Final Writeup

This project was very insightful and challenging, and taught me a lot about the importance of data collection and preprocessing. While I was initially inclined towards researching something related to the NBA, I wasn't sure where I would be getting the data from. Through meticulous research, I decided to use a package developed by fellow NBA fans called NBA API. This package allowed me to scrape any data I choose from NBA.com, and had a very extensive documentation that allowed me to fully understand how to collect the data I wanted. Before deciding the question I wanted to ask, I wanted to initially explore the data. I found that scraping the data was very challenging because of constant request timeouts. I was not able to navigate this and ultimately struggled loading datasets efficiently. Another approach I wanted to make was downloading the data as soon as I got it, but quickly realized that the data would be too large to download, meaning I would have to scrape it every time. Regardless, after exploring the api and the different data frames available, I realized that I wanted to explore the 3 point shot, after reading this article. I realized that there could be unique ways to visualize this data and wanted to take advantage of those methods. Therefore, the question I decided to explore was, "How has the 3 point shot impacted the way basketball is played in the NBA?". Instead of focusing on specific players, I felt it would be more relevant to focus on the performance of the league as a whole, and show how the 3 point shot has changed how basketball is played. I brainstormed how to best highlight this, and determined that the 4 categories I wanted to demonstrate this through were: Types of shot taken over time, how the ratio of 3 pointers made correlates to the total points scored for every season, how defenders impact the 3 point shot, and how the 3 point shot affects the way a position is played. After scouring through the datasets, I found the right dataframes to make the necessary visualizations. It was relatively easy (aside from the request timeouts I constantly dealt with) to create the visuals. One thing that was frustrating was I couldn't gather data from every year automatically, but needed to change the year every time I made a request. I used visuals and articles I found online to see the most popular kind of visuals used for the NBA. Filled in line charts, heatmaps, and stacked bar charts were all very common. I was able to make the visualizations relatively easily after preprocessing the data. The most challenging visual was the heatmap, because it required preprocessing to calculate the points per shot. I initially got the idea of points per shot from reading ESPN articles that commonly used it to show how effective a player is at defending the ball. I wanted to flip the script, and show how effective different shot distances are in relation to distance from the defender. I did essentially by dividing shots made with shots attempted for a given combination of shot distance and defender. Ultimately, it was a success. The other 2 visuals were relatively easy to implement, which I was happy about. Overall, I learned a lot about the process of creating an argument and supporting it with appropriate data and visuals. It is a lot more difficult than I thought it would be, and I found that oftentimes, it's very important to remember the question being asked. To elaborate, it's more important to focus on simplistic visuals that can be effectively interpreted than trying to get too specific and technical with data manipulation and visualization. Initially, I used a Tableau visual similar to the NBA shot chart, however the methods I used to create visuals involved a lot of specific inputs and calculations that made it too difficult to explain and consequently difficult to correlate to the main idea of the project. Through this, I learned it's more important to make a few solid, simple arguments rather

than make a single and very complex argument. In conclusion, I had a great experience working on this project and hope that it will improve my skills as a data scientist.