Scott Gill

Stat 426

Project Proposal

**What is the main question that you are trying to answer? Be as specific and detailed as possible.**

- What is the main factor in determining NFL TV Ratings/viewership? Most would tell you it has to do with the time slot of the game. Many would lead you to believe that it has to do with how big the market is for each of the teams participating. Others would lead you to believe it has more to do with the quality of teams playing. There are many theories out there, and I’d like to determine what is the most important factor in a specific TV Ratings model.

- Obviouslythere are some limitations to this study, since I’m assuming each Ratings model is different in itself. Creating a Ratings model is likely a machine learning exercise to begin with. But I hope to do a good job at weighing the many factors that go into every single game to determine what looks like “the best game.”

**Where are you getting your data? What does your data look like?**

- My first dataset will be from Kaggle, that has the site, teams, and result of every NFL game. I’ll also need to merge a couple of the other datasets that came with the large dataset, which includes features about each stadium as well as another that includes more features about each team.

- The hiccup is acquiring the correct TV Ratings dataset. I’ve found a few, nothing that I’ve settled on that completely work with my data. I’m trying to stay away from Kaggle as well since I’d like to achieve the maximum amount of points as possible for this project.

**What is the target?**

- It will end up being one of two variables – TV Ratings for an NFL game, or the amount of viewers for an NFL game

**What are the features?**

**-** I have dozens of features, here is a list of just a few of the key ones:

* Schedule\_date and Schedule\_season
* Schedule\_week and Schedule\_playoff
* Team\_home and score\_home
* Team\_away and score\_away
* Team\_favorite\_id, spread\_favorite, and over\_under\_line (gambling purposes)
* Stadium, weather\_temp, weather\_detail, etc.

**Are you planning on creating / engineering new features?**

- Definitely, there will be many things I’ve already thought of that will need to take place for proper analysis to occur. I’ll need to create a variable that includes the winning percentage for each team at the time of the game they were playing. Also a variable for whether the game was nationally broadcasted. Also I’ll need to add a variable for what number of game it is within their season. Perhaps another one is the number of wins back they will be from others in their division, which would signify importance. Lastly, I’ll need to just create a variable for the deficit of the score and take the absolute value of that.

These are lofty and ambitious goals, so I would love if you had any ideas to help with the creation of some of these.

**How many observations do you / will you have?**

**-** Currently the data has 12,935 observations, the number of how many games have been played over the course of NFL history. However, I will most likely be using only games that have taken place since 2000, since the television watching experience has changed so much then. At the very earliest I will be using games from 1980, when ESPN actually came around.

**How does the data that you are collecting help you answer your question of interest?**

I have, and should be able to create, just about every possible feature I would need that would be important for determining viewership of any NFL game. Whether it was a playoff game or not will make a big deal, since those are nationally broadcasted. There will be some limitations, but I feel like for the most part it should give us a relatively accurate, in depth look at the most important factors and some key relationships for identifying in the future.

If this is too ambitious, or you have any concerns about my efforts so far, please let me know because I’d love to sit down and hear your thoughts about it. Also, if you have any preferences/background with TV Ratings or where to go about finding the most reliable sources for them, I’d love for you to reach out.