Math 42 Project: NBA Season Prediction Model

Casey Tattersall, Taylor McAdams, William Sun

12/4/2022

Contents

1 Introduction

1.1 Project Background

Over the course of the long regular seasons in sports leagues like the NBA, fans and sportswriters have always loved to pass the time by ranking all of the league's teams from best to worst (and then arguing over each other's rankings). In the past, these rankings only used rudimentary statistics, and were instead heavily based on the writer's personal opinions for each team (using the so-called "eye test" as justification). In recent years, however, the invention of many new advanced statistics and the collection of massive amounts of data has allowed for a new ranking strategy to emerge: a mathematical model that can use all of this data to objectively rank the league's teams, free from any human bias - in theory, at least.

One example of a modeling-based approach to ranking teams that we used as inspiration for this project is FiveThirtyEight's NBA rankings model. They have made a name for themselves in recent years by producing incredibly accurate predictions and foreseeing the rise of underrated teams far before the general public. While their modeling structure is far beyond the scope of this project (using head-to-head matchup data for their ELO model and detailed individual player statistics for their RAPTOR player rating model), we found this subject very interesting, and decided to pursue it for our project.

FiveThirtyEight's predictions for the current NBA season using their ELO model:

			REGULAR SEASON				PLAYOFFS		
CURRENT RATING	TEAM	CONFERENCE	FULL-STRENGTH RATING	PROJ. RECORD	PROJ. POINT DIFF/G	CHANCE OF MAKING PLAYOFFS	FULL-STRENGTH RATING	CHANCE OF MAKING FINALS	CHANCE OF WINNING FINALS
1691	Celtics 17-4	East	1717	61-21	+7.3	>99%	1749	44%	30%
1582	Grizzlies 12-8	West	1666	53-29	+4.1	98%	1695	28%	13%
1620	Bucks 14-5	East	1658	53-29	+4.1	97%	1688	19%	11%
1573	Suns 14-6	West	1652	52-30	+4.5	97%	1678	22%	10%
1578	76ers 12-9	East	1639	50-32	+3.6	92%	1682	14%	8%
1583	Nuggets 13-7	West	1617	52-30	+3.3	96%	1654	16%	6%
1584	Warriors 11-10	West	1616	47-35	+2.3	87%	1654	13%	5%
1499	Heat 10-11	East	1612	46-36	+1.9	79%	1643	7%	3%
1588	Raptors 11-9	East	1616	48-34	+2.8	86%	1624	6%	3%
1560	Pelicans 12-8	West	1589	48-34	+3.0	86%	1606	7%	2%
1581	Mavericks 9-10	West	1586	45-37	+2.1	76%	1604	6%	2%
1534	Cavaliers 13-8	East	1581	48-34	+3.0	83%	1599	4%	1%
1432	Clippers 12-9	West	1568	42-40	-0.1	64%	1596	4%	1%
1556	Nets 11-11	East	1566	43-39	+1.0	62%	1598	3%	1%
1548	Hawks 11-10	East	1567	44-38	+0.9	67%	1580	2%	0.8%
1511	Timberwolves 10-11	West	1532	40-42	-0.6	45%	1539	1%	0.3%

For our project, we decided to create a simplified version of this model using team-wide statistics to predict what will happen in the current 2022-23 NBA season, which was about 25% complete at the conclusion of our project. To accomplish this, we decided to build a dataset containing a wide variety of statistics and choose a handful that had a moderate to strong correlation with winning (using linear regression models). Then, we would develop a formula that weighted all of these statistics and output a number that would represent the expected number of wins for that team in a full season (either 82 or 72 games, depending on the year).

1.2 Data Gathering

One unique advantage to conducting a data science project in the world of sports is the relatively easy access to extensive cleaned datasets. For our project, we used `basketball-reference.com` for all of our data. As a free source, it doesn't provide many of the cutting-edge aggregate statistics used in more advanced models, but it will work fine for our purposes.

For each of the five seasons that we analysed (from 2018 to 2022), we downloaded the data we needed into an Excel spreadsheet, and after some modifications there, we imported it into Python for further analysis:

```
In [4]: df_2021.head()
Out[4]:
```

	TEAM	PLAYOFFS	AGE	W	L	FG	FGA	FG RATE	3P	3PA	...	DEF EFG	DEF TOV	DEF DRB	DEF FT/FGA	DIST	LAYUP FREQ	CORNER FREQ	CORNER RATE	AST/TOV	Corner 3-pt Per Game
0	Atlanta Hawks	Made Playoffs	26.1	43	39	41.5	88.3	0.470	12.9	34.4	...	0.543	11.5	76.9	0.177	14.5	0.293	0.221	0.426	2.067227	3.238622
1	Boston Celtics	Made Final Four	26.1	51	31	40.7	87.4	0.466	13.2	37.1	...	0.502	12.5	77.3	0.183	14.8	0.334	0.231	0.389	1.823529	3.333769
2	Brooklyn Nets	Made Playoffs	29.1	44	38	42.0	88.4	0.475	11.5	31.7	...	0.521	11.7	75.1	0.201	13.9	0.304	0.235	0.393	1.794326	2.927653
3	Charlotte Hornets	Missed Playoffs	25.5	43	39	42.8	91.4	0.468	13.9	38.2	...	0.544	13.1	74.8	0.187	13.8	0.297	0.273	0.418	2.112782	4.359155
4	Chicago Bulls	Made Playoffs	26.3	46	36	41.7	86.9	0.480	10.6	28.8	...	0.541	11.9	78.3	0.199	13.9	0.364	0.287	0.372	1.867187	3.074803

2 Methodology

2.1 Exploratory Data Analysis

The first step in our project was to conduct exploratory data analysis on the various statistics in our dataset. We decided to use data from the most recent complete full season (2021-22) for this section.

