

PSTAT 100 Final Project

Lebron James: The New Michael Jordan?

Raymond Lee, Brandon Lee, Matthew Lee (3084852), Matthew Lee (3722287)

Author contributions

Raymond tidied the data, proposed questions to explore, and created graphics.

Brandon helped with coding, creating graphics, and writing the report.

Matthew Lee (3084852) helped with creating graphics, writing captions, and coding.

Matthew Lee (3722287) proposed questions to explore, created graphics, and helped with coding.

Abstract

This project looks at data from LeBron James's career in the NBA over the last 19 years and studies how his performance varies depending on various factors like his age and team. The aim of this project was to find out how well LeBron James performs against different teams in the NBA and to identify trends in LeBron James' performance throughout his career. Utilizing a correlation heatmap, bar plots, and scatter plots, we found that LeBron James performs extremely well against essentially every team in the NBA and that he performed the best while playing for Miami. We also found that there are diminishing returns for his performance if he plays for too long.

Introduction

Background

The NBA is home to many superstar athletes, but one player seems to stand out from the rest: LeBron James. As of writing this report in 2022, LeBron James has been in the NBA for 19 years and has achieved countless accolades. His long list of accomplishments includes: four NBA championships, four NBA MVP awards, four NBA Finals MVP awards, and two Olympic gold medals. He is currently also the highest in playoff points, the second highest career points, and the seventh highest in career assists. Although his stats and achievements indicate that he is one of the greatest basketball players of all time, it can still be disorienting and difficult to truly understand his performance throughout his career by examining his stats due to his long career and extensive accolades. Therefore, we wanted to

study his career and present it in a more digestible and insightful format to better understand LeBron.

Aims

We aimed to answer a few questions regarding LeBron's stats:

- How well does LeBron perform against each team in the NBA? Which team does he perform the best and worst against?
- Which of LeBron's stats contribute the most to winning a regular season game?
- Can we identify any trends in LeBron's performance?

By answering these questions, we hoped to demonstrate how valuable LeBron is to his team and gain insight into his career. We also hoped to discover whether age has managed to slow LeBron down at all. This analysis will also help us understand generally how a player can single-handedly impact a game and how certain stats can be more valuable in winning.

Materials and Methods

Datasets

The data consist of game-by-game stats throughout LeBron James's career. The data were obtained from <https://www.basketball-reference.com/>. According to SportRadar, the official statistics provider of the NBA, the data were collected by on-venue scouts at each game. The population consists of all regular season games that take place throughout LeBron James's career. Our sampling frame consists of all games that LeBron James played in from 2003 to 2022. The sampling mechanism is administrative data because the sample is the same as the sampling frame (census of the frame). Therefore, we do not have a scope of inference because the data do not represent a broader group, especially since we are only dealing with NBA games in which LeBron James was present. We most likely cannot extrapolate any conclusions for other players.

After removing games in which LeBron did not play and games that had missing values, the data set had 1230 observations and 43 variables. Of these 1230 games, 806 games were wins, and 424 games were losses. 468 of the wins were at home, and 338 of the wins were at away. 153 of the losses were at home, and 271 of the losses were at away. He has more wins and less losses at home, so this may be evidence of home court advantage.

Below is a description of all of the variables in the data set:

| Name | Variable description | Type |
|------|-------------------------------------|---------|
| g | number of game played in the season | numeric |
| year | year of game played | numeric |

| Name | Variable description | Type |
|-------------------|--|-----------|
| month | month of game played | numeric |
| age | age at which the game was played | numeric |
| location | court that the game was played on | character |
| team | team played for | character |
| opp | opponent team | character |
| diff | point differential at the end of the game | numeric |
| min | minutes played in game | numeric |
| sec | seconds played in game | numeric |
| fg | field goals | numeric |
| fga | field goal attempts | numeric |
| fg_percent | ratio of field goals made to field goals attempted | numeric |
| three_pts | 3-point field goals | numeric |
| three_pa | 3-point field goal attempts | numeric |
| three_pts_percent | ratio of 3-point field goals made to 3-point field goals attempted | numeric |
| ft | free throws | numeric |
| fta | free throw attempts | numeric |
| ft_percent | ratio of free throws made to free throws attempted | numeric |
| orb | rebounds grabbed on offense in game played, maintaining possession of the ball | numeric |
| drb | rebounds grabbed on defense in game played, gaining possession of the ball | numeric |
| trb | total rebounds | numeric |
| ast | passes to teammates that directly lead to a field goal or goaltending | numeric |
| stl | steals | numeric |
| blk | blocks | numeric |
| tov | loss of possession of the ball to the opposing team | numeric |
| pf | personal fouls | numeric |
| pts | points scored | numeric |
| gmsc | Game Score | numeric |
| pm | plus-minus | numeric |
| ts_percent | true shooting percentage | numeric |
| efg_percent | effective field goal percentage | numeric |
| orb_percent | offensive rebound percentage | numeric |
| drb_percent | defensive rebound percentage | numeric |
| trb_percent | total rebound percentage | numeric |
| ast_percent | assist percentage | numeric |

