



## Exploring NBA stats patterns

### ***Members:***

Hedi Mohamad Mahmood

Rasty Xalil Muhamad

Dana Fariq mirza

Saman Sardar Ali

### **Supervised by:**

Dr. Miran Taha Abdulla

# Introduction

**1- The National Basketball Association (NBA)** is a premier professional basketball league in North America, showcasing top-tier talent and fierce competition among 30 teams. Established in 1946, the NBA has become a global phenomenon, known for its high-flying athleticism, iconic players, and intense playoffs. With a rich history and a massive fan base, the league continues to captivate audiences worldwide, blending sportsmanship and entertainment on the hardwood.

**2- Analyzing data from 2012 to 2023** provides insights into the dynamic evolution of NBA gameplay and its impact on the sporting landscape. Over this period, statistical trends, technological advancements, and strategic innovations have reshaped the game. Fueled by advanced metrics, player analytics are integral in understanding and maximizing performance. From three-point shooting efficiency to player-tracking data, teams leverage analytics for informed decisions on recruitment, strategy, and in-game tactics.

Moreover, social media and digital platforms have transformed the NBA into a global phenomenon. Player personalities and off-court narratives play a significant role in shaping the league's appeal, fostering a unique connection between players and fans worldwide.

In terms of gameplay, the emphasis on three-point shooting has seen a remarkable surge, altering offensive strategies and redefining player roles. Versatility has become a prized asset, with players expected to

contribute in multiple facets of the game. This evolution is evident in stars and the depth and skill exhibited by entire team rosters.

The league's commitment to inclusivity and social justice has been a notable shift. NBA players and teams actively engage in addressing social issues, using their platform to advocate for change. This intersection of sports and societal impact adds depth to the NBA's identity.

As we extrapolate from the data, it becomes evident that the NBA is not just a sports league; it's a dynamic entity continuously adapting to the changing tides of the sporting landscape and societal expectations. The years from 2012 to 2023 mark a transformative era, and the NBA's ability to innovate and resonate beyond the court solidifies its status as a cultural phenomenon.

**3- In conclusion,** the data-driven analysis spanning from 2012 to 2023 vividly illustrates the NBA's remarkable metamorphosis. From the strategic integration of player analytics shaping team dynamics to the global embrace of social media elevating the league's visibility, each passing year has brought forth a new chapter in the NBA's narrative.

The three-point revolution has not only changed how the game is played but has also redefined player roles, placing a premium on versatility and long-range proficiency. Beyond the court, the NBA's commitment to social justice initiatives highlights a paradigm shift, emphasizing the league's role in addressing broader societal concerns.

The NBA's ability to adapt to evolving trends while staying true to its core essence as a basketball league underscores its resilience and cultural significance. Reflecting on the transformative years examined in the data, it's evident that the NBA is more than a sports league; it's a dynamic force shaping the intersection of sports, technology, and social impact on a global scale. The journey from 2012 to 2023 represents statistical trends and a captivating narrative of innovation, inclusivity, and the enduring spirit of competition.

## Methodology

The methods we used for making this project are varied, we used PYTHON programming language to develop the steps and a lot of its libraries to make the right graphs, steps, and statistics and to have a precise view of what has happened during this decade.

The libraries and their usages:

### **pandas (pd):**

Description: Provides data structures like DataFrame for efficient data manipulation and analysis.

Common Usage: Loading, cleaning, transforming, and analyzing structured data.

**seaborn (sns):**

Description: A statistical data visualization library based on matplotlib, making complex plots with concise syntax.

Common Usage: Creating attractive and informative statistical graphics.

**plotly.graph\_objects (go):**

Description: Part of the Plotly library for creating interactive plots and visualizations.

Common Usage: Building interactive and customizable graphs for data exploration and presentation.

**matplotlib.pyplot (plt):**

Description: A 2D plotting library for creating static, animated, and interactive visualizations in Python.

Common Usage: Generating various types of plots and charts for data visualization.

**Requests():**

Description: Allows sending HTTP requests and handling responses in Python.

Common Usage: Retrieving data from web APIs or making HTTP requests for various purposes.

**numpy (np):**

Description: A library for numerical operations, providing support for large, multi-dimensional arrays and matrices.

