Application of Big Data and Data Analytics:

Impact of Data and Statistics on Fantasy Basketball

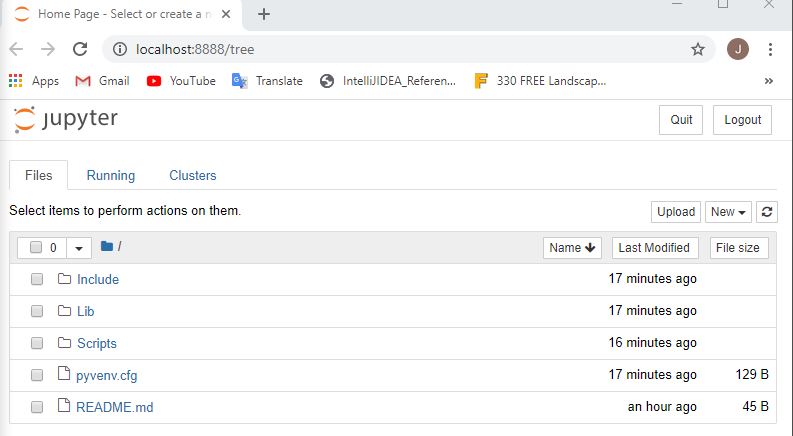
Progress Report: 03/01/2020

Jacob Mitchell

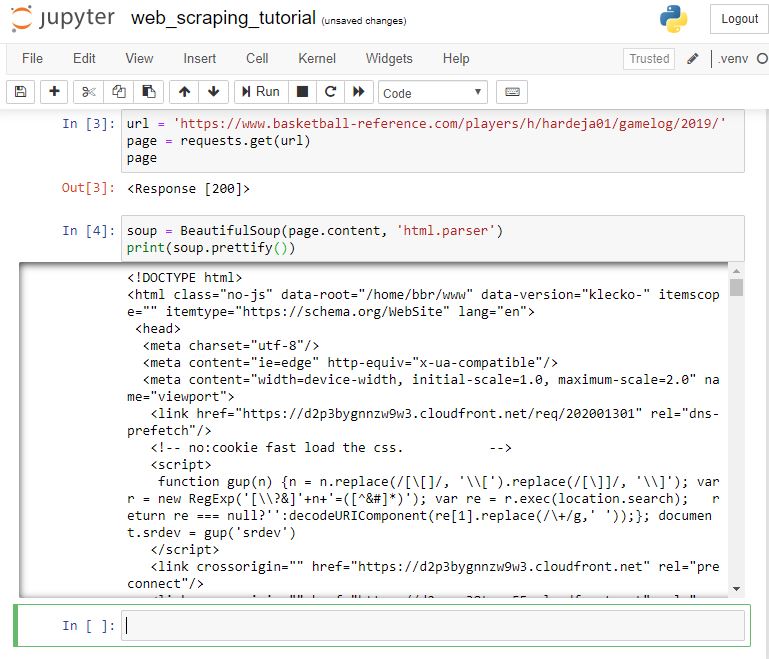
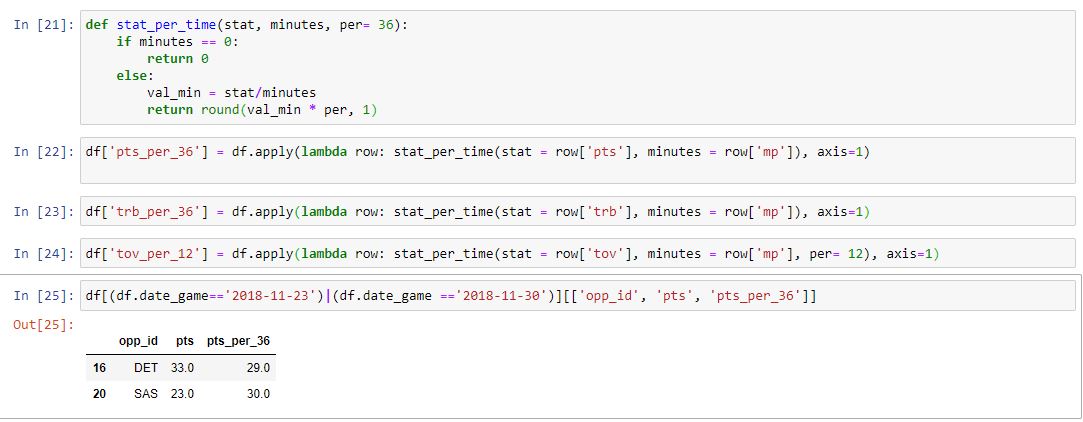
Overview: Last sprint, I had practiced using different tools in R studio to create a variety of different plots of random datasets to familiarize myself with these tools. This week, I’ve began working in the setting of data mining and tidying using real basketball statistics.

To this point, I’ve had no formal training or classes on web scraping or data science. As a result, I took to the Internet to find some tools to help gain a better understanding of the subject matter and the tools used in this setting. I found an introductory course on data analytics using python, jupyter, beautifulSoup, and Pandas in a basketball setting by Dan Watson, a medical data analyst who in his own time writes tutorials on data analytics for basketball.

Although not finished with the tutorial, I have logged roughly 8 hours of toying around with these newly introduced tools. Throughout the week, I have installed python on my machine, installed tools through python to allow me to analyze data, started a jupyter notebook to keep and organize my data, scraped data from basketball-reference.com, tidied up the data sets, and performed basic calculations on them. Screenshots of the work have been featured below.



My Jupyter notebook where my data and work are being stored.

My scrape of html from basketball-reference.com of James Harden’s stats sheets

Collecting stats based on different measurements.

I am by no means fluent in these tools yet, but this sprint has been a huge foundation on where this project will be going. Next sprint, I will be focusing on collecting more in depths stats and plotting data to find trends. I am on track to finish this project on time.