



NBA 96' – 2020 SEASON PLAYER INSIGHTS AND PREDICTIVE MODEL

DATA SCIENCE REPORT

JOEL DOMINGO | JOELDOMINGO117@GMAIL.COM



SUMMARY

This short report details observations pertaining to NBA player data between the '97-'98 and 2019-2020 seasons. The status of the NBA has been dubbed the 'small-ball' era, due to comparable changes of playstyle and the composition of the NBA player population. Based on observations from player data, we have seen shifts in what is considered the optimal individual height, weight, and skillsets of an individual, per position on the court. Downward trends in the average height and weight of drafted players are apparent in the early years of the 21st century, showing a shift towards athleticism and agility, to which the 'small-ball' playstyle revolves around. Trends also display a correlation between geographical origin of players, and their respective positions. Furthermore, we were able to use player statistics to develop a model which can predict ppg (points-per-game) of any specific player using a Linear Regression model, with an accuracy of 89%.



INTRODUCTION

SMALL-BALL ERA

In basketball, small ball is a style of play that sacrifices height, physical strength, and low post offense/defense in favor of a lineup of **smaller** players for speed, agility, and increased scoring (often from the three-point line).

THE SHIFT

The years in which implementation of the small ball style began to be implemented have been debated amongst NBA enthusiasts for some time, and its effectiveness. This report aims to use observable data to indicate when such a shift occurred, and the factors which contribute to such a push.

COMPARISON AND MODELLING

With such disparity in playstyle, technology, and player composition, we want to see if such factors contribute to comparable differences in performance between players of similar caliber. Specifically, we will be looking at player data pertaining to LeBron James and Vince Carter, both of whom were dubbed 'superstars' in their careers and hypothesize reasons for disparities in career performance over the years. Using the same data, as well as box-score statistics, we will be creating a model which will be able to predict a player's points per game using historical data.



Figure 1: LeBron James as part of the LA Lakers in 2021



THE DATA



CONTENT

- Player data including box scores and demographic information.



SOURCE

- Authored by Justinas Cirtautas.
- Found on Kaggle



VOLUME

- Contains 11,145 records.
- Each record contains 23 features

CONTENT

The dataset contains over two decades of data on each player who has been part of an NBA teams' roster. It captures demographic variables such as age, height, weight and place of birth, biographical details like the team player for, draft year, and season. In addition, it has box score statistics such as games played, average number of points, rebounds, assists etc.

SOURCE

The data has been captured using the NBA stats API and compiled by author Justinas Cirtautas. The dataset can be found at the following source:

- <https://www.kaggle.com/justinas/nba-players-data>

VOLUME

The data contains all data for the 1996-97 season all the way to the 2019-2020 season. It contains 11,145 records, each with 23 features. These features include:

- Player name
- Team
- Age
- Player height
- Player weight
- College
- Country
- Draft-year
- Draft-round
- Draft number
- Games played
- Points
- Rebounds
- Assists
- Net rating
- Rebound pct
- Usage percentage
- True shooting percentage
- Assist percentage
- Whether they were drafted or not
- Position



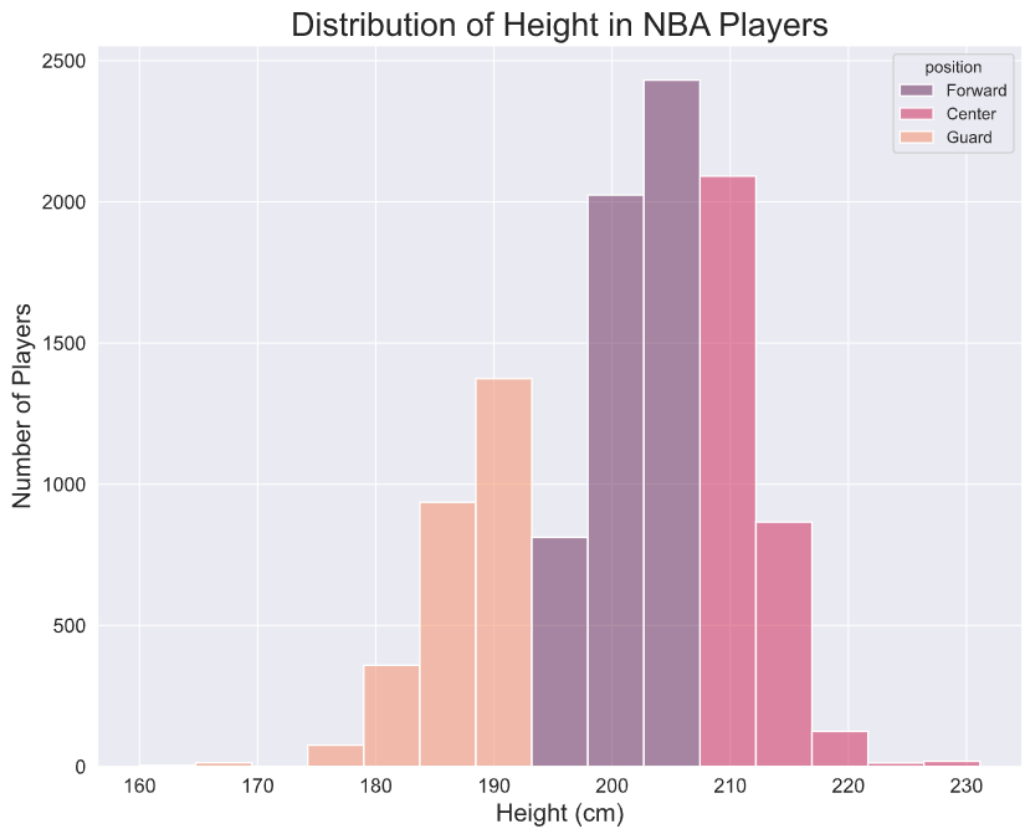
INSIGHTS

Player Population

The total player population since the '96-'97 season includes 2235 players. This includes those who may not have played in an official game but are included in the team roster.

