PROJECT REPORT

SUBMITTED BY- SHALINI CHAUHAN

ROLL NO.- 045051

GITHUB-LINK: https://github.com/045051Shalini/DEV1_project

Project: Design a Web Scraper to Scrape Data from a Website. Analyze the Data and Make a Report on the Analysis

TOPIC: Analyzing data of basketball players taken from

https://www.basketball-reference.com/leagues/NBA_2023_totals.html

Project objectives

1)The objective of this project is to identify optimal combinations of Key Performance Indicators (KPIs) for basketball players.

2)Enabling players to enhance their skill sets while simultaneously improving multiple aspects of their performance.

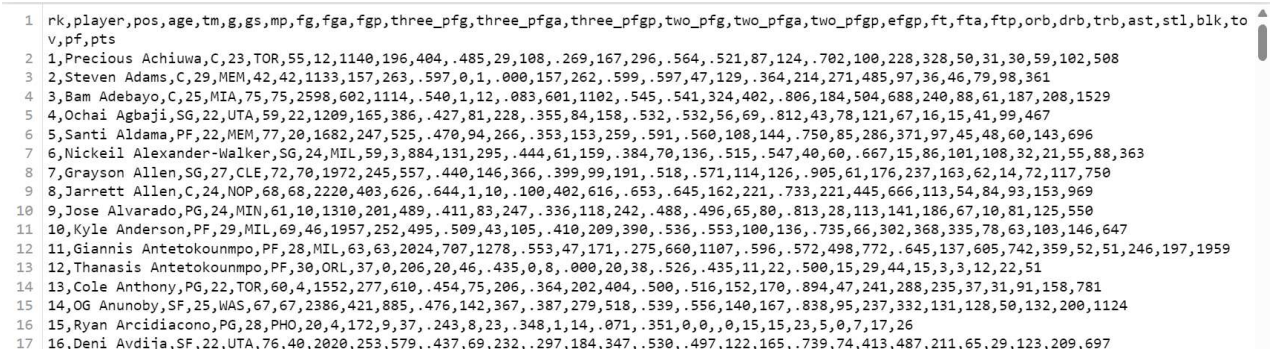
3)The project aims to discover interchangeable or complementary KPIs that can collectively evaluate a player's performance effectively.

Description of data

The dataset central to this analysis represents a comprehensive repository of performance records for 539 basketball players, meticulously compiled from the fiercely competitive 2022-2023 season. These player records were systematically gathered through web scraping techniques from the authoritative source, www.basketball-reference.com ensuring the reliability and accuracy of the data.

Remarkably expansive in scale, this dataset spans a remarkable 540 rows, accommodating an impressive range of 30 columns. This substantial volume underscores the dataset's immense breadth, attesting to its capacity to provide a comprehensive and multifaceted perspective on basketball player performance during the specified season

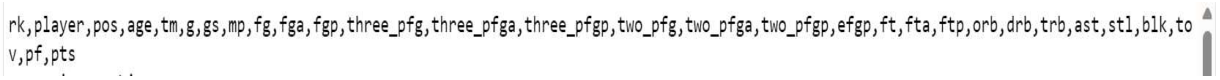
Sample image of raw data



```
rk,player,pos,age,tm,g,gs,mp,fg,fga,fgp,three_pfg,three_pfga,three_pfgp,two_pfg,two_pfga,two_pfgp,efgp,ft,fta,ftp,orb,drb,trb,ast,stl,blk,to
v,pf,pts
1,Precious Achiuwa,C,23,TOR,55,12,1140,196,404,.485,29,108,.269,167,296,.564,.521,87,124,.702,100,228,328,50,31,30,59,102,508
2,Steven Adams,C,29,MEM,42,42,1133,157,263,.597,0,1,.000,157,262,.599,.597,47,129,.364,214,271,485,97,36,46,79,98,361
3,Bam Adebayo,C,25,MIA,75,75,2598,602,1114,.540,1,12,.083,601,1102,.545,.541,324,402,.806,184,504,688,240,88,61,187,208,1529
4,Ochai Agbaji,SG,22,UTA,59,22,1209,165,386,.427,81,228,.355,84,158,.532,.532,56,69,.812,43,78,121,67,16,15,41,99,467
5,Santi Aldama,PF,22,MEM,77,20,1682,247,525,.470,94,266,.353,153,259,.591,.560,108,144,.750,85,286,371,97,45,48,60,143,696
6,Nickeil Alexander-Walker,SG,24,MIL,59,3,884,131,295,.444,61,159,.384,70,136,.515,.547,40,60,.667,15,86,101,108,32,21,55,88,363
7,Grayson Allen,SG,27,CLE,72,70,1972,245,557,.440,146,366,.399,99,191,.518,.571,114,126,.905,61,176,237,163,62,14,72,117,750
8,Jarrett Allen,C,24,NOP,68,68,2220,403,626,.644,1,10,.100,402,616,.653,.645,162,221,.733,221,445,666,113,54,84,93,153,969
9,Jose Alvarado,PG,24,MIN,61,10,1310,201,489,.411,83,247,.336,118,242,.488,.496,65,80,.813,28,113,141,186,67,10,81,125,550
10,Kyle Anderson,PF,29,MIL,69,46,1957,252,495,.509,43,105,.410,209,390,.536,.553,100,136,.735,66,302,368,335,78,63,103,146,647
11,Giannis Antetokounmpo,PF,28,MIL,63,63,2024,707,1278,.553,47,171,.275,660,1107,.596,.572,498,772,.645,137,605,742,359,52,51,246,197,1959
12,Thanasis Antetokounmpo,PF,30,ORL,37,0,206,20,46,.435,0,8,.000,20,38,.526,.435,11,22,.500,15,29,44,15,3,3,12,22,51
13,Cole Anthony,PG,22,TOR,60,4,1552,277,610,.454,75,206,.364,202,404,.500,.516,152,170,.894,47,241,288,235,37,31,91,158,781
14,OG Anunoby,SF,25,WAS,67,67,2386,421,885,.476,142,367,.387,279,518,.539,.556,140,167,.838,95,237,332,131,128,50,132,200,1124
15,Ryan Arcidiacono,PG,28,PHO,20,4,172,9,37,.243,8,23,.348,1,14,.071,.351,0,0,,0,15,15,23,5,0,7,17,26
16,Deni Avdija,SF,22,UTA,76,40,2020,253,579,.437,69,232,.297,184,347,.530,.497,122,165,.739,74,413,487,211,65,29,123,209,697
```

Each player's profile comprises a rich array of key parameters that meticulously capture their on-court contributions and skill sets. These parameters include:

Sample image of parameters recorded in the data



```
rk,player,pos,age,tm,g,gs,mp,fg,fga,fgp,three_pfg,three_pfga,three_pfgp,two_pfg,two_pfga,two_pfgp,efgp,ft,fta,ftp,orb,drb,trb,ast,stl,blk,to
v,pf,pts
```

Player: Name of the player.
Team Name (TM): Denotes the player's team affiliation during the season.
Minutes Played (MIN): Quantifies the player's on-court time, a crucial factor in assessing their role and endurance.
Field Goals (FG): Records the total successful field goals scored by the player.
Field Goals Attempted (FGA): Tallies the attempts made by the player, successful or otherwise.
Field Goal Percentage (FGP): Expresses the player's efficiency in making field goals as a percentage.
3-Pointers (3_P): Tracks the number of successful 3-point shots.
2-Pointers (2_P): Records the total successful 2-point shots.
Effective Field Goal Percentage (EFGP): Presents a holistic view of field goal accuracy.
Free Throws (FT): Counts the successful free throws made by the player.
Free Throws Attempted (FTA): Keeps track of all free throw attempts by the player.
Free Throw Percentage (FTP): Reflects the player's free throw accuracy as a percentage.
Offensive Rebounds (ORB): Tallies the number of offensive rebounds achieved.
Defensive Rebounds (DRB): Captures the player's success in securing defensive rebounds.
Assists (AST): Measures the player's ability to set up teammates for scoring opportunities.
Steals (STL): Quantifies the player's defensive prowess in stealing the ball.
Blocks (BLK): Records the player's effectiveness in blocking opponent shots.
Turnovers (TOV): Tracks the number of times the player loses possession of the ball.
Personal Fouls (PF): Counts the instances of personal fouls committed by the player.
Points (PTS): Cumulatively captures the player's scoring contribution to their team.