Common Usage: Scientific computing, mathematical operations on arrays, and data manipulation.

These libraries are commonly used in data analysis, visualization, and manipulation tasks in Python.

## Results and explaining codes:

```
df_cols = ['Year', 'Season_type'] + headers
df = pd.DataFrame(columns=df_cols)

season_types = ['Regular%20Season', 'Playoffs']
years = ['2012-13', '2013-14', '2014-15', '2015-16', '2016-17', '2017-18', '2018-19', '2019-20', '2020-21', '2021-22', '2022-23']

start_time = time.time()

for y in years:
    for s in season_types:
        api_url = 'https://stats.nba.com/stats/leagueLeaders?LeagueID=00&PerMode=Totals&Scope=S&Season='+y+'&SeasonType='+s+'&'
        r = requests.get(url=api_url).json()
        temp_df1 = pd.DataFrame(r['resultSet']['rowSet'], columns=headers)
        temp_df2 = pd.DataFrame({'Year': [y for _ in range(len(temp_df1))],
                                'Season_type': [s for _ in range(len(temp_df1))]])
        temp_df3 = pd.concat([temp_df1, temp_df2], axis=1)
        df = pd.concat([df, temp_df3], axis=0)
        print(f'Finished scraping data for the {y} {s}.')
        lag = np.random.uniform(low=5, high=40)
        time.sleep(5)

end_time = time.time()
elapsed_time = end_time - start_time
print(f'Total time taken: {elapsed_time} seconds')
df.to_csv('nba_player_data.csv', index=False)
```

Figure 1

We got the whole data by using pandas library to making a connection between each year's data. and saved it on a csv file and let the dataset to be read and worked on it.

```
#to get best correlation value and name of columns
```

```
max_corr_value=newdata.corr()[newdata.corr() <1].max().max()  
name_of_columns=newdata.corr()[newdata.corr() <1].unstack().idxmax()  
print(max_corr_value)  
print(name_of_columns)
```

```
0.993862190850016  
('FGM', 'PTS')
```

```
plt.scatter(newdata[name_of_columns[0]],newdata[name_of_columns[1]])
```

```
<matplotlib.collections.PathCollection at 0x1f549b3e880>
```

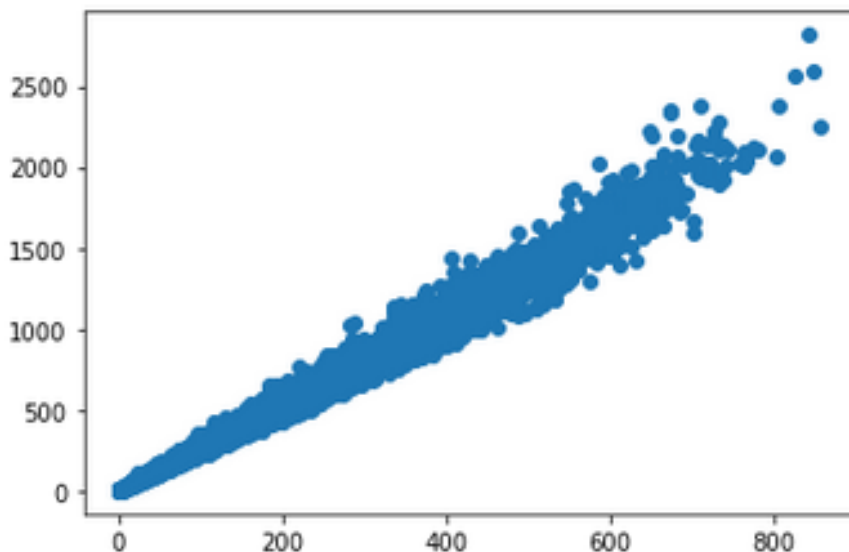


Figure 2,3

we wrote this code to find the best correlation between columns and showing the results in a scatter plot.



```
newdata.drop(columns=['RANK','EFF'], inplace=True)

newdata['season_start_year'] = newdata['Year'].str[:4].astype(int)

newdata['Season_type'].replace('Regular%20Season','RS', inplace=True)

rs_df = newdata[newdata['Season_type']=='RS']
playoffs_df = newdata[newdata['Season_type']=='Playoffs']

total_cols = ['MIN','FGM','FGA','FG3M','FG3A','FTM','FTA',
              'OREB','DREB','REB','AST','STL','BLK','TOV','PF','PTS']
```

