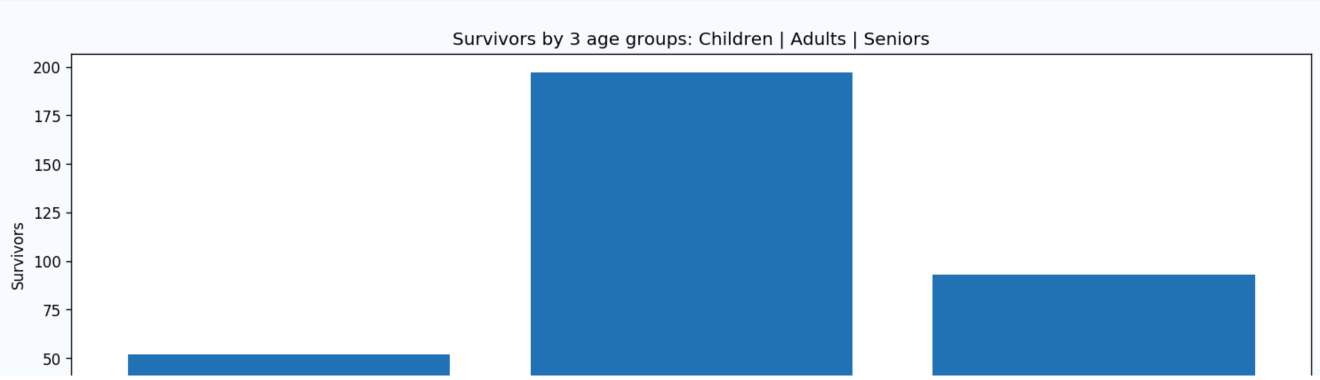
Percents of survival from within each group age

Out[58]:

Survived	0	1	total	Percent Survival	Normalized
Ages					
60-70	62	43	105	41	0.41
40-50	75	52	127	41	0.41
70-80	77	50	127	39	0.39
0-10	68	44	112	39	0.39
30-40	65	41	106	39	0.39
20-30	68	38	106	36	0.36
50-60	70	39	109	36	0.36
10-20	64	35	99	35	0.35

- From the crosstab table above we can see that the group age with the highest survival rate of 46% was the seniors' one (70-80) and the one with the lowest survival rate was the 10-20 group with only 33% survival rate.
We can see that the percentages of survivors from within each group varied, at most, in 13%. This doesn't seem odd and look more like a random distribution. Maybe, dividing the passengers' ages by different key will make things look different. Let's try and classify children as ones who are 14 years and younger; Adults from 15 to 60 and seniors from 60 and up and see if there is a meaningful difference in their survival rate:

Changing the age variable to 3 age groups



Drag and drop the 'Survived', 'Gender' and Class to the left column.

You can change the view to a bar chart and other visualizations under the drop down menu at the left

```
Percent of survivors from within each group age
=================================================
```

Out[61]:

Survived	0	1	total	Percent Survival	Normalized
Ages-1					
Senior	139	93	232	40	0.4
Adult	312	197	509	39	0.39
Child	98	52	150	35	0.35

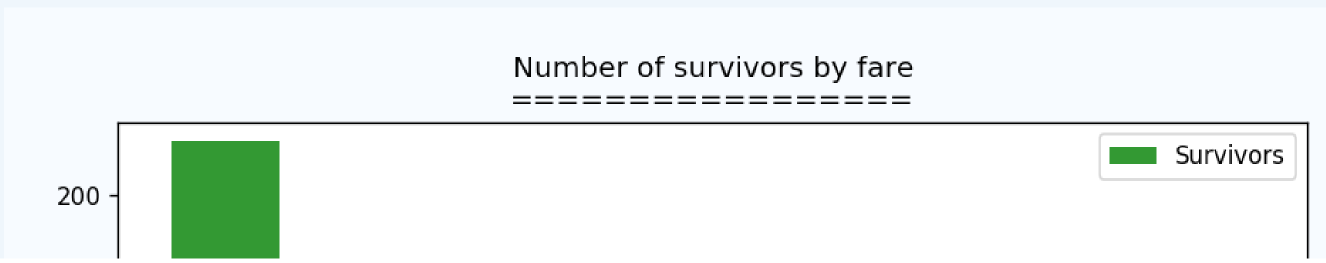
- We can see that there is about 10% difference between the survival rate of the adults and the two other age groups.
  Children in this analysis are considered to be 14 years and younger.
  If we were to change the max age of children to 18 it seems that will not make a significant difference in the Children's survival rate (39% survival rate instead of 42%).