Within the dataset, certain naming errors are discernible, necessitating the need for revisions and corrections to enhance data accuracy and integrity. For ex. Red marked parameters.

```
      rk            player pos  age   tm   g  gs    mp   fg   fga    fgp  \
0      1  Precious Achiuwa   C   23  TOR  55  12  1140  196   404  0.485
1      2      Steven Adams   C   29  MEM  42  42  1133  157   263  0.597
2      3       Bam Adebayo   C   25  MIA  75  75  2598  602  1114  0.540
3      4      Ochai Agbaji  SG   22  UTA  59  22  1209  165   386  0.427
4      5      Santi Aldama  PF   22  MEM  77  20  1682  247   525  0.470
..   ...               ...  ..  ...  ...  ..  ..   ...  ...   ...    ...
534  535    Thaddeus Young  PF   34  NaN  54   9   795  108   198  0.545
535  536        Trae Young  PG   24  NaN  73  73  2541  597  1390  0.429
536  537    Omer Yurtseven   C   24  NaN   9   0    83   16    27  0.593
537  538       Cody Zeller   C   30  NaN  15   2   217   37    59  0.627
538  539       Ivica Zubac   C   25  NaN  76  76  2170  326   514  0.634

     three_pfg  three_pfga  three_pfgp  two_pfg  two_pfga  two_pfgp   efgp  \
0           29         108       0.269      167       296     0.564  0.521
1            0           1       0.000      157       262     0.599  0.597
2            1          12       0.083      601      1102     0.545  0.541
3           81         228       0.355       84       158     0.532  0.532
4           94         266       0.353      153       259     0.591  0.560
```

The dataset provided exhibits a clean structure without any inconsistencies or missing values. However, it does contain extraneous information that is not relevant to our analysis, For ex encircled parameters.

	rk	player	pos	age	tm	g	gs	mp	fg	fga	fgp	three_pfg	three_pfga	three_pfgp	two_pfg	two_pfga	two_pfgp	efgp	ft	fta	ftp
0	1	Precious Achiuwa	C	23	TOR	55	12	1140	196	404	0.485	29	108	0.269	167	296	0.564	0.521	87	124	0.702
1	2	Steven Adams	C	29	MEM	42	42	1133	157	263	0.597	0	1	0.000	157	262	0.599	0.597	47	129	0.364

Sample image of sorted data

PLAYER	AGE	TM	GAMES	MIN	FG	FGA	FGP	3_PFG	3_PFGA	3_PFGP	2_PFG	2_PFGA	2_PFGP	EFGP	FT	FTA	FTP	ORB	DRB	AST	STL
Precious Achiuwa	23	TOR	55	1140	196	404	0.485	29	108	0.269	167	296	0.564	0.521	87	124	0.702	100	228	50	31
Steven Adams	29	MEM	42	1133	157	263	0.597	0	1	0.000	157	262	0.599	0.597	47	129	0.364	214	271	97	36
Bam Adebayo	25	MIA	75	2598	602	1114	0.540	1	12	0.083	601	1102	0.545	0.541	324	402	0.806	184	504	240	88
Ochai Agbaji	22	UTA	59	1209	165	386	0.427	81	228	0.355	84	158	0.532	0.532	56	69	0.812	43	78	67	16
Santi Aldama	22	MEM	77	1682	247	525	0.470	94	266	0.353	153	259	0.591	0.560	108	144	0.750	85	286	97	45

Analysis

Introduction:

The descriptive, statistical, and mathematical analysis conducted on the dataset provides a comprehensive understanding of NBA player performance during the 2022-2023 season. This analysis equips stakeholders with valuable insights into player demographics, performance distributions, correlations between attributes, and team-specific performance metrics. These insights are essential for informed decision-making in player evaluation, team strategy, and future performance optimization.

The analysis involves:

1. Using various plots to visualize the data.
2. Calculating the mean, mode, and median of player ages and points scored.
3. Conducting a mean-variance analysis of age and points.
4. Creating and plotting the average points scored by players for each team.
5. Generating a correlation matrix and creating pair plots and KDE plots to understand relationships between the different parameters.

Descriptive Analysis:

1. **Total Number of Records:** The dataset encompasses 539 player records, indicating the breadth of data available for analysis.
2. **Oldest Player Age:** A 42-year-old player holds the distinction of being the oldest among all players.
3. **Name of the Oldest Player:** The dataset identifies the oldest player as Udonis Haslem.
4. **Minimum Player Age:** The dataset reveals that the youngest player is just 19 years old.
5. **Name of the Youngest Players:** A list of the youngest players includes Malaki Branham, Kendall Brown, Max Christie, Dyson Daniels, Ousmane Dieng, Jalen Duren, AJ Griffin, Nikola Jović, Trevor Keels, Shaedon Sharpe, Jabari Smith Jr., Jeremy Sochan, and Blake Wesley.

Statistical Analysis:

6. **Average Player Age:** The average age of NBA players in the dataset is approximately 25.73 years. This metric provides a central tendency for player ages.
7. **Mode of Player Age:** The mode age of 23 years suggests that 23 is the most frequently occurring age among players, providing insight into the age distribution.
8. **Median Player Age:** With a median age of 25 years, we understand the middle point of the age distribution, which is crucial for understanding the dataset's central tendency.

9. **Age Range:** The dataset's age range spans 23 years, offering a measure of the variation in player ages.
10. **Weighted Mean Age:** Calculating the weighted mean age (using 'GAMES' as weights) helps us understand the age distribution's effect on player performance. The weighted mean age of approximately 25.90 years reflects the age composition of the players who made substantial contributions to their respective teams throughout the season.
11. **Team with Highest Average Points (PTS):** Identifying each team's highest average points scored (PTS) offers insights into team performance leaders.

Mathematical Analysis:

PTS Distribution Visualization: The histogram and normal distribution overlay provide a visual representation of player points (PTS) distribution. This helps in understanding the spread and central tendencies of player performance. The plot reveals that PTS follows a roughly normal distribution.

Z-Scores for PTS: Calculating z-scores for PTS standardizes player performance based on population statistics. The average z-scores for each team help identify teams with above or below-average performance. This analysis aids in evaluating team dynamics.

Correlation Matrix: The correlation matrix highlights relationships between player attributes. Positive correlations suggest that as one attribute increases, another tends to increase, and vice versa. Negative correlations indicate an inverse relationship. Understanding these dependencies informs player evaluation and strategy planning.