Figure 4

We used these lines of code to cleaning those data and columns that we don't need and replacing them with the right values.

```

change_df = newdata.groupby('season_start_year')[total_cols].sum().reset_index()
change_df['POSS_est'] = change_df['FGA']-change_df['OREB']+change_df['TOV']+0.44*change_df['FTA']
change_df = change_df[list(change_df.columns[0:2])+['POSS_est']+list(change_df.columns[2:-1])]

change_df['FG%'] = change_df['FGM']/change_df['FGA']
change_df['3PT%'] = change_df['FG3M']/change_df['FG3A']
change_df['FT%'] = change_df['FTM']/change_df['FTA']
change_df['AST%'] = change_df['AST']/change_df['FGM']
change_df['FG3A%'] = change_df['FG3A']/change_df['FGA']
change_df['PTS/FGA'] = change_df['PTS']/change_df['FGA']
change_df['FG3M/FGM'] = change_df['FG3M']/change_df['FGM']
change_df['FTA/FGA'] = change_df['FTA']/change_df['FGA']
change_df['TRU%'] = 0.5*change_df['PTS']/(change_df['FGA']+0.475*change_df['FTA'])
change_df['AST_TOV'] = change_df['AST']/change_df['TOV']

change_df

```

Figure 5

We classifying the whole data by year. And we found new data by percentage for showing more detailed graph.

For example : we have pts on the figure above, which means points that have been scored on a year. And the figure below shows how this point scoring have changed during the last decade, **each line on the plot specifies a column in our dataset.**

```
# Assuming you have change_df available
change_per48_df = change_df.copy()

# Transform data
for col in change_per48_df.columns[2:18]:
    change_per48_df[col] = (change_per48_df[col] / change_per48_df['MIN']) * 48 * 5

change_per48_df.drop(columns='MIN', inplace=True)

fig = go.Figure()
for col in change_per48_df.columns[1:]:
    fig.add_trace(go.Scatter(x=change_per48_df['season_start_year'],
                             y=change_per48_df[col], name=col))

fig.show()
```

