

# Exploratory Data Analysis of titanic dataset

from Kaggle [Titanic Data Set from Kaggle](#)

## General info about the data

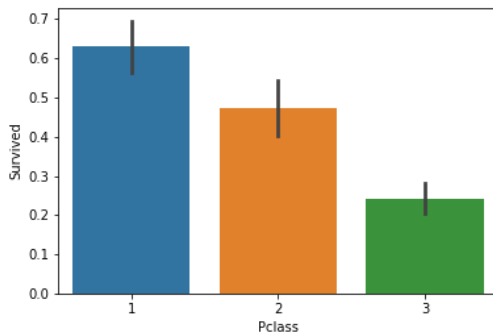
The sample of data is 891 passengers from the Titanic, passengers that could have been identified.

From the 12 variables available, we will focus on only 4 and an ID:

PassengerId, Survived (0=No, 1=Yes), Pclass (Passenger class 1,2,3), Sex (Male, Female), Age (in years, float)

**Missing data:** Age is not always indicated; this is in 19% of the cases. We decided to leave the missing data as unknown, as it could significantly change the result of the analysis if it was replaced by a random value.

## Is the survival rate associated to the class of passenger?

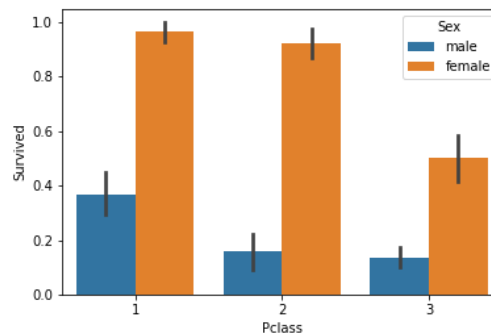
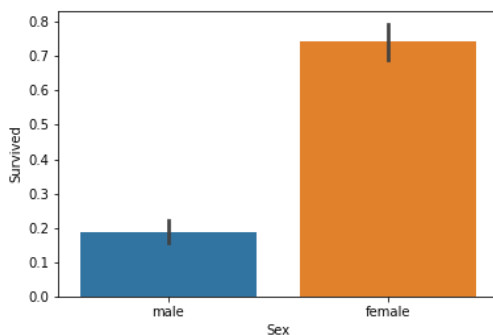


Created a barplot indicating the percentage of survivors per passenger class.

Over 60% of the passengers in 1<sup>st</sup> class survived, compared to only 24% in 3<sup>rd</sup> class.

We can conclude that the class has a relation with surviving the sinking.

## Is the survival rate associated to the gender?



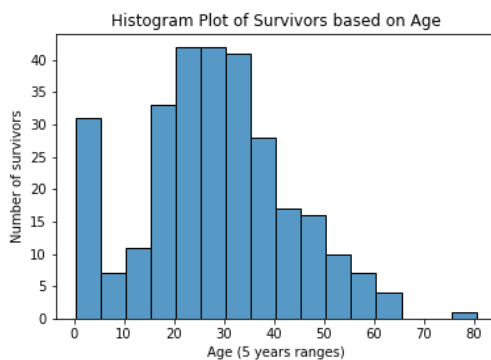
Created a barplot indicating the percentage of survivors for male and female passengers.

About 75% of the female passengers survived, for less than 20% of the male passengers. We can conclude that the gender has a relation with surviving the sinking.

Created an additional barplot (right) to see the relation between surviving rate, class and sex.

The female passengers in 3<sup>rd</sup> class had more chances to not survive than the women in the other classes.

## Is the survival rate associated to the age?



Created a barplot indicating the number of survivors per age range, each range is 5 years.

We can see that the high amount of survivors are in the first age range (0-5), the second group with highest rate is young adults, between 15 and 35 years old.

The shape of the plot is close to a normal distribution except the anomaly of the first age group.

We can conclude that the chances of surviving are related to the age.

For more information, all results and documents are available in [my public github repository](#)