For each statistic, we created a scatterplot with the stat on the x-axis and each team's total wins on the y-axis, fit a linear regression model to the data, and calculated the r^2 value. We were able to speed up this process using the following python script and looping over all columns in the dataset:

```
In [7]: # Function to make scatterplot and generate regression line and r^2 value given dataframe and statistic column

def test_statistic(df, stat, figsize = (8, 5), y = "W", playoffs = "PLAYOFFS", s = 50,
                   title = 'Default', xlabel = 'Default', ylabel = 'Default'):

    if title == 'Default':
        title = f'{stat} effect on Wins'
    if xlabel == 'Default':
        xlabel = stat
    if ylabel == 'Default':
        ylabel = 'Wins'

    plt.figure(figsize = figsize)
    scatter = sns.scatterplot(x = stat, y = y, data = df, hue = playoffs, palette = colors, s = s)
    scatter.set(title = title, xlabel = xlabel, ylabel = ylabel)

    x = df[stat].values.reshape(-1, 1)
    y = df[y].values
    model = LinearRegression().fit(x, y)
    plt.plot(x, model.predict(x), color = 'black', alpha = 0.6)
    r_2 = r2_score(y, model.predict(x))

    r2_dict[stat] = r_2

    plt.show()

    print(f'R-squared: {r_2}')
```

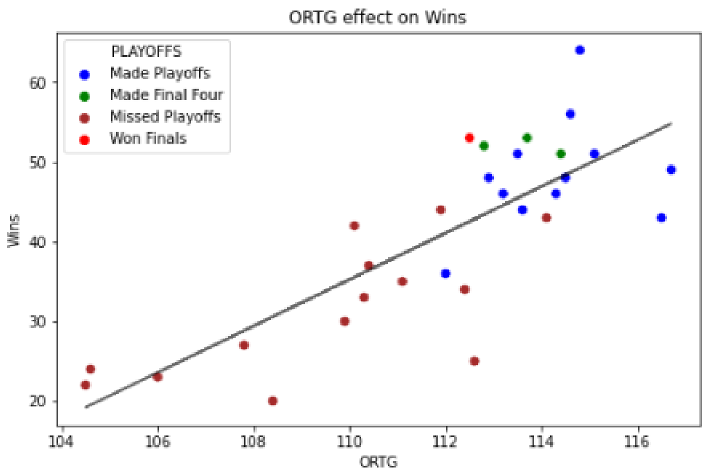
All graphs seen throughout this paper will either use this exact function, or a modified version of it. It is easy to see the general trend for each statistic in regards to wins, and one can also use the color scheme to see how the season's most successful teams (in regards to playoff success) performed. This difference between regular season wins and playoff success will be touched on later.

From here, we then began to create some new statistics by combining existing ones to try to find stronger correlations, an example of which can be seen in the following section.

2.1.1 Example: Net Rating

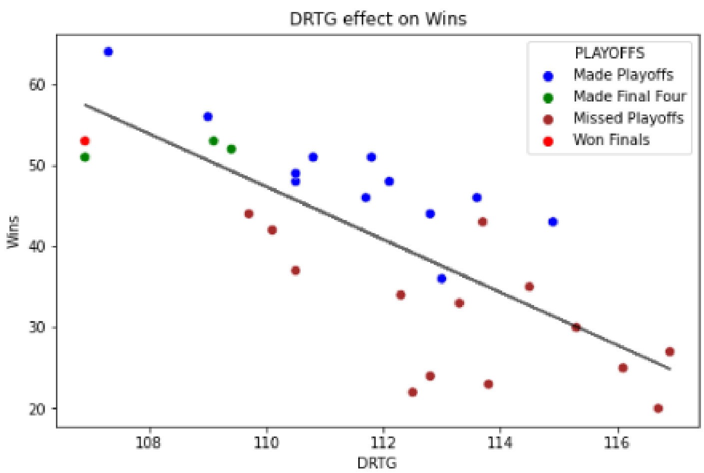
While the vast majority of statistics that we analysed had little to no correlation with winning, one stood out far above the rest: net rating (NRTG). This statistic is calculated by finding the difference between two statistics in our dataset: offensive rating (ORTG), and defensive rating (DRTG).

Offensive rating represents the average number of points a team scores per 100 possessions (higher is better). This had a pretty strong correlation with winning, with an r^2 value of 0.639:



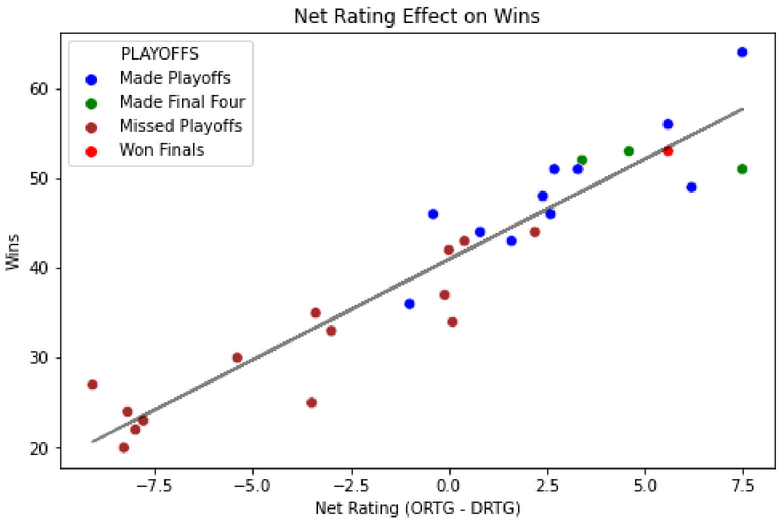

```
R-squared: 0.6389571497748235
```

Defensive rating represents the average number of points a team allows per 100 possessions (lower is better). This also had a pretty strong correlation with winning, with an r^2 value of 0.593:




```
R-squared: 0.5929527875025281
```

However, these numbers could be improved significantly by looking at net rating (ORTG - DRTG). Intuitively, this works because a team may be very good at offense (and has a great ORTG), but if they're very bad at defense (and has an awful DRTG), they still won't win a lot of games. However, if a team has a good net rating, then they must have a good offense and good defense. This can be seen with a much stronger r^2 value of 0.895:




```
R-squared: 0.8946227362831126
```

Net rating is such a strong predictor of a team's success that we decided to use it as the foundation of our model, and the majority of a team's score will be calculated from their NRTG.