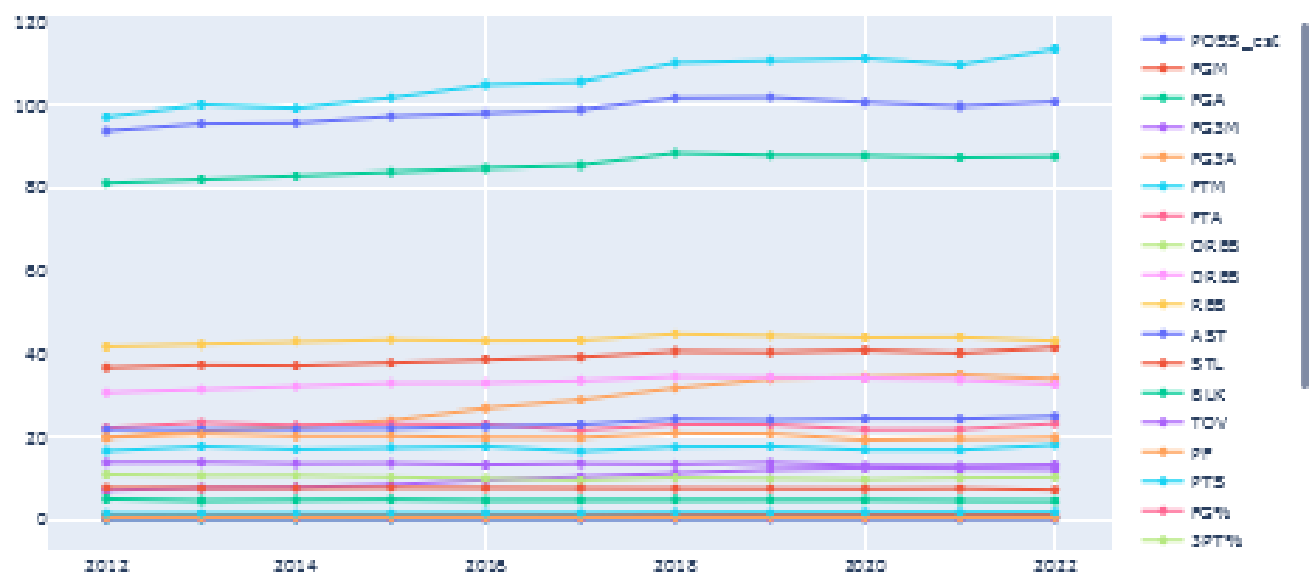


figure 6

in the figure above, we used **plotly library**, to showing the changes on each column's data during 10 years, and this library has a benefit, which is simplifying, and adjusting each column via a line and we can show which one should

be appeared if simply clicking on the button next to that specific line which each one of them has a different color.

For example, on the figure 7 which below that we only showed one column's value.

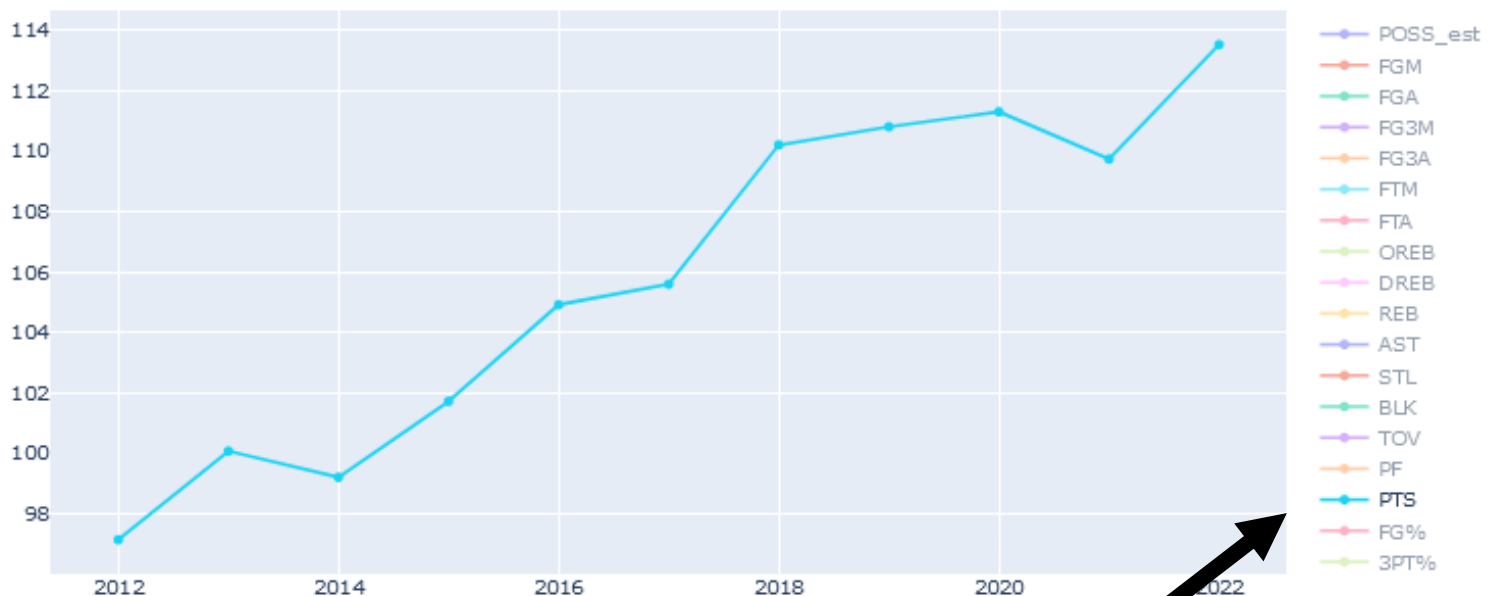


Figure 7

In the figure above, we specified one column only by using **plotly** library, which let us to do it.

predict for next five years

2023

2024

2025

2026

2027

Because positions per games from 2012 until now have been increasing and shooting three point shots and quality of shooting have increased over the years that tells us that the league is favoring offence more so with that result we could **predict that**

