

Assignment 3 - Part 2

The story about Titanic is well known and there is a big dataset with a lot of information about passengers (age, gender, class, survival rate...) that were on that boat. There are a lot of conclusions that can be made from there. In this project I am going to use statistical analysis and data visualization to provide answers to the next hypotheses:

- Determine if the survival rate is associated to the class of passenger
- Determine if the survival rate is associated to the gender
- Determine the survival rate is associated to the age

```
In [1]: import pandas as pd # importing libraries that I need
import seaborn as sns
from matplotlib import pyplot as plt
%matplotlib inline
```

```
In [2]: titanic = pd.read_csv("titanic.csv") # reading dataset and storing it under name titanic
```

For making conclusions about the association between two variables we can use functions such as corr() for calculating correlation, but also we can use graphs, plots, and charts to make our conclusions even more clear to readers by presenting them on visuals.

Association between survival rate and class of passenger

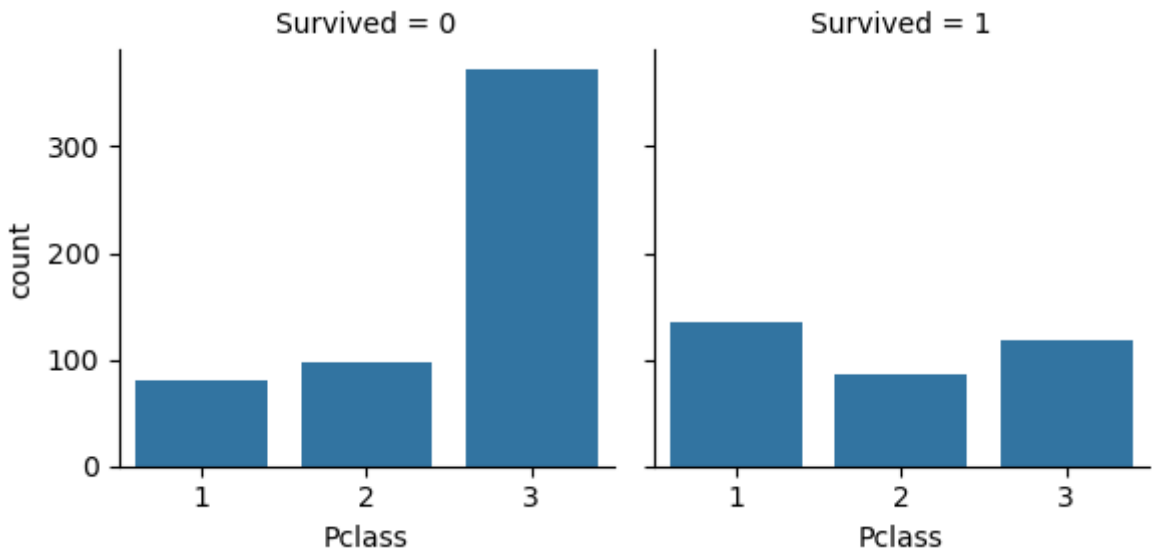
```
In [3]: titanic['Survived'].corr(titanic['Pclass']) # calculating correlation between survival rate and class of passenger
```

Out[3]: -0.33848103596101475

There is some negative correlation between survival rate and a class of passenger, so we can say that if a class of passenger has a higher value (that means lower class - 3 is lower class than class 1) there is a higher chance that survival rate will be lower (lower is 0) and that means in English that as a lower class makes higher chances not to survive. Let's see that on plots as well and make it more obvious.

```
In [6]: g = sns.FacetGrid(titanic, col="Survived")
g.map_dataframe(sns.countplot, x="Pclass")
```