2.2 Model Development

After repeating the previous steps for nearly 100 different statistics, we chose 5 to build our model around. In this section, we will go over each of the five statistics, and discuss why they were chosen and how much they will be weighted in the model.

2.2.1 Statistic 1: Net Rating

Model Weight: 70%

As seen previously, Net Rating is an extremely strong predictor of a team's success. We decided that a weight of 70% was best for our model, as it will provide a strong baseline for each team's score, but won't dominate the model too much. With the remaining 30%, we can use other statistics to develop a stronger correlation with winning than NRTG does by itself.

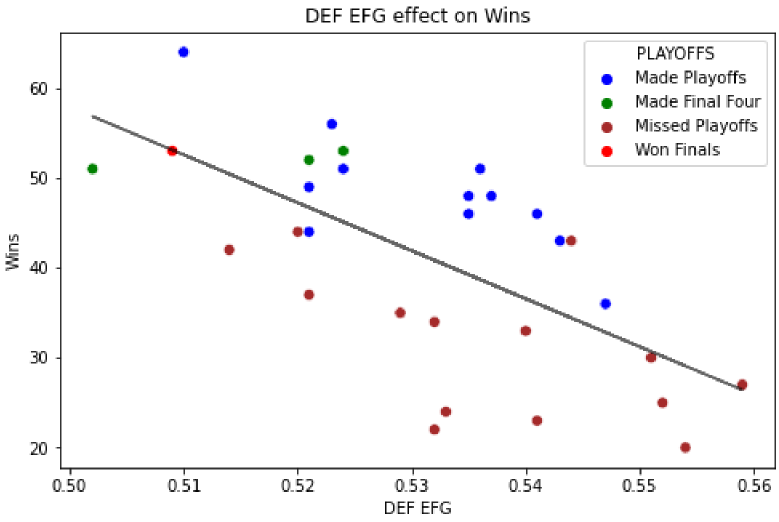
2.2.2 Statistic 2: Defensive EFG%

Model Weight: 13%

Along with the DRTG half of NRTG, Defensive EFG% is the only other defensive statistic in our model. This stat represents how well opposing teams shoot against the team's defense, adjusting for the difficulty of the shot (for example, making a 3-pointer increases the percentage more than making a 2-pointer would.

Since the statistic is a percentage of how many shots are made against our team, a lower score is better.

In 2021-22, Defensive EFG% had a moderate correlation with winning, with an r^2 value of 0.428:




```
R-squared: 0.42811115977657477
```

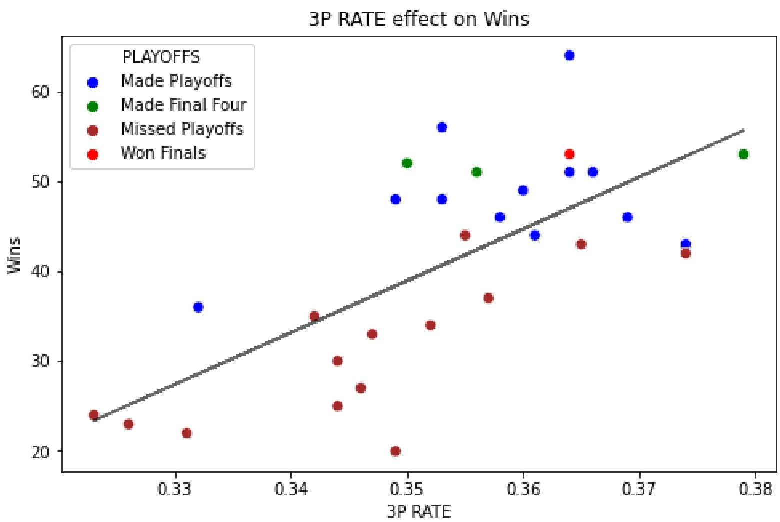
2.2.3 Statistic 3: 3-Point Percentage

Model Weight: 11%

3-Point Percentage refers to the percentage of three point shots a team makes. While this is by far the simplest statistic included in our model, we found that it had a surprisingly high correlation with winning.

While a great individual player can shoot over 40% from 3, it is incredibly rare for an entire team to shoot anywhere close to that high. Additionally, as 3-pointers have become a greater focus of the modern game, team 3-point percentages have significantly increased over the years.

In 2021-22, 3-Point Percentage had a moderate correlation with winning, with an r^2 value of 0.480:



