Analysis Write Up

1. Problem Summary

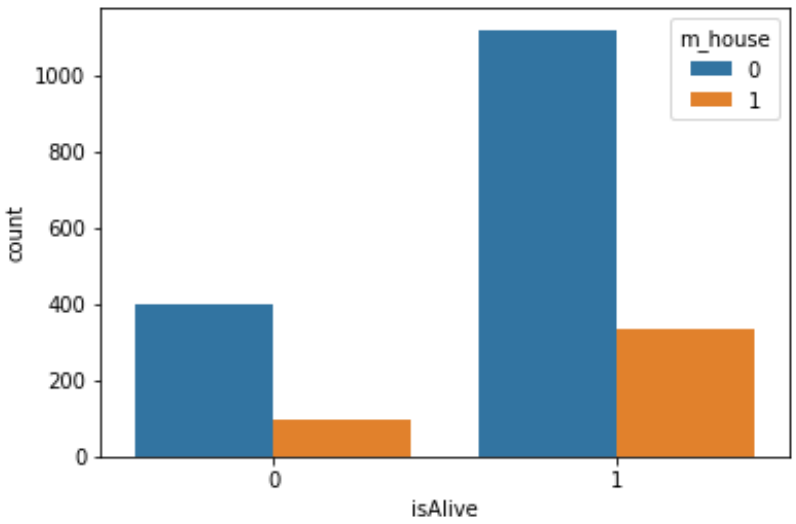
There are approximately 2,000 characters in *A Song of Ice and Fire* by George R.R. Martin. This book series was the inspiration for the HBO series *Game of Thrones*.

I was given the data set of the characters from the books and my tasks as a future master of analytics is to predict which characters in the series will live or die, and give data-driven recommendations for our client in Westeros on how to survive in Game of Thrones.

2. Key Insight and recommendation for action

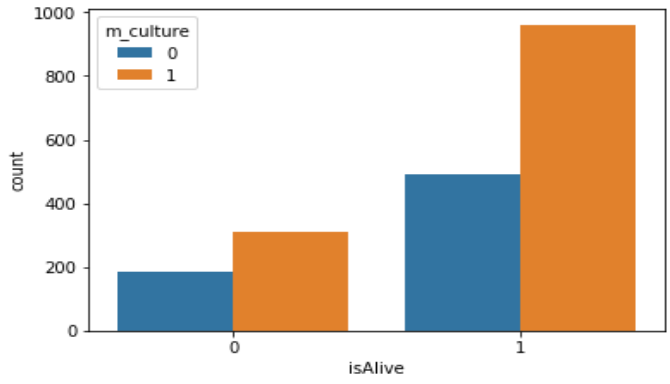
Overall, in the dataset, there are 75% of the people are alive, while the other 25% are dead. There are 25 different variables within the original dataset, which are: character number, name, title, male, culture, etc.

By doing the data exploration, I have come up with some insights that can help us to survive.