| Name | Variable description | Type |
|-------------|----------------------|---------|
| stl_percent | steal percentage | numeric |
| blk_percent | block percentage | numeric |
| tov_percent | turnover percentage | numeric |
| usg_percent | usage percentage | numeric |
| ortg | offensive rating | numeric |
| drtg | defensive rating | numeric |
| bpm | box plus-minus | numeric |

Here are some example rows of our data set. We split up the data set into multiple tables to show every column.

| | g | year | month | age | team | location | opp | diff | min | sec |
|---|-----|--------|-------|------|------|----------|-----|-------|------|------|
| 0 | 1.0 | 2003.0 | 10.0 | 18.0 | CLE | Away | SAC | -14.0 | 42.0 | 50.0 |
| 1 | 2.0 | 2003.0 | 10.0 | 18.0 | CLE | Away | PHO | -9.0 | 40.0 | 21.0 |
| 2 | 3.0 | 2003.0 | 11.0 | 18.0 | CLE | Away | POR | -19.0 | 39.0 | 10.0 |
| 3 | 4.0 | 2003.0 | 11.0 | 18.0 | CLE | Home | DEN | -4.0 | 41.0 | 6.0 |
| 4 | 5.0 | 2003.0 | 11.0 | 18.0 | CLE | Away | IND | -1.0 | 43.0 | 44.0 |

| | fg | fga | fg_percent | three_pts | three_pa | three_pts_percent | ft | fta | ft_percent |
|---|------|------|------------|-----------|----------|-------------------|-----|-----|------------|
| 0 | 12.0 | 20.0 | 0.600 | 0.0 | 2.0 | 0.0 | 1.0 | 3.0 | 0.333 |
| 1 | 8.0 | 17.0 | 0.471 | 1.0 | 5.0 | 0.2 | 4.0 | 7.0 | 0.571 |
| 2 | 3.0 | 12.0 | 0.250 | 0.0 | 1.0 | 0.0 | 2.0 | 2.0 | 1.000 |
| 3 | 3.0 | 11.0 | 0.273 | 0.0 | 2.0 | 0.0 | 1.0 | 1.0 | 1.000 |
| 4 | 8.0 | 18.0 | 0.444 | 1.0 | 2.0 | 0.5 | 6.0 | 7.0 | 0.857 |

| | pts | orb | drb | trb | ast | stl | blk | tov | pf | gmsc |
|---|------|-----|------|------|-----|-----|-----|-----|-----|------|
| 0 | 25.0 | 2.0 | 4.0 | 6.0 | 9.0 | 4.0 | 0.0 | 2.0 | 3.0 | 24.7 |
| 1 | 21.0 | 2.0 | 10.0 | 12.0 | 8.0 | 1.0 | 0.0 | 7.0 | 1.0 | 14.7 |
| 2 | 8.0 | 0.0 | 4.0 | 4.0 | 6.0 | 2.0 | 0.0 | 2.0 | 3.0 | 5.0 |
| 3 | 7.0 | 2.0 | 9.0 | 11.0 | 7.0 | 2.0 | 3.0 | 2.0 | 1.0 | 11.2 |
| 4 | 23.0 | 0.0 | 5.0 | 5.0 | 3.0 | 0.0 | 0.0 | 7.0 | 2.0 | 9.0 |

