Predicting whether the Los Angeles Lakers will have enough wins to guarantee them a spot in the 2022 – 2023 NBA Playoffs.

Purpose of the Project

The purpose of this project is to predict whether the Los Angeles Lakers will have enough wins to guarantee them a spot in the Playoffs by analyzing the current team's NRtg (net rating). This will allow the team to determine what changes, if necessary, in order to be guaranteed a spot in the Playoffs. Although this project looks specifically at the Los Angeles Lakers, the model can be implemented to any of the other 29 teams in the NBA (National Basketball Association).

About the Project

The NBA can be defined as having 2 distinguished periods throughout a season. The first, is the regular season which consists of 30 teams that play 82 games, and the second, a postseason tournament to determine the league champion called the Playoffs. Prior to the 2020 – 2021 season, teams that finished within the top 8 seeds in their respective conference, the East and the West, played in the Playoffs. However, with the introduction of the play-in tournament[1], the top 6 teams from each conference are now guaranteed a spot in the Playoffs while teams that finished the season between 7th and 10th place compete for the final 2 spots. This means that only 6 teams from each conference are guaranteed a spot in the Playoffs. Therefore, this analysis will look at the statistics of an NBA team to determine whether a team will achieve enough wins to finish the season within the top 6 places from each conference, guaranteeing the team a spot in the Playoffs.

This analysis will look specifically at the Los Angeles Lakers.

Key Questions

1. How many wins does the Los Angeles Lakers need in the regular season to be guaranteed a spot in the Playoffs?

[1] The play-ins, a preliminary tournament that determines the final two playoff seeds in the Eastern Conference and Western Conference, was introduced in the 2020 – 2021 season.

2. Which variable correlates the most to wins during the regular season?
3. How to calculate the NRtg (net rating) for a team?
4. Will the Los Angeles Lakers achieve enough wins to guarantee them a spot in the Playoffs?
5. What can the Los Angeles Lakers do to improve their regular season record to be guaranteed a spot in the Playoffs?

Key Insights

1. Teams in the Eastern Conference that are guaranteed a spot in the Playoffs average 51 wins in the regular season, and teams in Western Conference average 55 wins a season. The Los Angeles Lakers is part of the Western Conference. Therefore, the Los Angeles Lakers should have a goal of winning 55 games by the end of the season to be guaranteed a spot in the Playoffs.

2. The NRtg (net rating) correlates most with wins in the regular season. This means that the Los Angeles Lakers need to finish the season with a NRtg of around 6.3 to have 55 wins by the end of the season.

3. The NRtg (net rating) is the difference between the ORtg (offensive rating) and DRtg (defensive rating). So, by finding the variables that highly correlate with the ORtg and the DRtg, the NRtg can be predicted. The TS% (true shooting percentage) and PS/G (points scored per game) highly correlate with the ORtg. The O_eFG% (opponent field goal percentage) and PA/G (points allowed per game) highly correlate with the DRtg.

4. No, the Los Angeles Lakers will not have enough wins to be guaranteed a spot in the Playoffs for the 2022– 2023 season. The Los Angeles Lakers, with a NRtg of -1.7, is predicted to have 35 wins by the end of the 2022 – 2023 NBA season.

5. The Los Angeles Lakers should look to change the rotations of their current roster or bring in key players that will have a positive impact in order to improve the team's NRtg and as a result win more games to be guaranteed a spot in the Playoffs.

About the Data

The data contains season statistics for the 30 teams in the NBA (National Basketball Association) from the seasons that represent the Modern Era[2], years 2013 – 2022. The data that represents the 2023 season was used after the statistical model was created to predict the number of wins

[2] 2013 to present, a period characterized with more efficient shot selections by rewarding three-point shots compared to the two-point shots (https://content.iospress.com/articles/journal-of-sports-analytics/jsa200525)

the Los Angeles Lakers will have by the end of the season. I combined the data from the 2013 – 2022 season to create 3 different tables in Excel that contained information regarding the teams' standings at the end of the season, advanced team statistics, and per game statistics, using the data sources below. In R, the 3 tables were combined to create one large dataframe with a total of 58 columns.

