We are presented with a project to find a dataset and a website to answer a question. We want to look at ways to maximize money budget or ways to save, etc. Spencer and I want to look at a basketball data set because we play basketball. So, we are going to start looking at a dataset.

Nov. 13

We started looking at our dataset in order to find our question. We are using a basketball “Player of The Week” for the NBA in the last 30 or so years. We want to look at the more recent POTW winners. We want to know if the salary they are getting paid is justified based on our own observation. So to start, I looked at the the players with the most POTW awards and if you compare that to their salary, I’d say it will be a good question. Now our next step is finding the appropriate sites we need in order to find salaries.

Nov. 15

Started screen scraping on ESPN in order to find salaries. The question we are looking at is whether the salary that top players get paid are worthy given their accolades. We are using 2017-18 season to look at salaries and a dataset for player of the week. I don’t think we will need to go much farther than comparing salaries with accolades.

Nov. 21

We got our data scraped, will just have to clean and enter it into a dataframe. We aren’t as adept with Python so it is taking us more time for this process. We will have to start speeding up and get the website up soon. What we will have to do next is clean the data, combine it with our current dataframe. Then visualize our answer to find out what teams are getting the most out of their highest paid players.

Nov. 25

Just thinking about our question and giving background. With limited cap space in the NBA it’s important to save money and maximize their budget. As a fan, we want to know if the top players are worth these massive contracts or if they deserve more. Just from watching, we can say players like LeBron James and Stephen Curry deserve big contracts. We will see what the visualizations show.

Nov. 26

Gathered the appropriate data, but had trouble cleaning it. I am not sure what to do currently, but I plan on going in to Kent’s office and asking for some advice.

Dec. 1

We got our data frames to combine by adding the column names names to the original and then adding the data to it. There was some null values which shouldn’t happen, but I believe it’s due to repeats within the data. I think we will leave it for now because we can use the data given.

Dec. 3

I met with Kent today.

He told me to get more specific in grabbing the html tags to get what I want. Once I did that, it made my cleaning much easier.

I then could put it into a dataframe and then add the cleaning I need. Once I did that, I combined them by adding the columns to my first data frame, then adding the data.

Dec. 4th

We have begun our visuals and so far so good. Our data isn’t very complicated and deep se we are doing a lot of bar graphs, a couple of scatter plots for correlation, and a distribution plot. We have looked at overall award winners in the past two seasons. There were two clear winners in James Harden and DeMar DeRozan. This was a surprise but it was interesting.

The next step will try to find correlations with salary and awards. I will let you know how it goes.

Dec. 5th

We began correlations and found that there were some with salary and award wins. We found that the 5 highest paid players are also the 5 highest award winners. We looked at which positions win the award the most and found that SF and PGs win the most, and they are also paid the most. This was a very cool correlation to see. Heights and salaries don’t correlate and that’s okay, just wanted to see what it would look like, maybe a cluster and there isn’t. Surprisingly, shorter players win the award more, but is a little skewed given Stephen Curry and Damian Lillard knock it down because they are under 6’4.

We built the powerpoint and have to present tomorrow. A lot of groups have presented good content, but end up rambling. Ideally, Spencer and I can find a good mix between content and answering our question.

Dec. 6th

We gave our presentation and started our website. We struggled starting our website, but met with Kent in order to get where we need to go. We got that built and have been adding to it. So far it looks good, but can always get better.

We presented and it went well, we didn’t ramble. We could’ve added more content but now we know for next time to find a better mix. I am happy with the project. Given our experience with Python, it’s good that we figured out and found the answer to our question.

Dec. 7

I want to look further about those players who haven’t won and their salaries. Are they overpaid? Do overpaid players correlated to worse records? Are there more predictors for salary that we can quantify and find correlations with?

I thought moving this project forward allowed to me to think about more and more questions. So yes, we had a question to start, but by the end, we were on a completely different thought.