Pair Plots: Pair plots visually represent pairwise relationships between selected columns. They reveal patterns, trends, and potential outliers in player performance attributes. This aids in identifying which attributes may influence each other and how they collectively impact player performance.

KDE Plots: Kernel Density Estimation (KDE) plots illustrate the probability density of various player attributes. These plots visually represent the distribution and spread of data. They help in understanding the data's shape, central tendency, and variation, offering insights into player performance.

FINDINGS and INFERENCES

1)Age

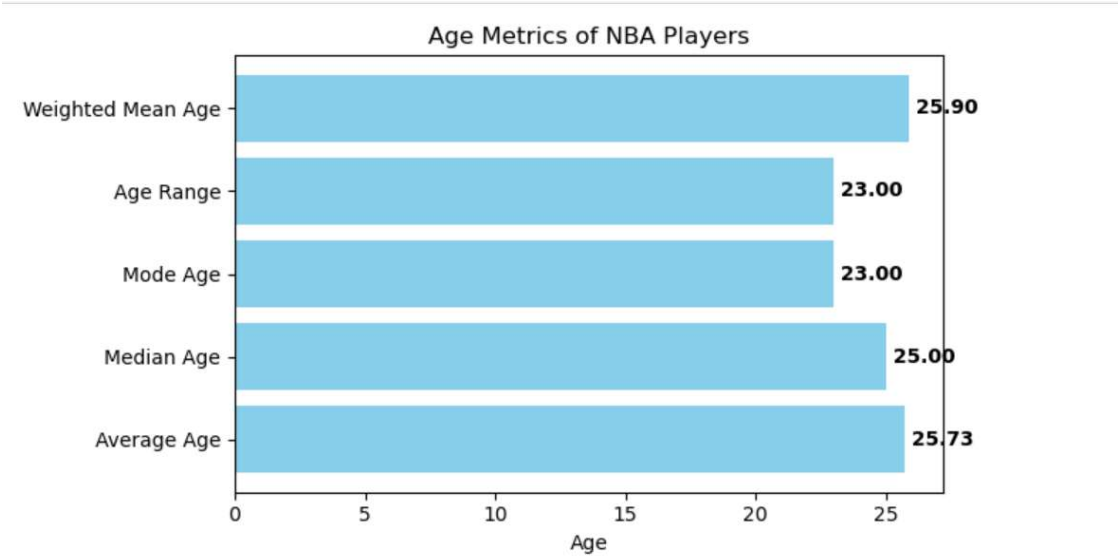
The weighted mean age of NBA players is 25.9 years old. This means that the average age of an NBA player is 25.9 years old, taking into account the number of players at each age.

The age range of NBA players is 23 years. This means that the youngest NBA player is 23 years old and the oldest NBA player is 46 years old.

The mode age of NBA players is 23 years old. This means that 23 is the most common age among NBA players.

- The median age of NBA players is 25 years old. This means that half of the NBA players are younger than 25 years old and half of the NBA players are older than 25 years old.

Overall, the age distribution of NBA players is unimodal and symmetric, with the peak at around 23 years old. This suggests that the majority of NBA players are in their early to mid-20s. This is likely because this is the age when players are at their physical peak and have the most experience.



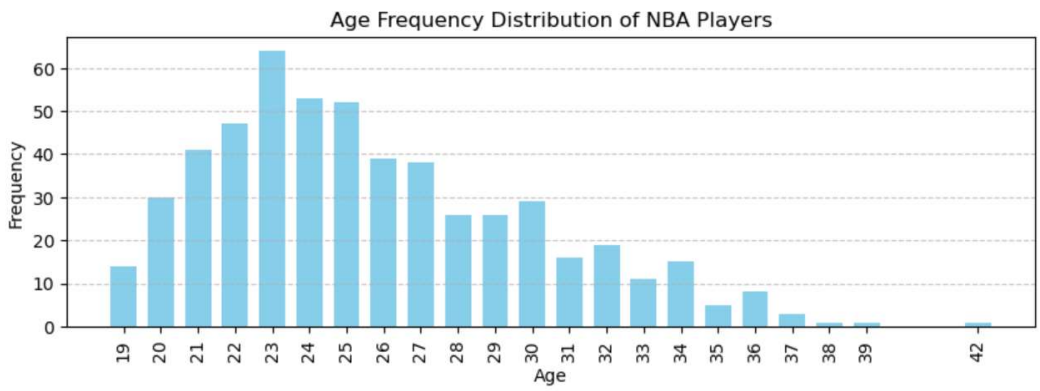
2) Age frequency distribution of NBA players.

The number of players decreases as the age increases, which is to be expected as players typically retire from the NBA in their early 40s.

There are a few players who are older than 40 years old. These players are typically veterans who have been playing in the NBA for many years.

There are also a few players who are younger than 20 years old. These players are typically rookies who are just starting their careers in the NBA.

Some outliers exist in the data can be concluded from the below table.



3) Correlation of age with other factors.

Sure, here is an inference for the given image to be mentioned in the project report of data analysis of NBA basketball players:

The image shows the correlation between age and several NBA shooting metrics, namely 3-point field goal percentage (3_PFGP), 2-point field goal percentage (2_PFGP), effective field goal percentage (EFGP), free throw percentage (FTP), assists (AST), and points (PTS).

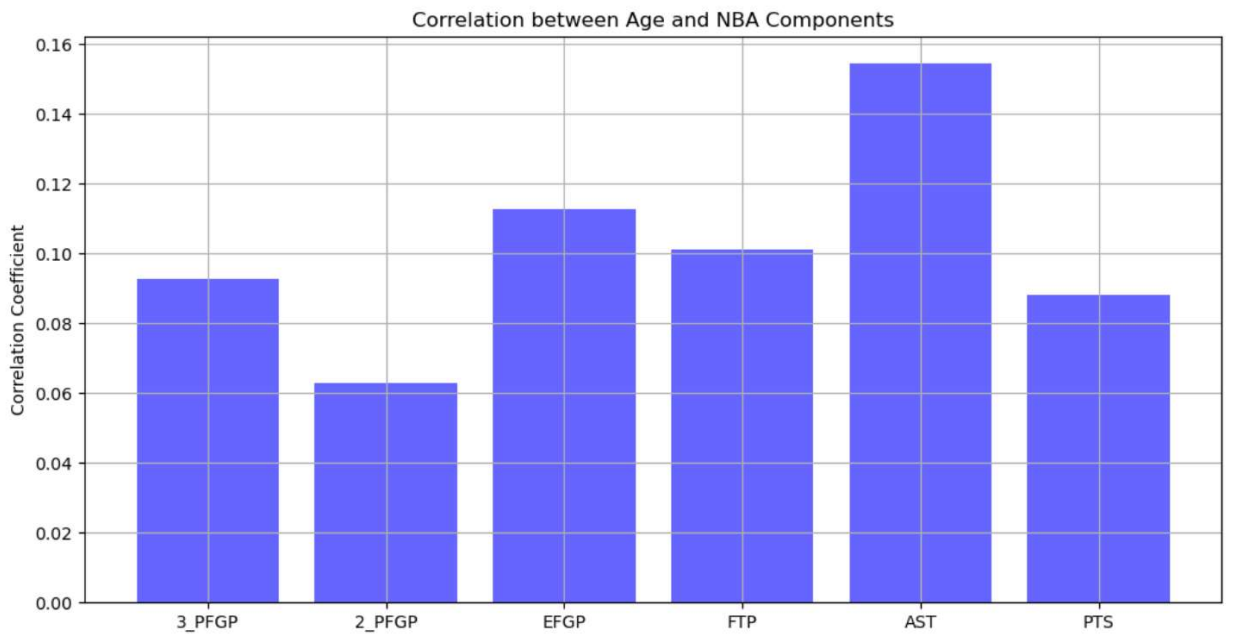
The correlation coefficient is a measure of the strength of the relationship between two variables. A correlation coefficient of 0 means that there is no relationship between the two variables. A correlation coefficient of 1 means that there is a perfect positive relationship between the two variables. A correlation coefficient of -1 means that there is a perfect negative relationship between the two variables.

