

NBA User Story 1 Report and Analysis

Introduction

This report analyzes data collected on NBA players' performance statistics collected between the years of 1950 and 2021.

Body

Data

This report was written using MyJupyterNotebooks. First I imported Pandas, Matplotlib, Seaborn, and Numpy to analyze and write reports on the data. Next I imported the NBA data from a .csv file. Then I requested the number of rows and columns (26176 and 20 respectively), the data types of each column (all were float except for the players' names and city(team)), and the column names.

Method

Since the client had requested an evaluation of the performance of the players, first I wanted to choose a metric that their performance could be measured by. Wikipedia's Basketball Statistics page (referenced June 7th 2024) provided the equation $(\text{Points} + \text{Rebounds} + \text{Assists} + \text{Steals} + \text{Blocks} + \text{Fouls Drawn}) - (\text{Missed Field Goals} + \text{Missed Free Throws} + \text{Turnovers} + \text{Shots Rejected} + \text{Fouls Committed})$ to calculate performance rating. I was missing the data points for "Shots Rejected", but I was able to make do with the others from the values included in my dataset. I replaced the null values with 0. Next, I calculated the value and added it to a new column which I named "Performance Index." Once I had calculated this value, I dropped all columns save Year, Player, Team, Minutes, Total games, and PI.

Next, I decided to analyze only the top 10 players (based on PI metric) to make the data easier to work with. I calculated the top 11 results and excluded Nikola Jokic, who had a very high PI rating but who had only so far played in the year 2021, so he was an outlier.

For each player in the top 10 list, I created a separate dataframe and then concatenated them together into a top 10 dataframe. I calculated the average PI score for each player, then added a column to the dataframe indicating whether that player had played above or below their personal average score that year (with a boolean 'Yes' or 'No').

Using this data, I mapped out each players' career on a line chart using their average as a benchmark.

In order to test whether the players' city had any correlation to their PI score, I also mapped out a line chart using the top 10 players' teams.

Results

When using the top 10 players' performance as a benchmark, I found that players on average hit their peak in the first 2 years of their careers. The first 10 years they typically play better than average, with each year past that falling lower on their performance index. I did not find significant evidence that the players' team had any correlation on their PI score.

Analysis

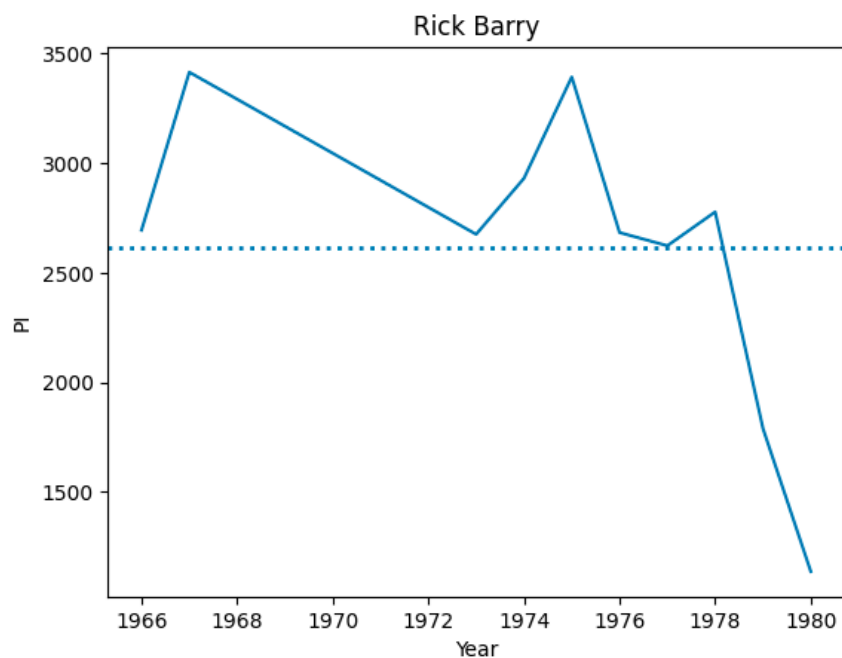
Finding the PI value:

```
#calculating our PI, Performance Index
df['PIA'] = df['PTS'] + df['REB'] + df['AST'] + df['STL'] + df['BLK']
df['PIB1'] = df['FTA'] - df['FTM']
df['PIB2'] = df['3PA'] - df['3PM']
df['PIB'] = df['PF'] + df['PIB1'] + df['PIB2'] + df['TOV']
df['PI'] = df['PIA'] - df['PIB']
```