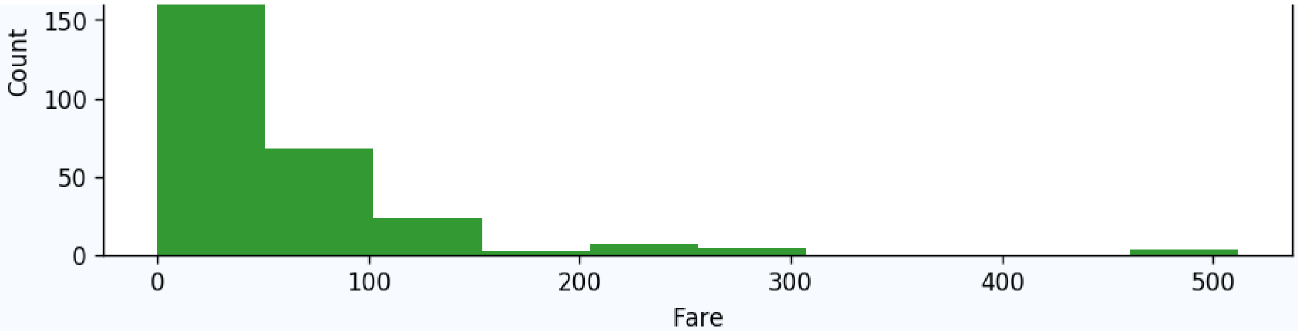
Top
Age

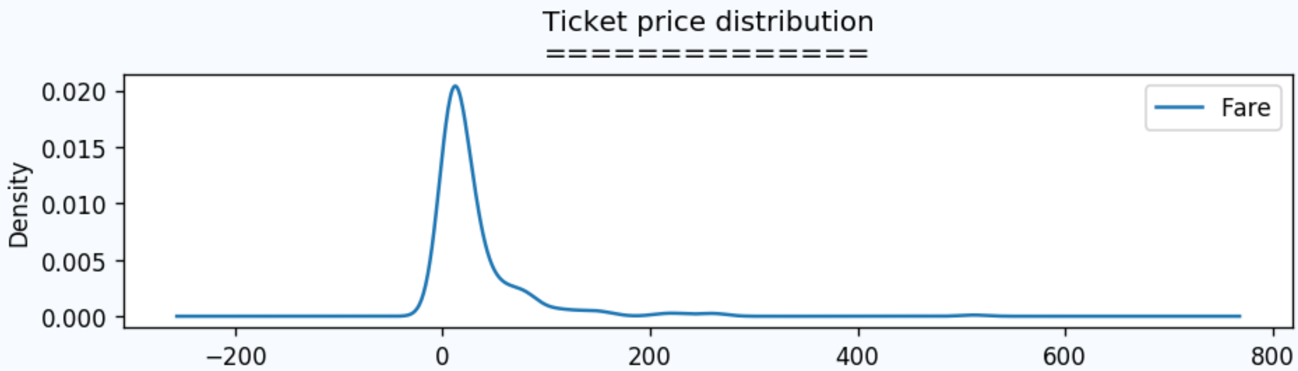
Survivors by Fare

Top

First, checking the distribution of all the tickets that were sold:





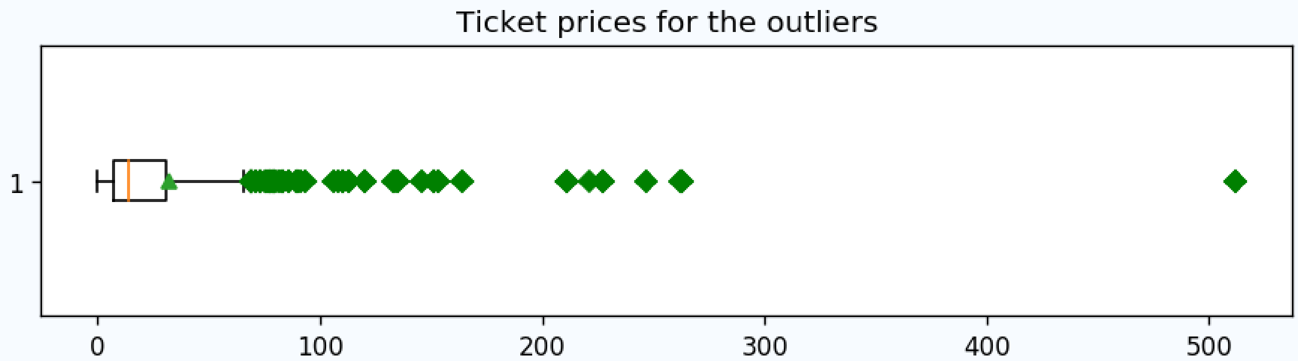


From the above plots, we see that most of the passengers paid anywhere between 0 and ~$50 for a ticket. Also, we can see bumps in the $200s and $500s ticket prices. Let's take a closer look at those numbers:

```
Basic statistics for the 'Fare' variavle
count    891.00
mean      31.79
std       49.70
min        0.00
25%        7.00
50%       14.00
75%       31.00
max      512.00
Name: Fare, dtype: float64
```

It seems that there is a huge difference between the max and the average prices of tickets. The standard deviation is bigger than the mean. There must be outliers, let's check if we can find them with a boxplot:

Outliers who paid more than $151 for their ticket



Most of the x axis above (showing the distribution of the ticket prices) is populated by outliers (in green). Next is a table with only the records of passengers who paid more than $151, which are the outliers.

```
Outliers' Fare variable numbers and basic statistics
=====================================================
count     29.00
mean     240.34
std      102.73
min      151.00
25%      164.00
50%      227.00
75%      262.00
max      512.00
Name: Fare, dtype: float64

Outliers - Ticket price and the number of people who purchases in this price
=========================================================================================
221    1
247    2
262    2
164    2
153    3
512    3
211    4
263    4
227    4
151    4
Name: Fare, dtype: int64

Survival rate for the outliers
===============================
Yes    0.69
No     0.31
Name: Survived_y_n, dtype: float64

Number of Outliers who survived
================================
Yes    20
No      9
Name: Survived_y_n, dtype: int64
```

Breaking down the numbers in the Fare variable, 69 percent of the passengers who paid more than 151 dollars for their ticket survived! In numbers, it is 20 passengers who survived and 9 who did not.

Also, the group that stands out most is the 512 dollars one: 3 passengers paid this sum of money, which is 128 times more expensive than the lowest price ticket ($4) and 16 times more than the median price.

Did those 3 passengers survive?

```
Top 3 most expensive ticket holders survival:
```

Out[67]:

	Survived	Class	Gender	Age	Fare	Sex	Survived_y_n	Ages	Survival
258	1	1	female	38	512	1	Yes	30-40	1

	Survived	Class	Gender	Age	Fare	Sex	Survived_y_n	Ages	Survival
258	1	1	female	38	512	1	Yes	30-40	1
679	1	1	male	9.6	512	0	Yes	0-10	1
737	1	1	male	9.2	512	0	Yes	0-10	1