Out[6]: <seaborn.axisgrid.FacetGrid at 0x1827ff82130>

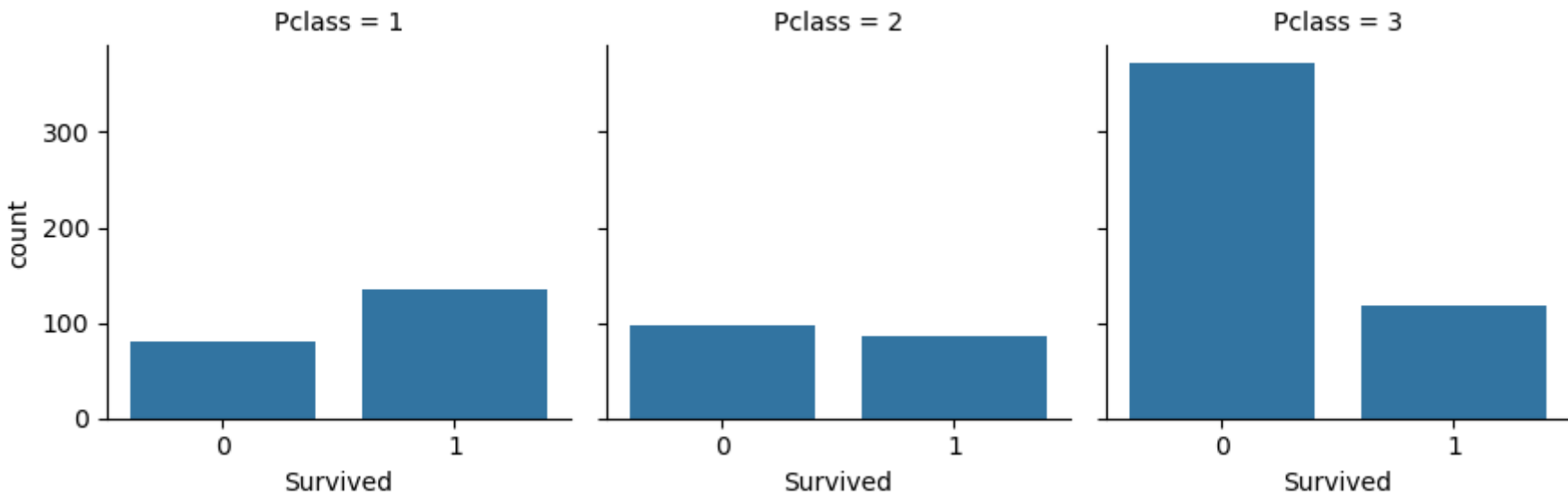


Here I made bar plots for comparing the number of passengers from each class in the group of the ones who survived, and the ones who didn't survive. We can see that in the group where are the ones who didn't survive, there is a higher number of passengers from third class, and in the survived group there are more passengers from class 1.

Now I will compare survival rates in separate classes. It looks more obvious that one who was in first class had a higher chance to survive than the one who was in second or third class

```
In [7]: g = sns.FacetGrid(titanic, col="Pclass")
g.map_dataframe(sns.countplot, x="Survived")
```

