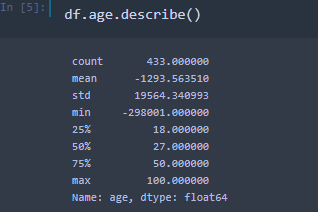
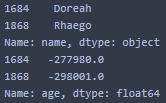
I used Game of Thrones dataset from Kaggle, one of the best TV series. The dataset contains data for 1946 characters. It indicates whether the character has died or not, and a number of variables, including the gender, nobility, etc. data for the character.

Data was not clean, so I had to do some changes. There were total 22108 NAs. It is always better to change the NA values that can be changed. Sometimes using a mean value for this, might work. When I looked at the age column, I saw something strange.

**Figure 1:**



Mean age was a negative value. So, I had to fix it. As you can see in the Figure 2 below, there were characters that have negative age. They were affecting the mean value.

**Figure 2:**

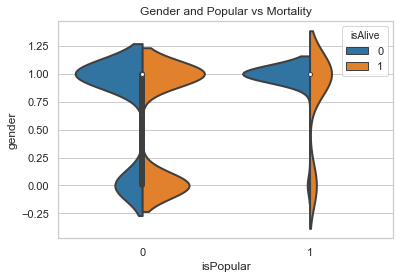
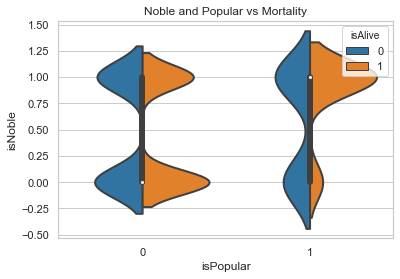
There are many ways to deal with NAN values. If there are few NAN values, it may be healthier to remove those values from the data and continue the analysis. However, there were too many NAN values in the data set we worked on. Filling the values that can be filled would be better in this case.

I researched the characters an what I found was; according to <https://gameofthrones.fandom.com/wiki/Doreah>, Doreah is actually around 25 and according to <https://gameofthrones.fandom.com/wiki/Rhaego>, Rhaego was never even born. I replaced the negative ages so that mean was fixed. However, there were still nulls in the data. I replaced them with mean. Other null values in the data was replaced with empty string. As a result of these processes, there are no NAs in the data.

By making exploratory data analysis, I will decide the variables I will put in my model.

Violin plots are almost the same as box plots. But they also show the probability density of the data at different values. Box plots show mean, median and interquartile ranges, the violin plots show the full distribution of the data.

**Figure 3-4-5-6:**



According to the violin plots:

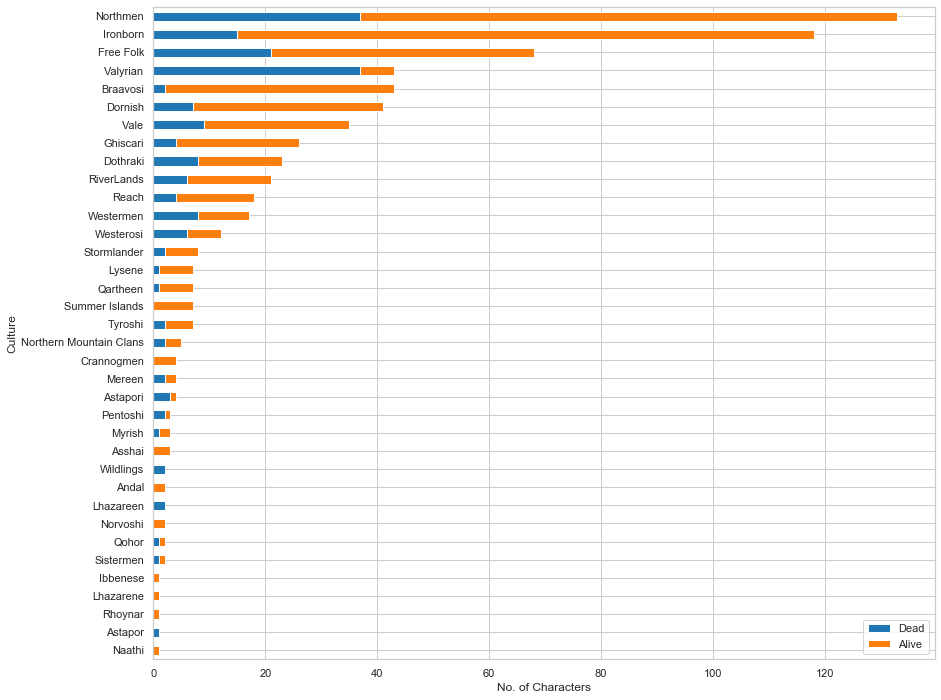
* Noble and popular characters are most likely to be alive,
* Popular characters that appearing in book1 most likely to be dead
* Not popular characters that appearing in book1 most likely to be dead
* Single and not popular characters are most likely to be alive
* Single and popular characters are most probably dead

So, if you are a popular, married and noble female character in the Game of Thrones Universe, you are most likely to be alive. Also, you have to appear in the book1.