Height and Weight Analysis

Average Height of NBA players	Average Weight of NBA players
200.81cm	100.64 kg



Observations

Looking at the distribution of player height, marked by respective positions, we see that forwards and guards share a similar distribution to each other. Teams opt to recruit the tallest individuals for their respective positions. However, this is not true for centers. We see that under 210 cm is the optimal height based on the composition of recruit height for the center position.



Is there a correlation between the height of a player and their country of origin?

Country	Average Height (cm)	Average Weight (kg)
USA	199.48	98.9
International	206.55	105.64

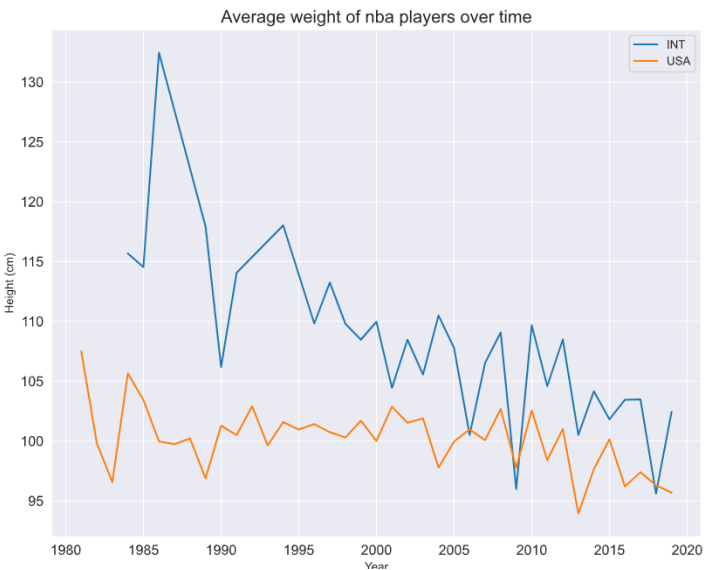
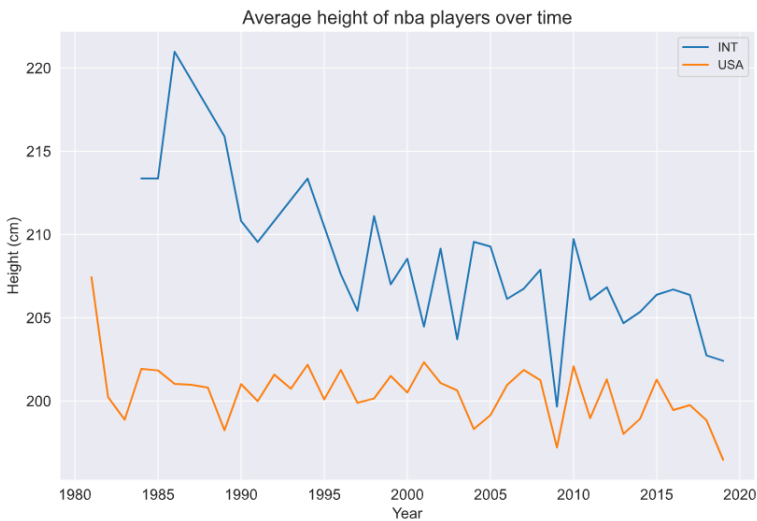
Observations

These figures can reflect several hypotheses.

1. NBA teams may have looked to recruit slightly bigger players internationally. The reason for this may be due to the performance of highly skilled international big-men in the last 2 decades such as Dirk Nowitzki (Germany), Nikola Jokic (Serbia) and Marc and Pau Gasol (Spain), all of which have been leaders on championship teams.
2. International players are not as biologically athletic as players born in the USA. Such players are known to be the most athletic in the league, with athletic features such as speed, lateral quickness and vertical leap being major factors in drafting a player.
3. Mix of both, or potential other reasons.
4. Neither, and the comparison difference is insignificant.

We can further observe these differences by looking at the average height of international and USA players over the years. Looking at the figures below, we see that many of the hypothesis' may have been true or false at different points in time.

- Larger big men may have been sought after internationally more often in the 90's era, since we see a big disparity between average height in that decade between USA born players and international players.
- We see a downward trend of heights and weights as time goes on, this may indicate that the NBA is moving towards a different style of play in the early 2000's in both USA born and international players. Instead of big men dominating the league, athleticism, game IQ, ball-handling, shooting, and speed are more valued, attributes which define 'small-ball'.



Changes in playstyle in the NBA

By adding positions to the dataset, we can see if these trends have affected the style of play on a position basis.