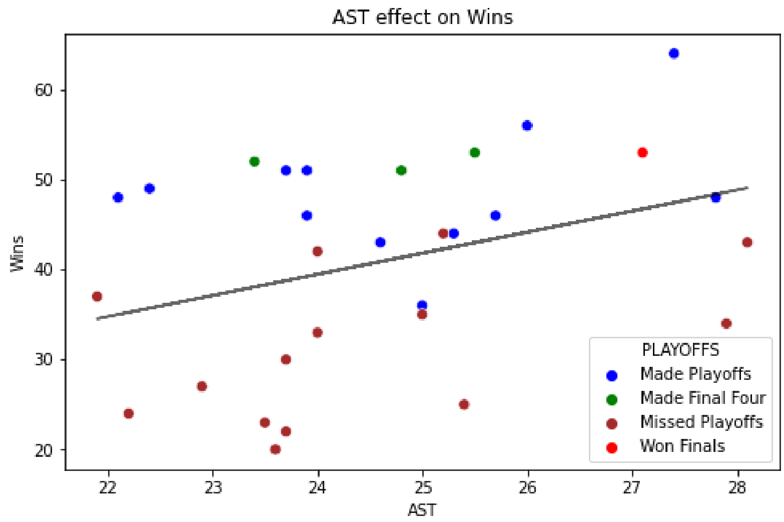

```
R-squared: 0.48007715010468777
```

2.2.4 Statistic 4: Assist/Turnover Ratio

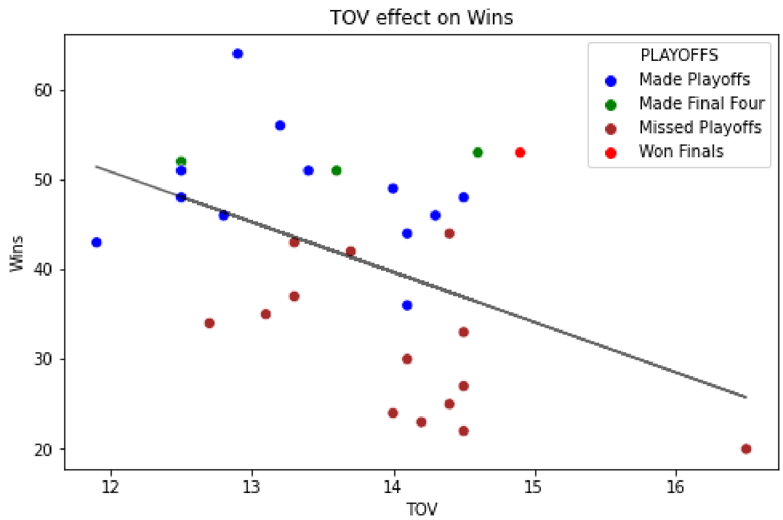
Model Weight: 4%

Assist/Turnover Ratio refers to the number of team assists per turnover. Since an assist comes as a result of a good pass, and a turnover often comes as a result of a bad pass, the AST/TOV ratio is known as a good estimate for how good a team is at passing.

Interestingly enough, assists and turnovers alone didn't have a very strong correlation to winning, with r^2 values of 0.125 and 0.207 respectively:

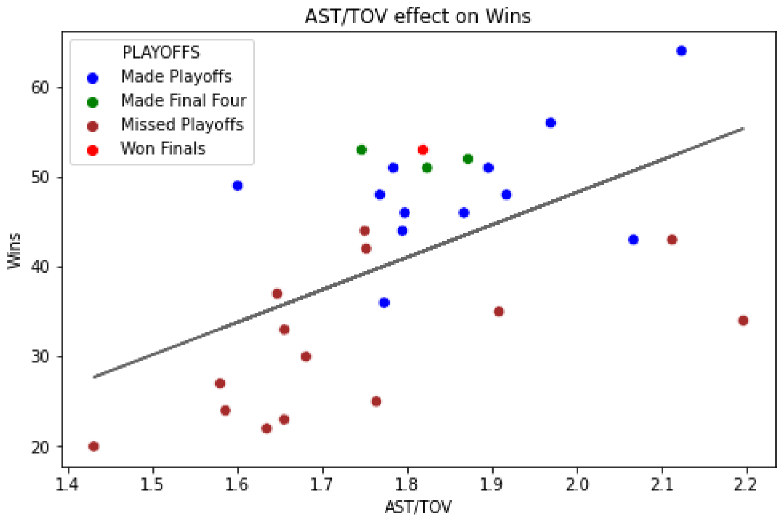



```
R-squared: 0.12462828425935257
```




```
R-squared: 0.20709105774634007
```

However, when combining the two into a ratio, the r^2 value increases to represent a weak to moderate correlation of 0.300:




```
R-squared: 0.30012218167053795
```

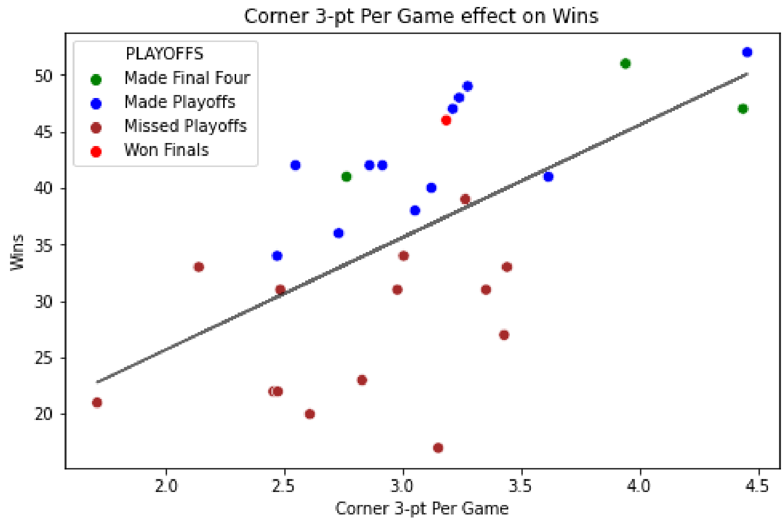
By taking magnitude out of the equation and looking at this data as a ratio, we were able to find a stronger correlation with winning. It appears that a team's passing efficiency (good passes per bad pass) is a better predictor of success than a team's passing frequency (total number of good passes).

2.2.5 Statistic 5: Avg. Corner 3-Point Shots Per Game

Model Weight: 2%

One shot at the forefront of modern data analytics in the NBA is the corner 3-pointer. Due to the shape of the court, this shot is slightly shorter than a normal 3-pointer (and therefore easier to make), but still rewards the team with three points when made. From a mathematical standpoint, no shot is more desirable than a proficient shooter taking an open corner 3-pointer, outside of an open dunk or layup.