The data preparation process is not over yet. When we look at the culture column, we see that there are different spelling types for the same culture group. For instance, Westermen takes 3 different values; westermen, westerman, westerlands. These name errors have been corrected.

After correcting these values, I wanted to find out the relationship between culture and survival.

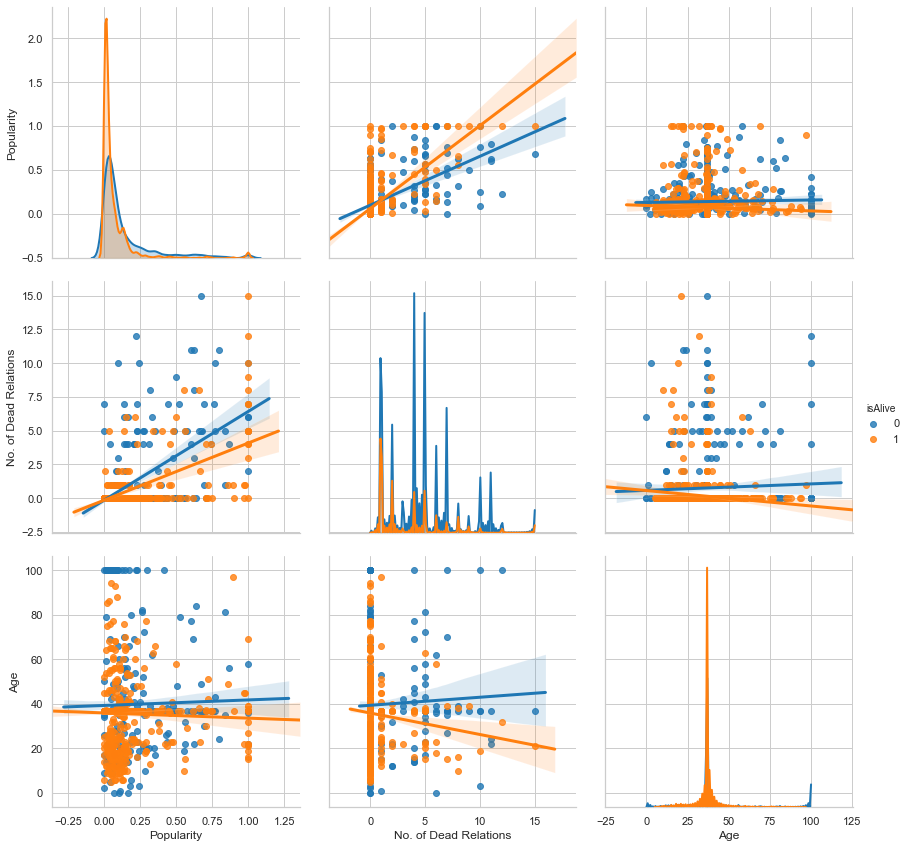
**Figure 7:**

****

We can see that Northmen and Ironborn are more than many in number. There is no death in Summer Islands and Crannogmen etc.

Also, I checked the relationship between popularity, number of dead relations, and age.

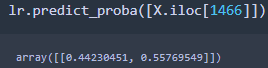
**Figure 8:**

****

There is nothing much to interpret from these graphs, but we can say that when popularity increases number of dead relations increased.

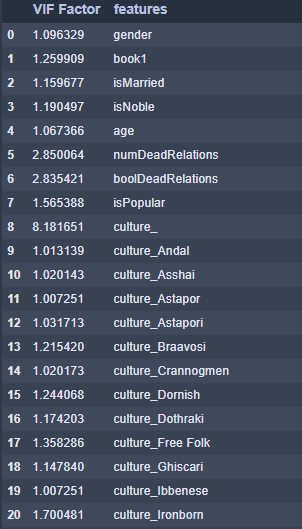
Therefore, for my model, I omitted several variables that were not required to estimate the character's survival rate. For example; S.No, plod, title, dateOfBirth, DateoFdeath, mother, father, heir,house, spouse, book2, book3, book4, book5, isAliveMother, isAliveFather, isAliveHeir, isAliveSpouse, popularity and name. I also removed the isAlive variable from the data so that I can use it for the response variable.

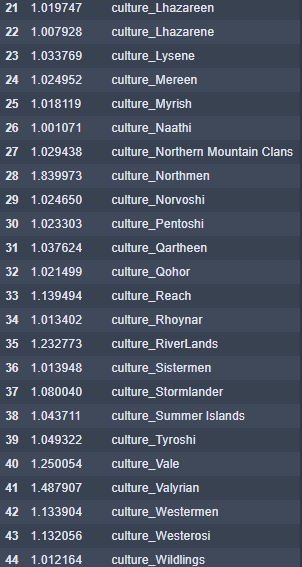
When I put the remaining variables into the model, the score of the model came to 0.776 which is pretty good actually. So, I made sure that I could put the character I wanted to guess into the model. I wanted to estimate the living rate of Arya Stark. While watching the series, my favorite character was already Arya Stark and I knew that she lived at the end of the series.

**Figure 9:**

According to my first model, Arya's chances of living is 55%.

I looked at the multicollinearity by looking vifs.

**Figure 10:**



Arya had the chance to live 0.557 by using the important variables I found by doing EDA. And my model has no multicollinearity problem.