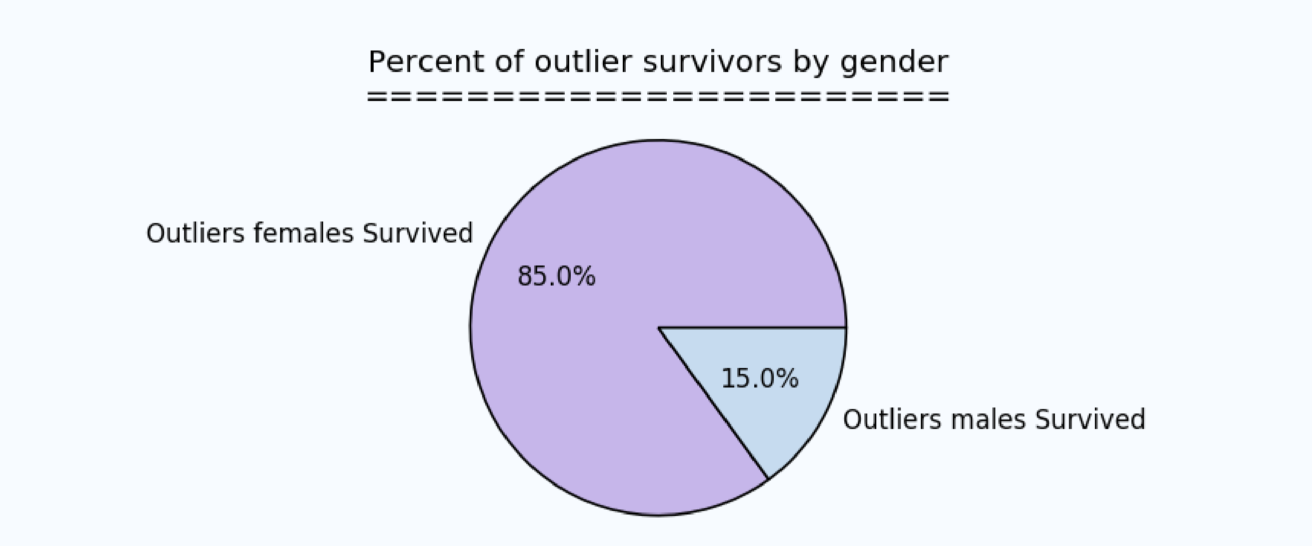
It seems that all 3 passengers, who were in their 30s, in first class and paid $512, survived. This is 100% survival. Nevertheless, this doesn't mean that there is a dependency between the ticket price and survival since there are only 3 items in this sample.

What about the rest of the passengers whose ticket price was more than 3 standard deviations above the average price? did their survival rate remain the same as the 'top-outliers' (100%)?

As it shows above under 'Survival rate for the outliers', 69% of the Outliers who paid more than $151 for a ticket survived. This is a higher rate than the rate of survivors in general (38%), and higher than the survival rate of females (65% from all survivors) and even higher than all survivors from the first class (63%) on board.

All outliers survivors

```
Outliers number of survivors by gender
==========================================
female    17
male       3
Name: Gender, dtype: int64


Outliers percent of survivors by gender
===========================================
female    0.85
male      0.15
Name: Gender, dtype: float64
```



85% of all survivors who paid more than $151 for their ticket were females.

Correlation between Fare and Age

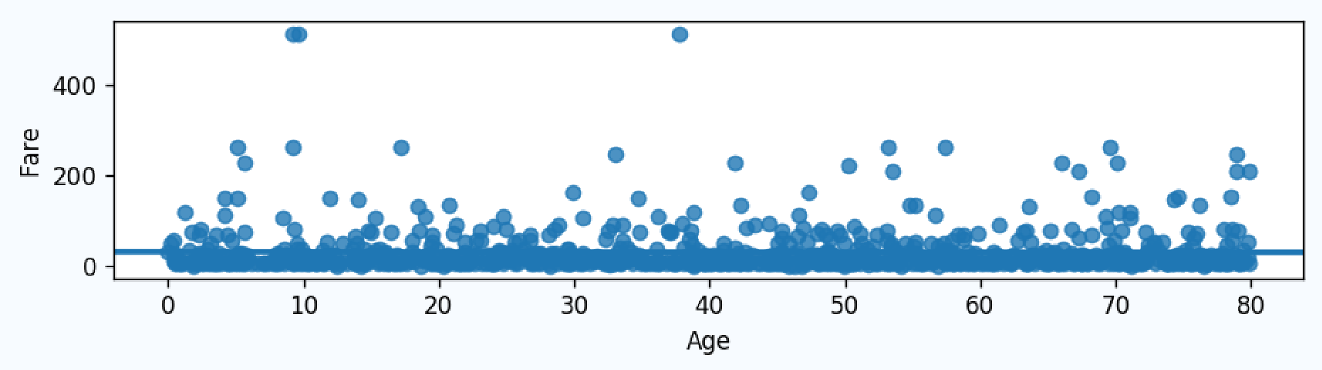
Correlation between Fare and Age

Correlation between Fare and Age
==================================
(451, 440)

- The 2 numbers above indicate a poor correlation between the two variables.

