Predicting the 2018-2019 NBA Regular Season using Pythagorean Wins and Monte Carlo Methods

Jeremy Lu

1. Introduction

The 2018-2019 NBA Regular Season began on October 16, 2018 and ended on April 10, 2019. It featured a total of 1230 games, with each of the 30 teams playing 82. At the season's conclusion, the Milwaukee Bucks finished as the top team, with a record of 60-22 (60 wins and 22 losses). The full standings are as follows:

Eastern Conference			
	W	L	Pct
Milwaukee Bucks	60	22	.732
Toronto Raptors	58	24	.707
Philadelphia 76ers	51	31	.622
Boston Celtics	49	33	.598
Indiana Pacers	48	34	.585
Brooklyn Nets	42	40	.512
Orlando Magic	42	40	.512
Detroit Pistons	41	41	.500
Charlotte Hornets	39	43	.476
Miami Heat	39	43	.476
Washington Wizards	32	50	.390
Atlanta Hawks	29	53	.354
Chicago Bulls	22	60	.268
Cleveland Cavaliers	19	63	.232
New York Knicks	17	65	.207

Western Conference			
	W	L	Pct
Golden State Warriors	57	25	.695
Denver Nuggets	54	28	.659
Portland Trail Blazers	53	29	.646
Houston Rockets	53	29	.646
Utah Jazz	50	32	.610
Oklahoma City Thunder	49	33	.598
San Antonio Spurs	48	34	.585
Los Angeles Clippers	48	34	.585
Sacramento Kings	39	43	.476
Los Angeles Lakers	37	45	.451
Minnesota Timberwolves	36	46	.439
Memphis Grizzlies	33	49	.402
New Orleans Pelicans	33	49	.402
Dallas Mavericks	33	49	.402
Phoenix Suns	19	63	.232

Table 1: Regular Season Standings

Additionally, all the statistics used in this project are courtesy of Basketball-Reference.

2. Simulating Games

2.1. Terminology

Before diving into the details of simulating, the NBA Season, some important statistical terms that are used must be defined:

- **Offensive Rating (ORtg)** - The number of points a team scores per 100 possessions. This means that the team with a higher ORtg will be considered, by this metric, to have a better offense. For example, a 110.3 ORtg means that a team scores 110.3 points every 100 possessions. During the 2018-2019 regular season, the mean team ORtg was 110.4. The Golden State Warriors had the league's top ORtg, of 15.9, while the New York Knicks had the league's worst ORtg of 104.5.

- **Defensive Rating (DRtg)** - The number of points a team allows scored per 100 possessions. This means that the team with a lower DRtg will be considered, by this metric, to have a better defense. For example, a 98.7 DRtg means that a team allows opponents to score 98.7 points every 100 possessions. During the 2018-2019 regular season, the mean team DRtg was 110.4. The Milwaukee Bucks claimed the league's top DRtg, with 105.2 while the Cleveland Cavaliers finished last with a DRtg of 117.6

- **Pythagorean Wins** - Gives an expected win percentage based on the number of points a team has scored and allowed. The exact formula is:

$$\frac{\text{Points scored}^c}{\text{Points Scored}^c + \text{Points Allowed}^c} \tag{1}$$

  where c is a the Pythagorean Exponent. This exponent is usually determined by selecting the value which minimizes the root mean squared error (RMSE) between Pythagorean Win percentage and a season's win percentage. Basketball-Reference uses an exponent of 14, so to be consistent. this project will use $c = 14$ as well.

2.2. Simulating Single Game Scorelines

Pythagorean Wins is traditionally a measure of how well a team will do across an entire season, playing different opponents. When applying this to an individual game, Pythagorean Wins, based on its design, will fail to capture the win probability against a single and unique opponent. To remedy this, we can use the following formula and apply it to single games, where p is the probability that a team beats its opponent:

$$p = \frac{\text{Pythagorean Win Percentage}}{\text{Pythagorean Win Percentage} + \text{Opponent Pythagorean Win Percentage}} \tag{2}$$

This gives us the probability that a team will defeat its opponent. Furthermore, in order to determine a winner using a computer simulation, we can then generate a random number r from 0 to 1. If r falls in the range $[0, p]$, then the team wins; and if it is in $(p, 1]$, their opponent wins.

