Chris Landaverde

CPS 3320: Python Programming

Project #1

**Project Description**

As an avid fan of the NBA, I’m constantly watching all teams face off against each other in an 82 game season plus the post-season or better known as the playoffs. With the game evolving and NBA players developing better skills both offensively and defensively, the question that will always arise is whether or not the NBA is more difficult now to play in than it ever was before. Since the offensive revolution beginning in the mid 2010’s with an unbelievable display of talent by the likes of Stephen Curry, LeBron James, Kevin Durant, etc., the NBA has seen a strong increase in scoring from individual players and teams. In this project we will be conducting an analysis of league averages from every NBA team and comparing it to a past season to see if there are any major differences.

**Hypothesis**

The hypothesis made and tested in this project is that NBA teams are averaging more total team points per game throughout the current 2022-2023 NBA season rather than the NBA’s previous 2021-2022 season.

**Data**

In this project, I utilized two datasets. One data set was of NBA league averages of the current 2022-2023 season and the second dataset was of NBA league averages of the 2021-2022 season. These datasets include statistical information such as games played, field goal percentages, field goal attempts, turnovers, offensive & defensive rebounding, and minutes played. I sourced my data from a highly credible NBA database. [www.basketball-reference.com](http://www.basketball-reference.com) is the #1 NBA statistical information distributor, referencing individual player, team, season, and historical statistics. Their datasets are constantly updated with every following NBA game that takes place. Some possible biases in this data are within the current 2022-2023 season, simply because the season has not been completed yet and there are still many games to play, however, the season is 3/4 way complete.

**Analysis/Methodology**

The first step of this project was to import the Pandas library and matplotlib library into my localhost Jupyter notebook. Within the same folder where my notebook is located are also my two data sets which will be loaded and read into our notebook using the pandas function df = pd.read(“TeamStats2023”) to create a data frame out of the excel dataset, the same had to be done for the 2021-2022 season stats by creating a data frame out of that excel dataset using df2 = pd.read(“TeamStats2022”) where I differentiated each data frame by labeling the other “2”. The main goal here is to analyze each team’s point averages per game. To do this, I had to create a view, naming it “points1” for the 2022-2023 NBA stats and a view called “points2” for the 2021-2022 NBA season, selecting two specific columns from each data set. I retrieved the column that displays each team and retrieved the column that displays each team points average for both data sets. For example, “points1 = df[[“Team”, “2023PTS”]]”.

Results:

|  |  |
| --- | --- |
| Graphical user interface  Description automatically generated | Table  Description automatically generated with medium confidence |

After obtaining these results, what I wanted to compute was the average points scored of ALL NBA teams for both seasons. This way I can get a better understanding of what scoring in the NBA is really like. To compute the mean, I defined the variable with a function for each individual season. “ptsavg1” represents the average of points scored by ALL NBA teams for the 2022-2023 season and “ptsavg2” represents the average of points scored by ALL NBA teams in the 2021-2022 season. By using the function “df[“2023PTS”].mean()” and “df[“2022PTS”]”.mean()”, I was able to get the mean of all NBA team point averages. By using the print function to display the “ptsavg1” and “ptsavg2”, the point average for each respective season was printed.

Results:

Text

Description automatically generated

To display both season point averages, I created a third data frame labeled as “df3” and merged both season point averages. By using “df3 = pd.concat([df[["Team", "2023PTS"]],df2["2022PTS"]], axis=1)”, I was able to successfully complete this task.

Results:

Table

Description automatically generated with medium confidence

**Conclusion/Results**

To better visualize the results to see a significant rise in the average points teams are scoring in the NBA per season, I created a bar graph with the help of our the matplotlib library. To do this, I took the existing data frame with both season average (df3) and set the bar graph X axis to represent the Teams and I set the Y axis to represent points averaged for both individual seasons, displaying two bars for each NBA team, each bar representing their respective NBA season.

Results:

Chart, bar chart

Description automatically generated

Now that we have this visual representation of the average points scored by each individual NBA team per the 2022-2023 NBA season and for the 2021-2022 NBA season, it is clear that there has been a rise in the number of points scored on average by each team in the current 2022-2023 NBA season than points scored on average by each team in the 2021-2022 NBA season. Thus, supporting my initial hypothesis. An even better and deeper dive into this analysis could include analyzing data sets of NBA team season averages of the past decade to further support my hypothesis, by using a larger sample size of data, I could display and show the continuous rise of team scoring to an even further extent.