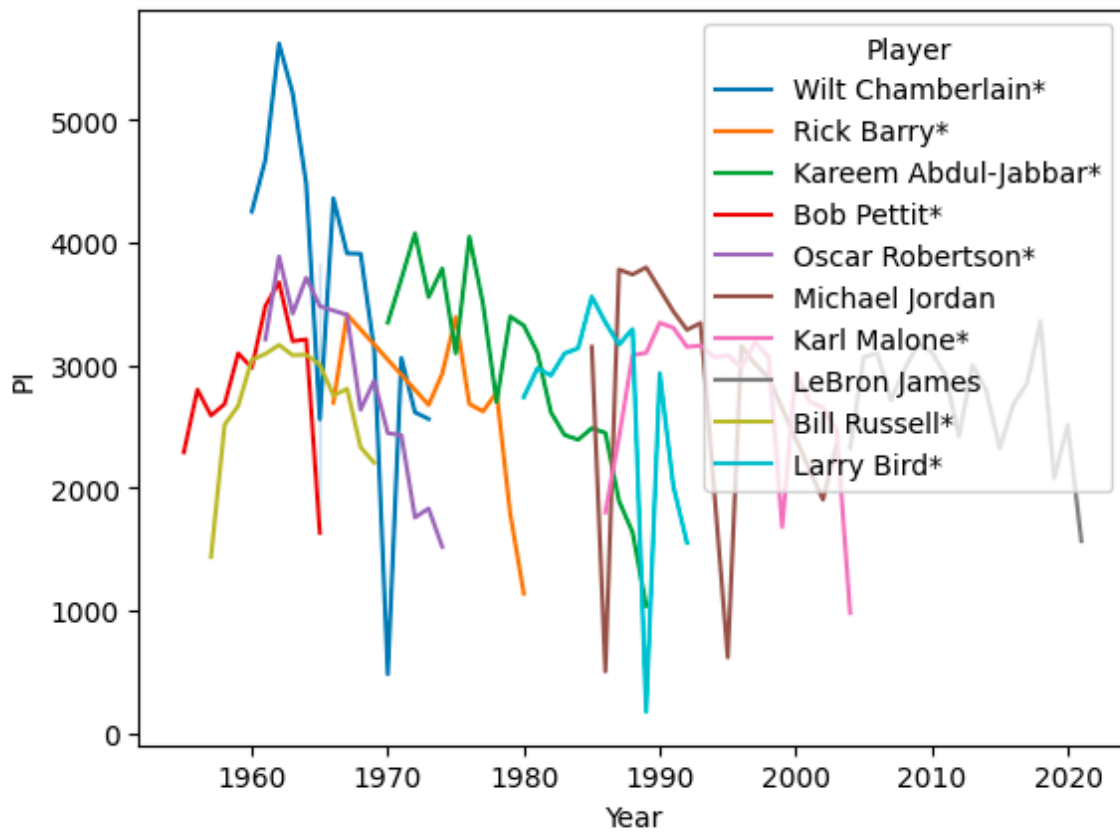
```
df.head()
```

	Year	Player	Team	GP	MIN	FGM	FGA	3PM	3PA	FTM	...	REB	AST	STL	BLK	PTS	PIA	PIB1	PIB2	PIB	PI
0	1950	Curly Armstrong	FTW	63.0	0.0	144.0	516.0	0.0	0.0	170.0	...	0.0	176.0	0.0	0.0	458.0	634.0	71.0	0.0	288.0	346.0
1	1950	Cliff Barker	INO	49.0	0.0	102.0	274.0	0.0	0.0	75.0	...	0.0	109.0	0.0	0.0	279.0	388.0	31.0	0.0	130.0	258.0
2	1950	Leo Barnhorst	CHS	67.0	0.0	174.0	499.0	0.0	0.0	90.0	...	0.0	140.0	0.0	0.0	438.0	578.0	39.0	0.0	231.0	347.0
3	1950	Ed Bartels	TOT	15.0	0.0	22.0	86.0	0.0	0.0	19.0	...	0.0	20.0	0.0	0.0	63.0	83.0	15.0	0.0	44.0	39.0
