LeBron James' Performance and Efficiency by Minutes Played Per Game

A statistical analysis of LeBron's in-game performance when playing high, low, and average minutes per game.

Click here to view the full Github repository with data sources and Jupyter Notebook

By Zach Cornelison

This project serves as basic exploratory data analysis of LeBron James' game performance as it relates to his minutes played. The main question I am seeking to answer is, "Does LeBron play better or worse when he plays more minutes?" Through data collection and cleaning of advanced statistics from multiple sources, I used analytics and visualizations to find an answer. The results led me to the conclusion that on a whole-game basis, LeBron James plays better when playing more minutes. However, when breaking the statistical whole-game results down to per-minute results, he actually plays worse per-minute as more minutes are played. A personal goal of this project was to demonstrate my developing skills in Pandas, Jupyter Notebooks, Data Analysis, and Data Visualization.

Problem

With LeBron James in his 17th season of his NBA career, his health and stamina are a key concern of his coaches and teammates. Keeping a player free of injury while maximizing their on-court potential is essential. In an ideal world, LeBron would play all 48 minutes of the game without rest, since his knowledge of the game and ability to perform almost always exceed all other nine players on the court. However, players sometimes have bad games or choke under pressure. Coaches must make important decisions regarding how much to rest a player without it being detrimental to the team's success. This project seeks to shed light on this problem of how many minutes LeBron should play by analyzing his performance. I wanted to see if LeBron's minutes played has an outstanding impact on his performance by looking at performance metrics fluctuations when playing a below average, average, or above average amount of minutes. These findings would be of interest to any basketball fan who is curious how LeBron's time on the court affects his play.

Data

To answer my question, I gathered basic and advanced performance data on every game LeBron has played in throughout his NBA career. The basic performance data comes from the game logs for Lebron James on basketball-reference.com. This source contains game log data for each game of each season LeBron has played in, making retrieval relatively easy. Once in a text file, I used Pandas to read the dataframe into my Jupyter Notebook to begin cleaning. This dataframe contained 1303 rows--each representing one game--and 30 columns, each representing various metrics of the game. Some of the key metrics included minutes played, points, assists, rebounds (offensive and defensive), turnovers, plus/minus, and gamescore. A glossary defining all statistics used will be provided at the end of this project.

For the advanced performance data, I used stats.nba.com. I first downloaded the data into an excel file. I then cleaned the data using Tableau Prep Builder and joined it with the basic performance data to obtain one full dataset with all attributes for each game. Before joining, this dataset included 1303 rows (one for each game) and 18 columns. The metrics included in this dataset (pre-join) included offensive and defensive ratings, assist metrics, shooting metrics, and player impact metrics, to name a few. The final dataframe, including basic and advanced metrics and after dropping unnecessary columns, was 1303 rows by 45 columns. *Note the final dataset was not used until after initial cleaning was done to the basic stats dataset in Python and the advanced stat dataset in Tableau.*

Data Cleaning and Preprocessing

I first began by cleaning the basic stats dataset. There were a few attributes that were not needed for the purpose of this analysis. Stats on each game like Rank (RK--the numbered game of each season), Age (Lebron's age for the given game), and Games Started (GS-1 if started, 0 if benched) were not relevant, so I dropped each of these columns from my dataframe.

After checking the data types of each column, I noticed almost all of the columns were labeled as “objects" rather than “floats” or “ints.” I started by converting the “Date” column to the appropriate datetime type in order to sort my dataframe in ascending order, from LeBron’s first game to his most recent.

The “Points” column had some string values indicating games in which LeBron did not play, causing the column to be of data type “object” instead of “int.” These values were “Inactive,” “Did Not Play,” “Did Not Dress,” and “Not With Team.” After looking closely at these values I determined to drop the four different strings indicating a lack of play and removed these rows entirely from my dataframe. This removed 98 rows from the dataframe, taking the size from 1303 to 1205. Now the points column was correctly identified as an integer data type.

Next I transformed the minutes from “MM:SS” format to a decimal value in order to be used for analysis. By splitting the string value for each game and applying a lambda

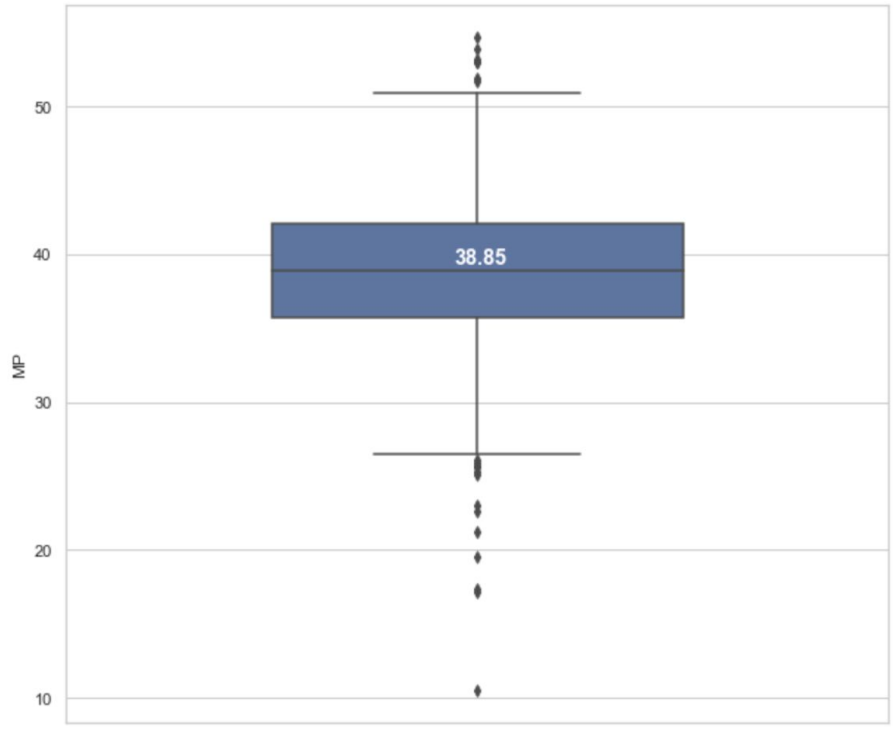
function, I was able to transform each “Minutes Played” value into a floating point decimal of the correct data type.

No other drops or specific changes were needed for the remaining 21 columns other than changing the data types. I used a conversion dictionary to change each respective column from type “object” to the correct type “int” or “float.”

Once each column had the correct data type, I checked for any null values. The only columns that contained nulls were 3P% and FT%. After closer inspection of each occurence of a null value, I noticed that they resulted from games where LeBron had zero three-point or free throw attempts. Since it is impossible to divide by zero, these games recorded a null value for these instances. I decided to leave the rows with null values in, since dropping or changing these values would unnecessarily affect the entire dataset.

Since my analysis was to be based on games played at, below, and above the average range of minutes, I decided I would drop games where LeBron played less than half of a full 48-minute game since these occurrences are usually due to in-game injuries. Using a box plot visualization (pictured below), I noticed his “minutes played” column had some outliers. A few outliers (maybe 3-4) were very near the end of the tail -- around the 25 minute mark-- and the rest were further away. I chose 25 minutes as the cutoff point and removed these 7 rows from the dataset. The final row count for this basic stats dataframe was 1198 rows.

```
Out[41]: Text(0.005, 39.35, '38.85')
```



I added a “Season” column to denote which season each game occurred in. Doing so allowed me to view various statistics against the progression of his career in a season-by-season format, instead of by date. Since my dataframe did not contain this attribute, I researched the start and end dates of each season of Lebron James’ career, from the 2003-2004 season to the current 2019-2020 season. I then created a function that looped through the “Date” attribute for each row and assigned a season number based on the labeled range of season dates I manually defined. I then prepended this “Season” column to my dataframe that would indicate which season (1-17) any given game occurred on.

