Nba Analysis 2023 Full Description

Overview

I have always been a fan of NBA basketball and am seeking out to answer a few questions about the 2023 season. This short analysis explores some trends about scoring in the nba. I take a look at each team's top scorer, their contribution to the team's overall scoring, and the relationship between 3pt shots and scoring. I also see if there are correlations between these trends and a teams win percentage.

Datasets used

To begin I am using the dataset 2023_nba_player_stats dataset from Kaggle, created by user: AMIRHOSSEIN MIRZAEI.
Here is a link to the dataset:
https://www.kaggle.com/datasets/amirhosseinmirzaie/nba-players-stats2023-season

For the team stats I am using the dataset NBA Team Stats from gaggle user:
MICHAEL H .
Here is a link to the dataset:
https://www.kaggle.com/datasets/mharvnek/nba-team-stats-00-to-18

Scope of Project

The scope of this project will include an analysis of the following for the 2023 NBA season:

- The top scoring player from each team and their impact on the team's whole scoring.
    - The relationship of the above to team wins
- The relationship between 3 point shots and team overall scoring.
    - The relationship of the above to team wins

The scope will not include the following:

- Any analysis of historical data outside of the 2023 NBA season.
- Any Analysis beyond the scope of what is listed above.
    - More analysis of the 2023 NBA season is to be revisited in a separate project.

Tools used

- For Data cleaning I am using Google Sheets
- For Analysis I am using BigQuery for SQL
- For Visualization I am using Tableau

Data Cleaning

- 1. Made copies of the two datasets to preserve the raw files.
- 2. Removed Duplicates with Google Sheets functionality
- 3. Trimmed Whitespace
- 4. Correct formatting such as changing columns to percentages to keep data consistent
- 5. Created a column in the teams dataset called team_abbr. This is in order to have the team values in each dataset match. For this I used a vlookup in the teams sheet. I used a player from each team to map to the team and pulled in the team abbreviation from the player sheet. I then copied only the values into a new column, and deleted the vlookup column.
    - This task could've been done manually but I felt using the vlookup would give me good practice

Analysis

- For the SQL queries please refer to the SQL queries document where each query is written fully.
- I first upload the two cleaned csv files into BigQuery where I will be writing my SQL queries.
- I preview each dataset and then begin to write my first query.
- I will first explore the first question in the scope:
- *The top scoring player from each team and their impact on the team's whole scoring.*
    - Players points per game (PPG):

Row	PName	POS	Team	PPG
1	Joel Embiid	C	PHI	33.07575757575...
2	Luka Doncic	PG	DAL	32.39393939393...
3	Damian Lillard	PG	POR	32.17241379310...
4	Shai Gilgeous-Alexander	PG	OKC	31.39705882352...
5	Giannis Antetokounmpo	PF	MIL	31.09523809523...
6	Jayson Tatum	SF	BOS	30.06756756756...
7	Stephen Curry	PG	GSW	29.42857142857...
8	Kevin Durant	PF	PHX	29.06382978723...
9	LeBron James	SF	LAL	28.90909090909...