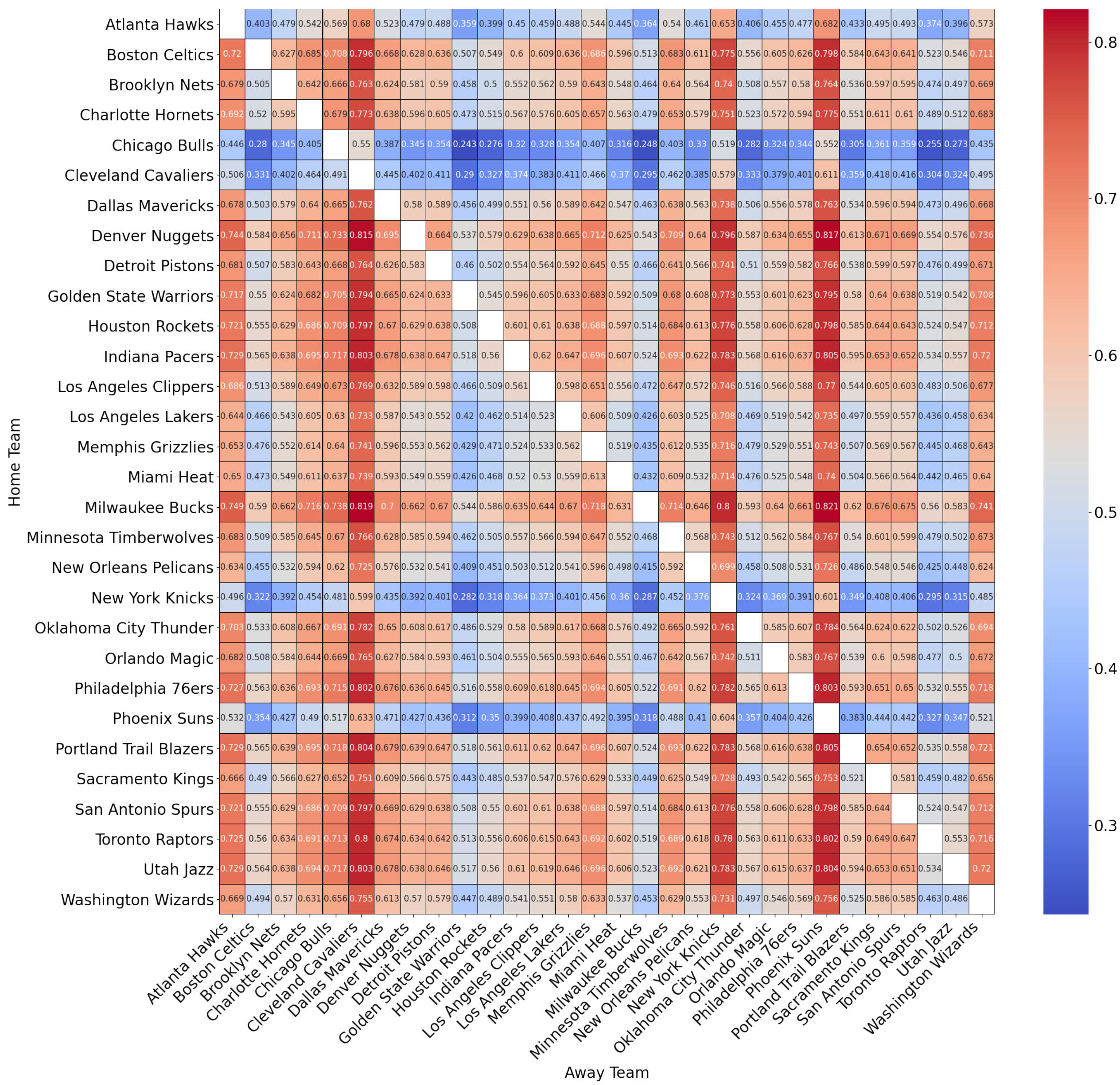
Another issue that arises while using Pythagorean Wins is that it is calculated using total points scored and total points allowed. However, this does not account for teams which play at different paces; some teams could score more points simply because they had more possessions, not because their offense was better. By using Offensive and Defensive Rating, which are per 100 possession metrics, we can better quantify offensive and defensive performances. Thus Pythagorean Win Percentage in this case will be calculated as

$$\frac{\text{ORtg}^c}{\text{ORtg}^c + \text{DRtg}^c} \tag{3}$$

We can also use Home/Away ORtg and DRtg to get Home and Away Pythagorean Win Percentages, in order to account for home court advantage.

As an example, if the league's best team by record, the Milwaukee Bucks, played against the league's worst team by record, the New York Knicks, in Milwaukee, the Bucks would have an expected win percentage of $\dfrac{0.815}{0.815+0.204} \approx 0.7998$. Simulating this match up 10,000 times using the aforementioned random number generation method, we see that the Bucks win against the Knicks 7894 times, or a win probability of 0.7894, which is relatively close to the expected value. In actuality, the Bucks defeated the Knicks twice at home (and also twice away).

Using this method, the win probability for every matchup is shown below:



To read this table, the home win probability for a Home vs. Away matchup corresponds to its cell. To find the away win probability, you can read this table from Away vs. Home and subtract that probability from 1. For example, using this matrix, the Bucks have a $0.8 \times 100 = 80\%$ chance of defeating the Knicks at home, and the Knicks have a

$(1 - 0.8) \times 100 = 20\%$ chance of defeating the Bucks on the road.