Data Sources:
https://www.basketball-reference.com/leagues/NBA_2023.html
https://www.basketball-reference.com/leagues/NBA_2022.html
https://www.basketball-reference.com/leagues/NBA_2021.html
https://www.basketball-reference.com/leagues/NBA_2020.html
https://www.basketball-reference.com/leagues/NBA_2019.html
https://www.basketball-reference.com/leagues/NBA_2018.html
https://www.basketball-reference.com/leagues/NBA_2017.html
https://www.basketball-reference.com/leagues/NBA_2016.html
https://www.basketball-reference.com/leagues/NBA_2015.html
https://www.basketball-reference.com/leagues/NBA_2014.html

Tools Used:
I used R Studio to create the statistical models using these packages:

1. tidyverse (dplyr, ggplot2)
2. ggpubr

In Depth Analysis

1. How many wins does the Los Angeles Lakers need in the regular season to be guaranteed a spot in the Playoffs?

To determine how many wins the Los Angeles Lakers need in the regular season to be guaranteed a spot in the Playoffs, I need to determine how many games, on average, a team that is guaranteed a spot in the Playoffs win. This will act as the target for the Los Angeles Lakers to have by the end of the season.

Since there are 2 conferences, the East and the West, I split up the teams to their respective conferences and created 2 new data frames that contain all the data from teams that finished within the top 6 in previous seasons. Although there is a column that represents that number of wins each team has, it wouldn't be wise to average those values because there were 2 seasons (2019 – 2020 and 2020 – 2021) that had less games due to COVID-19. So, I calculated what percentage of games teams that finish in the top 6 of their conference win.

Win Percentage = (Total Wins / Total Games) x 100

Then, I took the avereage win percentage for each conference by dividing the sum of the win percentages by the number of teams that made the Playoffs from the 2013 – 2022 seasons (54 teams from the West and 54 teams from the East), then multiplied it by 82. This allows us to see how many games a team will need to win in a regular 82 games season to be guaranteed a spot in the Playoffs.

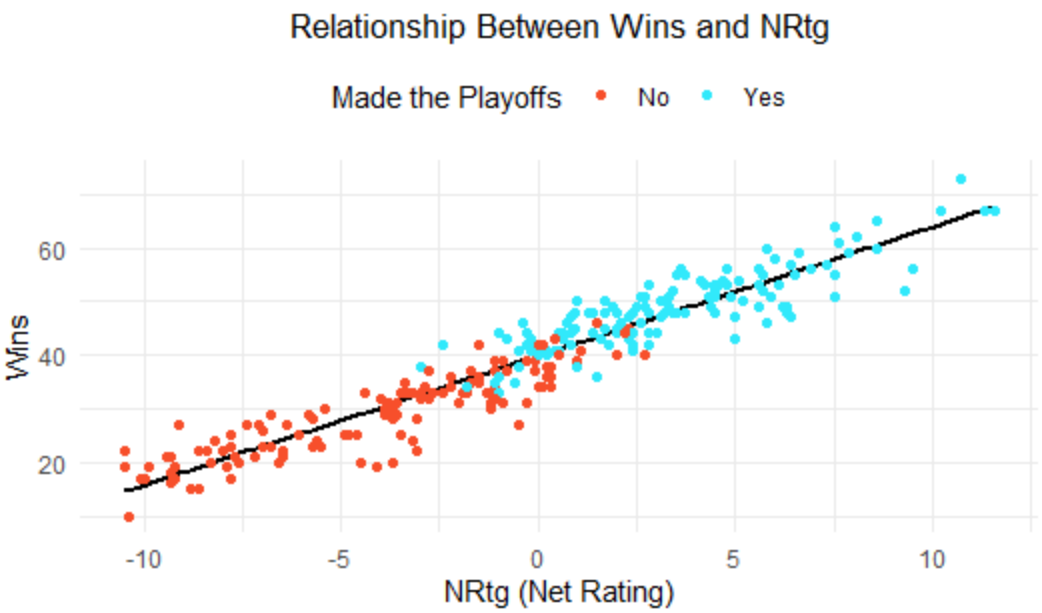
Average Number of Wins = (Total of Win Percentage x 54) / 82

From the results, we see that teams in the East win, on average, 51 games while teams in the West win, on average, 55 games in the regular season. Therefore, the Los Angeles Lakers should have a goal of winning 55 games by the end of the season to be guaranteed a spot in the Playoffs.