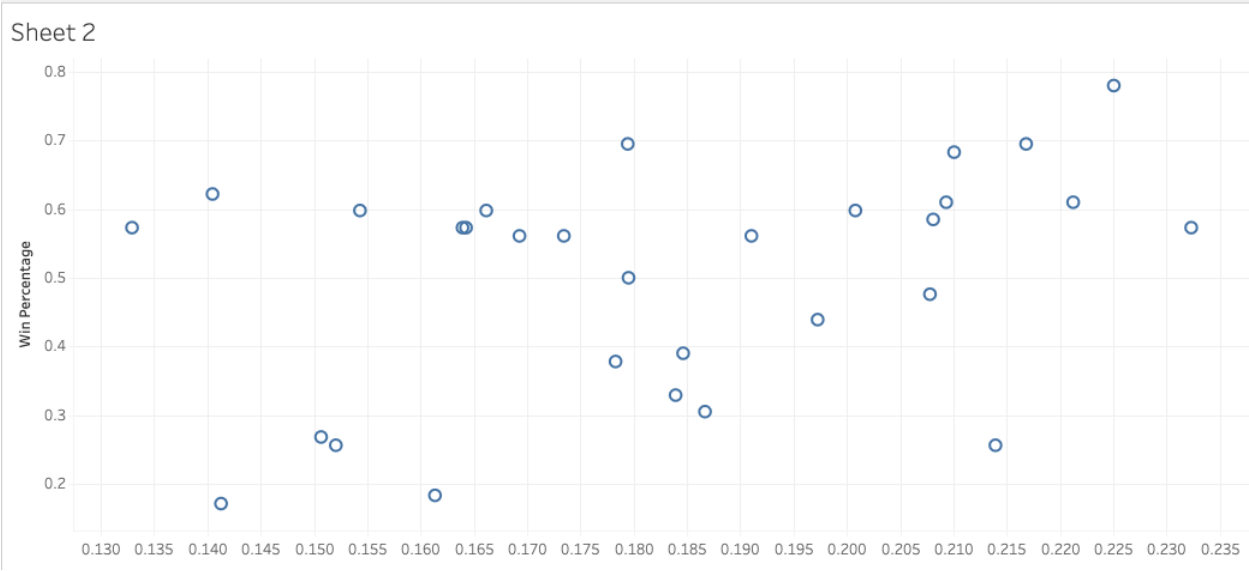
- Each player's contribution to the total team's scoring (last column)

Row	PName	PTS	team	points	player_point_contr_p
1	Joel Embiid	2183	PHI	9397	0.232308183462...
2	Jayson Tatum	2225	BOS	9887	0.225042985738...
3	Luka Doncic	2138	DAL	9664	0.221233443708...
4	Shai Gilgeous-Alexander	2135	OKC	9847	0.216817304762...
5	Damian Lillard	1866	POR	8722	0.213941756477...
6	Anthony Edwards	1946	MIN	9264	0.210060449050...
7	Julius Randle	1936	NYK	9249	0.209319926478...
8	Donovan Mitchell	1922	CLE	9236	0.208098744045...
9	Zach LaVine	1913	CHI	9206	0.207799261351...
10	Giannis Antetokounmpo	1959	MIL	9756	0.200799507995...

- What is the correlation between a top players percentage of team scoring to win percentage?
  - Note: Since BigQuery does not allow temp tables without paying for a subscription, I downloaded the results of the last query and re uploaded them for a new dataset.
  - First we can look at each teams top scorer contribution and their win percentage

Row	Team	win_percentage	top_scorer_contribut
1	PHI	0.573	0.232308183462...
2	BOS	0.78	0.225042985738...
3	DAL	0.61	0.221233443708...
4	OKC	0.695	0.216817304762...
5	POR	0.256	0.213941756477...
6	MIN	0.682999999999...	0.210060449050...
7	NYK	0.61	0.209319926478...
8	CLE	0.585	0.208098744045...

The next step is to visualize this information to make it easier to interpret.