Out[71]:

<matplotlib.axes._subplots.AxesSubplot at 0x21b03a14eb8>



- This graph shows the regression line almost flat (almost 0), which means that there is no correlation between those 2 variables. The price of the ticket was not dependent on the age of the passenger. Some young passengers in their 20s paid as much as older people in their 70s.

Correlation between Fare and Class

Correlation between Fare and Class
==================================
(249, 642)

- We can see a much stronger correlation between the fare passengers paid and the class they were in, than with the Age they were.

[Top](#)
[Fare](#)

Summary

[Top](#)

Discussion

* This project does not include the crew members on board and their survival statistics. The scope of this analysis is limited to the passengers only.

The luxury steamship RMS Titanic sank in the North Atlantic Ocean in the early morning hours of 15 April 1912 while carrying 891 passengers (577 males and 314 females). Passengers were divided to 3 different Classes, where third class composed the majority of passengers (more than 50% were from the third class (491 compare to 400 from both first and second classes)). The Titanic passengers, who's ages ranged from less than 1 year to almost 80, paid anywhere between $512 per ticket to not paying at all.
Important to note that the original dataset was missing 177 records of the Age variable. Random numbers were introduced instead of the empty cells in the dataset in order to be able to do calculations that included the Age of passengers.

From 981 passengers 342 survived and 549 did not. This is about 40% of the population on the Titanic that survived. For every 10 people who survived 16 perished.

Taking into consideration the above analysis and given data, females survival from the entire population was almost twice as that of males (65%/35%). Moreover, females' survival rate from only women passengers was 74% compare to only 18% for men. Being a woman, one had 4 times more chance to survive on the Titanic in its first and only voyage.

First Class passengers survived disproportionally to their number from the population. They had 63% survival rate compare to 47% for Second Class and 24% for Third Class. Clearly being a First Class member gave one a better chance to survive. In Second Class the difference in survival rate was 4 times in favor of women (70/17). And in Third Class women survived 1.6 more times than men (72/47). By the numbers of gender survival and class we can see that women survived more than men in all classes. The highest rate of survival for women by class was for the ones in the second class with 80%, follwed by 67% for the first class and 61% for the third class. Class seem to did not matter as much as gender for survival. Unless the difference is not statistically significant different, which will be interesting to check with a statistical test.

There were 29 passengers who bought a significantly more expensive ticket than the rest of the passengers for more than 151 dollars and with average of 240 dollars per ticket. Maximum price of ticket purchased was 512 dollars. The survival rate of this group was 69% men and women together, which are 20 survivors out of 29. From those 20 (probably rich) survivors 85% were women.

Analyzing the age groups, it doesn't seem to be that age affects someone survival rate significantly. A statistical test should be done to prove this last point.

* Conclusion: So, who had the best chances to survive? Females on the Titanic had the best chance to survive, eapecially ones in Scond Class. The chance for women will increase to 85% if one pays more than 151 dollars for the ticket.

* Further interesting analysis: Did young females have better chances to survive than young males? Did males and females paid the same amount for their tickets?

[Top](#)

[Summary](#)

Sources

[Top](#)