While many teams are still stuck in the past, a few have begun emphasizing this shot in their game plans, designing plays to generate open corner threes and even signing players to focus entirely on shooting corner 3s and playing defense. This can be seen in the graph below, as while there isn't a very strong correlation with winning ($r^2 = 0.35$), a few teams shoot more corner 3s than everyone else - and all of them are incredibly successful at the moment. For this reason, we decided to give a small bonus to teams that shot a lot of corner 3s in our model.




```
R-squared: 0.35018765881148994
```

2.2.6 Generating Team Score

Using these five statistics, it's now time to calculate each team's score. To do this, we use the following formula:

$$0.70 * \frac{NRTG + 12.5}{25} + 0.13 * \frac{DEFEFG - 0.5}{0.06} + 0.11 * \frac{3PRATE - 0.31}{0.07} +$$
$$0.04 * \frac{AST/TOV - 1.5}{0.6} + 0.02 * \frac{CORNER3/G - 2.5}{1.5}$$

The formula assigns each team a score between 0 and 1 for each of the five statistics by finding where it fell on the range of scores over the past five seasons in a linear normalization. For example, NRTG tends to be distributed between -12.5 and +12.5 (with most teams near the middle of that range), so a team with a NRTG of 0 would receive a NRTG score of 0.5.

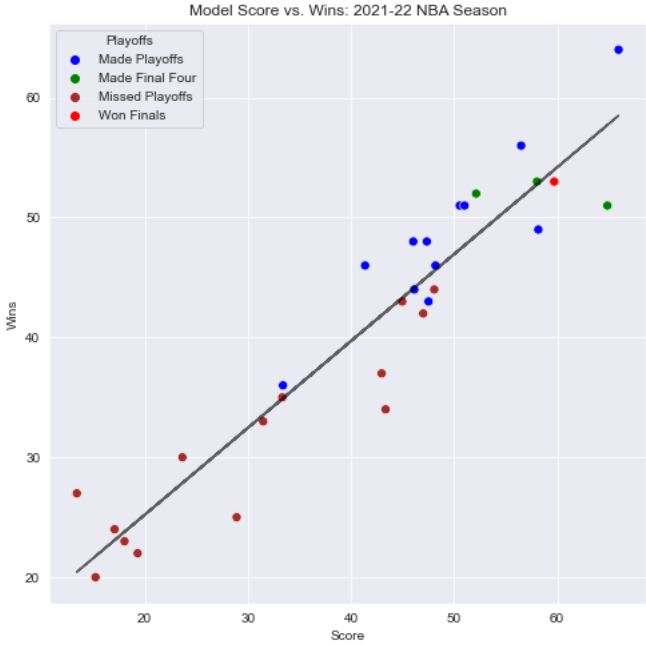
Then, each statistic score was multiplied by the model weight of that statistic, and all of this was added together. Continuing the previous example, since NRTG had a weight of 0.7, a team with a NRTG of 0 would contribute 0.35 to its final score from NRTG alone.

We then multiplied each team's score, which is roughly distributed between 0 and 1, by the number of games in a season (which is 82 in a normal season, but was 72 in the covid-shortened seasons of 2019-20 and 2020-21) to get an estimated number of wins for each team.

3 Discussion of the Results

3.1 2021-22 Season (Original Data)

After constructing the model, we began by testing it on the 2021-22 season, using the same data that we conducted our EDA on. This generated the following results:




```
R-squared: 0.894503871749798
```

Team	Wins	Rank	Score	Modeled Rank	Difference
Phoenix Suns	64	1	65.9608	1	1.9608
Memphis Grizzlies	56	2	56.5062	6	0.5062
Golden State Warriors	53	3	59.7124	3	6.7124
Miami Heat	53	3	58.0478	5	5.0478
Dallas Mavericks	52	5	52.1356	7	0.1356
Boston Celtics	51	6	64.8702	2	13.8702
Philadelphia 76ers	51	6	51.0122	8	0.0122
Milwaukee Bucks	51	6	50.5448	9	-0.4552
Utah Jazz	49	9	58.1708	4	9.1708
Toronto Raptors	48	10	46.0348	16	-1.9652
Denver Nuggets	48	10	47.3550	13	-0.6450
Chicago Bulls	46	12	41.3608	20	-4.6392
Minnesota Timberwolves	46	12	48.1914	10	2.1914
Brooklyn Nets	44	14	46.1332	15	2.1332
Cleveland Cavaliers	44	14	48.0766	11	4.0766
Atlanta Hawks	43	16	47.5190	12	4.5190
Charlotte Hornets	43	16	44.9688	17	1.9688
Los Angeles Clippers	42	18	47.0024	14	5.0024
New York Knicks	37	19	42.9680	19	5.9680
New Orleans Pelicans	36	20	33.3986	21	-2.6014
Washington Wizards	35	21	33.3248	22	-1.6752
San Antonio Spurs	34	22	43.3534	18	9.3534
Los Angeles Lakers	33	23	31.4716	23	-1.5284
Sacramento Kings	30	24	23.6324	25	-6.3676
Portland Trail Blazers	27	25	13.3906	30	-13.6094
Indiana Pacers	25	26	28.8968	24	3.8968
Oklahoma City Thunder	24	27	17.0560	28	-6.9440
Detroit Pistons	23	28	18.0154	27	-4.9846
Orlando Magic	22	29	19.2946	26	-2.7054
Houston Rockets	20	30	15.2028	29	-4.7972