2. Which variable correlates the most to wins during the regular season?

The Los Angeles Lakers need to win 55 games in the regular season to be guaranteed a spot in the Playoffs. In order to determine which statistic most correlates with wins, I performed a Pearson correlation test between wins and 46 other variables that represent the offensive statistics and defensive statistics from all 30 teams in the association, across 9 seasons. From these calculations, the NRtg (net rating) has the highest correlation with 0.95. The NRtg is the difference in the score per 100 possessions as a combination of 5 players. It will display the quality of play and the quality of the team.

I created a linear regression model to tell us approximately the NRtg a the team needs to win 55 games in the regular season; therefore, securing a spot in the Playoffs.



Call:
lm(formula = W ~ NRtg, data = df)

Residuals:

Min	1Q	Median	3Q	Max
-11.6338	-2.3582	0.1805	2.6591	9.0763

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	39.83787	0.23705	168.06	<2e-16	***
NRtg	2.40815	0.04943	48.72	<2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.895 on 268 degrees of freedom
Multiple R-squared: 0.8986, Adjusted R-squared: 0.8982
F-statistic: 2374 on 1 and 268 DF, p-value: < 2.2e-16

From the graph we see that teams with a positive NRtg are more likely to win; therefore, make the Playoffs.
The linear regression shows that the results are highly significant.

Our regression equation for wins is:

$$\text{Wins} = 39.83787 + 2.40815(\text{NRtg})$$

The Los Angeles Lakers need to average 55 wins during the regular season. So, the formula to finding the minimum net rating is:

$$55 = 39.83787 + 2.40815(NRtg)$$
$$NRtg = 6.296173$$

This means that the Los Angeles Lakers need to finish the season with a NRtg of around 6.3 to have 55 wins by the end of the season.

3. How to calculate the NRtg (net rating) for a team?

The NRtg is the difference between the ORtg (offensive rating) and the DRtg (defensive rating).

$$NRtg = ORtg - DRtg$$

Since the NRtg is a point differential per 100 possessions, all teams' statistics are based on the same number of possessions creating an equal playing field. Therefore, the NRtg allows a respective team to have a general comparison of their team's performance to other teams. A team will have a positive NRtg if the team won more games than lost.

So, How does a team achieve a positive NRtg?
A team will need a high ORtg and a low DRtg.

To determine which statistics have a statistical relationship with the ORtg, I performed another Pearson correlation test. I compared the ORtg from all 30 teams, across 9 seasons. The same was done for the DRtg.

Variables highly correlated to the ORtg:

- TS% has a positive correlation of 0.903844287
    - TS% = a measure of shooting efficiency that takes into account 2-pointers, 3-pointers and free throws
- eFG% has a positive correlation of 0.879423785
    - eFG% = Statistics adjusted for the fact that a 3-point field goal is worth one more point than a 2-point field goal percentage

Variables highly correlated to the DRtg:

- OeFG% has a positive correlation of 0.903844287
    - OeFG% =Opponent's statistics adjusted for the fact that a 3-point field goal is worth one more point than a 2-point field goal percentage
- PA/G% has a positive correlation of 0.879423785

- PA/G% = Opponent's points per game

Now we can use these correlations to predict the ORtg and DRtg.