For the intended analysis, I needed to bin, or categorize, each game Lebron played as either “Below,” “Average,” or “Above,” — meaning significantly high or low in number of minutes played. I had initially planned on using the min and max values taken from the summary statistics of the data as the designation points of each bin. However, after binning under these values and counting the number of games in each bin, I noticed there was a heavy imbalance in the distribution of games. The “Average” bin contained 845 games, while the “Below” and “Above” bins contained 244 and 109 games, respectively. To fix this imbalance, I used the minimum, first quartile, third quartile, and maximum values from the summary statistics instead. Not only did these values seem more realistic to use, but they also yielded a much more balanced distribution of games to each bin. The bin values were 597 for “Average,” 301 for “Below,” and 300 for “Above.”

Lastly, the two metrics I was particularly interested in from this dataframe were “GameScore” and “Plus/Minus,” since these stats are pseudo-advanced beyond individual metrics such as points or assists. I wanted to view these stats on a per-minute basis, so I created two new columns called “GmSc/min” and “PM/min.” Each column was constructed by taking the respective attribute (either GameScore or Plus/Minus) and dividing each value in the row by the minutes played, resulting in a per-minute value for both stats. For instance, if one row contained 40 for the “minutes played” value and 20 for “GameScore”, then the resulting “GmSc/min” column value was now 0.5. This was done for every row in the dataframe.

With all of the cleaning and preprocessing complete for the basic stats dataframe, I exported this dataframe as an excel file in order to be joined with the advanced stats dataset in Tableau Prep.

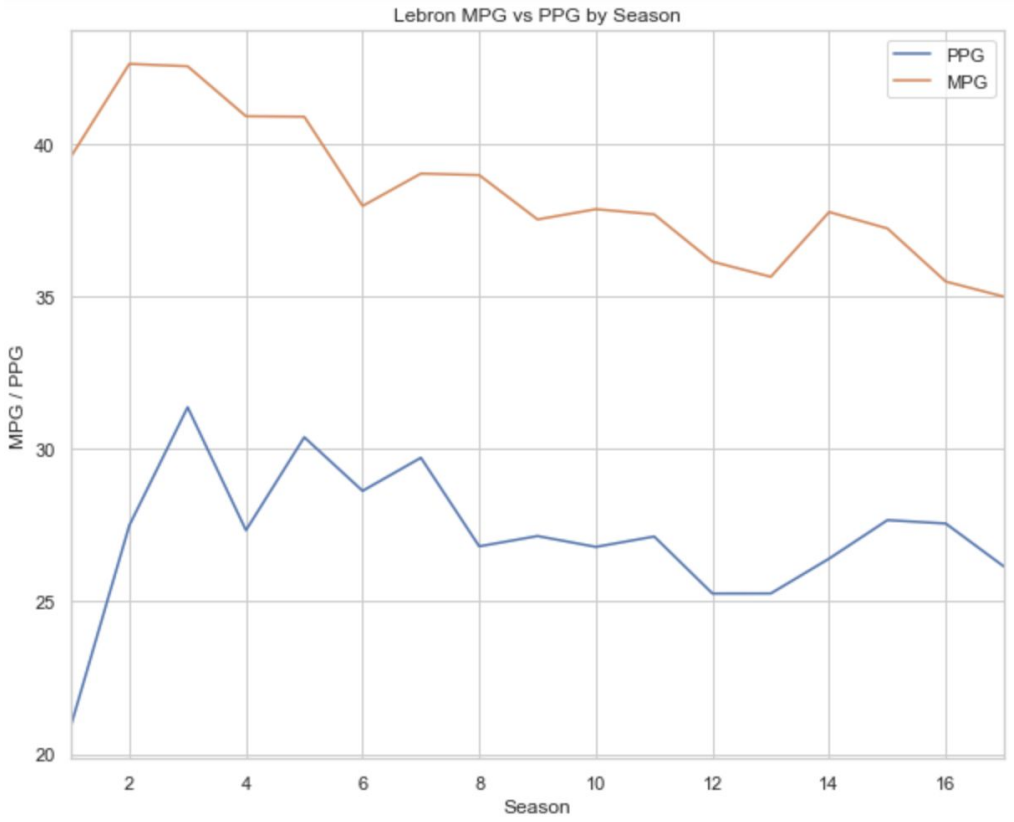
I imported the advanced dataset using Tableau Prep Builder, along with the newly cleaned basic stats dataset to join them. Prior to the join, I performed minimal cleaning on the advanced stats data. I converted the “Matchup” column into a new column with only the dates of each game and renamed it. I was then able to join the two datasets on this “Date” column, dropping the 7 mismatched games from the advanced stats dataset.

The newly-combined, final dataset was the 1198x45 size I referenced earlier and was ready for analysis.

Data Analysis and Visualization

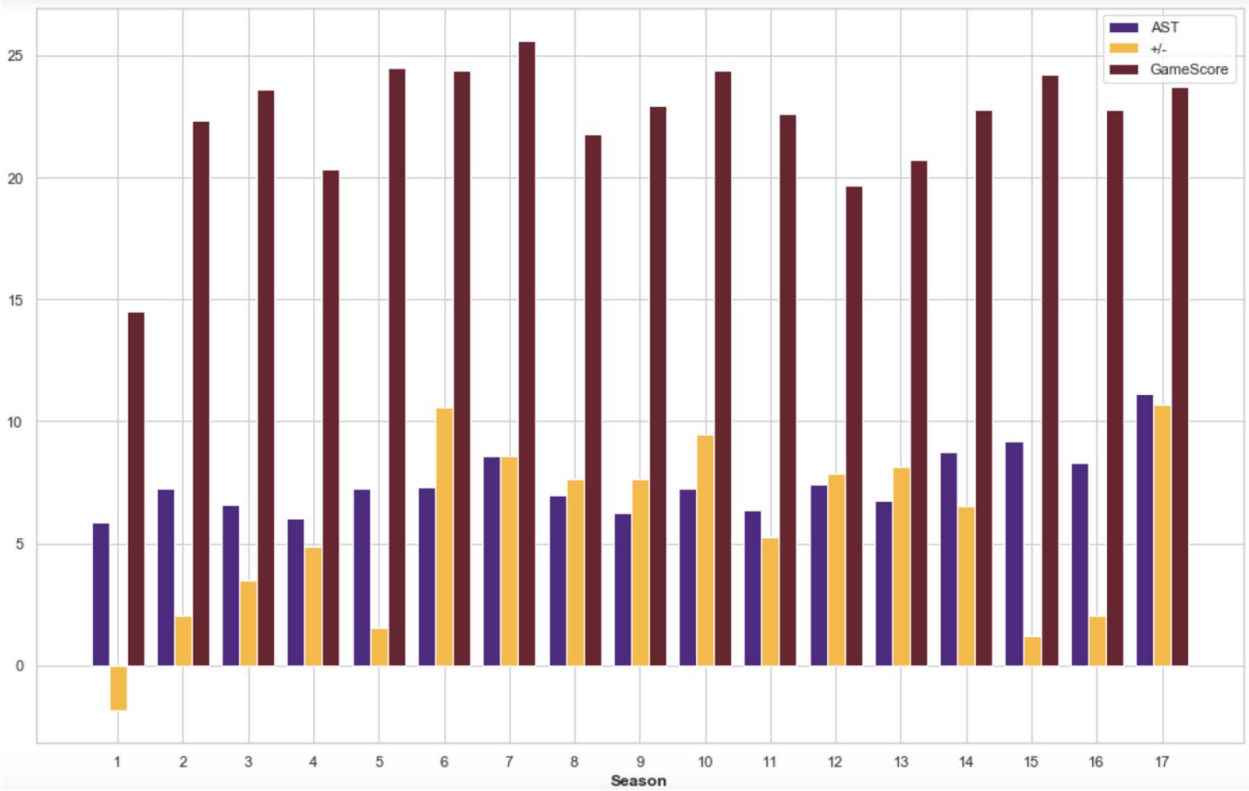
Curious how LeBron's basic performance stats have changed as his career has progressed, I analyzed these trends first. I began by creating a new dataframe containing aggregated basic stats averaged on each season. The dataframe was indexed by season and contained columns for minutes per game (MPG), points per game (PPG), turnovers per game (TovPG), assists per game (AstPG), Plus/Minus per game (PMPG), and GameScore per game (GSPG). The range of turnovers per game LeBron averaged did not change much, ranging from 2-4, so I decided to leave this stat out of any visualizations.

