Application of Big Data and Data Analytics:

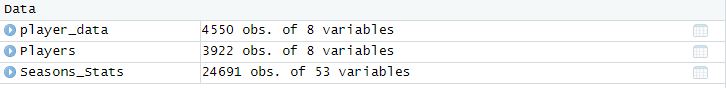
Impact of Data and Statistics on Fantasy Basketball

Progress Report: 02/08/2020

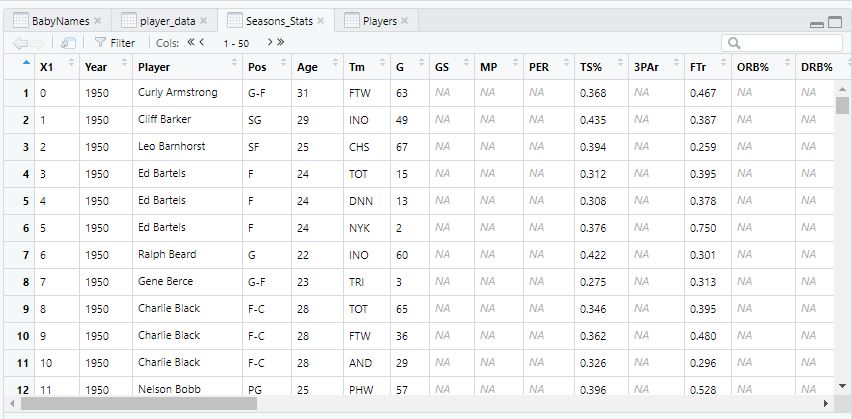
Jacob Mitchell

Objective: The goal of this research project is to gain a deeper understanding of Big Data, Data Analytics and Statistics to analyze the impact that they can have on making educated predictions or decisions. To do this, I am collecting flat data files from sources on the web and using R and Python to gather information about the given data sets.

To see how this information can positively influence decision making, I am going to use data and find trends and see how certain teams and players generally do in certain settings to predict fantasy basketball scores for teams, players, and predict overall games. To compare the results of this, I am including other members to partake in this experiment as well. The main variable of this experiment will be the knowledge of the subject matter which will be split into four main categories: not knowledgeable in basketball or statistics, knowledgeable about basketball but doesn’t read the data, not knowledgeable about basketball but reads the data, and knowledgeable about basketball and also reads the data. Seeing how other players score based on their knowledge of the sport and their awareness of the data trends will allow us to see how useful the data is to make educated decisions and predictions.

The Process: Firstly, I’ve collected a few different data sets to start learning the software I am using. Pictured below are the data sets I’ve gathered from Kaggle.com, a website that has a large community that posts datasets and is provide, compare, analyze, and share their findings on a vast variety of datasets. 

These data files contain stats on all players from the years 1950-2017, from field goal percentage (FG%), to position played (Pos). I have yet to plot anything using these data sets yet, but the data contained can be seen below.



Next Sprint: I have found another useful external resource to dive further down the analyzing of this and other similar datasets. A blog called “Hardwood Convergence” by Dan Watson, a health care data analyzer who does basketball analytics, introduces these concepts to users who haven’t had the same experience with data. This tutorial goes into scraping data from the web to using jupyter, python, and pandas to work with the data sets we scrape and tidy. I intend to use this tutorial to further my abilities in the subject matter and use these tools to start making predictions with. I will also set up a fantasy league where the draft for fantasy basketball league is done by people who fall into each of the previously defined cases.

The research: To supplement my individual research and experiment, I am also intending on doing some research on some academic journals to see the impact of data analytics in other settings and another to introduce how data analytics are used in professional sports settings. I have not narrowed down what articles I intend on using quite yet so they are not included here. However, I should at least know what I will be referencing by the next sprint as this sprint was mostly to collect datasets and get comfortable in an analytics environment.

I am at a comfortable pace to complete this project in a timely manner.