There are 3 categories for positions in the NBA, usually dictated by height due to different mobility of different body-types. These are as follows:

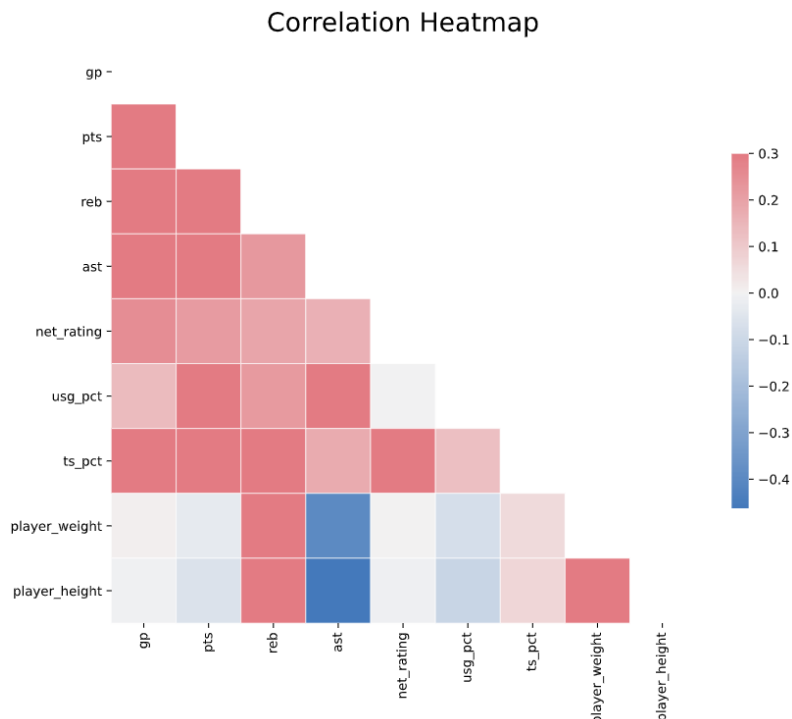
- **Bigs/Centers:** They often play the center position, as they are the tallest. Their advantage in height means their main role is for rebounding, rim-protection, and blocking.
- **Forwards:** These players are not as tall as centers but are usually bigger and more physical than guards. They can score well and are good defensive players.
- **Guards:** These are smaller players regarding height and weight, and often see the ball in their hand for most of the time. Offense is run by them and due to their size, they are usually the fastest.

Within the data, we have separated positions based off height.

- ≥ 206 cm: Center
- ≤ 195 cm: Guard
- > 195 & < 206 cm: Forward

Correlation Observations

From the correlation heatmap below, we observe that height and weight is highly correlated with number of rebounds. This suggests that centers and forwards have advantages when it comes to crashing the boards and getting rebounds. However, we see that both assists, and usage percentage are negatively correlated (usage percentage refers to the '*percentage of team plays used by a player while he was on the floor*' [www.basketball-reference.com])). This makes sense because it is often the guards who are mostly in possession of the ball during offensive plays.