4	1950	Ed Bartels	DNN	13.0	0.0	21.0	82.0	0.0	0.0	17.0	...	0.0	20.0	0.0	0.0	59.0	79.0	14.0	0.0	41.0	38.0

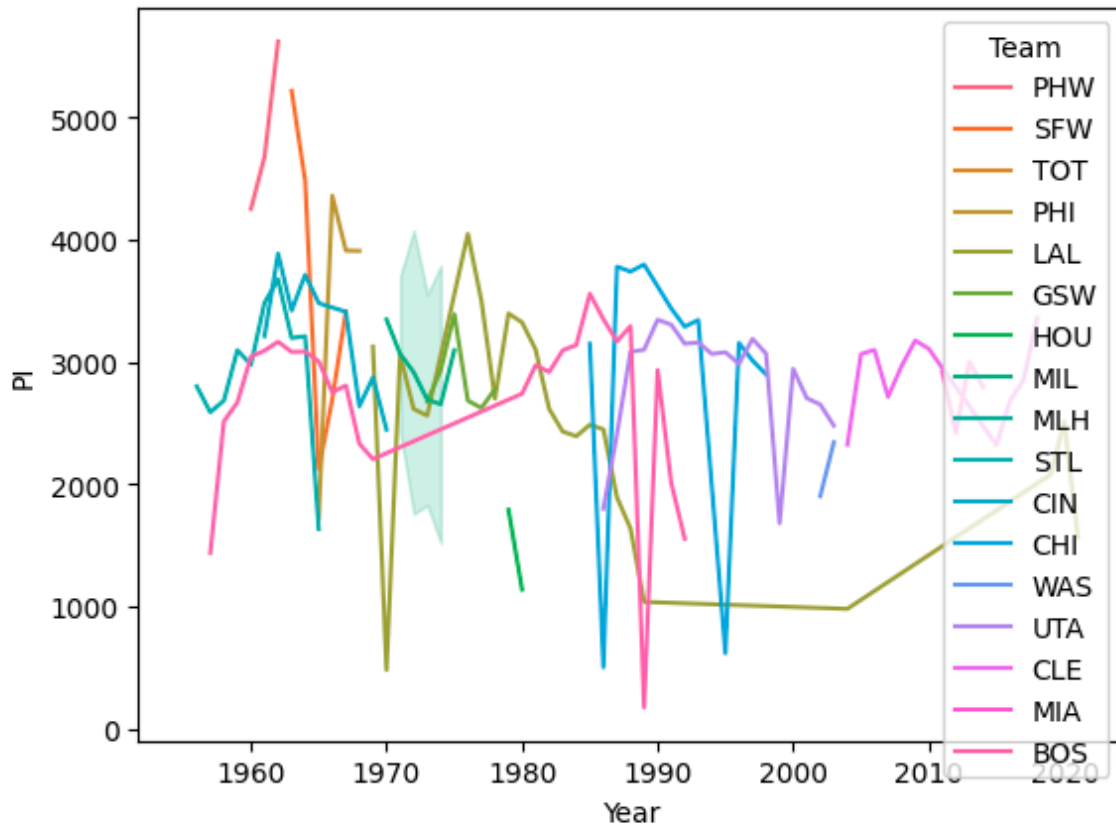
An example of a player's score:



The analysis of all top 10 players:



The analysis of each players' team:



Conclusion

Team has little correlation on a player's PI score (calculated by (Points + Rebounds + Assists + Steals + Blocks + Fouls Drawn) – (Missed Field Goals + Missed Free Throws + Turnovers + Fouls Committed)). In general, the best years of a players' career are the first 10, and especially the first 2. After 10 years, contract extensions are risky.