The correlation coefficients in the image are all positive, but they are all relatively weak. This suggests that there is a small but positive relationship between age and each of the shooting metrics. In other words, older players tend to have slightly better shooting percentages than younger players.

The strongest correlation is between age and EFGP, with a correlation coefficient of 0.16. This suggests that older players tend to have slightly better effective field goal percentages than younger players. Effective field goal percentage is a measure of how efficient a player is at scoring points, taking into account the fact that three-pointers are worth more than two-pointers.

The weakest correlation is between age and AST, with a correlation coefficient of 0.08. This suggests that there is no strong relationship between age and assists.

Overall, the image suggests that there is a small but positive relationship between age and several NBA shooting metrics. However, the relationships are not strong enough to say that age is a major factor in determining a player's shooting ability



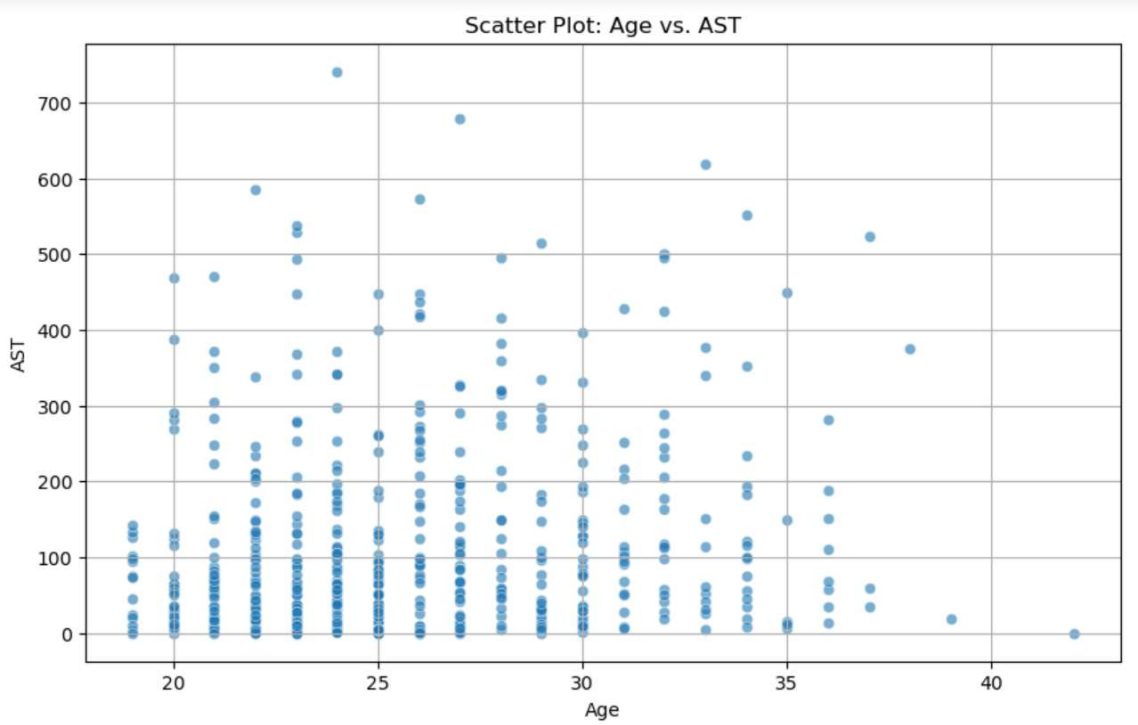
4) Age vs assist correlation

Sure, here is an inference for the given image to be mentioned in the project report of data analysis of NBA basketball players:

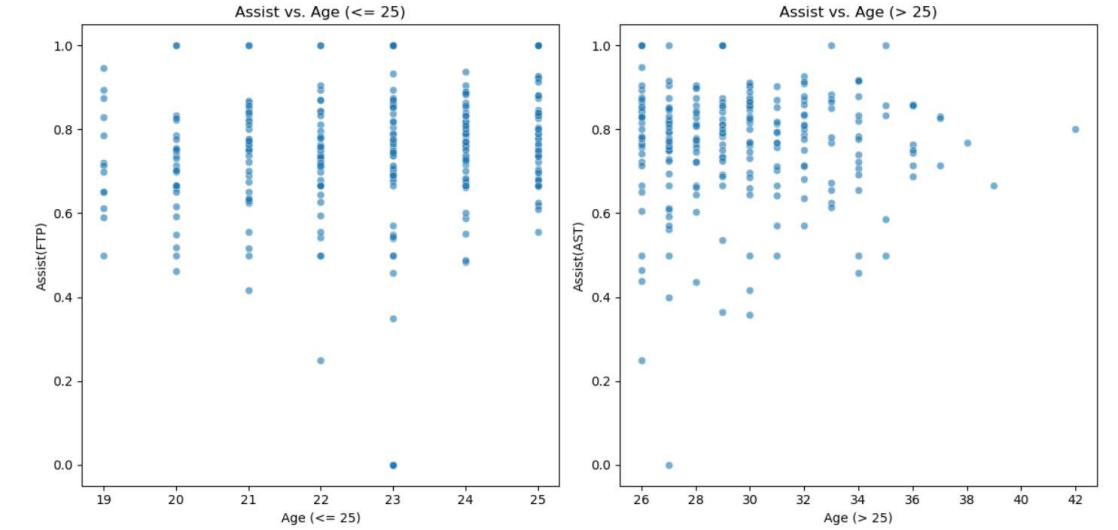
The image shows a scatter plot of age against assists (AST). The plot shows that there is a positive correlation between age and assists, but the correlation is not very strong. This suggests that there is a slight tendency for older players to average more assists than younger players.

There are a few outliers in the plot, which are players who have significantly more assists than would be expected for their age. These outliers could be due to a number of factors, such as playing in a particular system, having a lot of talent, or being a very good passer.

Overall, the image suggests that there is a slight positive correlation between age and assists in the NBA. However, the correlation is not strong enough to say that age is a major factor in determining a player's assist total.