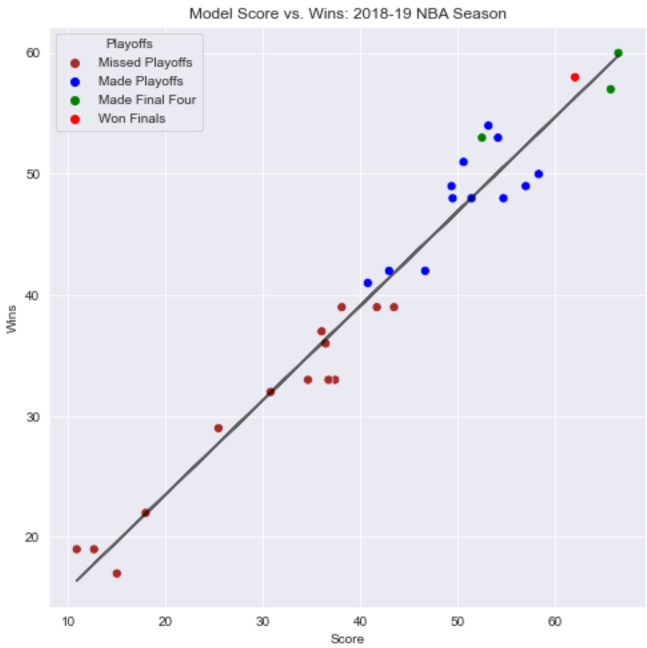
While the correlation between our model's score and regular season wins was about what we expected ($r^2 = 0.895$), there were a few interesting things we noted.

First, our model overrated all of the teams that made deep playoff runs to the conference finals (the last four teams in the playoffs) including the Celtics (+13.9, by far the biggest positive outlier) and the Warriors (+6.7).

Since many teams don't try their hardest to win every regular season game, but give everything they have to win in the playoffs, an argument can be made that regular season wins is just another statistic that can be used to estimate how good a team is. This suggests that our model could be better at finding the best teams than regular season wins alone, but further analysis would need to be done to confirm that.

3.2 2018-19 Season

The 2018-2019 season provided the best results out of all the seasons we studied, with a correlation of 0.958 between wins and model predicted wins. Nearly all playoff teams were correctly predicted, and the league's best teams again set themselves apart.



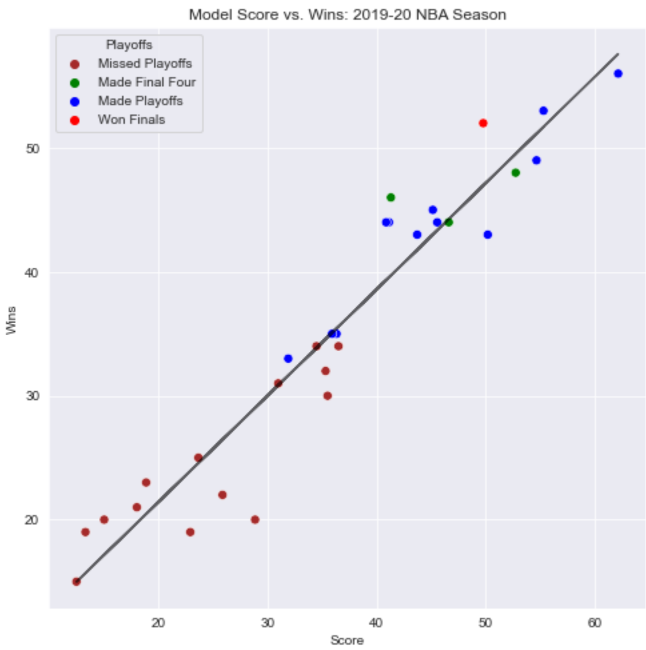

```
R-squared: 0.9580910796022633
```

	Team	Wins	Rank	Score	Modeled Rank	Difference
16	Milwaukee Bucks*	60	1	66.5184	1	6.5184
27	Toronto Raptors*	58	2	62.0658	3	4.0658
9	Golden State Warriors*	57	3	65.7312	2	8.7312
7	Denver Nuggets*	54	4	53.1688	8	-0.8312
24	Portland Trail Blazers*	53	5	52.5046	9	-0.4954
10	Houston Rockets*	53	5	54.1528	7	1.1528
22	Philadelphia 76ers*	51	7	50.6104	11	-0.3896
28	Utah Jazz*	50	8	58.3348	4	8.3348
1	Boston Celtics*	49	9	57.0064	5	8.0064
20	Oklahoma City Thunder*	49	9	49.3804	13	0.3804
12	Los Angeles Clippers*	48	11	49.4952	12	1.4952
11	Indiana Pacers*	48	11	54.7104	6	6.7104
26	San Antonio Spurs*	48	11	51.4304	10	3.4304
2	Brooklyn Nets*	42	14	42.9598	16	0.9598
21	Orlando Magic*	42	14	46.6662	14	4.6662
8	Detroit Pistons*	41	16	40.7786	18	-0.2214
15	Miami Heat	39	17	43.4682	15	4.4682
3	Charlotte Hornets	39	17	38.0972	19	-0.9028
25	Sacramento Kings	39	17	41.7052	17	2.7052
13	Los Angeles Lakers	37	20	36.0308	23	-0.9692
17	Minnesota Timberwolves	36	21	36.4408	22	0.4408
6	Dallas Mavericks	33	22	37.4084	20	4.4084
18	New Orleans Pelicans	33	22	36.7360	21	3.7360
14	Memphis Grizzlies	33	22	34.6286	24	1.6286
29	Washington Wizards	32	25	30.7910	25	-1.2090
0	Atlanta Hawks	29	26	25.4364	26	-3.5636
4	Chicago Bulls	22	27	17.9498	27	-4.0502
5	Cleveland Cavaliers	19	28	10.8650	30	-8.1350
23	Phoenix Suns	19	28	12.6444	29	-6.3556
19	New York Knicks	17	30	14.9896	28	-2.0104

3.3 2019-20 Season

The 2019-20 season was an interesting one. This was the first of two shortened seasons caused by the Covid-19 Pandemic, with only 72 regular season games played. Additionally, this year's playoffs were played in a "bubble" with no fans in attendance, leading to many fascinating statistical outliers. One of these was the fact that many teams shot the ball far more efficiently than they had in the past, likely due to not having to deal with crowd noise during games and travel-related distractions between games.

