Kamal Kakish

CPS 3320 Python Programming

Project 1

Background:

In the world of professional supports, statistics are the main driver in decision making and storytelling. Sometimes we may even notice things with our eyes that we want to verify by analyzing historical data. I consider myself a big fan of sports and in this case, we are going to dive into some questions I have pertaining to basketball.

Hypothesis:

The same way that all living lifeforms are subjected to evolution, sports are too. Basketball players have become faster, stronger, and more skilled during the years. While players have evolved to become more competitive, so have analytics. Analytics show that it is more efficient to shoot 3-point shots more frequently than it is to take mid-range shots. This theory was first put into action by the Houston Rockets and Golden State Warriors from 2015-2019. The emergence and success of great shooting inspirations such as James Harden, Klay Thompson and Steph Curry leads me to believe that drafted players will shoot more threes at a higher percentage as great 3-point shooting is becoming more sought after by NBA front offices. Additionally, I also hypothesize that with the increase of 3-point field goal percentage, there will be a decrease in overall field goal percentage. This is because players that shoot threes often might not be shooting regular mid-range shots and layups at a high frequency or percentage. The closer you are to the rim dictates the higher the field goal percentage (historically). Therefore, I

believe that players are shooting 3-point shots more frequently and this is lowering their overall field goal percentage while increasing their 3-point percentage.

Data:

The data that I needed to answer my questions is historical data of NBA players that were drafted. Kaggle had this data for me at: https://www.kaggle.com/benwieland/nba-draft-data.This data was curated at from https://www.sports-reference.com. This data corresponds to the pick with which a player was selected in the NBA draft. The data includes career statistics such as minutes, points, rebounds, assists and averages associated with each (per game). It also includes statistics related to shooting percentage, win share percentage, and BPM (box plus-minus). There is also categorical data such as draft position, draft year, and player url. For a full reference please see *draft-data-20-years.csv.* I made sure there was no possible biases in place that may influence my results. These biases included players that were drafted that never played. These values were defined as null.

Analysis/Methodology:

First thing I had to do was import this data.csv file into our data frame. For this I used *pandas*. I began exploring my data once it was loaded to the data frame. I started to think of other sub-problems that I might want solved while investigating the data, such as discovering leaders in statistical categories. To solve my initial problem, I first found the mean field goal percentage (FG%) of each group by draft year. Upon receiving the results, I also wanted to validate a sample of my results. I exported the draft class of 1990 to another csv and validated the values I received from my python program by comparing them to results I got from my excel formula. Results were valid. I proceeded to do the same for 3-point percentage(3PT%) and validated my results for that category as well. Once my statistics were validated, I used

matplotlib.pyplot to plot my results. This provided a visual representation of the statical trends of both. I used a line plot to illustrate the trends:
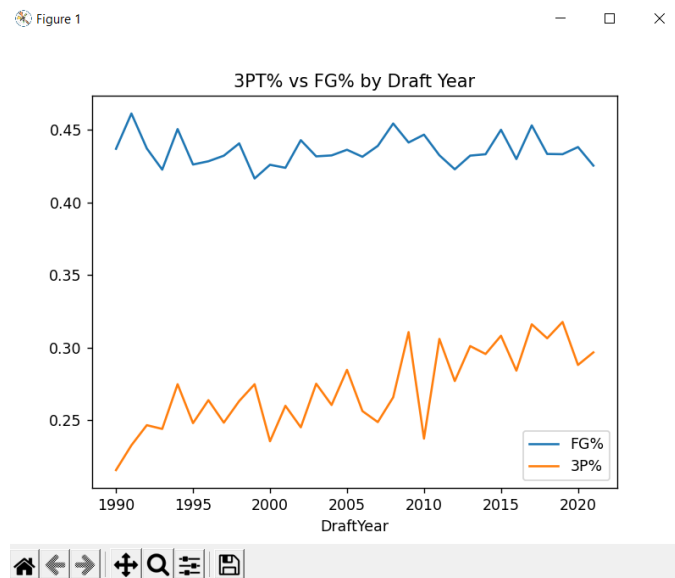

Figure: 1

In **Figure 1** you can clearly see that the 3-point percentage of players has increased over the timeline of our data. The overall field goal percentage of players seems to decrease during this time frame as well, showing opposite behavior to that of 3-point percentage (although less significant).

Conclusion:

Upon analyzing my results, I find my hypothesis to be validated. The hypothesis stated that players are shooting threes more frequently and this is lowering their overall field goal percentage while increasing their 3-point percentage. The results show that the average 3-point percentage of players has increased exponentially during the timeline observed while the overall field goal percentage has decreased. Something I also noticed was that there was a sharp increase in 3-point percentage and a sharp decrease in overall field goal percentage since 2015 which was

the emergence of the Golden State Warriors dynasty and rise of the fiercely competitive Houston Rockets. These team followed a high volume 3-point shooting model. Because of their success I believe it caused a trend across the league. I can confidently say these were amongst key contributors to the trends I observed in my data.

Project reflection:

This project particularly piqued my interest due to my passion for sports and statistics. It seems like there is a positive correlation between the advancement of sports and statistics and we can thank the advancement in modern day computing for this relationship. I wanted to solve this problem because I observed this change by watching sports my whole life and I wanted to see if the statistics backed my observations. During the development process I needed to research ways to load my data, parse my data, perform mathematical operations, and visualize my data in a meaningful way. Pythons' libraries were very useful to me as that is how I accomplished my data handling and visualization goals. I did not have to pivot from the goal of my project at all, in fact, I thought about sub problems and solved those problems as well! I wanted to find the leader in each main statistical category, and I did that during the additional time I had. My main take aways from this experience is that my interest in data science has grown even more. As long as I keep having questions, I will always look for answers.