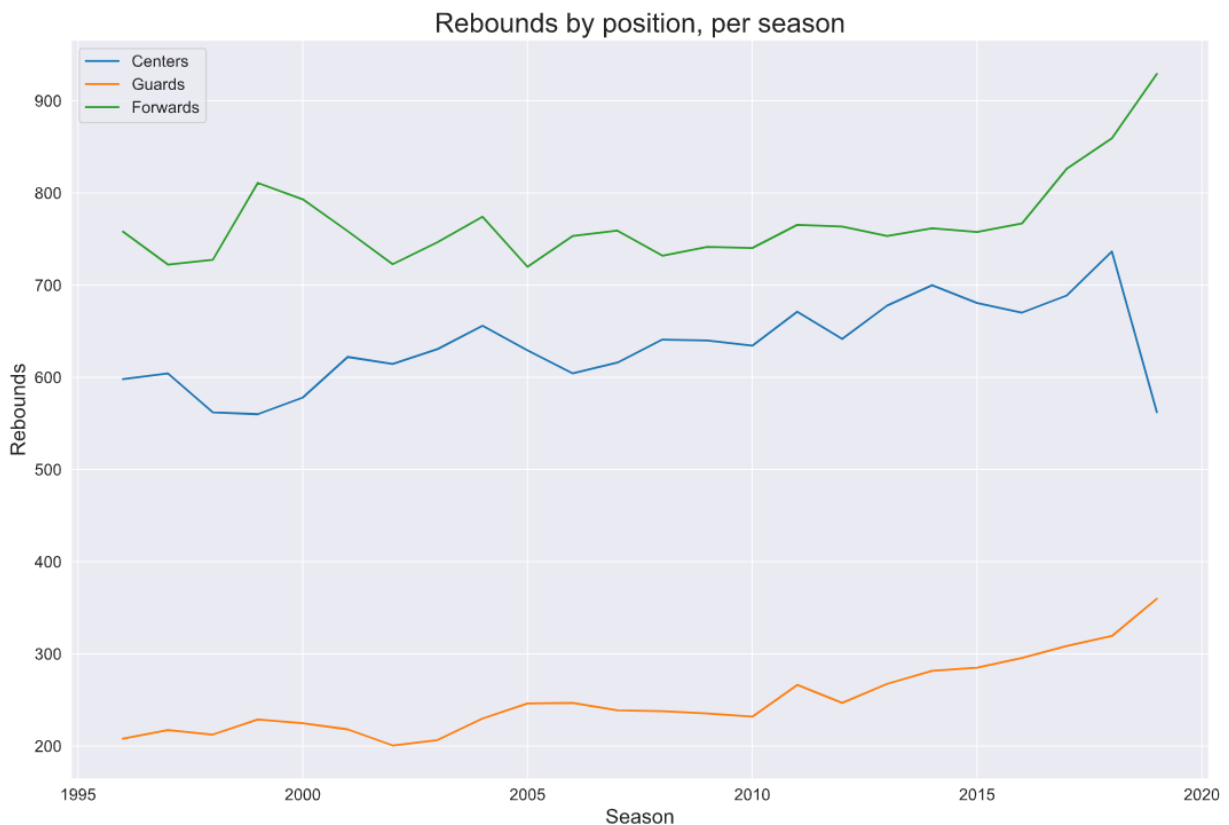


Position Observations - Rebounds

From the trends shown below, we can see a steep decline in total rebounds by centers (tallest players) in recent years. At the same time, we see a significant increase in total rebounds by forwards and guards in a similar time frame. This may suggest that the smaller players are becoming athletic enough to contest the rebounds against the larger players, and in the same way, larger players are not rushing for rebounds in the same way they used to.

Furthermore, forwards are having an easier time rebounding. This may be explained by our our definition of positions in our data, this may indicate that the average height of people who play center is reducing. Players who are officially classed as centers on the NBA roster may not be required to be as tall as center players in previous generations.

We see a clear example of this in the 2020 season. In January 2020, the Houston Rockets took to the extreme, defeating many teams, including the Dallas Mavericks at 128-121 without a player listed as taller than 6ft 6in. According to records, it was the first time an NBA team had done so since 1963, demonstrating the evolution of playstyle in the modern era towards small ball styles of play.



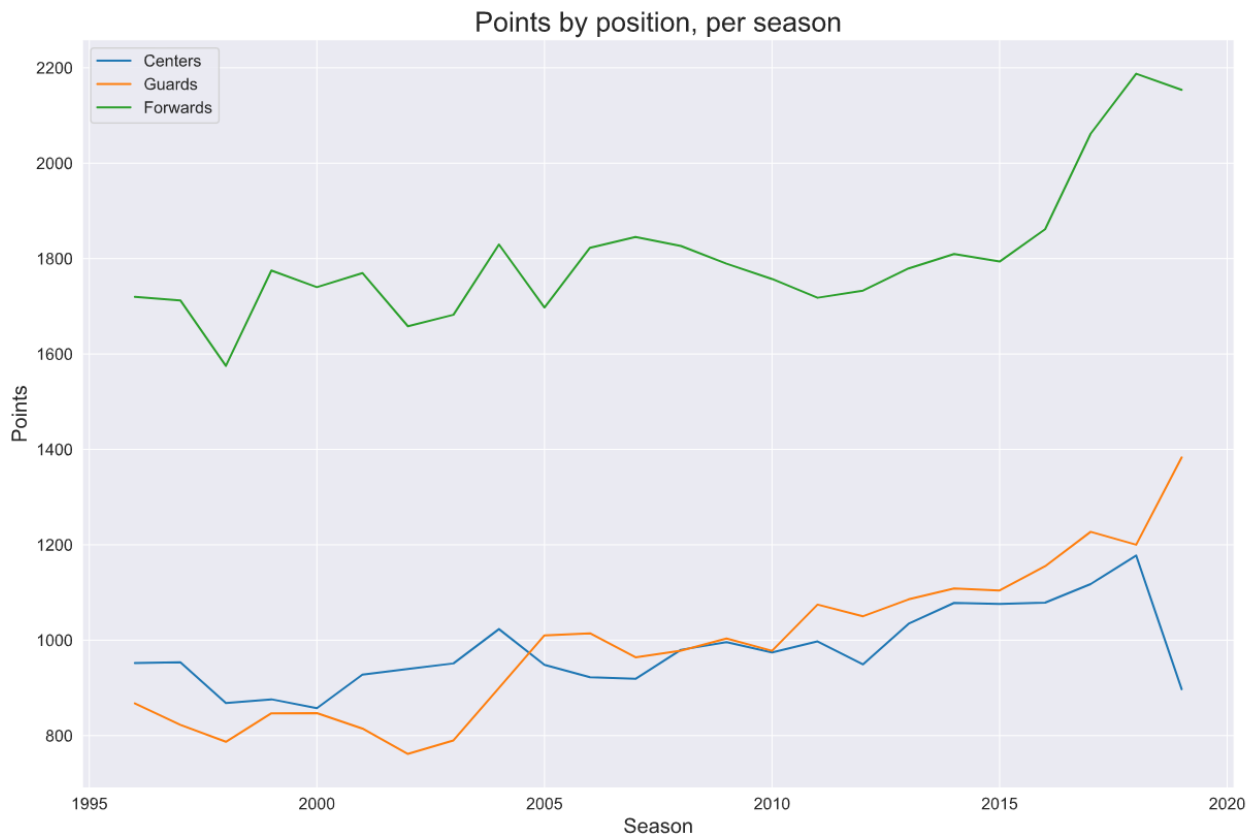


Position Observations - Scoring

Looking at points, we see that during the 90's and early 2000's, offence revolved around getting the ball to the centers for them to dominate the paint and get easier points. However, we see a change in trends from 2005 and onwards, where the points for guards see an upwards trend, eclipsing the points for big men. This indicates that offense is starting to move towards being more guard focused. i.e. ranged shooting, isolation basketball etc.