I first created a simple line graph to view the relationship between LeBron's minutes per game and points per game by each season (pictured below). From Season 3 to the current season, the graph appears to show a trend that as Lebron James's MPG value has decreased, his PPG has as well. The spikes in PPG also appear to be consistent with MPG. This trend is probably not too surprising for anyone who has played or watched basketball,. The more time you have to play, the more chances you have to score, and the more points you will have by the end of the game. So it is not too surprising to see that seasons in which Lebron has played more minutes per game on average, he has also scored more points per game on average.



Plotting each remaining stat from the dataframe on a line graph showed that there were no real trends across LeBron's career. Using a bar chart to group these stats into one

graph, it is clear these stats are not affected by LeBron's aging and tenure in the league (pictured below). Knowing the seasons would not have a huge effect on these stats allowed me to confidently continue deeper analysis of performance metrics with respect to minutes per game.



Shooting

Beginning with LeBron's shooting, I used the minutes played binned ("MP-binned") column to analyze how well LeBron shoots when playing high, low, or average minutes. I created a new dataframe consisting of MP-binned and 3P% of size 1198x3. Using the "group-by" function, I grouped the data according to the MP-binned column and averaged out the 3P% for each range. The resulting dataframe showed that when playing below-average minutes, LeBron shot 33.2% from three. When playing an average range of minutes, that percentage dropped to 29.8%. For above-average minute games, LeBron picked his shooting back up to 31.7%. These results show there's no major difference in Lebron's 3-Point shooting percentage at various durations of play. This is not overly surprising, as he is a decent 3-point shooter who doesn't usually need long to get into a good rhythm. The 3.3% drop in shooting when Lebron plays average minutes compared to below average can likely be explained by the more minutes played equating to more shots taken; when a player is only shooting 33% from 3, meaning he will make 3 out of every 10 shots taken, his percentage is likely to drop after taking more shots, unless he stays a consistent 1/3 made three pointers with every 3 additional shots he takes. Still, this drop is not very significant.

Assists, Turnovers, Fouls

Next, I gathered the assists, turnovers, and fouls columns and performed the same grouping and averaging done above. The resulting dataframe is shown below. Again, there was no significant change. As LeBron plays more minutes, the number of assists increases. This is to be expected as more time on the court yields more opportunities to make good plays. The same holds true for personal fouls. There is an interesting stat within turnovers (TOV) though—as LeBron plays above-average minutes, his turnover rate slightly decreases. This is unexpected given the general trend of having more time to play would yield more opportunity to turn the ball over, especially given how often LeBron James has the ball in his hands being the best player on his team at all times.

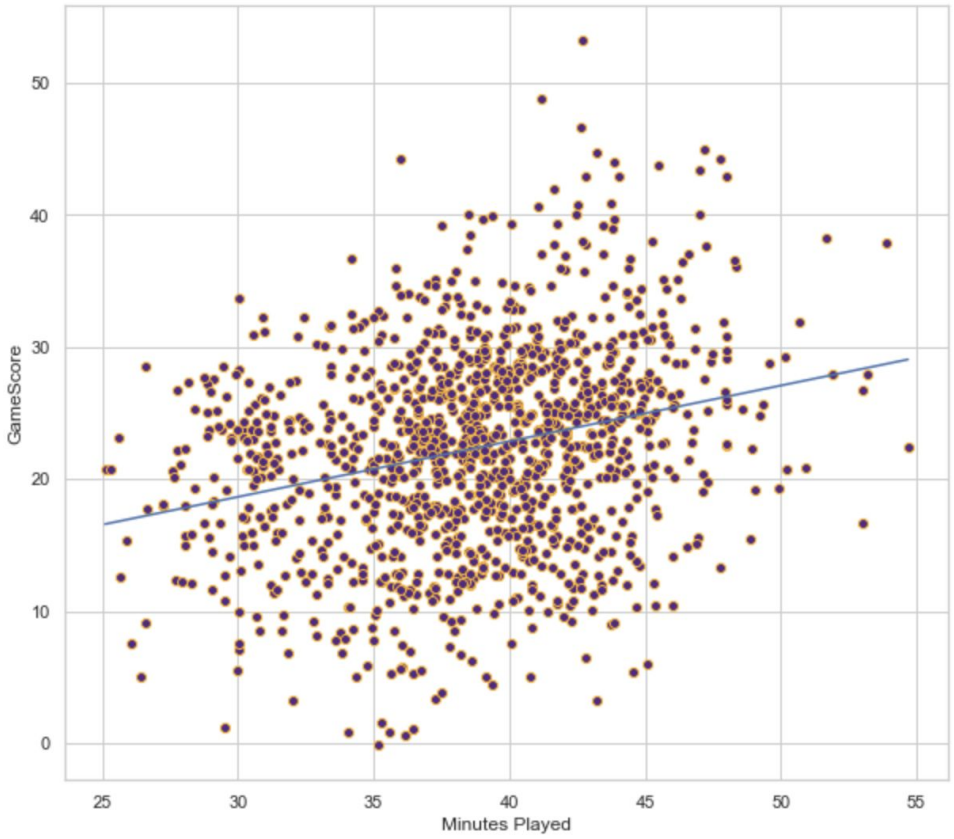
MP-binned	AST	TOV	PF
Above	7.683333	3.590000	2.173333
Average	7.407035	3.624791	1.865997
Below	6.581395	3.112957	1.478405