Out[7]: <seaborn.axisgrid.FacetGrid at 0x18201e63940>

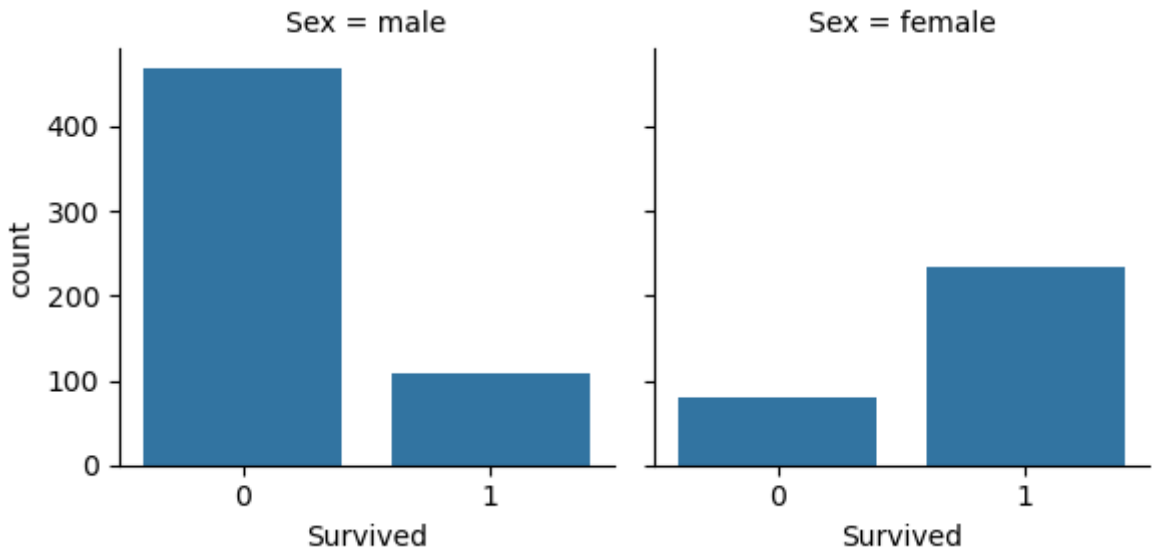


Association between survival rate and gender of passenger

If we compare survival(1) and not survival (0) beyond male passengers we can see that much more male passengers didn't survive. On the other hand, if we look at the same thing for female passengers, we can see that more females survived (there are more 1 than 0). From a comparison of these two bar graphs, we can conclude that the survival rate is somehow associated with gender.

```
In [8]: g = sns.FacetGrid(titanic, col="Sex")
g.map_dataframe(sns.countplot, x="Survived")
```

Out[8]: <seaborn.axisgrid.FacetGrid at 0x182020a7640>



Association between survival rate and age of passenger

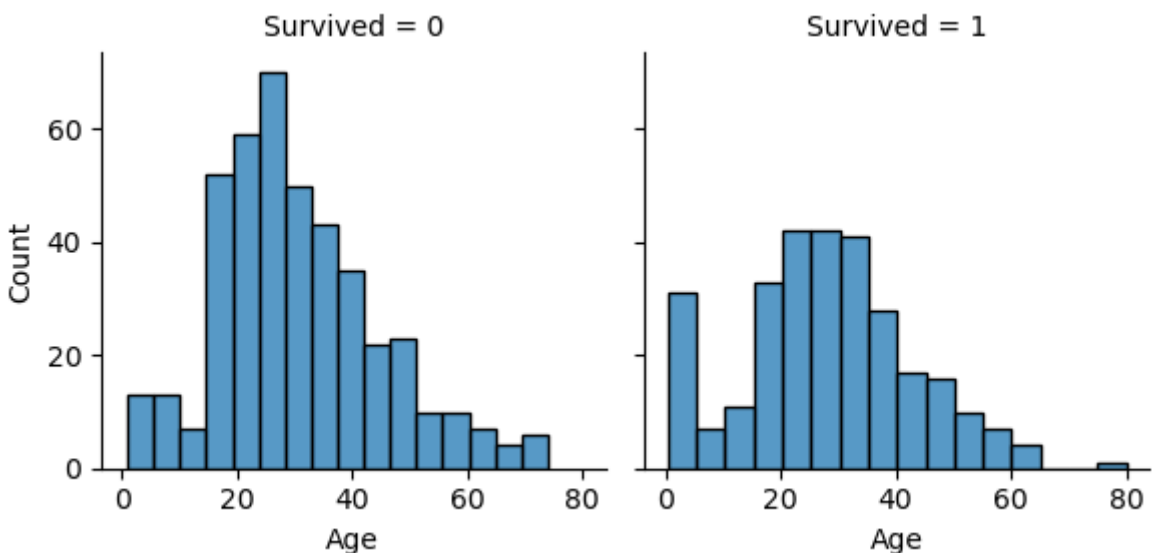
```
In [9]: titanic['Survived'].corr(titanic['Age']) # calculating correlation between survival rate and age of passenger
```

Out[9]: -0.07722109457217764

There is some small negative correlation between survival rate and a class of passenger, so we can say that if the passenger is older there is a higher chance that the survival rate will be lower (lower is 0) - not to survive. In this case, the correlation is not that high but there is some correlation between those two variables. Let's see that on plots as well.

```
In [10]: g = sns.FacetGrid(titanic, col="Survived")
g.map_dataframe(sns.histplot, x="Age")
```

Out[10]: <seaborn.axisgrid.FacetGrid at 0x1820211ef10>



From the histograms also we can conclude the same as from calculating correlation, but the association between those two variables is not as obvious as it was the case with survival rate and gender or survival rate and passenger class.