

Riley Fiske

Dr. Gregory Tanner

DATA-318

18 February 2022

Project 1 Proposal

For project one, I would like to circulate it around a dataset of a Star Wars interest survey¹ that was put on by Fivethirtyeight.com, which is affiliated with ABC News. The survey sought out information such as if the respondent had seen any of the (at the time) six Star Wars films, if they would consider themselves a fan, which films they had seen, their ranking of all six films, the favorability status of 14 well-known characters, information about the individuals' sex, age, income, education, and region, and a few other fun questions. I would like to see if an individuals' favorite film, or favorability of each film, can be predicted using the individuals' personal information. It would be interesting to see if the situation you are in as a person has an impact on what Star Wars film stuck with you the most. To do this, I plan to use both linear and logistic regression, and to have a training set that is run through the data, then to see if it can accurately predict the test set's favorability of each film using the regression model selected. This will then be evaluated using a confusion matrix to see the sensitivity, specificity, and accuracy of the data to see if it was successful or not. There are many different directions that could be taken with this data set, however this one seems the most intriguing to me and I look forward to seeing the result of it.

¹ Flowers, A (2014) Star Wars Survey [Source Code]