Rain Lasch
INFO 2201
Prof. Zietz
12 November 2021

Final Project Proposal

For my final project, I want to build a program that tells a story about how NBA players' twitter performance affects both their performance on the court and their salary. I think this is an interesting question to explore since social media is often used as a means of advancing business with increased social influence, and I think NBA players may undergo a similar process in order to increase their value in the player market. Many players, for example, are extremely active on social media, and there have been claims around certain players' social media activity like that Kevin Durant has more tweets than career points. LeBron James similarly tweets on a regular basis and is the most followed player in the NBA. Both players are some of the most paid in the league at 5th and 6th respectively, both making over $40M in salary for the 2021-22 season. While the two players are undoubtedly two of the most skilled basketball players in history, I think players' increased social media over time may contribute to increased game performance and salary as the players become central influencers for NBA franchises.

For the design of the project, I will collect data on NBA players' tweets, seasonal stats, and salaries. For the tweets, I have accessed Twitter's API with an app that I have created in their developer portal, and I was able to complete a request to the API using application-only authentication with a bearer token that the developer portal provided. This authentication process allows me to use read-only access to get public Twitter users' tweet timelines, and these timelines can be used to get ranges of tweets for specific periods of time (like a season), mentions, number of tweets, and tweet content like text and images. The authentication process was a bit tough to figure out, but after reading the documentation and consulting StackOverflow, I was able to get it working. I plan to, at the minimum, get the number of tweets that particular players have tweeted over the course of a season or two. This will allow me to compare the number of tweets and overall Twitter activity for a user between two seasons, and this

comparison can show whatever differences in performance or salary occurred along with changes in social media activity.

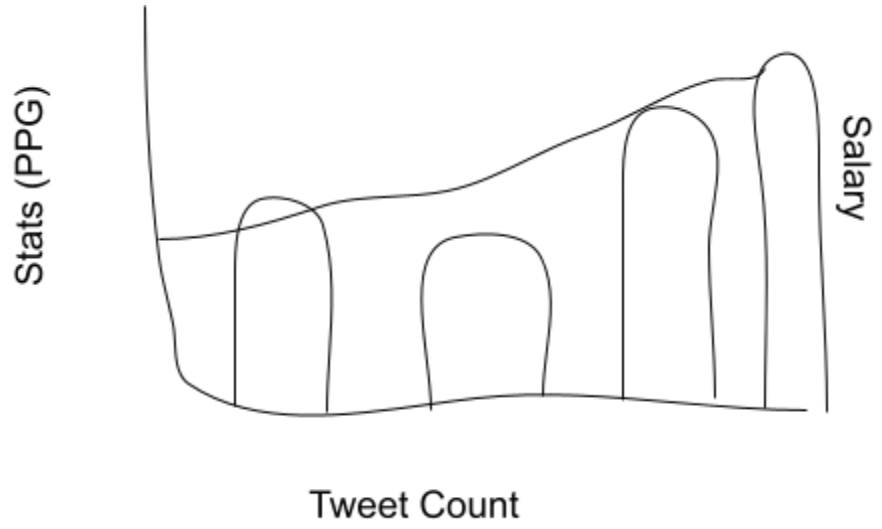
For the NBA statistics, I've found an API called *balldontlie* which gives player statistics for games in a season or simple season averages. This API is very interesting as there was no signup required, yet it has tons of information on players' stats for individual games in a season, their current team, specific teams for each season, and other useful NBA information. I plan to request information from this API regarding player game stats and season averages for different seasons. With this information, I can get players' performance stats for entire seasons, but I can also get data on their performance for individual games within a season. With players' seasonal performance stats, I could compare these along with the yearly tweet count and their salaries for two seasons, and with game data, I may be able to analyze how twitter activity affects game performance within a season, too.

For the player salaries, I got a CSV file of NBA player salaries from 1985 to 2018 (2017-18 season). So, for this project I will stick to stats and tweets within this time range as well. To go along with this, there is a players CSV as well that tells what player name goes with what player ID. So, with this data, I will use the players CSV to figure out the player ID for whichever player is requested (i.e. a request of LeBron James will find his ID to use in the salaries dataset). Then, having found the player ID, I will find their salaries for different years, and I can then visualize this in various different ways, and after getting the data coded together correctly, I plan to add some visualizations that show how athletic performance and player salaries have varied over time with respect to their Twitter activity.

Some roadblocks I expect to encounter include getting the data from APIs to work with pandas CSV data. I am not quite sure how I will apply the CSV values to API data at the moment, so I will definitely have to figure out how to interact with both kinds of data at the same time. Another potential problem could be the complexity of the Twitter API, but there is plenty of documentation and example code to work through the various Twitter endpoints. Visualization will also pose another challenge as I think any graphs I create would require two y-axes. I could create multiple graphs to display two different effects on salary and game performance, but I think a visualization with both of these components would be much more informative. Finally, there may be some problems with running Python functions and code in the correct order, for that was a major problem to take care of in my earlier work with the World Bank API. With all

these obstacles in my mind, I think I can manage to put together a nice final product with these data.

Expected Output:



Source Links

Salaries and Players Data Site
https://www.r-bloggers.com/2020/08/nba-salaries/

Data.World Workspace
https://data.world/rainer27/twitter-nba-analysis/workspace/file?agentid=datadavis&datasetid=nba-salaries&filename=players.csv

balldontlie API
https://www.balldontlie.io/?shell#introduction

Twitter API
https://developer.twitter.com/en/docs/twitter-api

StackOverflow Bearer Token Authorization
https://stackoverflow.com/questions/21651846/oauth-access-token-request-twitter-api-and-oauth-verifier-field/21653087#21653087