Griff Prime

330 Project Proposal

For my project, I plan on looking at two different NBA datasets. One of them is going to list what each NBA player's salary is, and the other gives their statistics from the 2017-2018 season. I was motivated to look at data from the NBA because I am a sports fan and wanted to do a project related to one of the major team sports. I was able to find free data available from the NBA most easily and therefore choose it.

I am looking to answer questions based on how different salaries impact how well NBA players perform during a season (2017-2018). The specific questions I plan on answering right now are:

- *Do NBA players that receive higher salaries have better statistics during a season?*
- *How does a players age impact how much they make, on average?*
- *Does a teams success during a season correlate to the money that the team spends on its players?*

I plan on accessing one dataset from a downloaded CSV file, and another from a website using HTML to create a table. The descriptions of each dataset are below.

Dataset 1:

- *Social Power NBA*
- *1.4 MB*
- *https://www.kaggle.com/noahgift/social-power-nba*
- *CSV*
- *I downloaded the CSV file and plan on using it*

Dataset 2:

- *2017-2018 NBA Player Stats*
- *209 Rows*
- *https://www.nbastuffer.com/2017-2018-nba-player-stats/*
- *HTML format*
- *Using website URL and HTML*

To manipulate the data to get the answers that I need, I'm first going to need to get both of the datasets into Pandas. After, I am going to need to use a join or merge to combine the two tables on the player's names and then I will be able to answer my research questions. At this point, I will have the player's statistics combined with their salary and other information from dataset 1 and will be able to make findings and visualizations from them.

I could create a scatter plot of player success based on their statistics relative to how much money they make to see if there is a correlation. I plan on creating other visualizations depending on which finding I determine to be statistically significant.