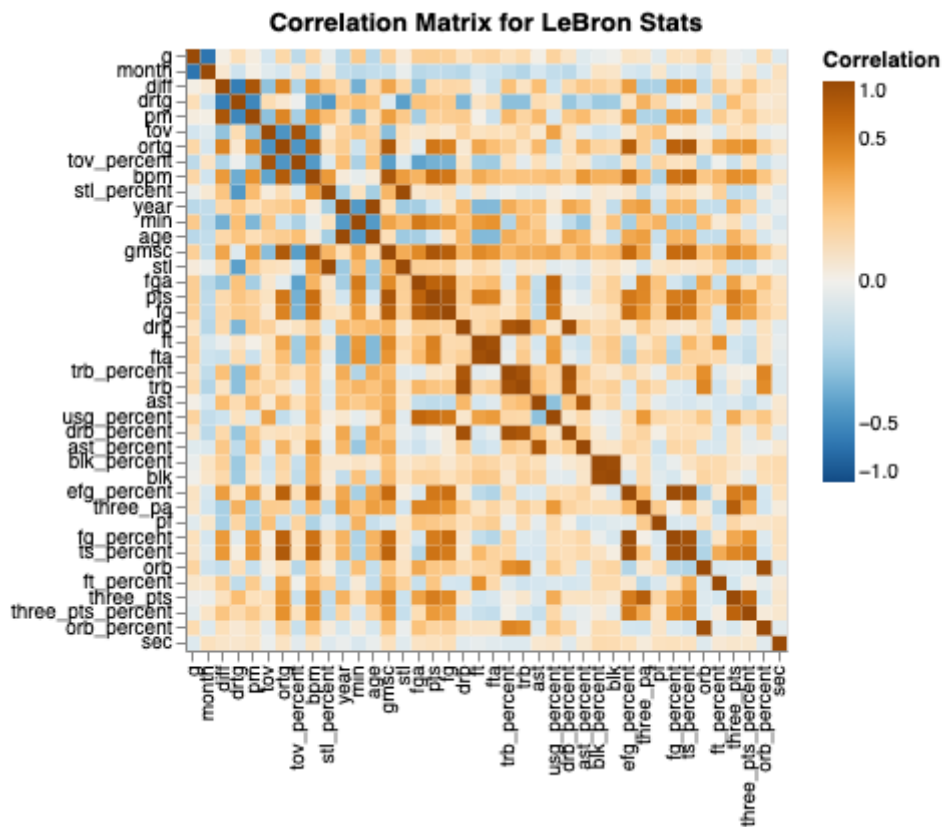
| | ts_percent | efg_percent | orb_percent | drb_percent | trb_percent | ast_percent | stl_percent | blk_percent | tov_percent | usg_percent |
|---|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 0 | 0.586 | 0.600 | 6.0 | 10.9 | 8.6 | 39.1 | 5.0 | 0.0 | 8.6 | 25.6 |
| 1 | 0.523 | 0.500 | 5.6 | 22.5 | 14.9 | 33.9 | 1.2 | 0.0 | 25.8 | 28.5 |
| 2 | 0.311 | 0.250 | 0.0 | 13.3 | 6.6 | 23.6 | 2.7 | 0.0 | 13.4 | 18.2 |
| 3 | 0.306 | 0.273 | 4.9 | 21.1 | 13.1 | 27.8 | 2.5 | 5.1 | 14.9 | 13.7 |
| 4 | 0.546 | 0.472 | 0.0 | 12.4 | 6.3 | 12.0 | 0.0 | 0.0 | 24.9 | 30.7 |

| | ortg | drtg | bpm | pm |
|---|-------|-------|------|-------|
| 0 | 123.0 | 112.0 | 10.8 | -9.0 |
| 1 | 89.0 | 97.0 | 1.0 | -3.0 |
| 2 | 74.0 | 112.0 | -6.5 | -21.0 |
| 3 | 85.0 | 95.0 | 2.0 | -3.0 |
| 4 | 85.0 | 112.0 | -7.3 | -7.0 |