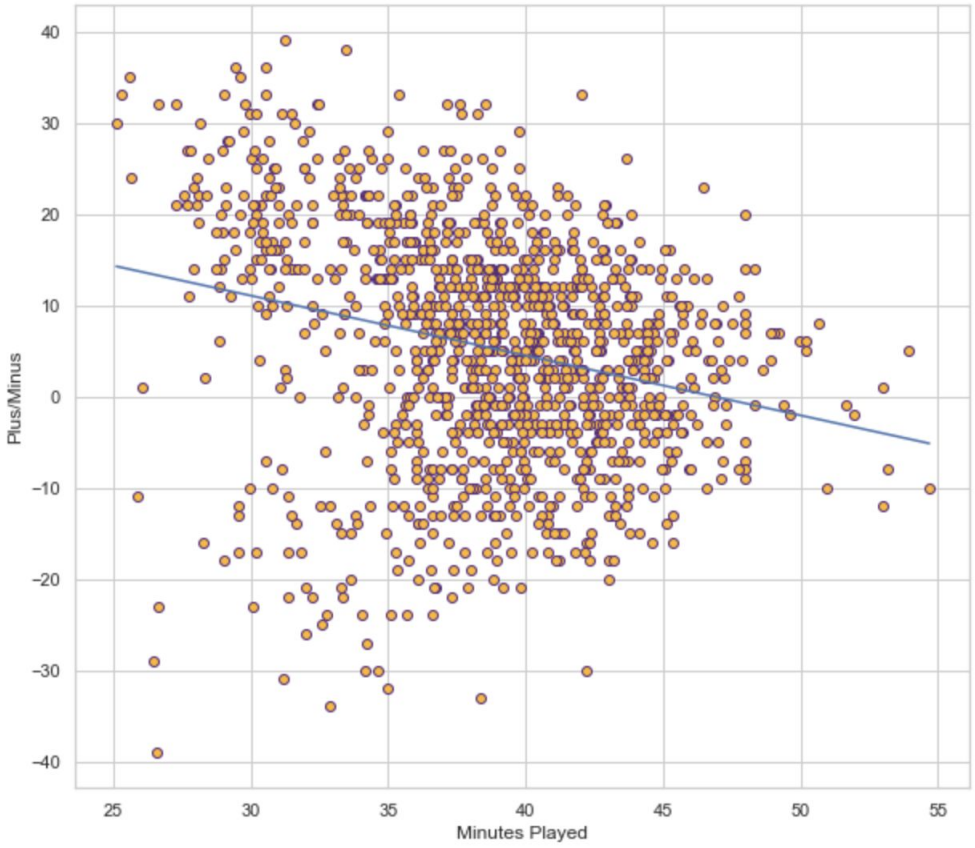
GameScore and Plus/Minus

The two remaining basic performance stats of interest are GameScore and Plus/Minus. GameScore is an all-encompassing stat that is derived from eleven different measures—seven offensive and four defensive. Plus/Minus measures a player's benefit to the team's score while on the floor. These performance stats will give greater insight into LeBron's overall performance. Using the same group and average procedure, the data revealed a clear positive relationship for GameScore and a negative relationship for Plus/Minus with respect to the ranges of minutes played.

MP-binned	GmSc	+/-
Above	24.950000	1.920000
Average	22.411893	4.554439
Below	19.522259	10.325581

To further analyze these trends, I created a scatter plot for both stats with each point representing one game. The first scatterplot—showing GameScore against minutes played—shows a weak positive correlation, but still trending upward. The second—showing Plus/Minus against minutes played—shows a weak negative correlation, with a decrease in spread as minutes played increases.





Starting with GameScore, there is an increase in overall performance as LeBron plays more minutes. Since this stat takes field goal attempts, fouls, free throws missed, and turnovers into account, playing more minutes is not always going to yield a higher GameScore. For instance, a player can play more minutes and have more points per

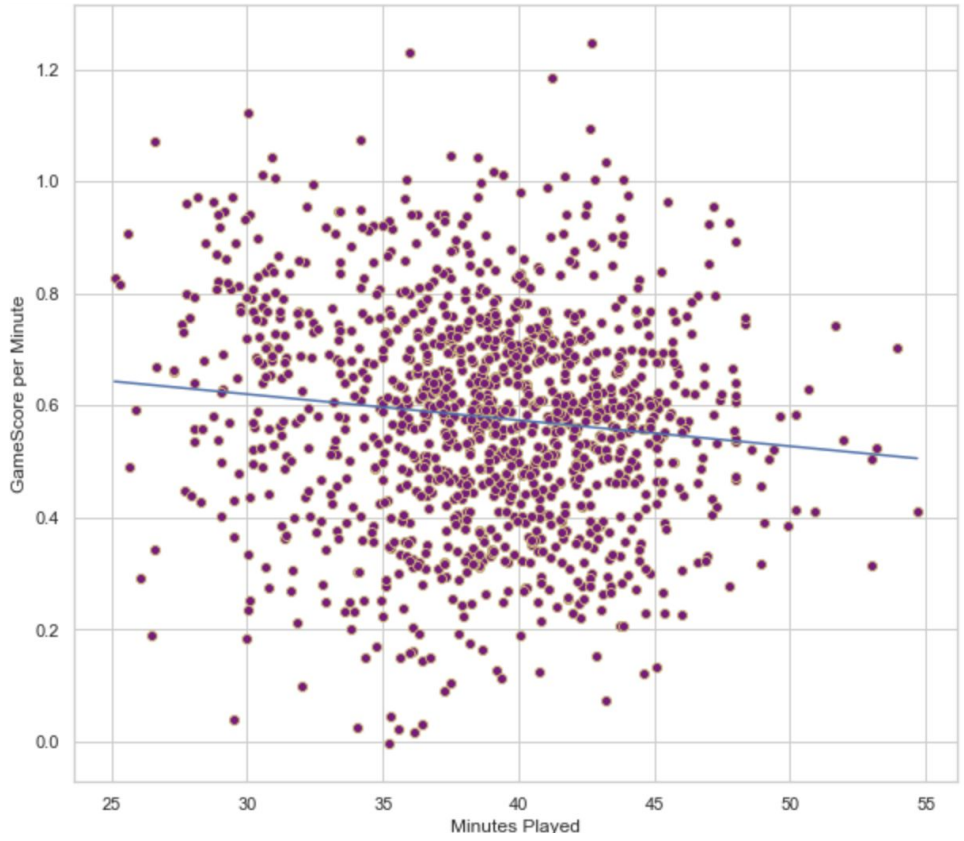
game on paper, but their shooting percentage might drop with more minutes if they are only making a small percentage of the shots they take. However, since GameScore increases with each bin of minutes played, it appears that LeBron, on average, plays better when he plays more minutes.

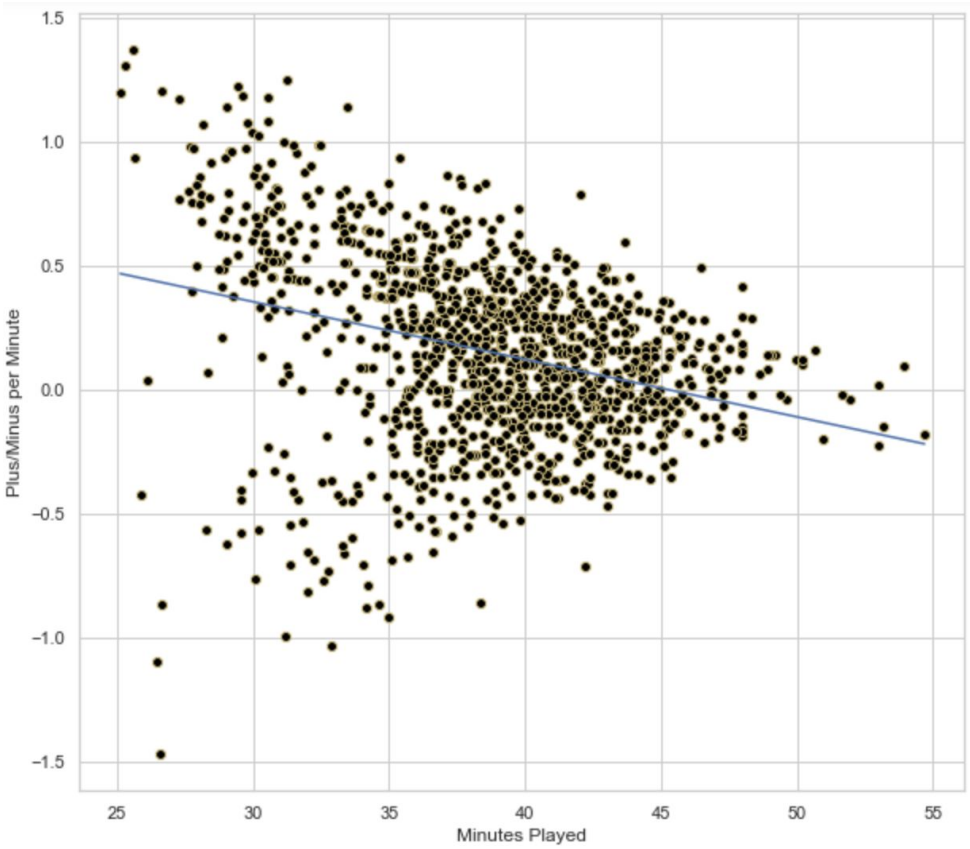
The Plus/Minus (+/-) results are interesting, but explainable. The results show a very high plus/minus for games in which LeBron played a “below” average amount of minutes. A +/- of 10 is a pretty remarkable average for 301 games. The “average” bin results are still impressive as, on average, LeBron's team scores over four and a half more points than the other team when he is on the floor. In “above average”-minute games, , LeBron’s +/- drops to 1.9. Initially, this would seem to indicate that Lebron's performance—or contribution to the team's performance—declines when playing more. This is most likely the result of playing better competition. Lebron is likely to play more minutes in closer games and less in blowouts. Since +/- is indicative of a team's margin while a certain player is on the court, and this range of Lebron playing most of the game has a low +/-, it would follow that the margin is lower because the games are closer. The converse is true for the "below" bin of minutes for Lebron. In games where Lebron's +/- is 10, he is likely going to play less since his team is out-scoring the other team by at least 10 points when Lebron is on the floor.