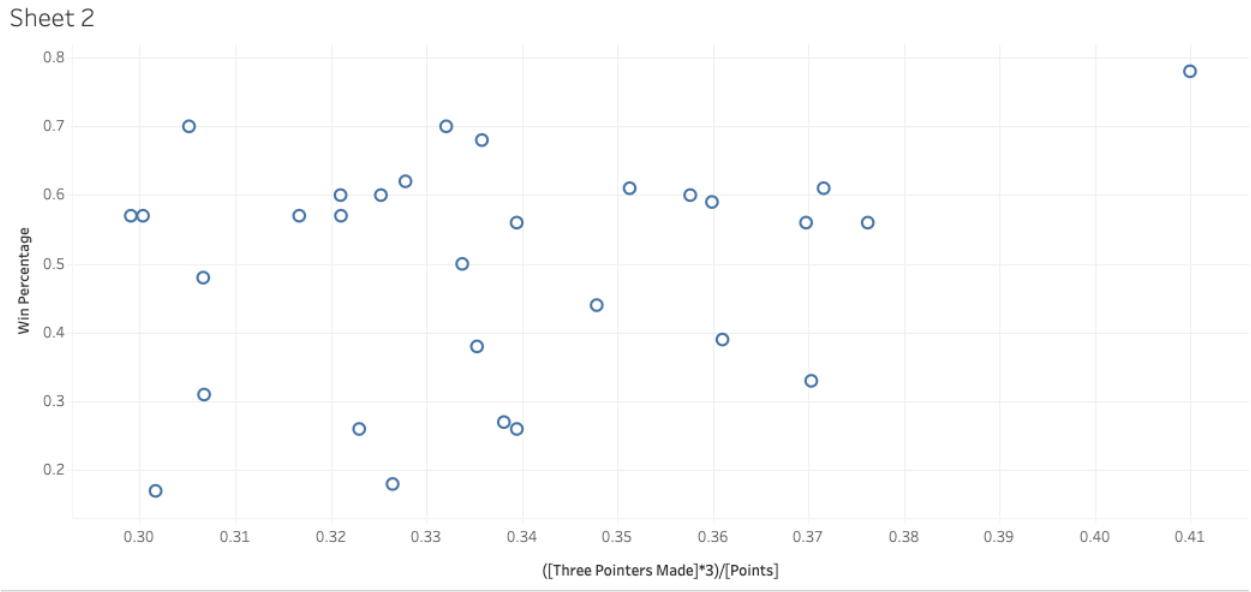


As we can see here there is not a strong correlation.

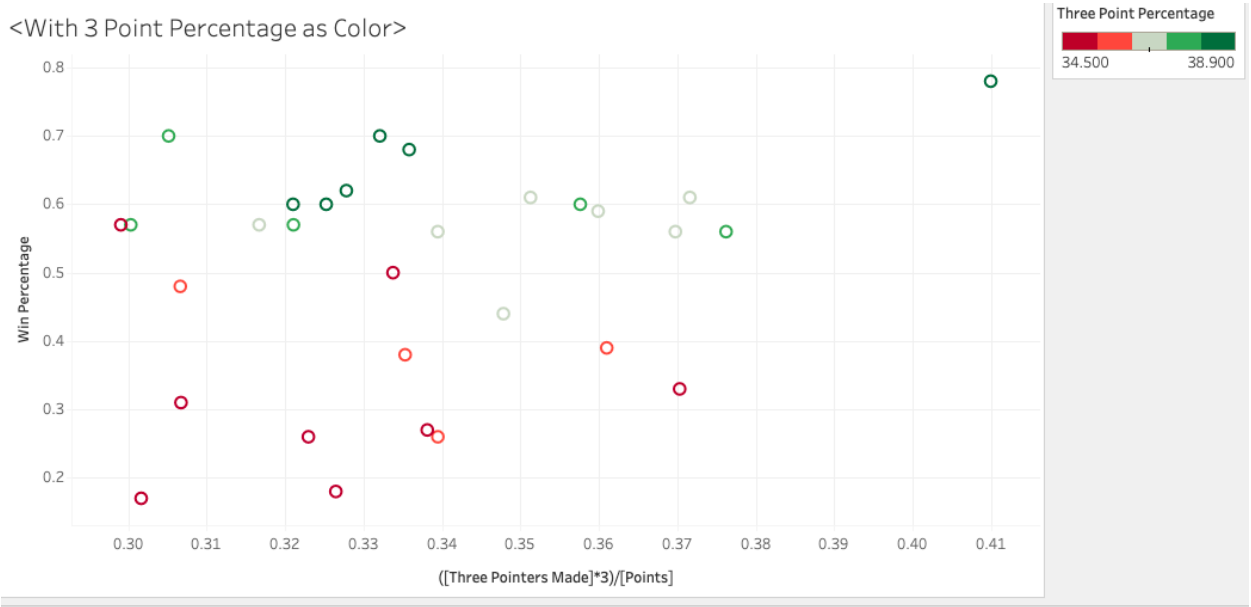
- Question 2 in the scope:
  - *The relationship between 3 point shots and team overall scoring.*
    - Percentage of scoring from 3 pointers (last Column)

Row	Team	points_from_3pt	points	percentage_of_total
1	Boston Celti	4053	9887	0.409932234246...
2	Golden State Warriors	3633	9657	0.376203789996...
3	Dallas Mavericks	3591	9664	0.371585264900...
4	Memphis Grizzlies	3213	8677	0.370289270485...
5	Sacramento Kings	3534	9558	0.369742623979...
6	Brooklyn Nets	3267	9050	0.360994475138...
7	Cleveland Cavaliers	3324	9236	0.359896058899...
8	Milwaukee Bucks	3489	9756	0.357626076260...
9	New York Knicks	3249	9249	0.351281219591...

