Group 2 – Project 1 Proposal

Gage Anderson, David Ferguson, Brandon Ibarra, Quintele Jackson

Topic: NFL Betting Data

Our exploratory data analysis will be over NFL Scores and Betting Data from 1999 through 2019, with the dataset found on Kaggle. Three datasets are given: NFL Stadiums, NFL Teams, and Spreadspoke Scores (the betting data). These datasets provide information on historical NFL game data, a list of every NFL team that has played a game during the time chosen (useful for tracking teams that have moved cities), and every stadium that has hosted an NFL game during the time chosen.

Inspiration for this project was found from our groups common interest in football, specifically the NFL. An analysis by Ty Walters[1] provided an example of how the data could be cleaned up (i.e., games listed on the wrong date) as well as how to deal with certain null values.

To analyze the data, we will merge the three data sets. All game data outside of the 1999-2019 time period given will be dropped from the table as well as any playoff or neutral site games. Any teams that moved cities/states will have their games scores aggregated under the current team's name and location. Certain columns will need to have their types adjusted (i.e., game date on the Spreadspoke_scores.csv file is listed as objects instead of Date Times).

[1] https://www.kaggle.com/code/twalters20/nfl-betting-model

Group 2 – Project 1 Proposal

Gage Anderson, David Ferguson, Brandon Ibarra, Quintele Jackson

Some possible research questions we could ask would be:

A. Do certain stadiums have more of a “home field advantage” than others?
B. How often do the favored teams win?
C. Which type of stadium has more average points scored: indoors (dome) or outdoors?

A possible thing we may predict with a regression model would be how likely it is for a team that is favored to beat the spread. Our color scheme will use red (#D50A0A), white (#FFFFFF), and blue (#013369) to go along with the logo colors of the NFL.

As for Roles & Responsibility, these listed are preliminary and are subject to change as new realizations or obstacles arise during the project:

- Gage: Cleaning the data and merging the tables together; regression analysis
- David: Research Question A
- Quintele: Research Question C
- Brandon: Research Question B

The entire group will work on the presentation slides and the final paper.