Commenting on the large differential between centers/guards and forwards, this may also be a result of our sub-setting of position. As previously mentioned, it may not be a true indication of position. For example, Australian NBA star Ben Simmons is listed at 208cm, yet plays at the guard position. Due to the rising importance and prevalence of athleticism as a factor in recruiting, we can assume that many optimal guards are taller and therefore be listed by our definition into the forward position.

In our early code, we established that a forward was between 195cm and 206 cm. However, the most prolific guards are known to be in this forward height range, due to being bigger in size, yet having the skillset, speed, and agility of typical guards, thus having a physical advantage over their smaller matchups. This can be exemplified in the cases of Michael Jordan and Kobe Bryant, both of whom played guard positions, yet are listed at 198cm. So, in this case, they are listed as forwards in our data.

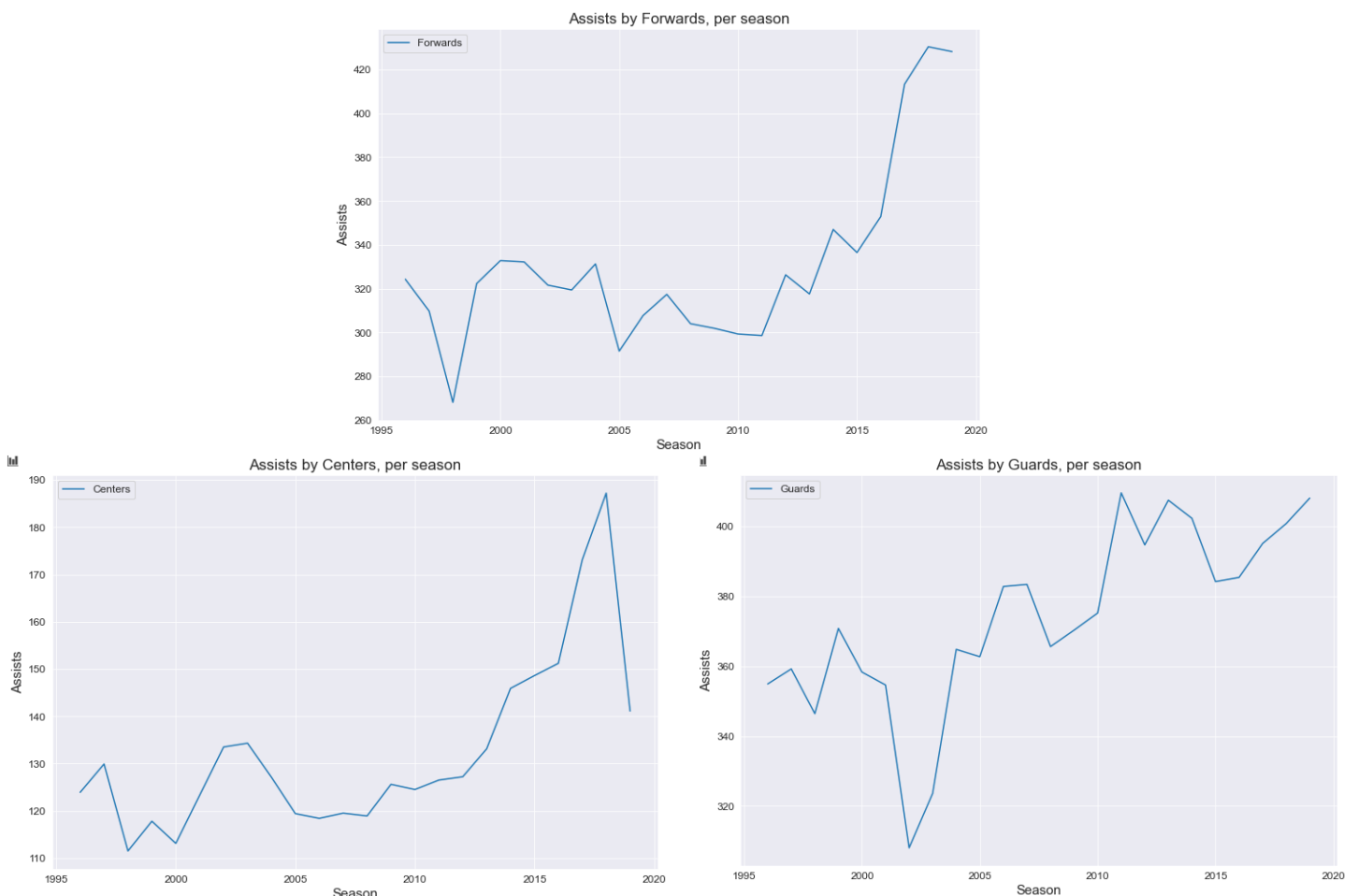


Position Observations - Assists

For centres, we see that they are passing more often. This further supports that NBA teams are looking to run plays and score off guards and their shooting ability. This is especially evident after 2010 after Miami Heat Coach employed this style during the 2011-12 season, labelling it 'pace and space'. After its success, NBA teams evaluated the efficiency of this style to be super-effective and began employing similar tactics.

This effect is most noticeable in the sharp increase after 2015, after the Golden State Warriors (led by point guard Steph Curry, arguably the best shooter of all time), employed this to an even greater extent, relying on shooting and spacing the floor (rendering big men ineffective). The large dip in the last few years can be attributed to the increasing amount of high scoring players who play guard and forward positions being taller and more athletic (e.g. Kevin Durant (208cm), Ben Simmons (208cm), Giannis Antetokounmpo (211cm)) but are classified in our data as centres.