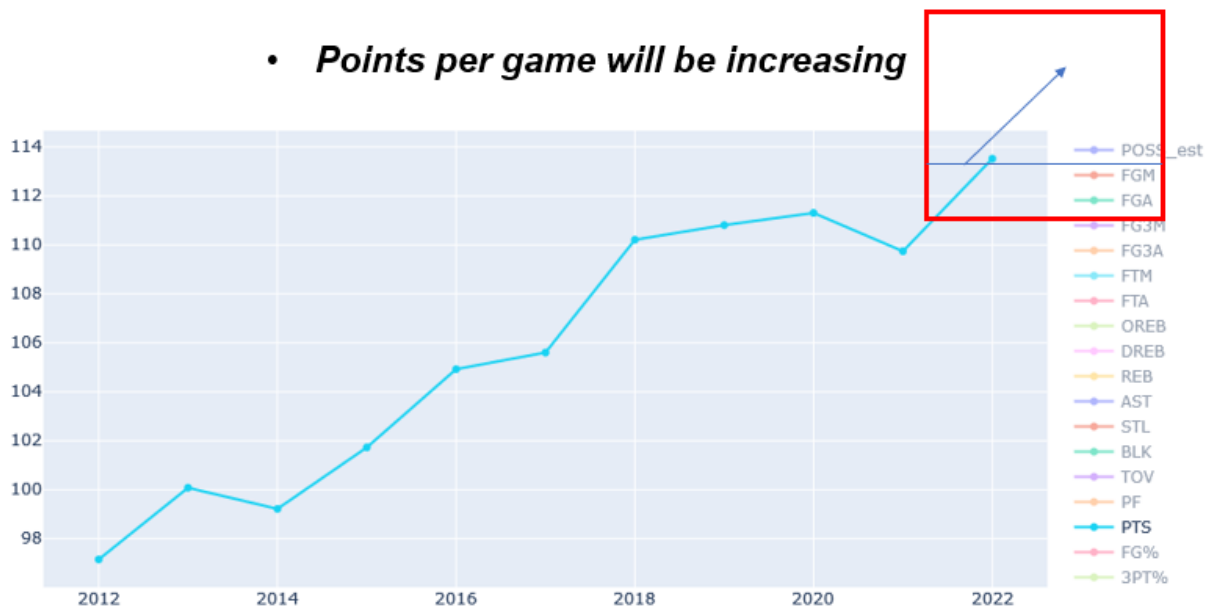


figure 8

we estimated that if the league continue to play like this the points per game for each team will be higher or stay high.