To further analyze these two stats, I used the per-minute columns previously created to see how LeBron performs according to the GameScore and Plus/Minus measures on a per-minute basis for each range. The results were extremely interesting, showing a negative relationship for both GameScore and Plus/Minus with respect to the binned minutes played.

MP-binned	GmSc/min	PM/min
Above	0.558132	0.042840
Average	0.575634	0.117816
Below	0.609833	0.331071

As minutes played increases, GameScore per minute shows a slight decrease, only dropping 8.5%. Plus/Minus per minute shows a much larger decrease, dropping 87.1%. Using two more scatterplots, I wanted to see if the visualizations confirmed these changes from whole game to per-minute performance for each bin.





The decrease in GameScore is likely due to the fact that in closer games in which LeBron plays more minutes, the game tends to slow down and possessions become longer as teams look to get better shots each play. Consequently, there are less plays/minute and therefore less activity from LeBron. This is an interesting trend that I

will explore further to see if it holds for the advanced statistics. Seeing as the scatter plot is very weakly negatively correlated, I do not believe that LeBron necessarily plays worse on a per-minute basis in higher minute games, but rather the slowing of plays/minute with increasing minutes is driving these values down.

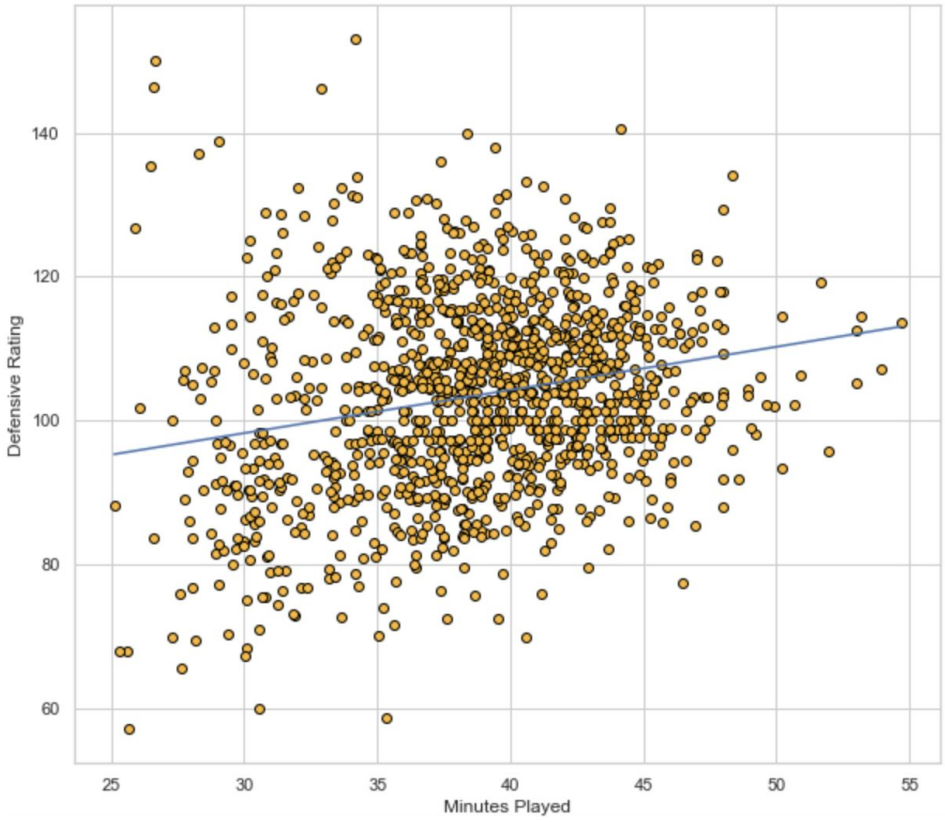
With Plus/Minus, there is a clear decrease when looking at averages and per minute statistics. Both scatterplots for Plus/Minus (whole game and per-minute) show a nearly identical, arrow-shaped pattern. As more minutes are played, the Plus/Minus performance shows less spread but still a negative trend. As previously stated, this is likely due to closer games with tougher competition.