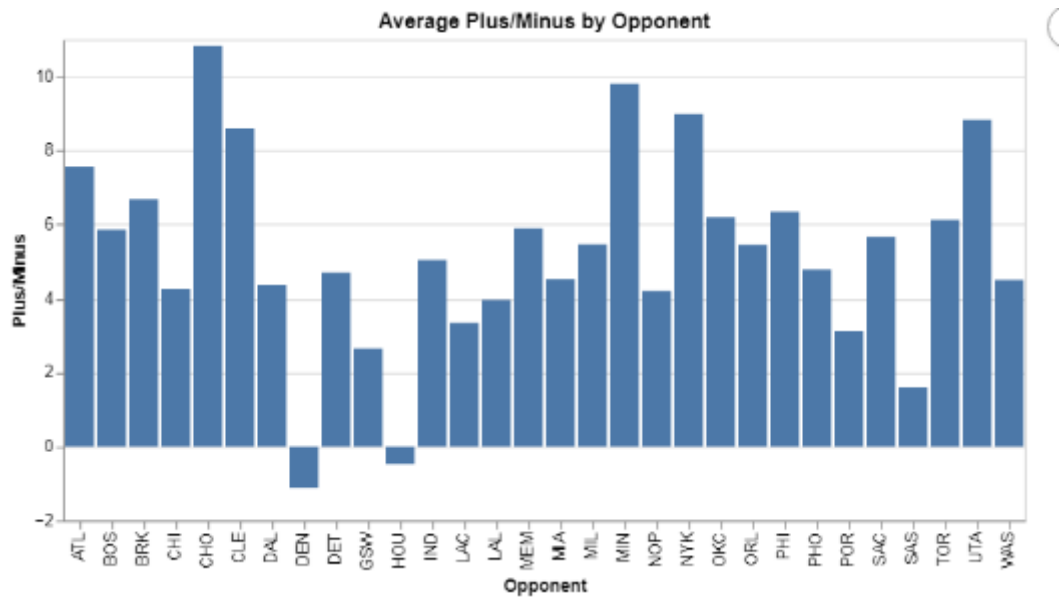
Methods

We first tidied the data and then created graphics that displayed LeBron's performance in a more digestible manner. These graphics involved variables that are believed to be most influential in determining the outcome of a game. These include age, minutes, plus/minus, true shooting percentage, location, and Game Score. We created tables, bar plots, and scatter plots to present our findings. Bar plots allow us to easily compare values and scatter plots allow us to easily identify any trends across time or values of other variables. We grouped the data either by year, team, or opponent to provide more context to the data. This also allows us to clearly visualize any trends in his performance throughout his career. We also created a correlation matrix to see which stats are most correlated with each other and which stats are most correlated with winning a game.

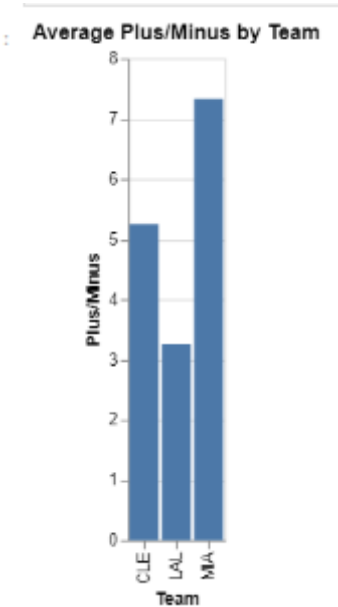
Results



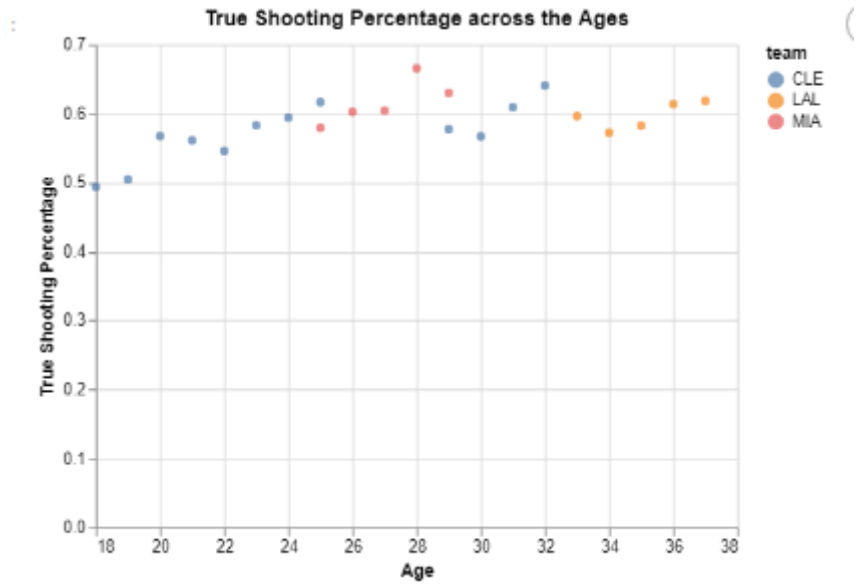
A heatmap of all numeric variables. Point differential is moderately positively correlated with true shooting percentage.



Bar plot of average point differential against each opponent when LeBron is in the game. His team is ahead the most against Charlotte and behind the most against Denver.