The Los Angeles Lakers won the finals this year despite our model only ranking them 6th - the lowest ranking of any champion that we studied. One reason for this is that the Lakers weren't a great defensive or 3-point shooting team during the regular season (34.9%), which our model didn't like, but they improved significantly in the playoffs, especially from their star players.



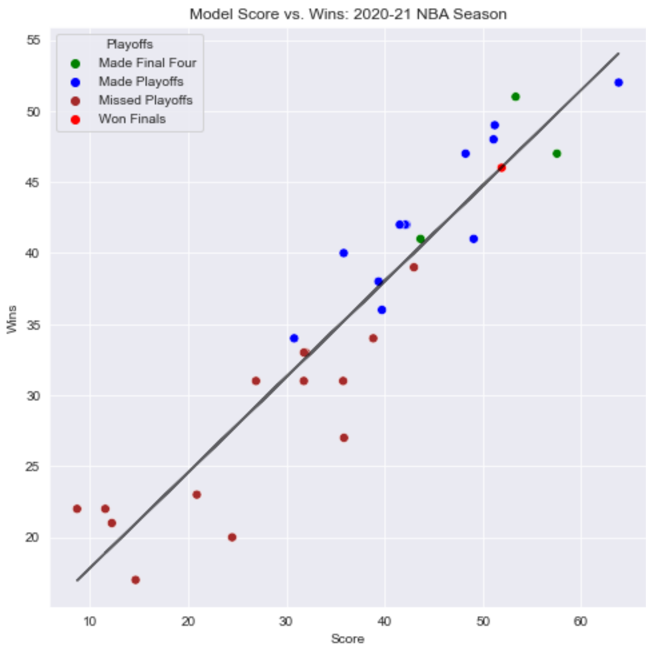

```
R-squared: 0.9200875384236137
```

	Team	Wins	Rank	Score	Modeled Rank	Difference
16	Milwaukee Bucks*	56	1	62.1216	1	6.1216
27	Toronto Raptors*	53	2	55.2816	2	2.2816
13	Los Angeles Lakers*	52	3	49.7592	6	-2.2408
12	Los Angeles Clippers*	49	4	54.6408	3	5.6408
1	Boston Celtics*	48	5	52.7328	4	4.7328
7	Denver Nuggets*	46	6	41.3064	11	-4.6936
11	Indiana Pacers*	45	7	45.1440	9	0.1440
10	Houston Rockets*	44	8	41.1264	12	-2.8736
15	Miami Heat*	44	8	46.6056	7	2.6056
28	Utah Jazz*	44	8	45.5472	8	1.5472
20	Oklahoma City Thunder*	44	8	40.8672	13	-3.1328
6	Dallas Mavericks*	43	12	50.1768	5	7.1768
22	Philadelphia 76ers*	43	12	43.7112	10	0.7112
24	Portland Trail Blazers*	35	14	35.9136	16	0.9136
2	Brooklyn Nets*	35	14	36.3240	15	1.3240
23	Phoenix Suns	34	16	36.5040	14	2.5040
14	Memphis Grizzlies	34	16	34.4880	19	0.4880
21	Orlando Magic*	33	18	31.9032	20	-1.0968
26	San Antonio Spurs	32	19	35.3232	18	3.3232
25	Sacramento Kings	31	20	30.9816	21	-0.0184
18	New Orleans Pelicans	30	21	35.5032	17	5.5032
29	Washington Wizards	25	22	23.6736	24	-1.3264
3	Charlotte Hornets	23	23	18.8856	26	-4.1144
4	Chicago Bulls	22	24	25.8840	23	3.8840
19	New York Knicks	21	25	18.0360	27	-2.9640
8	Detroit Pistons	20	26	28.8648	22	8.8648
0	Atlanta Hawks	20	26	15.0408	28	-4.9592
17	Minnesota Timberwolves	19	28	22.9320	25	3.9320
5	Cleveland Cavaliers	19	28	13.3272	29	-5.6728
9	Golden State Warriors	15	30	12.4992	30	-2.5008

3.4 2020-21 Season

Nothing too special happened in the 2020-21 season regarding our analysis. While the correlation between wins and model predicted wins wasn't as strong as the previous two years, nearly all playoff teams were still predicted correctly.

In fact, the only incorrect playoff prediction was the Golden State Warriors, who received a high score from our model (predicted 42.99 wins, actual 39 wins) likely due to our model recognizing their exceptional 3-point shooting ability, but failing to recognize all of the team's weaknesses.




```
R-squared: 0.8833750678771104
```