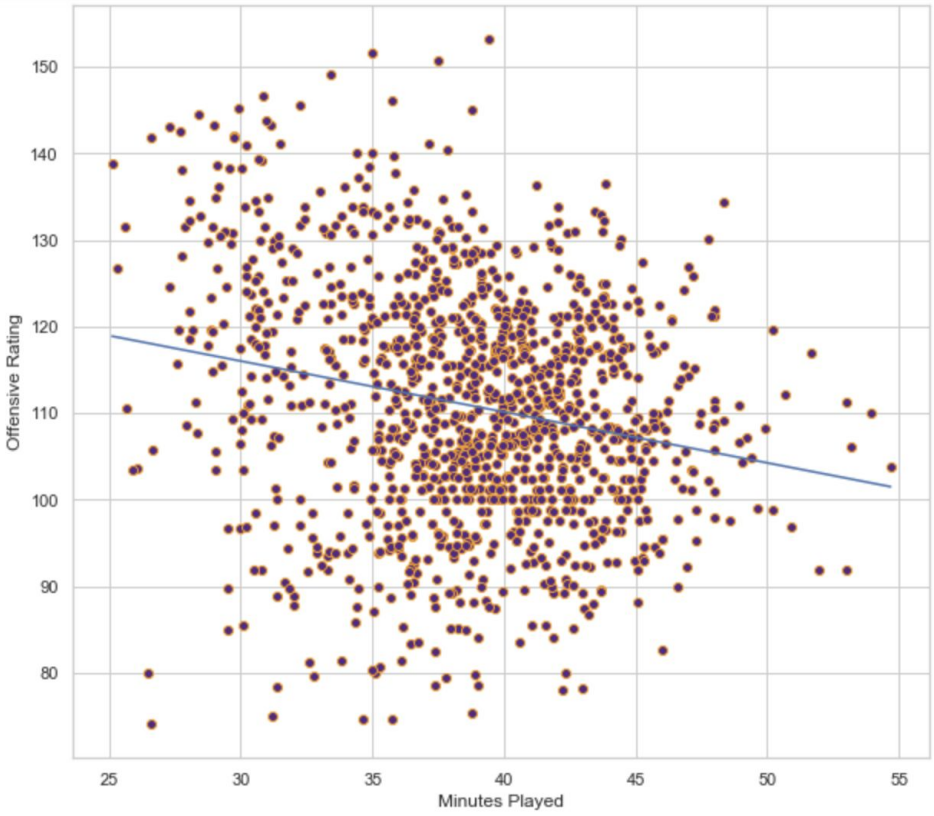
Advanced Statistics

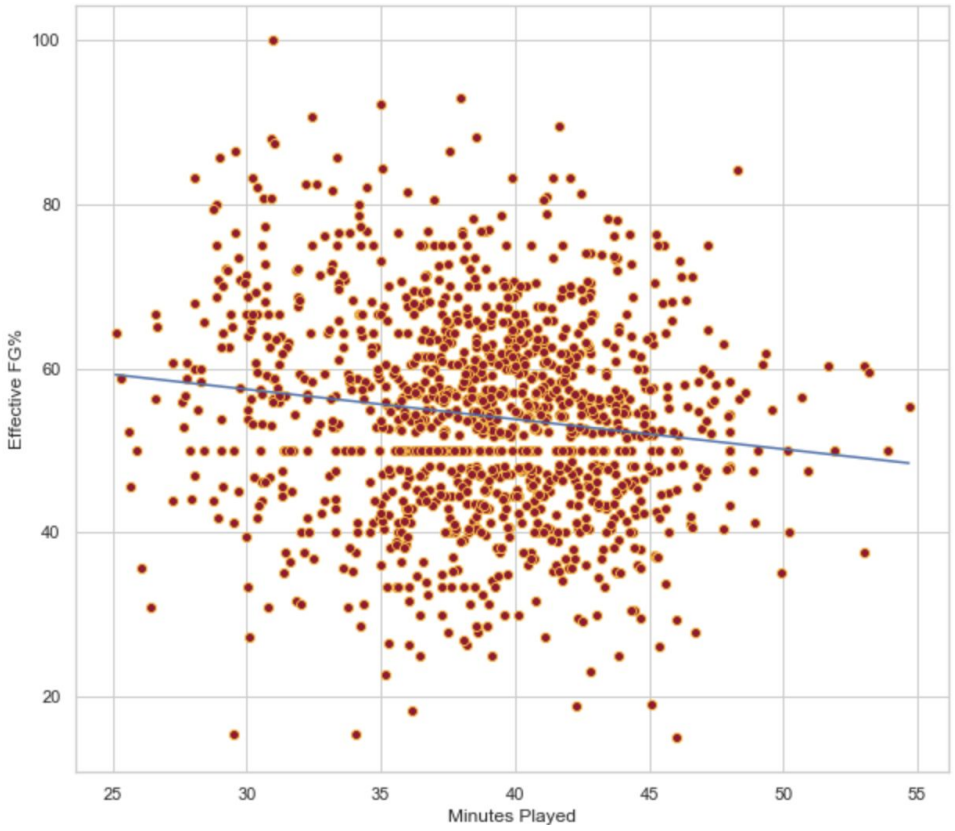
Using advanced performance statistics, I can drill down to LeBron's offensive and defensive performance both on a whole game and per-minute basis to determine if these trends are consistent, as well as causation for them. The stats of interest here are offensive rating (OFFRTG), defensive rating (DEFRTG), net rating (NETRTG), effective field goal % (EFG%), true shooting % (TS%), and player impact estimate (PIE). In short, these stats are measures of offensive, defensive, shooting, and overall performance. To begin, I created a dataframe with the same grouping and average procedure with each of the above stated stats (pictured below).

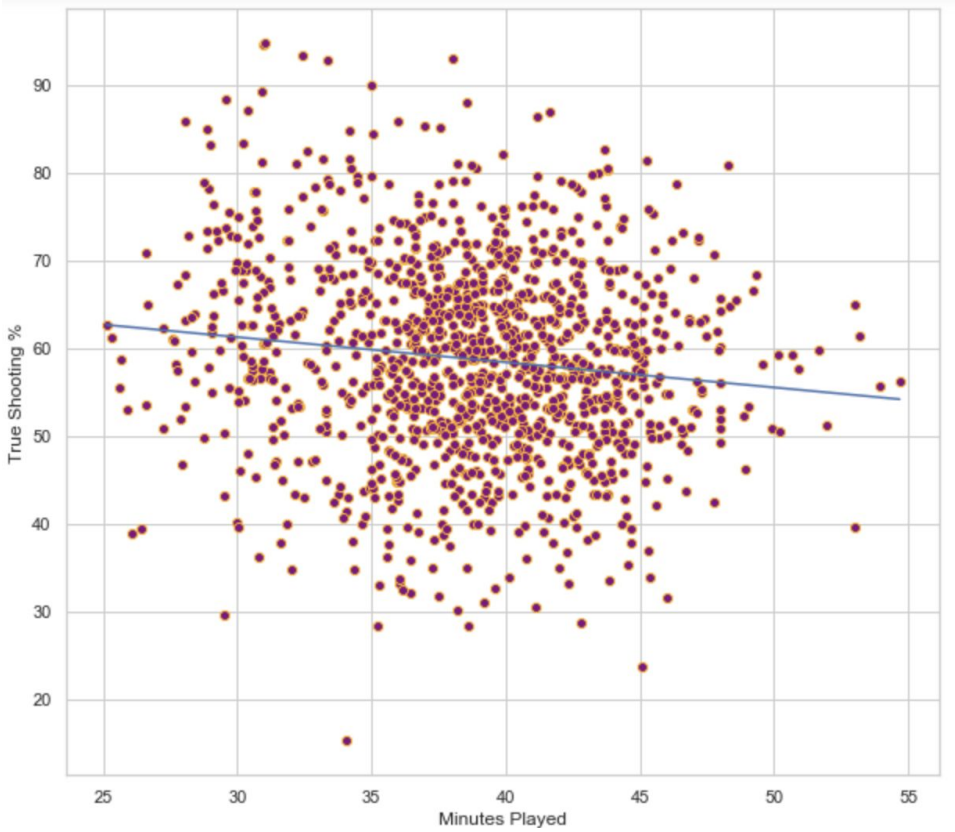
MP-binned	OFFRTG	DEFRTG	NETRTG	EFG%	TS%	PIE
Above	108.197333	106.431333	1.764000	51.814333	57.021667	19.307333
Average	110.001843	104.222446	5.779062	54.129983	58.684422	19.410385
Below	115.418272	98.859136	16.561794	56.974086	60.945847	20.079402

