Final Project Report

For my final project, I scraped NBA data. I noticed some upward trends in the points scored per game per season. To determine why this was happening, I wanted to understand the effects of rule changes in the NBA on the number of points scored per game. I was familiar with the website Basketball Reference and decided to scrape data from this website using the Python package called Beautiful Soup. I focused on critical metrics like points per game (PPG), effective field goal percentage (Effective FG%), 3-point percentage (3 point percentage), and 3-pointers attempted (3 pointers attempted). I then used regular expressions to extract relevant data and placed it into a data frame. I also scaled the dataset to be able to look at it altogether.

Using matplotlib.pyplot to create a bar graph showing the average PPG, computed using numpy calculated over a for loop, for each of the selected NBA periods (1946-2023). Notable rule changes, such as introducing the 24-second shot clock and lane modifications, are highlighted on the graph. I then generated individual scatter plots, using plotly.express, for PPG, Effective FG%, 3 point percentage, 3 pointers attempted, and Field Goal Attempts (FGA) per season. Each plot incorporates trend lines and vertical lines, using plotly.graph\_objects, to outline significant rule changes. Additionally, I scaled variables (PPG, Effective FG%, 3 point percentage, 3 pointers attempted), using Pandas pd.to\_numeric to convert text to numbers, in a combined scatter plot to quickly compare the trends.

I looked further at the impact of 3-pointers on the points per game score. I noticed a drastic increase in the number of 3-pointers attempted starting in 2012, so I calculated correlation coefficients, using Pandas.corr function, for 3-point attempts, 3-point percentage, and effective FG% regarding PPG. The heatmap, generated using seaborn and matplotlib.pyplot, visually represents these correlations, with redder positive values indicating positive correlations. The analysis showed a positive correlation between all variables, but the strongest correlations were between the PPG, 3 pointers attempted, and Effective FG%. The NBA has placed a stronger emphasis on the 3-point shot than in the past which resulted in increased scoring. Further statistical analysis, such as regression modeling, can offer a more in-depth understanding of these relationships and potential predictors of PPG in the NBA.