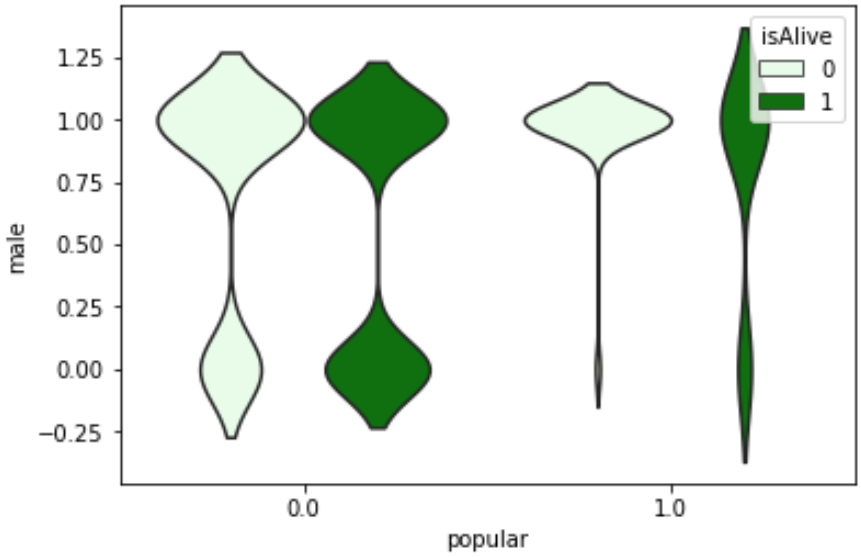


First of all, we can see that if we serve a house, there are a higher chance to survive. The chart above shows that people who have a house to serve, which is in blue, are more likely to survive. So my first recommendation for our client is to find a house to serve, that way they are a bit safer since they are under protection of a big house.

Culture, on the other hand, doesn't help keep us being alive in the Game of Throne world.

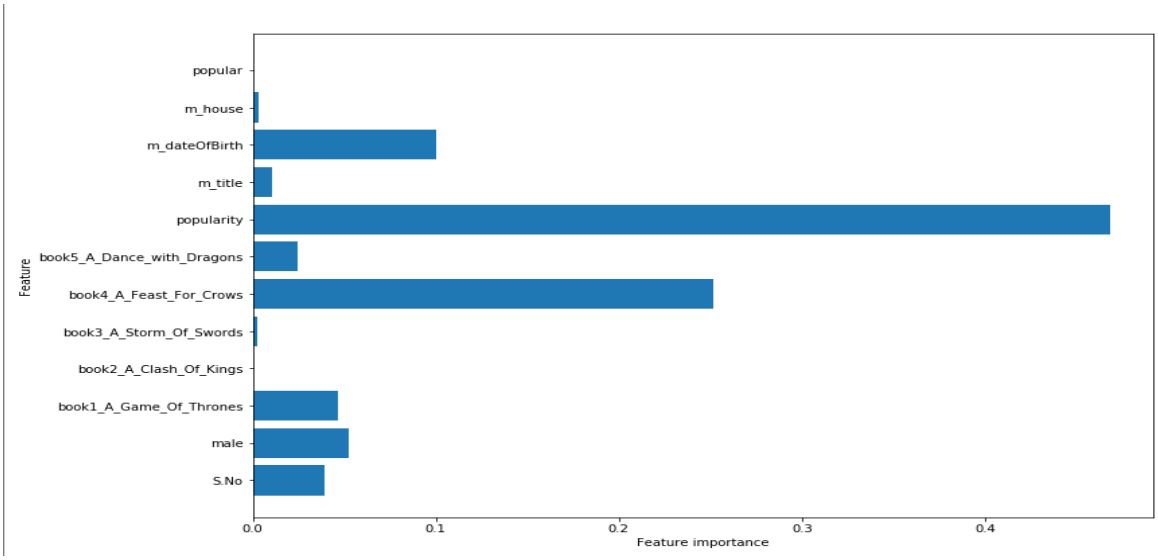


Even if you belong to a culture group, there is still a high chance that you die since there are more people who doesn't belong to any culture group survived (the one in orange).



By checking the popularity and gender of the character, we can see that female has a higher chance of surviving than men. Moreover, the unpopular character survived much more than the popular ones, which mean that in order for our client to survive, they should keep a low profile, disguise as a female person try to avoid being mentioned in the book. By doing so, you will have a much higher chance to survive.

By checking the feature importance, we can see that there are several factors that contribute to the survival in Game of Throne and I have discussed on what we should do with these variable with the analysis above.



The best result for my predicting model was with the Random Forest model, which had the training AUC score of 0.892 and the test AUC score of 0.846. And the Cross validation value was 0.82

Bibliography

Jeffreylancaster/game-of-thrones. Retrieved March 19, 2019, from https://github.com/jeffreylancaster/game-of-thrones

Game of Thrones Wiki. (n.d.). Retrieved March 19, 2019, from https://gameofthrones.fandom.com/wiki/Game_of_Thrones_Wiki