Harley Ryan

December 7th, 2024

DS 2004

CFB Case Study Rubric

Why am I doing this? This is an opportunity for you to collect, synthesize, and analyze real world data to solve a real world problem, that problem being how TV networks determine the teams for nationally televised games. You are going to have the opportunity to practice your skills with APIs, data cleaning, and data synthesis in order to perform your analysis. You will also have the opportunity to hone your statistical skills by using a probit regression model in this case study.

What am I going to do? You will take the code in the GitHub repository provided and use it to query the collegefootball.com API for the necessary data. You will use that data, along with the Reddit flair data that is also in the GitHub repository to create a database that incorporates Reddit users associated with each team along with stats from the collegefootball.com API database. Then, looking at the week by week game data, you will build a probit model that will be able to predict what factor contributes the most to a CFB game being broadcast on a flagship network, whether that is team performance or fanbase size. Ultimately you will deliver a pdf with an output showing the results of your probit model, along with predictions for if various games would be on national TV.

Tips for Success:
- Use CFBD's Elo score as a proxy for team performance. This is a far better metric than something like record, as it is a consistent, week by week metric that crosses over between seasons and also factors in the quality of a team's opponents. Documentation on the Elo system has been included in the GitHub repository if you are unfamiliar with it.
- Remember that Reddit flairs are a static metric, unlike Elo or performance based metrics. They have been provided for you in CSV form, as the Reddit API is rather unfriendly and not the best.
- Review probit models. This is a great model that allows you to predict the likelihood of something. While you don't need to understand all of the math behind the probit model to complete this assignment, you should review the concepts presented in the article in the GitHub.
- Remember to consider both teams in the game. Broadcasters consider both the home and away teams when deciding what games to televise.

Spec Category	Spec Details
Formatting	- Goal: Make your submission as readable and understandable as possible to graders The format of your submission should be a PDF with any conclusions that you made from the analysis. This should be submitted along with any code or other output that you create in a GitHub repo
Conclusion File	- Goal: Explain your findings to the graders and make sure they know that you successfully completed the assignment Your repo should include a conclusion file with some of the findings of the probit model that you built. Remember to also include a few predictions of your own for if games with various statistics will be on national television or not.
Code	- Goal: Allow others to replicate your findings and help graders understand your results Your repo should include any code that you ran to get your results. Make sure to add comments for any additional code that you wrote, and credit any outside sources that you used.