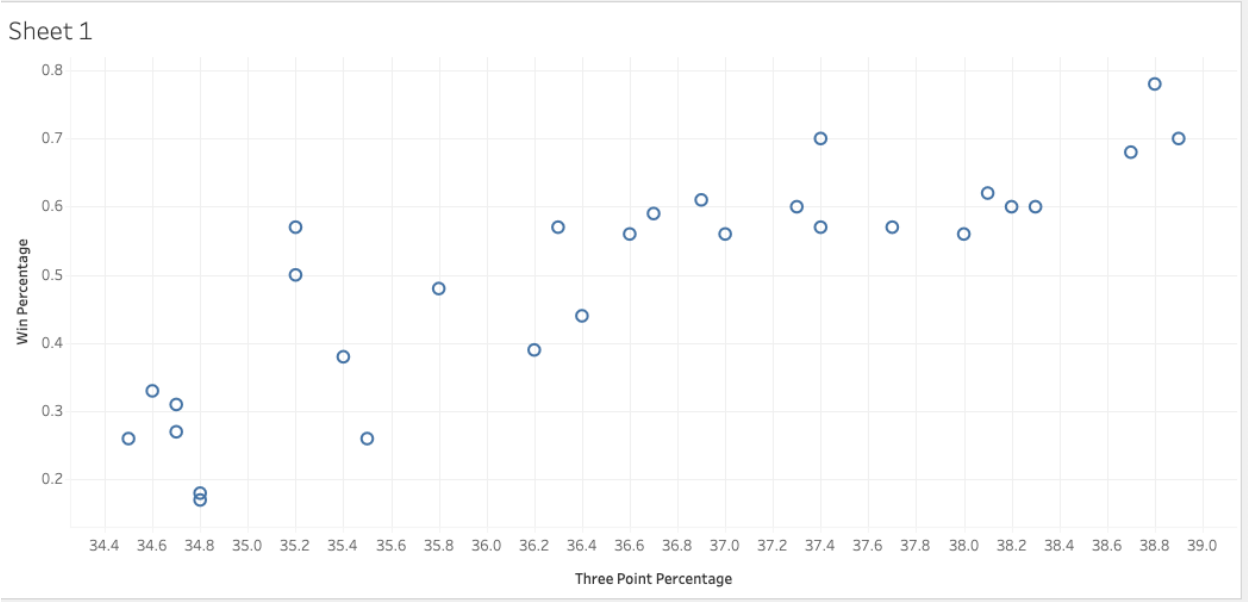
- Do teams with more points coming from 3 pointers have a better win percentage? The correlation is not clear from the table. Let's visualize this.



- There does not seem to be a strong correlation. Let's add in some coloring for 3 point percentage. Let's color the range of 3 point percentage from red (worst) to green(best). The scale is shown below.



- From This chart we see that just because a team has a lot of points from 3 point attempts, does not mean that they have a good percentage. Let's take a look at the correlation between 3 point percentage and win percentage:



- Here we see a much stronger correlation. We can see that teams with a higher 3 point percentage tend to have a better win percentage.

<u>Analysis Results</u>

- In this short analysis I set out to answer some questions about scoring and win percentage in the nba. From the analysis I have found that:
    - Teams whose top player contributes more to their overall scoring do not necessarily have a better win percentage.
        - Further question: What if we were to look at top 2 or 3 players, not just the top player.
    - Teams with more points coming from 3 pointers do not necessarily have better win percentage. However, there is strong evidence to suggest that teams with a higher 3 point percentage tend to have a better win percentage.
        - Further question: What other stats correlate with win percentage?

Final Thoughts

The NBA keeps track of numerous stats and has many professionals working on finding trends. My analysis only focused on a few of these aspects. With my analysis I do not intend to replicate any of the NBA analytics, nor either confirm or contradict any of their findings. However, through this exercise I have seen a glimpse of the world of analytics and sports. This topic is very interesting to me and I look forward to exploring it in more detail in future projects.