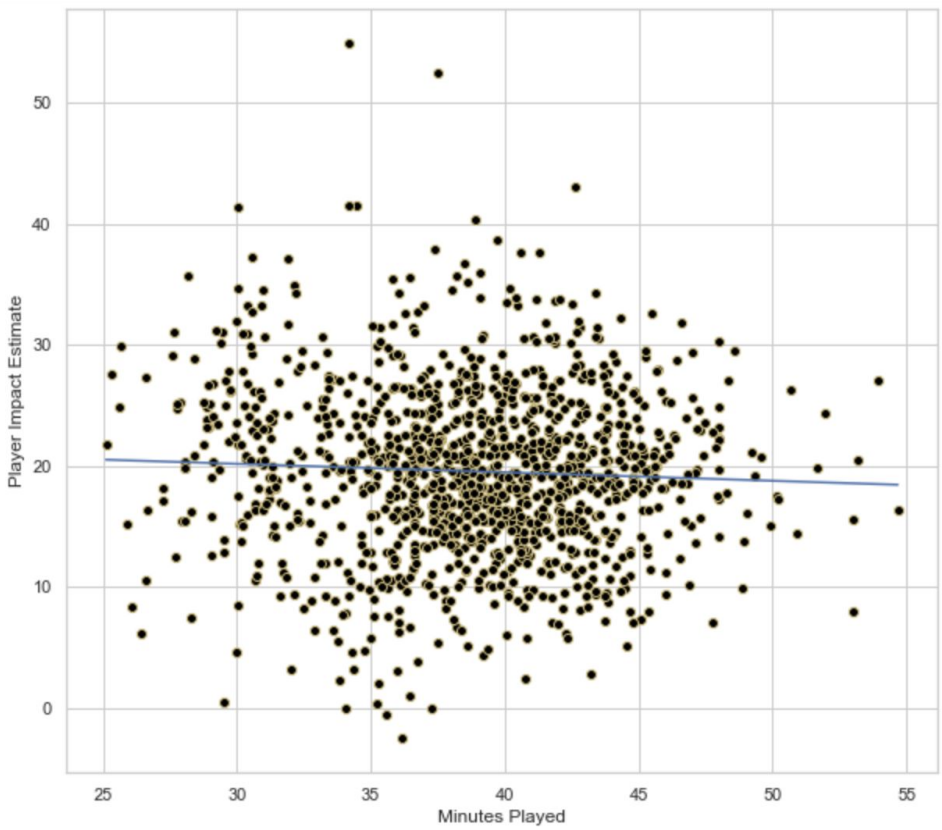
From these initial results, it seems that on average LeBron's offensive performance declines while his defensive performance improves. His shooting performance, represented by effective field goal % and true shooting %, declines as well. His player impact estimate declines, but only slightly. Seeing as offensive rating, defensive rating, and field goal shooting all have the largest shifts, I will explore these particular statistics further by analyzing them against both a whole game and a per-minute basis.


Defensive Rating
140
120
100
80
60
25
30
35
40
45
50
55
Minutes Played


Offensive Rating
150
140
130
120
110
100
90
80
25
30
35
40
45
50
55
Minutes Played


Effective FG%
100
80
60
40
20
25
30
35
40
45
50
55
Minutes Played

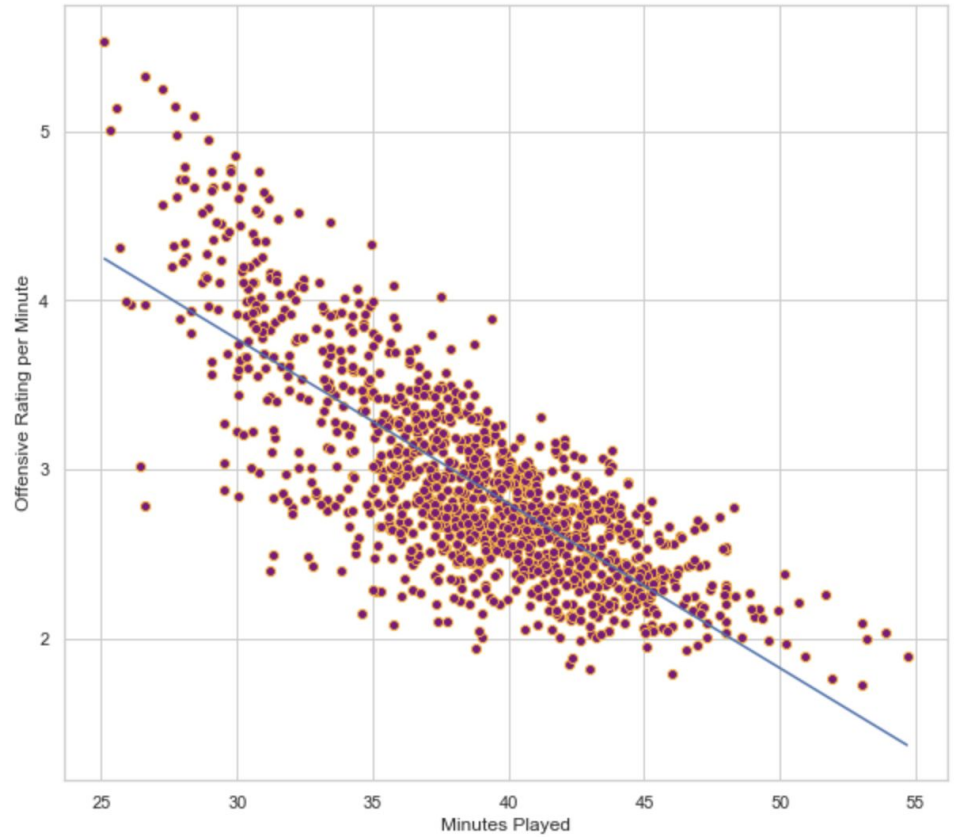

True Shooting %
90
80
70
60
50
40
30
20
25
30
35
40
45
50
55
Minutes Played

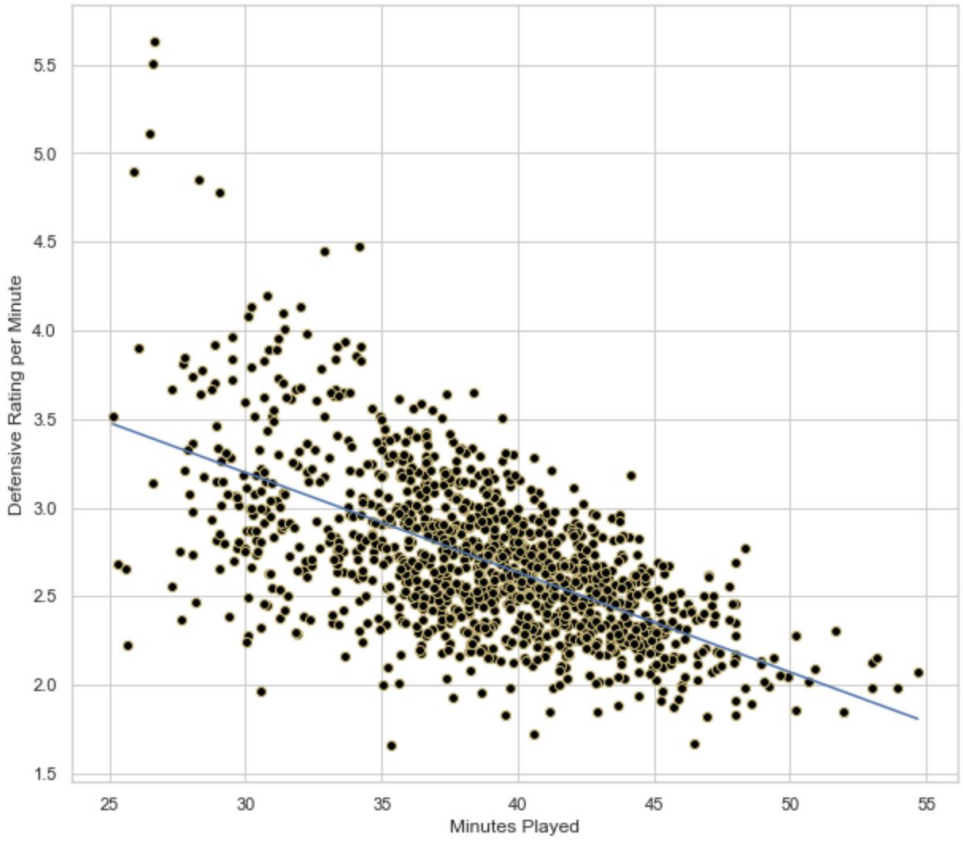


From the scatter plots above, offensive and defensive rating show the strongest trends of a respective decrease and increase in performance as minutes played increases. Both shooting metrics also show a negative relationship with minutes played, but are not strongly correlated.