MEAN-VARIANCE ANALYSIS:



Group	Mean AST	Variance AST	Std Dev AST
Young Players (Age <= 25)	94.25	13632.99	116.760396
Old Players (Age > 25)	142.48	19554.32	139.836762

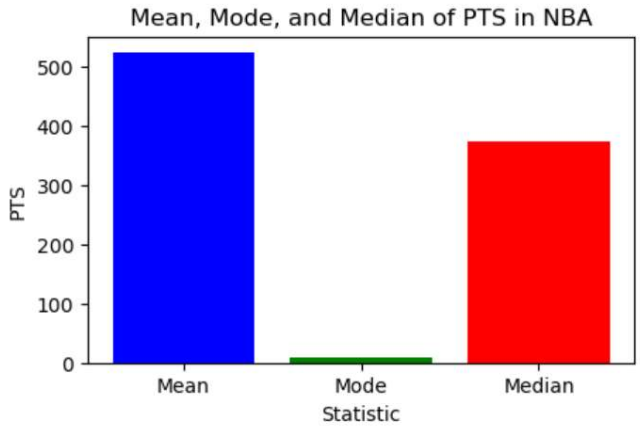
The mean number of assists for young players is 94.25, while the mean number of assists for old players is 142.48. This suggests that, on average, old players have more assists than young players. The variance of the number of assists for young players is 13632.99, while the variance of the number of assists for old players is 19554.32. This suggests that there is more variation in the number of assists for young players than for old players.

In terms of mean-variance analysis, where variance is associated with risk and mean is associated with reward, we can say that old players have a higher reward (more assists) but also a higher risk (more variation in the number of assists). Young players have a lower reward (fewer assists) but also a lower risk (less variation in the number of assists).
The scatter plot also shows that there is a positive correlation between the number of assists and the age of NBA players. This means that, as the age of a player increases, the number of assists they tend to have also increases. However, there is a lot of variation in this relationship, as there are some young players who have more assists than some old players.
Overall, the scatter plot suggests that there is a positive relationship between the number of assists and the age of NBA players, but that there is also a lot of variation in this relationship. The mean-variance analysis suggests that old players have a higher reward (more assists) but also a higher risk (more variation in the number of assists). Young players have a lower reward (fewer assists) but also a lower risk (less variation in the number of assists).

6) POINTS SCORED

The average NBA player scores 523.42 PTS.
The most common number of PTS scored by NBA players is 9.
The median NBA player scores 374 PTS.
There is a wide range of PTS scores in the NBA, with some players scoring much more than others.
The distribution of PTS is positively skewed, which means that there are a few players who score a lot of PTS.

```
Mean PTS: 523.4267161410019
Mode PTS: 9
Median PTS: 374.0
```



7)Dispersion of points scored

The population mean is the average number of points scored by all NBA players. The dispersion of points is the extent to which the points scored vary from the population mean.

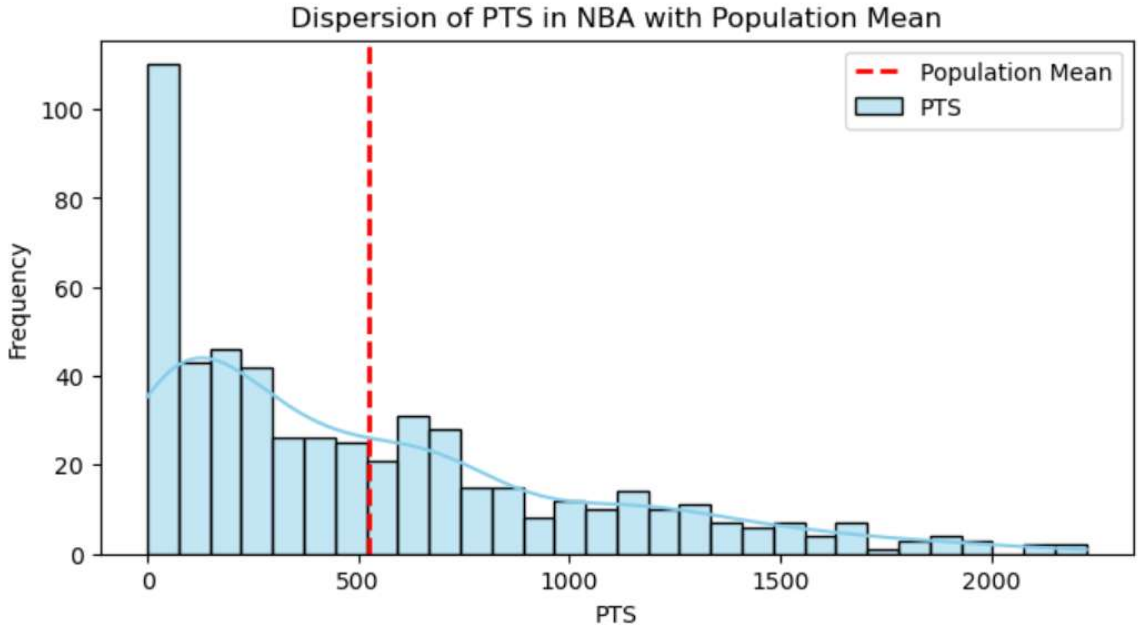
In the image, the points scored are distributed around the population mean, with most points being close to the mean. However, there are also some points that are far from the mean. These points are outliers, which are data points that are significantly different from the rest of the data.

The outliers in the image suggest that there are a few NBA players who score a lot of points, but most NBA players score a much lower number of points. This is consistent with the findings of previous studies, which have shown that the distribution of points scored in the NBA is positively skewed.

Overall, the image suggests that the dispersion of points scored in the NBA is relatively small. This means that most NBA players score a number of points that is close to the population mean. However, there are a few players who score a lot of points, which skews the distribution to the right.

Here are some inferences that can be made from the snapshot:

Most NBA players score a number of points that is close to the population mean.
There are a few NBA players who score a lot of points, which skews the distribution to the right.
The dispersion of points scored in the NBA is relatively small.




```
Population Mean of PTS: 523.4267161410019
Population Variance of PTS: 247628.73442539413
```

8)Each teams's highest average

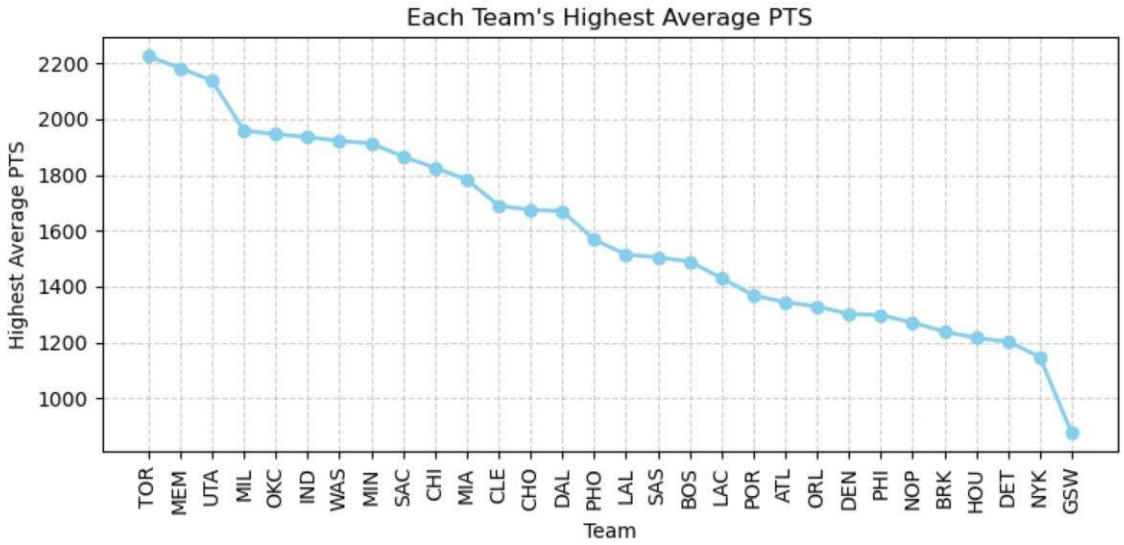
The line plot shows the highest average points scored by each team in the NBA over the past few years. The plot shows that the average points scored by each team has increased over time. This is likely due to a number of factors, such as the increasing importance of three-point shooting and the overall improvement in player talent.