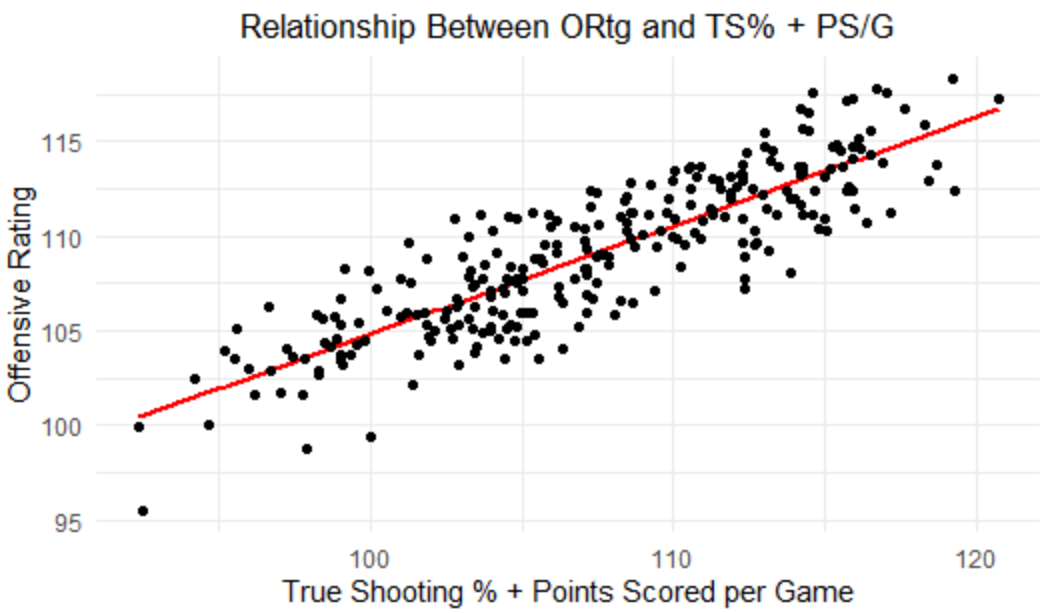
ORtg (Offensive Rating) Prediction
First, I fitted the variables into a multivariate regression model to understand how these variables combined affect the ORtg of a team.

```
Call:
lm(formula = ORtg ~ TS. + eFG., data = df1)

Residuals:
    Min      1Q  Median      3Q     Max
-4.5668 -1.1866 -0.1145  1.1928  5.8777

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   12.390      3.121   3.970 9.25e-05 ***
TS.          195.288     24.274   8.045 2.83e-14 ***
eFG.         -22.341     22.608  -0.988    0.324
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.72 on 267 degrees of freedom
Multiple R-squared:  0.8176,  Adjusted R-squared:  0.8162
F-statistic: 598.4 on 2 and 267 DF,  p-value: < 2.2e-16
```

From the regression model, we see that the eFG% is not significant, meaning that there is no statistical relationship between the eFG% and ORtg.

Therefore, I created another multivariate regression model substituting the eFG% with the PS/G (points scored per game), the next highest correlation to ORtg.

PS/G has a positive correlation of 0.858321231.




```
Call:
lm(formula = ORtg ~ TS. + PS.G, data = df1)

Residuals:
    Min      1Q  Median      3Q     Max
-4.3440 -1.0509 -0.0255  1.0124  4.5629

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  19.41223    2.56342   7.573 5.98e-13 ***
TS.         117.25455    8.04738  14.571  < 2e-16 ***
PS.G          0.23086    0.02836   8.140 1.51e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.542 on 267 degrees of freedom
Multiple R-squared:  0.8533,  Adjusted R-squared:  0.8522
F-statistic: 776.7 on 2 and 267 DF,  p-value: < 2.2e-16
```

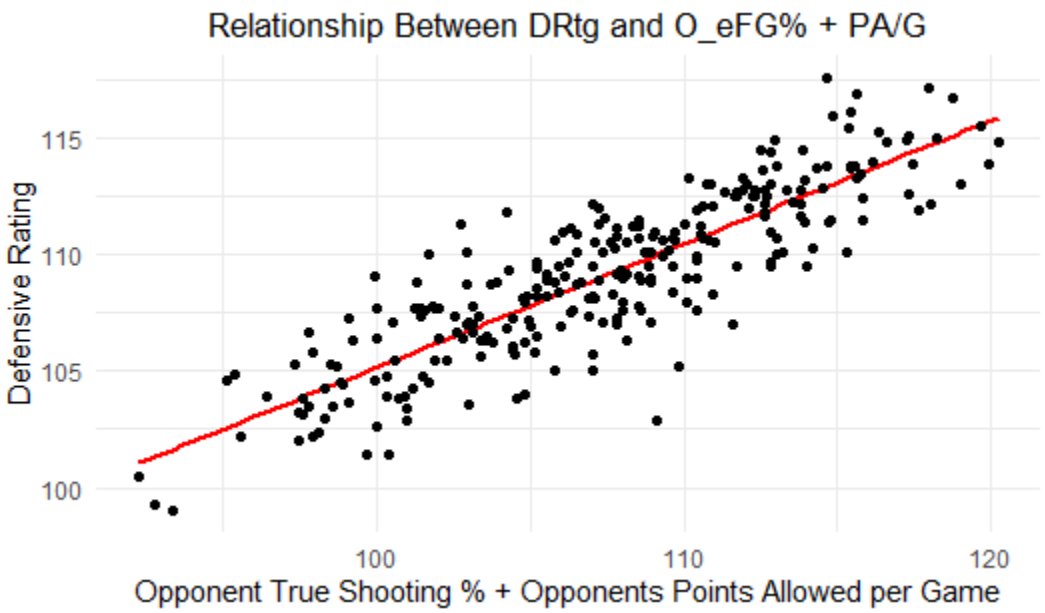
From the model, we see that both variables are significant, and teams with a higher TS% and PS/G will have a higher ORtg.