	Team	Wins	Rank	Score	Modeled Rank	Difference
28	Utah Jazz*	52	1	63.8424	1	11.8424
23	Phoenix Suns*	51	2	53.3520	3	2.3520
22	Philadelphia 76ers*	49	3	51.2208	5	2.2208
2	Brooklyn Nets*	48	4	51.0984	6	3.0984
7	Denver Nuggets*	47	5	48.2472	8	1.2472
12	Los Angeles Clippers*	47	5	57.5568	2	10.5568
16	Milwaukee Bucks*	46	7	51.9192	4	5.9192
24	Portland Trail Blazers*	42	8	41.5440	13	-0.4560
6	Dallas Mavericks*	42	8	42.2280	11	0.2280
13	Los Angeles Lakers*	42	8	42.0912	12	0.0912
0	Atlanta Hawks*	41	11	43.6608	9	2.6608
19	New York Knicks*	41	11	49.0752	7	8.0752
15	Miami Heat*	40	13	35.8344	18	-4.1656
9	Golden State Warriors	39	14	42.9912	10	3.9912
14	Memphis Grizzlies*	38	15	39.3840	15	1.3840
1	Boston Celtics*	36	16	39.7296	14	3.7296
11	Indiana Pacers	34	17	38.8440	16	4.8440
29	Washington Wizards*	34	17	30.7872	23	-3.2128
3	Charlotte Hornets	33	19	31.9896	20	-1.0104
26	San Antonio Spurs	33	19	31.7880	21	-1.2120
18	New Orleans Pelicans	31	21	31.7736	22	0.7736
4	Chicago Bulls	31	21	35.7768	19	4.7768
25	Sacramento Kings	31	21	26.8992	24	-4.1008
27	Toronto Raptors	27	24	35.8632	17	8.8632
17	Minnesota Timberwolves	23	25	20.8728	26	-2.1272
5	Cleveland Cavaliers	22	26	11.5488	29	-10.4512
20	Oklahoma City Thunder	22	26	8.6760	30	-13.3240
21	Orlando Magic	21	28	12.2184	28	-8.7816
8	Detroit Pistons	20	29	24.4728	25	4.4728
10	Houston Rockets	17	30	14.6376	27	-2.3624

3.5 Current 2022-23 Season: First 20 Games

Now, what we've all been waiting for: what will the model predict for the current NBA season?

At the time of writing, each team has played approximately 20 out of their 82 games, so the data we'll be working with is a bit limited. Plenty of weak teams get off to strong starts before falling off later in the season, while plenty of strong teams start off slowly before figuring everything out. Additionally, many mediocre to bad teams will realize that they aren't going to go on a deep playoff run as the season continues, causing them to start "tanking" (trading away their good players and losing on purpose to secure a higher draft pick the following year). Since our model can't predict that, we have to interpret the results while knowing which teams are decent now, but could start tank later in the year (such as the Pacers, Jazz, and Thunder).

Our model predicted the following final standings for the Eastern Conference:

TEAM	REC	PCT
Boston Celtics	66-16	0.805
Cleveland Cavaliers	61-21	0.744
Milwaukee Bucks	54-28	0.659
Philadelphia 76ers	54-28	0.659
Brooklyn Nets	48-34	0.585
Indiana Pacers	44-38	0.537
Chicago Bulls	39-43	0.476
Atlanta Hawks	38-44	0.463
Toronto Raptors	37-45	0.451
Washington Wizards	37-45	0.451
Miami Heat	35-47	0.427
New York Knicks	31-51	0.378
Charlotte Hornets	26-56	0.317
Orlando Magic	21-61	0.256
Detroit Pistons	17-65	0.207

And the Western Conference:

TEAM	REC	PCT
Phoenix Suns	64-18	0.780
New Orleans Pelicans	56-26	0.683
Utah Jazz	46-36	0.561
Golden State Warriors	45-37	0.549
Memphis Grizzlies	44-38	0.537
Denver Nuggets	44-38	0.537
Dallas Mavericks	43-39	0.524
Sacramento Kings	42-40	0.512
Los Angeles Clippers	42-40	0.512
Portland Trail Blazers	39-43	0.476
Minnesota Timberwolves	36-46	0.439
Oklahoma City Thunder	35-47	0.427
Los Angeles Lakers	31-51	0.378
Houston Rockets	21-61	0.256
San Antonio Spurs	8-64	0.098

The model predicts an eventual finals matchup between the Boston Celtics and Phoenix Suns, who have set themselves apart from the other 28 teams.

Additionally, it expects the San Antonio Spurs to have one of the worst seasons of any team in NBA history. Given that the Spurs started to tank well before the season even began by trading their best player to the Hawks, this prediction isn't as outlandish as it seems.

When comparing these results to the current NBA standings, we see that our model believes teams like the Warriors and Cavaliers will improve throughout the year (which makes sense, as they are strong defensive and 3-point shooting teams) while teams like the Miami Heat and Toronto Raptors will decline.

4 Conclusion

Overall, we are happy with how our model performed given its limited input. It is difficult to predict who will win an NBA game given only team-wide statistics, as there are so many other factors at play.

If we wanted to improve the model, we would pursue one of the following:

1. **Include individual statistics, not just team-wide statistics.**

   The modern NBA is dominated by a select few star players, especially at the end of close games. Teams that have one or two superstars and a bunch of decent players tend to be more successful than teams that have a bunch of good players, but no true star. However, since our model only looks at team statistics, it would prefer the latter team.

2. **Incorporate Head to Head Data**

   Incorporating head to head data (which teams beat which other teams) into our model would likely improve its predictions. FiveThirtyEight's ELO model is primarily focused on this data, and it has been very successful.

3. **Use More Advanced Statistics**

   As previously mentioned, we were limited to relatively simple statistics in our project. There are many advanced statistics being developed by data scientists working for teams and sportsbooks, and it would be interesting to see how our model performs using those. Unfortunately, they are not freely available online.

5 References

All data and information used in our project was acquired from the links below.

2018-19 NBA Season Data:
https://www.basketball-reference.com/leagues/NBA_2019.html

2019-20 NBA Season Data:
https://www.basketball-reference.com/leagues/NBA_2020.html

2020-21 NBA Season Data:
https://www.basketball-reference.com/leagues/NBA_2021.html

2021-22 NBA Season Data:
https://www.basketball-reference.com/leagues/NBA_2022.html

2022-23 NBA Season Data:
https://www.basketball-reference.com/leagues/NBA_2023.html

FiveThirtyEight Predictions:
https://projects.fivethirtyeight.com/2023-nba-predictions/

GRADESCOPE SUBMISSION NOTE: The final drafts of our datasets and EDA/analysis, in addition to our presentation slides, are included in the github repository below. Much of our preliminary analysis and unused data was not saved.
https://github.com/Kc1227/Math-42-Final-Project