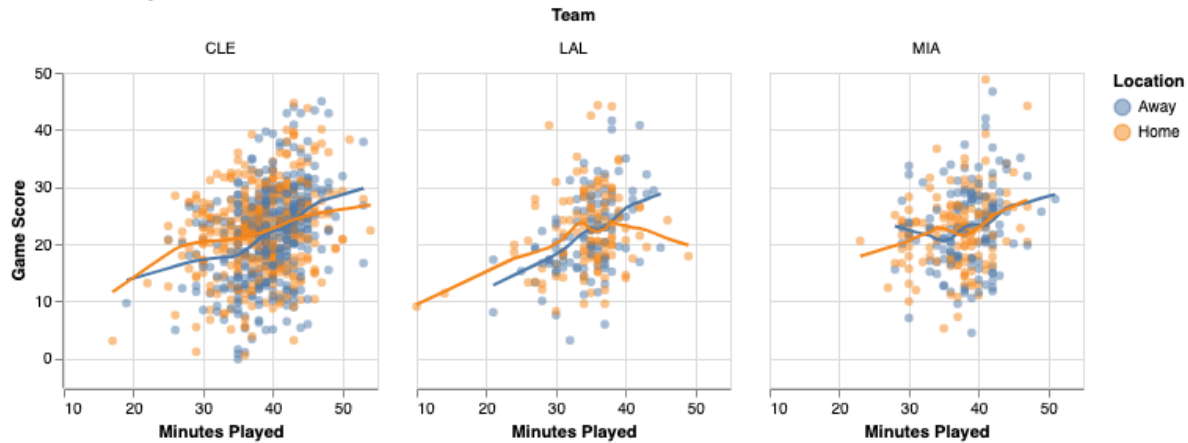


Bar plot of average point differential when LeBron is in the game for each team he has played for. The lead that his team has when he is playing was highest during his time in Miami (7.32 points).



Scatter plot of LeBron's true shooting percentage against age. LeBron's highest true shooting percentage was .665 when he was on the Miami Heat at age 28.

Game Score by Minutes



Scatter plot of LeBron's Game Score for each game against minutes played with a LOESS line. It is also split into each team he has played for, and the points are colored by location. For Cleveland, his Game Score drops after playing about 43 minutes at home games. For the Los Angeles Lakers, his Game Score drops after playing about 37 minutes at home games. For Miami, there does not seem to be a significant difference in Game Score across minutes played or location.

Discussion

First, LeBron and his team have a lead against most opponents when he is playing. Even when they are behind, they are only down by approximately 1 point or less on average. Essentially, his presence is associated with winning against every team. Second, the point differential while LeBron is on the court was highest when he played for Miami. LeBron's

time in Miami is considered to be his prime years, so this finding makes sense. Next, we have concluded that age does not seem to have a significant effect on LeBron's true shooting percentage, which is a measure of scoring efficiency that takes 2-pointers, 3-pointers, and free throws into consideration. His scoring efficiency has remained quite high over the years. Even at age 37, his true shooting percentage was .618, which is higher than much of his younger years. Also, the correlation heatmap showed that point differential is moderately positively correlated with true shooting percentage, which means that the lead increases as LeBron scores more efficiently.

While playing in home games for Cleveland and the Los Angeles Lakers, LeBron's Game Score noticeably drops when he plays past approximately 43 minutes and 37 minutes respectively. There seems to be a point of diminishing returns for his performance in home games as he plays more minutes. However, his Game Score still increases for away games for Cleveland and the Los Angeles Lakers past these amounts of playing time. His Game Score does not differ significantly between home and away games as he plays more minutes for Miami. Overall, he seems to perform better at away games even as he plays longer.

There are some caveats to address. First, observations with missing values were omitted from the data set. This may have exaggerated (or even underestimated) some findings. Second, the data set does not include some other useful and interesting player stats such as player efficiency rating (PER) and data regarding shot charts. PER is a per-minute rating of a player's performance that adds up all of a player's positive stats and subtracts all of the negative stats. It allows us to compare players despite a difference in minutes played between them. Shot charts provide detailed information on the shot that was attempted. This includes type of shot, area, and distance. Analyzing these variables in addition to the ones in this project may have provided more insight into LeBron's performance and value. Finally, we believe that our findings were eye-opening only because LeBron has been the best and most influential player on his teams (and one of the best in the league). Our conclusions may have been more neutral if we had used the data of a player who does not have as big of a role on his team.

Future analyses or projects should attempt to compare LeBron's stats to other players of his caliber. This would provide insight into exactly how well he is performing and how he compares to the greatest players of all time. Discovering why LeBron and his team performs relatively worse against Denver and Houston would also be an interesting follow-up. It may also be interesting to analyze playoff game data and compare them with the current findings. Playoff games tend to be more competitive and volatile, so the findings might be more intriguing.