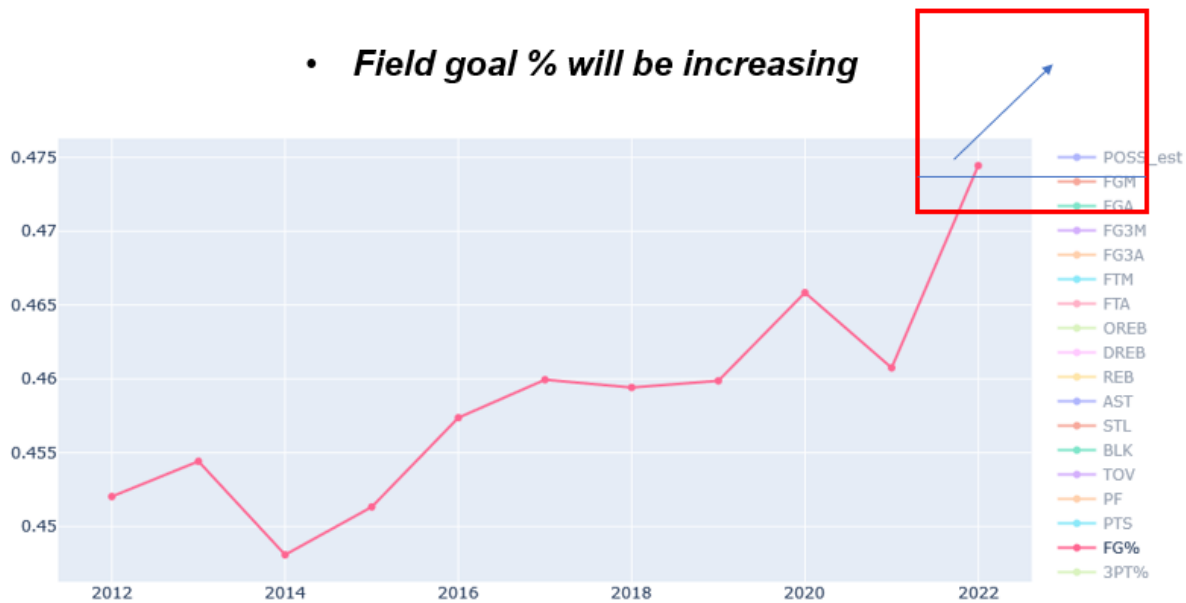


Figure 9

This figure (figure 9) shows the field goal percentage over the years if we watch 2012 we see that the percentage is near %45 but fast forward to 2022 it near %47, which means that the shot quality taken over the years have getting better and better.

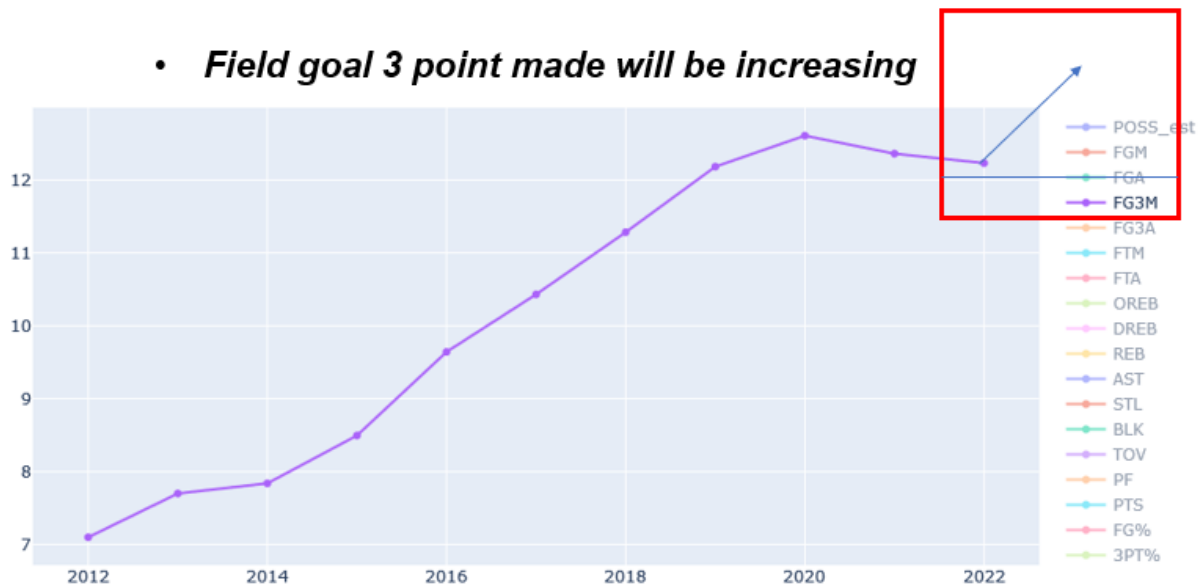


Figure 10

This figure shows that over the years three point shot that were made in a game have been increasing, so as of this result we could say the years to come the three point shot quality would get better and better so the three point made will increase.

predict for next five years

2023

2024

2025

2026

2027

Because the league nowadays is more focused on offence the defensive quality in the league has been decreased so based on that we could predict that:

- ***Block per game will be decreasing or stay low***

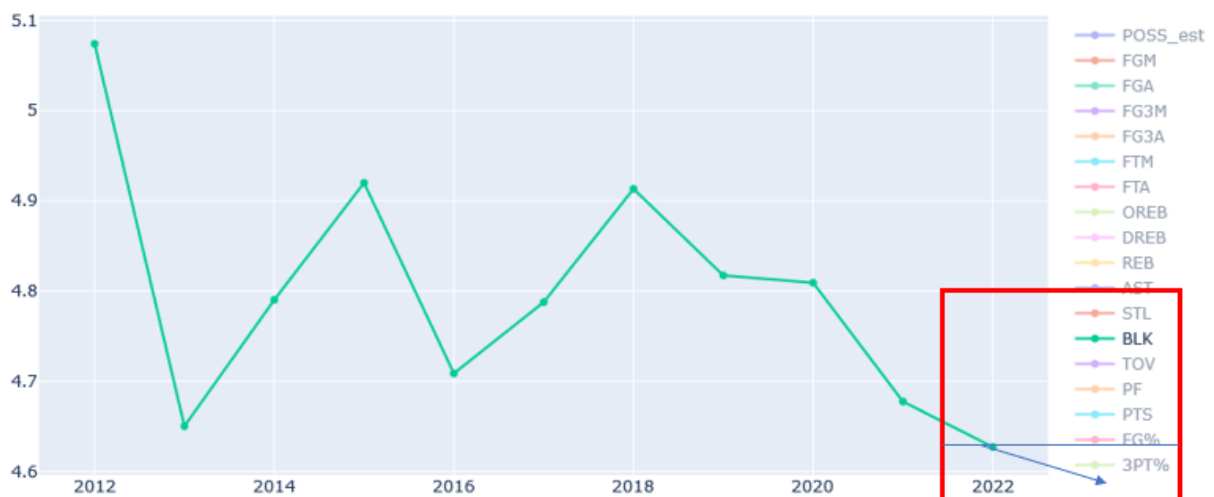


Figure 11

we could see that because of the down fall of the defense in nba we see that block per game have been decreased.



- *steal per game will be decreasing or stay low*

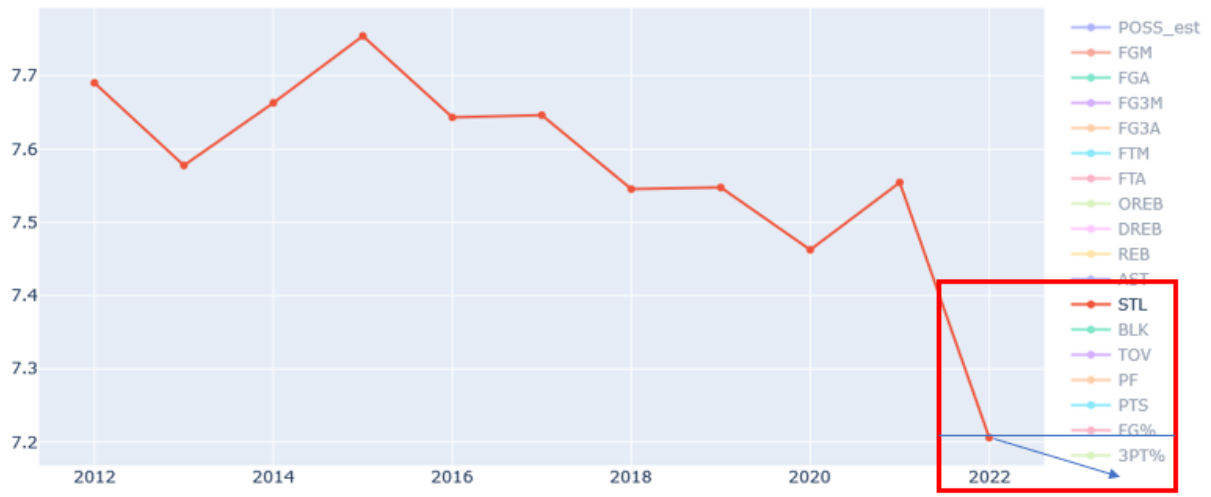


Figure 12

As of the other aspects of defense we could see this plot about steal per game that have been decreasing over the years so we could predict that if the league continues like this style of defending the steals per game will be lower or stay low for years to come.

## Reference:

- **Official NBA Website:**

- The NBA's official website (<https://www.nba.com/>) provides a wealth of information, including player statistics, team standings, schedules, and more.

- **Basketball-Reference:**

- Basketball-Reference (<https://www.basketball-reference.com/>) is a popular and comprehensive basketball statistics website. It offers detailed statistics on players, teams, and historical data.

- **ESPN:**

- ESPN's NBA section (<https://www.espn.com/nba/>) provides news, scores, schedules, and a variety of statistics related to NBA teams and players.

- **NBA Stats:**

- The NBA Stats website (<https://stats.nba.com/>) is an official NBA platform that offers in-depth statistical analysis, including advanced metrics, player tracking, and team performance.