Overall, we see that assists have sharply gone up 2015 onwards, even by guards. This further emphasises shifting away from isolation playstyle, and towards small-ball playstyle, involving drawing in players and passing it to the open man for an open shot (from 3pt range). Assists growing from forwards also suggests that not only are the passing to guards, but to center-classified players. This indicates that taller players are adjusting to playstyle and learning how to shoot the ball from range.



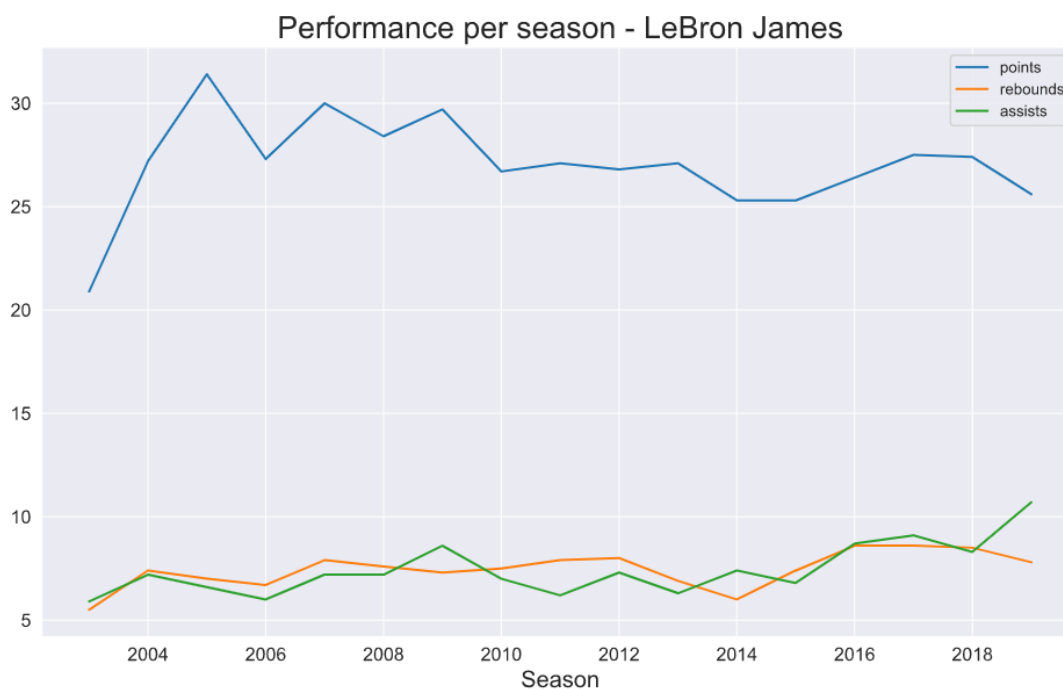
CASE STUDY: LEBRON JAMES

Context

LeBron James has been known as one of the best players of all time. Part of this stems from the fact that he has been indestructible, having been playing in the NBA since 2003 (16 seasons). Using the data, we're able to track his historic performance and observe if there are any signs of diminishing performance.

Career Performance (gathered from dataset)