$$ORtg = 19.41223 + 117.25455(TS\%) + 0.23086(PS/G)$$

This formula will be used to determine the Los Angeles Lakers's ORtg.

DRtg Prediction

Next, I needed to create another multivariate regression model to understand the relationship between the O_eFG% and PA/G to the DRtg of a team.




```
Call:
lm(formula = DRtg ~ O_eFG. + PA.G, data = df1)

Residuals:
   Min     1Q  Median     3Q    Max
-3.5615 -0.7996 -0.0682  0.8688  3.1836

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  28.47503    1.95587   14.56   <2e-16 ***
O_eFG.      112.27513    6.46080   17.38   <2e-16 ***
PA.G          0.21032    0.02259    9.31   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.243 on 267 degrees of freedom
Multiple R-squared:  0.8803,  Adjusted R-squared:  0.8794
F-statistic: 981.9 on 2 and 267 DF,  p-value: < 2.2e-16
```

From the model, we see that both variables are significant, and teams with a higher O_eFG% and PA/G will have a higher ORtg.

$$DRtg = 28.47503 + 112.27513(O\_eFG\%) + 0.21032(PA/G)$$

This formula will be used to determine the Los Angeles Lakers's DRtg.

Now that I know which variables can be used to calculate the ORtg and DRtg, I can predict the NRtg for the Los Angeles Lakers; therefore, concluding whether they will finish the regular season with a guaranteed spot in the Playoffs.

4. Will the Los Angeles Lakers achieve enough wins to guarantee them a spot in the Playoffs?

Finally, I can predict the Los Angeles Lakers's NRtg, which will allow us to predict how many wins they will have by the end of the season, using these 2 formulas:

1. ORtg = 19.41223 + 117.25455(TS%) + 0.23086(PS/G)
2. DRtg = 28.47503 + 112.27513(O_eFG%) + 0.21032(PA/G)

The current season's NRtg is necessary to predict the total number of wins the team will achieve.
Statistics from the current (2022 – 2023) season:
ORtg predictors:

TS% = 0.557
PS/G = 116.9

DRtg predictors:

O_eFG% = 0.536
PA/G = 117.9

$$NRtg = ORtg - DRtg$$

$$NRtg = [19.41223 + 117.25455(TS\%) + 0.23086(PS/G)] - [28.47503 + 112.27513(O\_eFG\%) + 0.21032(PA/G)]$$

$$NRtg = [19.41223 + 117.25455(0.557) + 0.23086(116.9)] - [28.47503 + 112.27513(0.536) + 0.21032(117.9)]$$

$$NRtg = -1.740679$$

Now that the NRtg is predicted, the number of wins a team will have by the end of the season can also be predicted by using the first regression model I created:

$$Wins = 39.83787 + 2.40815(NRtg)$$

Wins = 39.83787 + 2.40815(-1.740679)

Wins = 35.64605

The Los Angeles Lakers, with a NRtg of -1.7, is predicted to have 35 wins by the end of the 2022 – 2023 NBA season. The Los Angeles Lakers will not have enough wins to be guaranteed a spot in the Playoffs for the 2022– 2023 season.

5. What can the Los Angeles Lakers do to improve their regular season record to be guaranteed a spot in the Playoffs?

The Los Angeles Lakers currently have an NRtg of -1.740679, which is short from the estimated NRtg of 6.3 needed to be guaranteed a spot in the Playoffs. Since the NRtg equals the ORtg subtracted by the DRtg, they should look to find methods to increase their ORtg or lower their DRtg. Their current ORtg is 111.7105 and DRtg is 113.4512. This can be done by changing the rotations of their current roster or by bringing in key players that will have a positive impact in order to improve the team's NRtg and as a result win more games to be guaranteed a spot in the Playoffs.