To further analyze these two statistics of offensive and defensive rating, I looked at these two as per-minute stats, as done previously with GameScore and Plus/Minus.

MP-binned	OFFRTG/min	DEFRTG/min
Above	2.426607	2.387182
Average	2.831269	2.681331
Below	3.616967	3.082405





Both offensive rating and defensive rating show a strong, clear negative relationship to minutes played on a per-minute basis. Offensive rating decreases at a sharper rate of 32.9%, while defensive performance decreases at 22.4%. These results are consistent with the whole-game results—the sharper decline in offensive performance here holds

true that LeBron's offensive play worsens with more minutes played; his defensive play may worsen on a per-minute basis due to more minutes played, but not as sharply as the offensive performance's decline. As seen in the whole-game results, his defensive rating still increases, coinciding with the increase in the whole-game GameScore increase seen earlier.

Results and Conclusion

The four key statistics of GameScore, Plus/Minus, Offensive Rating, and Defensive Rating reveal some interesting trends in Lebron James' game when comparing against his minutes played per game. GameScore and Plus/Minus showed that, when averaged on a whole game basis, LeBron's GameScore increased and Plus/Minus decreased for each bin of minutes played. For each game as a whole, LeBron's overall performance, measured by GameScore, increased when playing more minutes. His Plus/Minus decreased with each bin, most likely as a result of competing in games that are more competitive with closer scores to the opponents. Seeing as his overall performance increased while the plus/minus decreased indicates that LeBron plays well in tighter games, which is why he is regarded as one of the best basketball players alive. However, when analyzing these stats on a per-minute basis, his GameScore actually slightly decreased, along with Plus/Minus, as more minutes were played. This is likely due to the fact that in closer games, teams take longer possessions to get better looks at scoring, drawing out the game and decreasing the number of possessions per minute. As possessions per minute decrease, LeBron has less opportunities to make statistical increases in performance, thus lowering his GameScore per minute.

Further analyzing these trends, the data showed that LeBron's offensive performance may be to blame for his decrease in overall performance on a per-minute basis. When looking at this statistic on a whole game basis, LeBron's offensive rating drops roughly 6.3% from 115.42 to 108.20 when going from below average to above average minutes played. His defensive performance saw a larger shift of a 7.7% increase from 98.86 to 106.43. However, from a per-minute standpoint, his offensive rating sees a much sharper drop of 32.9% from 3.62 to 2.43, and defensive rating 22.4% drops from 3.08 to 2.39 per minute.

Analyzing each game as a whole, LeBron plays better all-around in games in which more minutes are played, likely due to his increased defensive performance. While his offensive ratings drop slightly, his defensive ratings pick up the slack and allow for him to contribute to a greater overall statistical performance. However, analyzing each game on a per-minute basis, LeBron's statistical performances on offense and defense drop as more minutes are played. This is likely due to the game being drawn out by longer possessions with more focus on making the best possible plays out of each possession. Less chances to make a play with more minutes being played will yield a statistical decrease in performance per-minute. As the data showed with GameScore, this

decrease is only slight on a per-minute basis, affirming the drawing out of the game is likely to blame.

Glossary of Statistics Used

Listed below are all 45 columns from the “full_data” dataframe--column name in dataset, full name, and definition.

'Season' -- Season -- Numerical label of which season, 1-17, a game occurred in
'G' -- Season Game -- Numerical label of the number of game (1-82) in a given season that LeBron has played in
'Tm' -- Team -- Denotes which team LeBron was currently playing for
'@' -- At -- Denotes whether the game was home (null value) or away (‘@’)
'Opp' -- Opponent -- The opponent LeBron’s team was playing against for that game
'Result' -- Result -- Denoted by either “W” or “L” followed by the margin of victory or loss in parentheses
'MP' -- Minutes Played -- How many minutes LeBron played in a game
'FG' -- Field Goals -- Number of made field goals
'FGA' -- Field Goal Attempts -- Number of field goals attempted
'FG%’ -- Field Goal Percentage -- Made field goals divided by field goals attempted
'3P' -- Three Pointers -- Number of made three pointers
'3PA' -- Three Pointers -- Number of three pointers attempted
'3P%' -- Three Point Percentage -- 3P divided by 3PA
'FT' -- Free Throws -- Number of made free throws
'FTA' -- Free Throw Attempts -- Number of free throws attempted
'FT%' -- Free Throw Percentage -- FT divided by FTA
'ORB' -- Offensive Rebounds -- Number of offensive rebounds captured
'DRB' -- Defensive Rebounds -- Number of defensive rebounds captured
'TRB' -- Total Rebounds -- ORB plus DRB
'AST' -- Assists -- Number of assists earned
'STL' -- Steals -- Number of steals earned
'BLK' -- Blocks -- Number of blocks earned
'TOV' -- Turnovers -- Number of turnovers earned
'PF' -- Personal Fouls -- Number of fouls committed
'PTS' -- Points -- Number of points scored
'GmSc' -- GameScore -- An overall measure of a player’s performance, derived from the following equation: (Points x 1.0) + (FG x 0.4) + (FGA x -0.7) + ((FTA-FT) x -0.4) + (ORB x 0.7) + (DRB x 0.3) + (STL x 1.0) + (AST x 0.7) + (BLK x 0.7) + (PF x -0.4) + (TOV x -1.0).
'+/-' -- Plus/Minus -- The point differential when a player or team is on the floor (player’s team’s points scored minus opponent team’s points scored)
'MP-binned’ -- Minutes Played Binned -- User-created column to categorize the MP for a given game as below, above, or average

'GmSc/min' -- GameScore per Minute -- User-created column calculated the GameScore per minute for a given game, derived by dividing GameScore by MP
'PM/min' -- Plus/Minus per Minute -- User-created column calculated the Plus/Minus per minute for a given game, derived by dividing Plus/Minus by MP
'OFFRTG' -- Offensive Rating -- Measures a team's points scored per 100 possessions. On a player level this statistic is team points scored per 100 possessions while he is on court
'DEFRTG' -- Defensive Rating -- The number of points allowed per 100 possessions by a team. For a player, it is the number of points per 100 possessions that the team allows while that individual player is on the court
'NETRTG' -- Net Rating -- Offensive rating minus defensive rating
'AST%' -- Assist Percentage -- The percentage of teammate field goals a player assisted on while he was on the floor
'AST/TO' -- Assist to Turnover Ratio -- The number of assists for a player or team compared to the number of turnovers they have committed
'AST RATIO' -- Assist Ratio -- Assist Ratio is the number of assists a player averages per 100 possessions used
'OREB%' -- Offensive Rebound Percentage -- The percentage of available offensive rebounds a player or team obtains while on the floor
'DREB%' -- Defensive Rebound Percentage -- The percentage of available defensive rebounds a player or team obtains while on the floor
'REB%' -- Rebounding Percentage -- The percentage of available rebounds a player or team grabbed while on the floor
'TO RATIO' -- Turnover Ratio -- The number of turnovers a player or team averages per 100 possessions used
'EFG%' -- Effective Field Goal Percentage -- Measures field goal percentage adjusting for made 3-point field goals being 1.5 times more valuable than made 2-point field goals
'TS%' -- True Shooting Percentage -- A shooting percentage that factors in the value of three-point field goals and free throws in addition to conventional two-point field goals
'USG%' -- Usage Percentage -- The percentage of team plays used by a player when he is on the floor
'PACE' -- Pace -- The number of possessions per 48 minutes for a team or player
'PIE' -- Player Impact Estimate -- Measures a player's overall statistical contribution against the total statistics in games they play in

[Github Repository]