Points per season	Rebounds per season	Assists per season
27.06	7.45	7.45

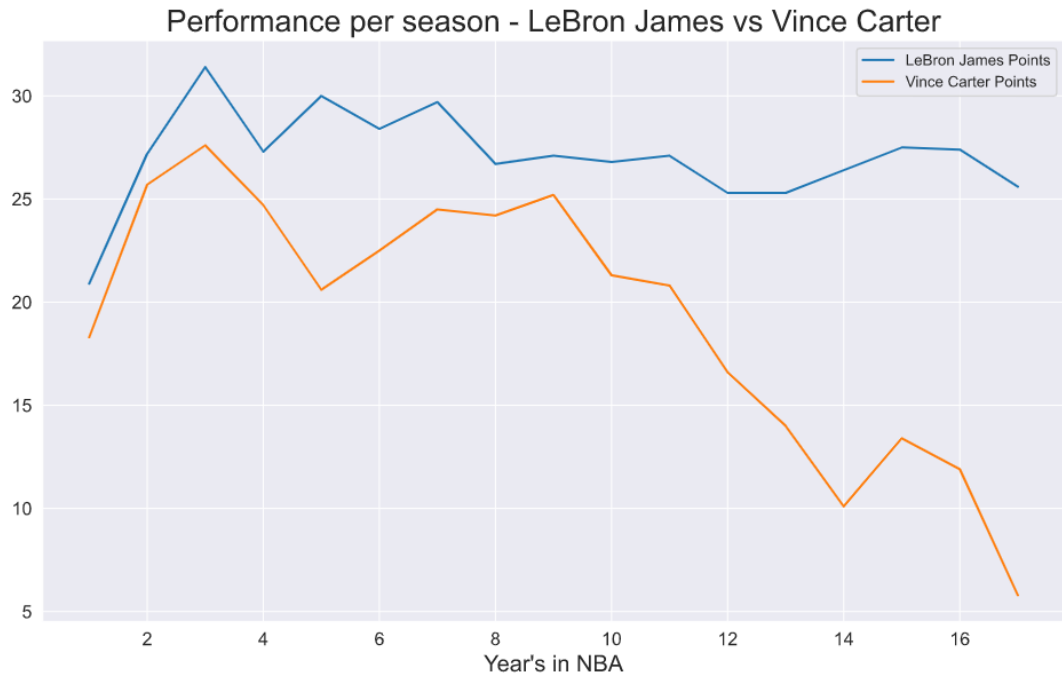


Observations

We can see that as soon as LeBron entered the league in 2003, his performance was exceptional. As a rookie, he averaged more than 20 points a game. This continued for a number of years as he developed. Furthermore, the data shows that he was able to maintain a reputation as an all-round player, with high assists and rebounds. Throughout his career (16 seasons), we see that he has been able to maintain these per-game averages. To put this into perspective, let's look at a similar case where someone was in the league for an extended period: **Vince Carter**, who was in the league for 22 seasons (drafted in 1998 – ended career 2020 season).



LeBron James vs Vince Carter Career Performance

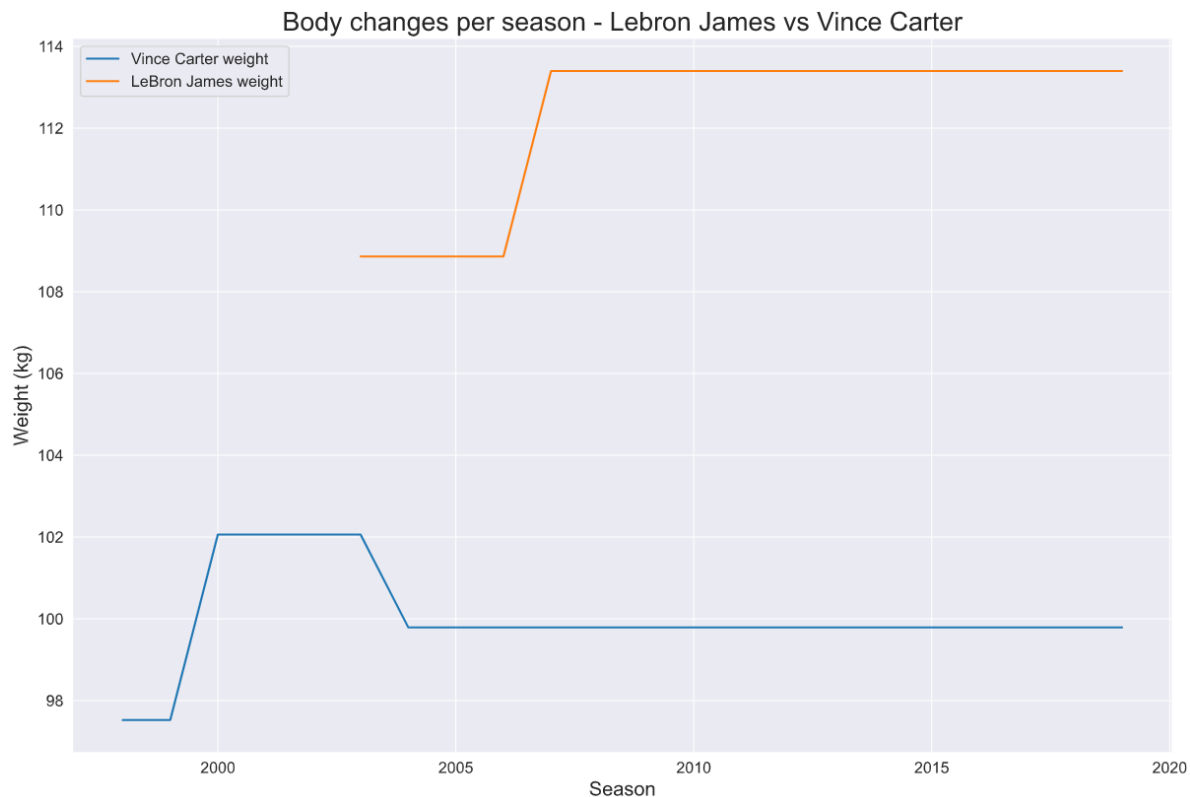


Observations

When observing the career progression in terms of points, we see that Carter’s performance record was comparable to LeBron’s earlier career, both averaging near 20 points in the rookie season, climbing by similar amounts in subsequent seasons. However, Vince Carter’s (coined the best dunker of all time) performance seemed to have dropped after his 9th season in the NBA, contrasting to LeBron, who in his 9th season, had shown no significant decline in his performance.

Hypothesis: How has LeBron been able to maintain his performance over his career?

One popular hypothesis has been his ability to maintain his body and adapt it through intense training and maintenance routines. In 2018, a report by ESPN and Business Insider revealed that LeBron reportedly spends a staggering \$1.5 million per year to take care of his body (source: <https://www.businessinsider.com.au/how-lebron-james-spends-money-body-care-2018-7?r=US&IR=T>).



Observations

We can see that both Vince and LeBron adapted their bodies to face the competition in the NBA, previously being college and high school students respectively. Both players gain a considerable amount of weight for an athlete, most likely to put on muscle to transition from tertiary competition to facing fully-grown athletes at a professional level. This is often the case with all NBA players. However, we see that after Vince Carter's peak performance years, his weight dropped, and maintained until the end of his career. This contrasts to LeBron, who has been able to maintain a higher density physique to this date. This further supports the hypothesis, training to counter the fragility that comes with age. This may explain how he is able to play basketball at a high level for so many years.