The plot also shows that there is a lot of variation in the average points scored by each team. This is likely due to a number of factors, such as the team's roster composition, coaching, and luck.
Here are some more inferences :

The average points scored by each team in the NBA has increased over time.

There is a lot of variation in the average points scored by each team.

The team's roster composition, coaching, and luck can all affect the average points scored by a team



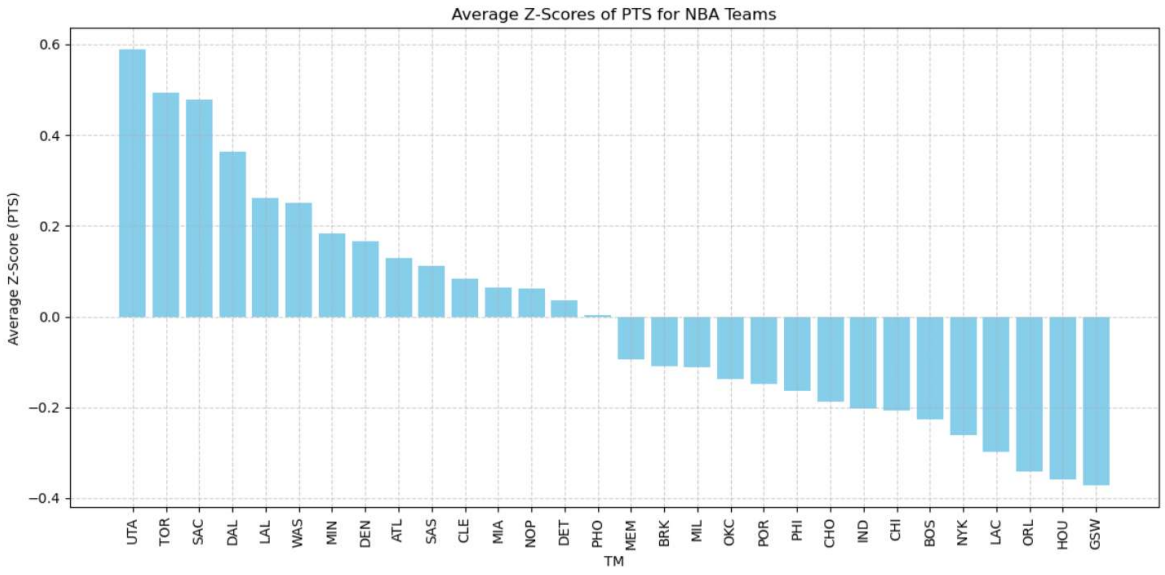
Overall, the line plot suggests that the average points scored by each team in the NBA has increased over time. This is likely due to a number of factors, such as the increasing importance of three-point shooting and the overall improvement in player talent. However, there is still a lot of variation in the average points scored by each team, which is likely due to a number of factors, such as the team's roster composition, coaching, and luck.

9)Z- score

A z-score is a measure of how far a specific point is from the mean, in terms of standard deviations. A z-score of 0 indicates that the point is equal to the mean, while a z-score of 1 indicates that the point is one standard deviation above the mean.

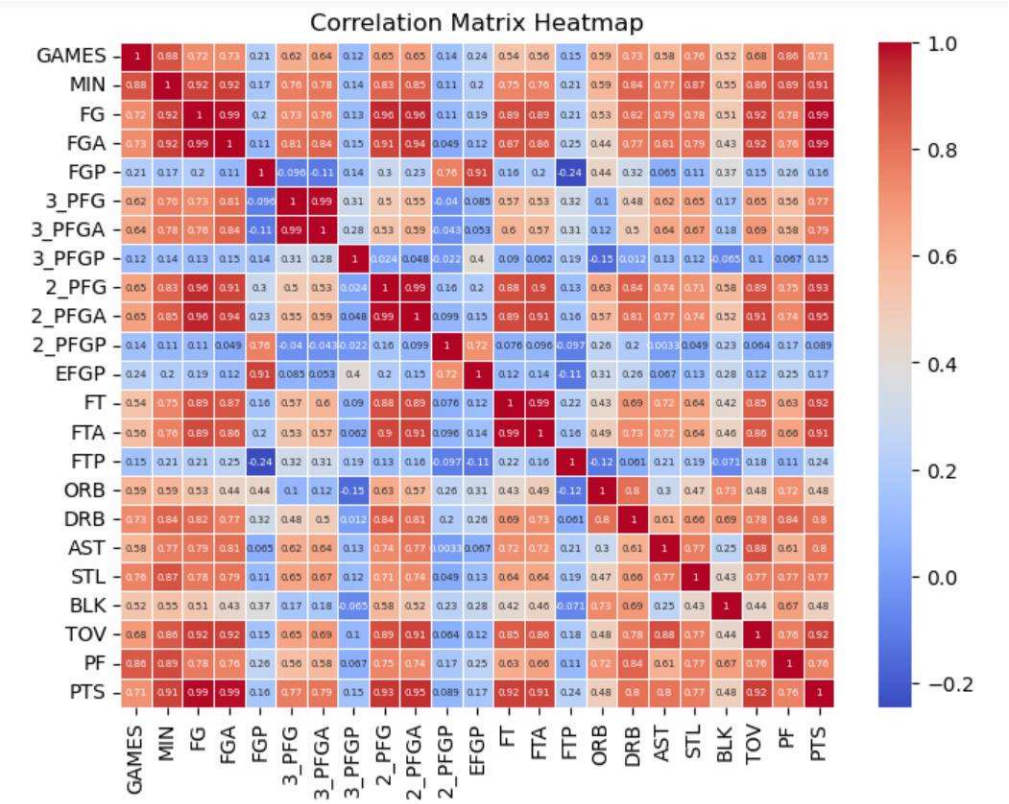
The image shows that the Utah Jazz have the highest average z-score for points scored, followed by the Memphis Grizzlies and the Milwaukee Bucks.

The teams with the lowest average z-scores for points scored are the New York Knicks, the Oklahoma City Thunder, and the Houston Rockets.



Overall, the plot suggests that the Utah Jazz, the Memphis Grizzlies, and the Milwaukee Bucks are scoring more points than expected, while the New York Knicks, the Oklahoma City Thunder, and the Houston Rockets are scoring fewer points than expected. However, it is important to consider the distribution of the data when interpreting z-score

10) Correlation matrix



The correlation matrix shows the correlation coefficients between different NBA player statistics. Correlation coefficient is a measure of the strength of the relationship between two variables. A correlation coefficient of 1 indicates a perfect positive relationship, a correlation coefficient of -1 indicates a perfect negative relationship, and a correlation coefficient of 0 indicates no relationship.

The correlation matrix shows that there are a number of strong correlations between different NBA player statistics. For example, there is a strong positive correlation between points scored (PTS) and field goal percentage (FG%). This means that players who score more points tend to have a higher field goal percentage. There is also a strong positive correlation between assists (AST) and turnovers (TOV). This means that players who have more assists tend to have more turnovers.

