Stat 386 Documenting

Collecting the data: Looked up Quora and Reddit to find the best website for NBA statistics which lead me to the website that has the API of every NBA player and stat you could think of from each season. This was a breeze since all it took was a quick google search. I always knew I wanted to do a data analysis project on the NBA as well.

EDA: I wanted to create a function that generated a distribution and table of every numerical stat in the NBA so I can get an visual and numerical description of the NBA official leaders. It took me a really long time to create this function, so I was really happy when it finally worked. I then asked Samantha and Christian what I should put in my blog for the immense amount of data I’ve created, and they said that I should see which stats correlate with one another. I found a video that shows how to do a correlation matrix and table, which came in super handy. This helped me show which stats can be misleading and which stats can be connected to one another. At first, I wanted to compare every numerical column with each other to see visually if there was a linear relationship between the variables. However, after talking about it with Christian, we’ve concluded that I should actually compare the statistics of the top 10 NBA players. The reason why was to see how the game has changed, considering how the 3 point shot has been a common theme amongst rising NBA talent, and that more players have been scoring efficiently due to changes to the referee rule book. Me and Christian were also talking about the recent load management star players from high caliber teams partake of due to energy conservation and injury prevention for the playoffs, so we wanted to see if there was a difference in the number of games played by the top 10 players. I found a visualization online that showed ticket sales between the year 2001 and 2020, which was perfect for my blog (I did nothing to create the visualization, just thought it would be interesting to add there).

Stream lit Dashboard: This was probably my favorite part of the project, despite the amount of pain and suffering this has caused me. I pretty much googled and chat gpt’d everything, and had to be very specific with my searches. For example, I had to search up how to code for stream lit on a function that gives me a bunch of plots, but the more I coded, the more I got used to the stream lit syntax. My stream lit requirements text file would still not work, so I could not use seaborn or matplotlib. One thing I had to do was to ask chat to convert all of the seaborn and matplotlib code I had into stream lit and plotly.express syntax. I’ve come to learn to love stream lit as my team lead who is also in this class (Braden) wants to use it for our upcoming data project for BYU Continuing Education quality assurance. One thing that was a little easier to do however was when I was getting the data of players per decade, since I followed the syntax of the lab we did in class when we did streamlit practice.

Feedback: I received a lot of feedback from several classmates and friends. They all said that I could use some more pictures for my data collection since people can’t see my screen and won’t know what I’m talking about. Therefore, I took screenshots of what I did as I collected the request URL from the NBA statistics website. I also was told that I should add more functionalities to my dashboard to meet the full tier 1 requirements. I was able to do so fortunately and now have 3 fully functioning stream lit dashboard tools.