Page Name | URL

- Udacity Nano degree Pandas Series (and many other Udacity's course materials that are not cited here) https://classroom.udacity.com/nanodegrees/nd002/parts/0021345403/modules/3176718735
- Visualizing the distribution of a dataset http://seaborn.pydata.org/tutorial/distributions.html
- Pandas options and settings http://pandas.pydata.org/pandas-docs/stable/options.html
- pandas.DataFrame.round http://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.round.html
- How to make a pandas crosstab with percentAges? http://stackoverflow.com/questions/21247203/how-to-make-a-pandas-crosstab-with-percentAges
- Pandas: add crosstab totals http://stackoverflow.com/questions/26932682/pandas-add-crosstab-totals
- color example code: named_colors.py http://matplotlib.org/2.0.0b4/examples/color/named_colors.html
- Pandas - replacing column values http://stackoverflow.com/questions/31888871/pandas-replacing-column-values
- Pandas: Replacing column values in dataframe http://stackoverflow.com/questions/23307301/pandas-replacing-column-values-in-dataframe
- How To Convert Data Types in Python 3 https://www.digitalocean.com/community/tutorials/how-to-convert-data-types-in-python-3
- Setting the Title, Legend Entries, and Axis Titles in matplotlib https://plot.ly/matplotlib/figure-labels/
- count the frequency that a value occurs in a dataframe column http://stackoverflow.com/questions/22391433/count-the-frequency-that-a-value-occurs-in-a-dataframe-column
- Working with missing data http://pandas.pydata.org/pandas-docs/stable/missing_data.html
- pandas.DataFrame.dropna http://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.dropna.html
- Count of unique value in column pandas http://stackoverflow.com/questions/41665659/count-of-unique-value-in-column-pandas
- Create a Column Based on a Conditional in pandas https://chrisalbon.com/python/pandas_create_column_using_conditional.html
- Basic Data Plotting with Matplotlib Part 3: Histograms https://bespokeblog.wordpress.com/2011/07/11/basic-data-plotting-with-matplotlib-part-3-histograms/
- seaborn.FacetGrid http://seaborn.pydata.org/generated/seaborn.FacetGrid.html
- seaborn.factorplot http://seaborn.pydata.org/generated/seaborn.factorplot.html#seaborn.factorplot
- pandas.Series.value_counts http://pandas.pydata.org/pandas-docs/stable/generated/pandas.Series.value_counts.html
- Plotting multiple different plots in one figure using Seaborn http://stackoverflow.com/questions/38082602/plotting-multiple-different-plots-in-one-figure-using-seaborn
- Resizing matplotlib figure with set_fig(width/height) doesn't work http://stackoverflow.com/questions/31841289/resizing-matplotlib-figure-with-set-figwidth-height-doesnt-work
- Pandas error - invalid value encountered http://stackoverflow.com/questions/30519487/pandas-error-invalid-value-encountered
- How do I use matplotlib autopct? http://stackoverflow.com/questions/6170246/how-do-i-use-matplotlib-autopct
- How to normalize index http://stackoverflow.com/questions/21247203/how-to-make-a-pandas-crosstab-with-percentages
- Random row selection in Pandas dataframe http://stackoverflow.com/questions/15923826/random-row-selection-in-pandas-dataframe
- Matplotlib screenshots https://matplotlib.org/users/screenshots.html
- Smooth histogram from data in column in Pandas DataFrame? http://stackoverflow.com/questions/35590727/smooth-histogram-from-data-in-column-in-pandas-dataframe
- Find the unique values in a column and then sort them http://stackoverflow.com/questions/32072076/find-the-unique-values-in-a-column-and-then-sort-them

- pylab_examples example code: boxplot_demo.py https://matplotlib.org/examples/pylab_examples/boxplot_demo.html
- Matplotlib boxplot without outliers https://stackoverflow.com/questions/22028064/matplotlib-boxplot-without-outliers
- Pandas HTML Output Conditional Formatting - Highlight cell if value in range https://stackoverflow.com/questions/37638402/pandas-html-output-conditional-formatting-highlight-cell-if-value-in-range
- 'Could not interpret input' error with Seaborn when plotting groupbys https://stackoverflow.com/questions/32908315/could-not-interpret-input-error-with-seaborn-when-plotting-groupbys
- Udacity x^2 test https://classroom.udacity.com/courses/ud201/lessons/1331738563/concepts/189540023092
- How to Include image or picture in jupyter notebook https://stackoverflow.com/questions/32370281/how-to-include-image-or-picture-in-jupyter-notebook
- Better Plotting In Python With Seaborn https://robinsones.github.io/Better-Plotting-in-Python-with-Seaborn/
- How to reference a IPython notebook cell in markdown? https://stackoverflow.com/questions/28080066/how-to-reference-a-ipython-notebook-cell-in-markdown
- https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chi2.html

Top
Sources

Navigate back to:

1. Sources
2. Summary
3. Fare
4. Age
5. Gender
6. Class
7. Analyzing the data
8. Discovering the data
9. Top