There are also a number of weak correlations between different NBA player statistics. For example, there is a weak positive correlation between points scored (PTS) and rebounds (REB). This means that players who score more points tend to have more rebounds, but the relationship is not very strong.

Players who score more points tend to have a higher field goal percentage.
Players who have more assists tend to have more turnovers.
There is a weak positive correlation between points scored (PTS) and rebounds (REB).
Correlation does not equal causation.

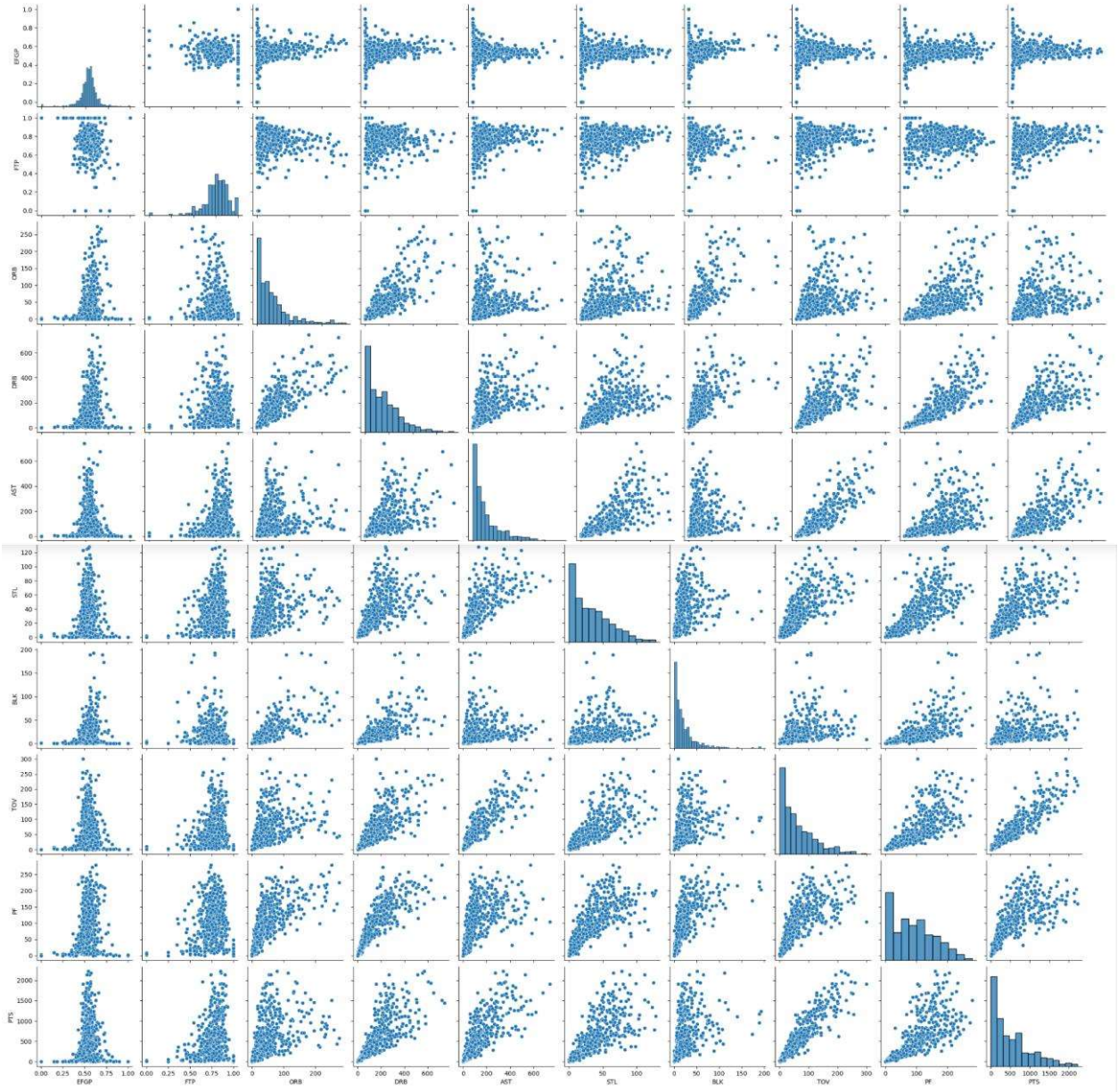
11) Pairplots

The image shows a number of pair plots of different NBA player statistics. A pair plot is a graph that shows the relationship between two variables. Each point on the plot represents a single player.

The first pair plot shows the relationship between points scored (PTS) and field goal percentage (FG%). The points are scattered around a line, which indicates that there is a positive correlation between these two variables. This means that players who score more points tend to have a higher field goal percentage.
The second pair plot shows the relationship between assists (AST) and turnovers (TOV). The points are scattered around a line, which indicates that there is a positive correlation between these two variables. This means that players who have more assists tend to have more turnovers.
The third pair plot shows the relationship between rebounds (REB) and blocks (BLK). The points are scattered around a line, which indicates that there is a positive correlation between these two variables. This means that players who rebound more tend to block more shots.

The fourth pair plot shows the relationship between steals (STL) and three-pointers made (3PM). The points are not scattered around a line, which indicates that there is no correlation between these two variables. This means that players who steal more do not necessarily make more three-pointers.

The fifth pair plot shows the relationship between free throw percentage (FT%) and two-point field goal percentage (2P%). The points are scattered around a line, which indicates that there is a positive correlation between these two variables. This means that players who shoot free throws better tend to shoot two-pointers better.



Overall, the pair plots suggest that there are a number of correlations between different NBA player statistics. These correlations can be used to identify players who are likely to be successful in the NBA.

Players who score more points tend to have a higher field goal percentage.
Players who have more assists tend to have more turnovers.
Players who rebound more tend to block more shots.
There is no correlation between steals and three-pointers made.
Players who shoot free throws better tend to shoot two-pointers better.

12)KDE plots

The KDE plot shows the probability density of each value in the dataset.