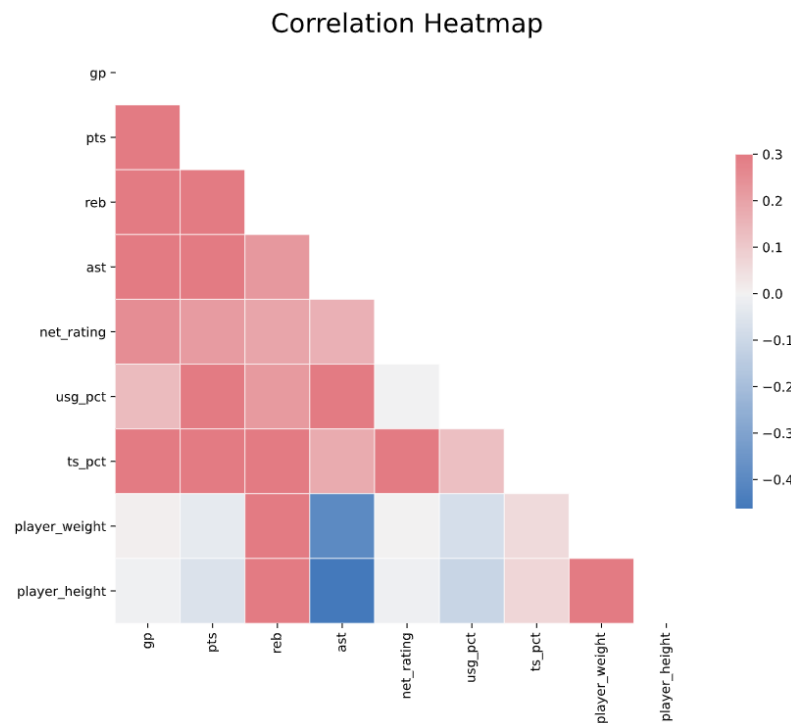
PREDICTIVE MODELING: POINTS PER GAME

Goal

Using box-scores from the provided dataset, we want to create a machine learning model which can predict LeBron's points per game for a season. By adjusting the learning data to a per-game basis, the model should be able to predict LeBron's statistics for his next few games. However, since the per-game data is not in the scope of the dataset, we will limit it to season average.

Coefficients

Looking at the correlation heatmap, the best numerical predictors for ppg include rebounds, assists, net rating, usage percentage and true shooting percentage.

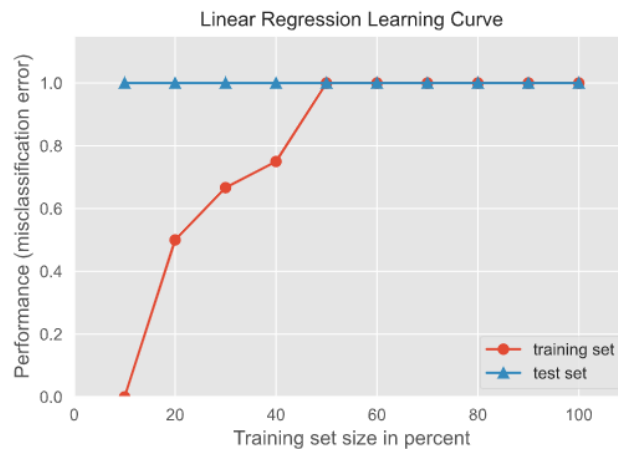


Feature	Coefficient
Gp	0.54
Reb	0.62
Ast	0.65
Usg_pct	0.63
Ts_pct	0.38
Ast_pct	0.32

Model

For our model, we will be using the Linear Regression model from sklearn. The model will undergo a train test split of 33%, with a random state of 7.

Learning Curve



We see in the learning cycle, the model begins with high amounts of variance, and quickly reduces it as it learns the data. In the end, our model produced an r2 score of 0.89.

Model	Score
Linear Regression	0.89

Use Case

To provide a practical example of our model, we will predict LeBron's current season points per game using his average box scores from 37 (total) games played this season. The following stats can be found on <https://www.basketball-reference.com/players/i/jamesle01.html> :

Feature	Statistic
Rebounds	7.3
Assists	7.8
Usage Percentage	0.317
True Shooting Percentage	0.595
Assist Percentage	0.392

Prediction

Y predict	Value
Points per game	25.4

LeBron's predicted ppg for this season is 25.4 ppg. As of the 37 games played this season, his true ppg is 25.6, that's only a 0.2-point discrepancy.



CONCLUSIONS

From the analysis and our modelling, the following conclusions come forth:

- The NBA is becoming more athletic.
- Play style is becoming less big-men oriented (small-ball)
- This shift to small-ball began being tested in team-systems in the mid 2000's but were more consistently implemented after 2010.
- More teams implemented the style of system after viewing the success of Miami and Golden State in 2011 and 2015 respectively.
- Players are becoming more skilled and relying less on pure physicality, and more on shooting, ball-handling, etc.
- Trends indicate that the best future NBA players will be able to play multiple positions.
- Bigger players with the ability to also shoot from range will dominate the league.
- These attributes are what NBA scouts are looking for when drafting.
- LeBron James will continue to dominate the league for the next few years, showing no significant decline in performance in recent years.
- Players should focus their training on body maintenance to increase longevity in the league.
- LeBron's predicted ppg for the 2021 season is 25.4 ppg.