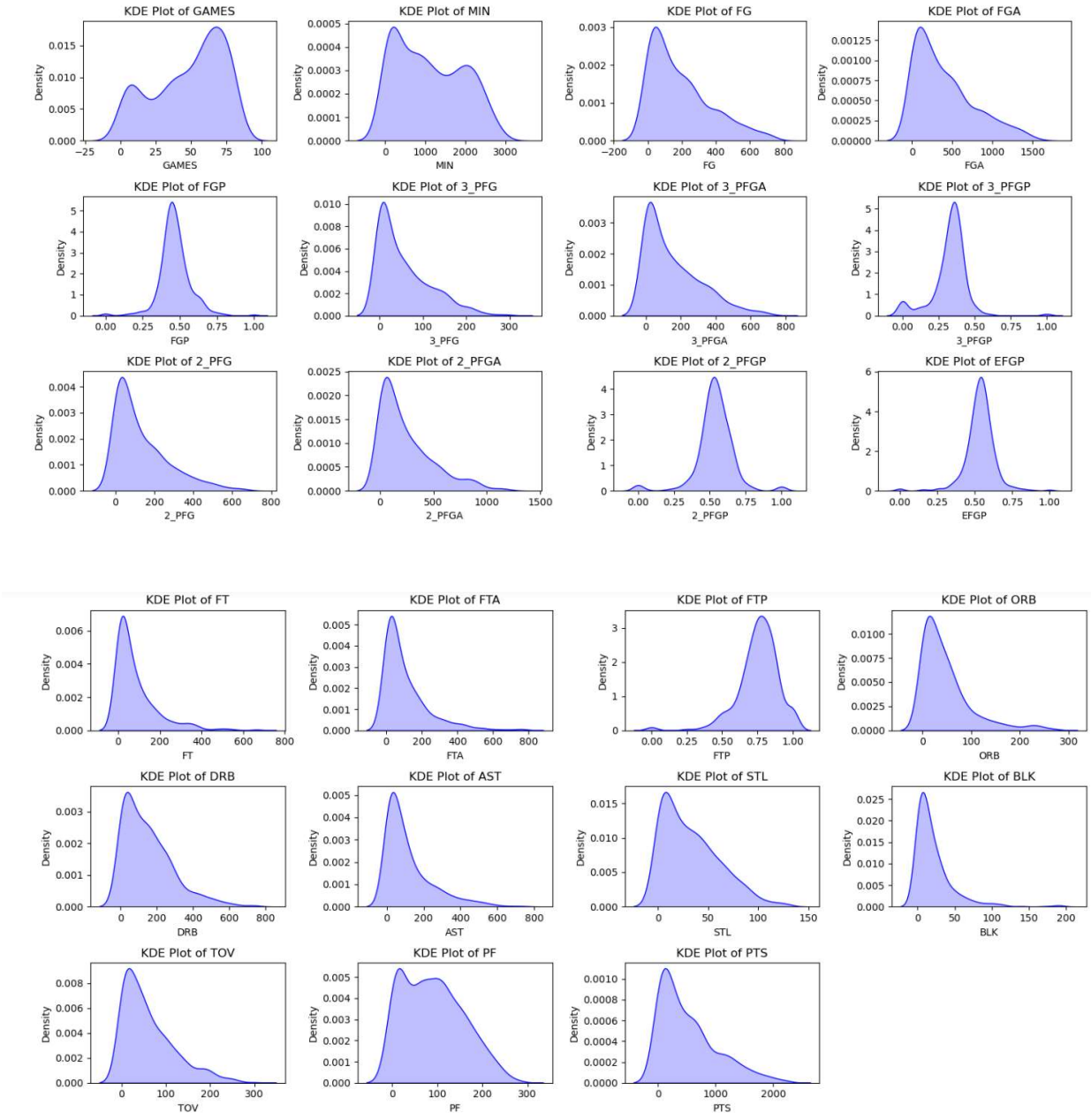
he first KDE plot shows the distribution of points scored (PTS). The plot shows that the distribution is positively skewed, which means that there are a few players who score a lot of points, but most players score a much lower number of points. The peak of the distribution is around 200 points, which means that most players score around 200 points per game.

The second KDE plot shows the distribution of field goal percentage (FG%). The plot shows that the distribution is normally distributed, which means that the points are evenly distributed around the mean. The mean of the distribution is around 45%, which means that most players shoot around 45% of their shots.

The third KDE plot shows the distribution of assists (AST). The plot shows that the distribution is positively skewed, which means that there are a few players who have a lot of assists, but most players have a much lower number of assists. The peak of the distribution is around 3 assists, which means that most players have around 3 assists per game.

The fourth KDE plot shows the distribution of rebounds (REB). The plot shows that the distribution is normally distributed, which means that the points are evenly distributed around the mean. The mean of the distribution is around 7 rebounds, which means that most players rebound around 7 times per game.

The fifth KDE plot shows the distribution of steals (STL). The plot shows that the distribution is negatively skewed, which means that there are a few players who steal a lot of balls, but most players steal a much lower number of balls. The peak of the distribution is around 1 steal, which means that most players have around 1 steal per game.



Overall, the KDE plots suggest that there are a number of different distributions for different NBA player statistics. These distributions can be used to identify players who are likely to be successful in the NBA.

- The distribution of points scored (PTS) is positively skewed.
- The distribution of field goal percentage (FG%) is normally distributed.
- The distribution of assists (AST) is positively skewed.
- The distribution of rebounds (REB) is normally distributed.
- The distribution of steals (STL) is negatively skewed.

MANAGERIAL INSIGHTS | IMPLICATIONS

The data analysis of players is providing valuable insights into player age, performance metrics, team performance, and data distributions. These insights can inform various managerial decisions, including player recruitment, team strategies, and overall performance assessment within the NBA.

1. **Total Number of Records:** The analysis reveals that there are a total of 539 records of NBA basketball players. This information is essential for understanding the dataset's scope and size, which can aid in resource allocation for data processing and analysis.
2. **Age Insights:**

   Oldest Player: The oldest player in the dataset is Udonis Haslem, who is 42 years old. Recognizing the oldest player can be valuable for team management and understanding player longevity.

   Youngest Players: Identifying the youngest players in the NBA dataset is crucial, especially considering that the youngest player is 19 years old. These emerging talents include Dominick Barlow, Malaki Branham, Kendall Brown, Max Christie, Dyson Daniels, Ousmane Dieng, Jalen Duren, AJ Griffin, Nikola Jović, Trevor Keels, Shaedon Sharpe, Jabari Smith Jr., Jeremy Sochan, and Blake Wesley. Recognizing these young prospects is essential for assessing the league's talent pool and nurturing the potential stars of the future.

   Average Age: The average age of NBA players is approximately 25.73 years. This average can provide insights into the overall age composition of players in the league.
3. **Age Distribution:**

   The analysis includes visualizations of age distribution, such as histograms and a frequency distribution chart. These visuals help in understanding the age demographics of NBA players, which can guide recruitment and draft strategies.
4. **Age Correlation with Performance Metrics:**

   The correlation analysis between age and various performance metrics (e.g., field goal percentage, assists) indicates how age relates to player performance. For instance, there is a positive correlation between age and assists, suggesting that older players tend to have more assists. These insights can influence player selection and strategy.

5. **Age vs. Assists Scatter Plot:**

   The scatter plot illustrating the relationship between age and assists allows managers to identify trends. In this case, it is evident that players below 25 and players above 25 have distinct trends in their assists, indicating potential differences in playing styles and roles.

6. **Age Group Comparison:**

   Separating players into age groups (e.g., <= 25 and > 25) and analyzing their performance metrics (e.g., assists) can help in making targeted decisions based on player age. For instance, older players have a higher mean number of assists, but a lower variance compared to younger players.

7. **Highest Average Points by Team:**

   Identifying the team with the highest average points scored (PTS) can assist team management in recognizing which teams are performing well offensively. This information can be used to evaluate team strategies and player contributions.

8. **PTS Dispersion and Normal Distribution:**

   The analysis of points scored (PTS) includes dispersion visualization and a comparison with a normal distribution. Understanding the distribution of points scored is crucial for assessing the consistency and performance of players in the league.

9. **Average Z-Scores of PTS by Team:**

   Calculating and visualizing the average z-scores of points (PTS) by team allows managers to evaluate how each team performs relative to the league average. This insight can guide team-specific strategies and player acquisitions.

10. **Correlation Matrix and Pair Plots:**

    The correlation matrix and pair plots provide a comprehensive view of how different performance metrics are related. Managers can use these visuals to identify strong correlations or patterns, helping in player evaluation and strategy development.

11. **KDE Plots for Performance Metrics:**

    Kernel Density Estimation (KDE) plots for various performance metrics provide a detailed understanding of their distributions. These plots are valuable for assessing the spread and concentration of performance values among players.