3. Results

3.1. Overall Simulated Team Performances

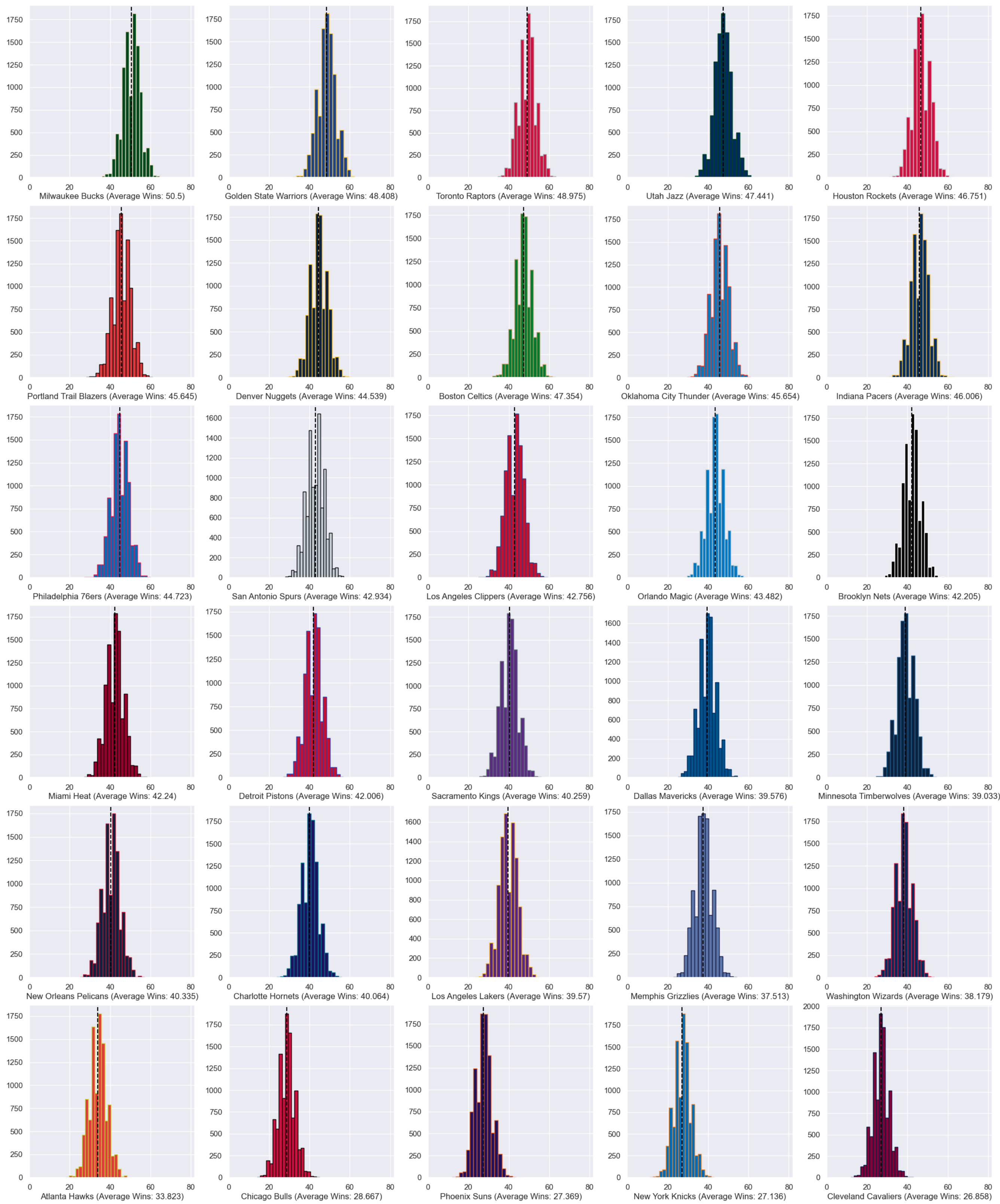
	2018-2019			Simulated					
	W	L	Pct	W	L	Pct	Δ W	95% Interval	(Max W, Min W)
Milwaukee Bucks	60	22	.732	51	31	.622	-9	(42, 59)	(68, 34)
Toronto Raptors	58	24	.707	49	33	.598	-9	(40, 57)	(65, 33)
Philadelphia 76ers	51	31	.622	45	37	.549	-6	(36, 53)	(60, 27)
Boston Celtics	49	33	.598	47	35	.573	-2	(39, 56)	(63, 32)
Indiana Pacers	48	34	.585	46	34	.561	-2	(37, 54)	(63, 29)
Brooklyn Nets	42	40	.512	42	40	.512	0	(33, 51)	(60, 27)
Orlando Magic	42	40	.512	43	39	.524	+1	(35, 52)	(59, 28)
Detroit Pistons	41	41	.500	42	40	.512	+1	(33, 51)	(60, 27)
Charlotte Hornets	39	43	.476	40	42	.488	+1	(32, 49)	(56, 22)
Miami Heat	39	43	.476	42	40	.512	+3	(34, 51)	(60, 27)
Washington Wizards	32	50	.390	38	44	.463	+6	(30, 47)	(55, 22)
Atlanta Hawks	29	53	.350	34	48	.415	+5	(25, 42)	(50, 18)
Chicago Bulls	22	60	.268	29	53	.350	+7	(21, 37)	(45, 14)
Cleveland Cavaliers	19	63	.232	27	55	.329	+8	(19, 35)	(43, 11)
New York Knicks	17	65	.207	27	55	.329	+10	(19, 35)	(43, 12)

Table 2: Eastern Conference Results

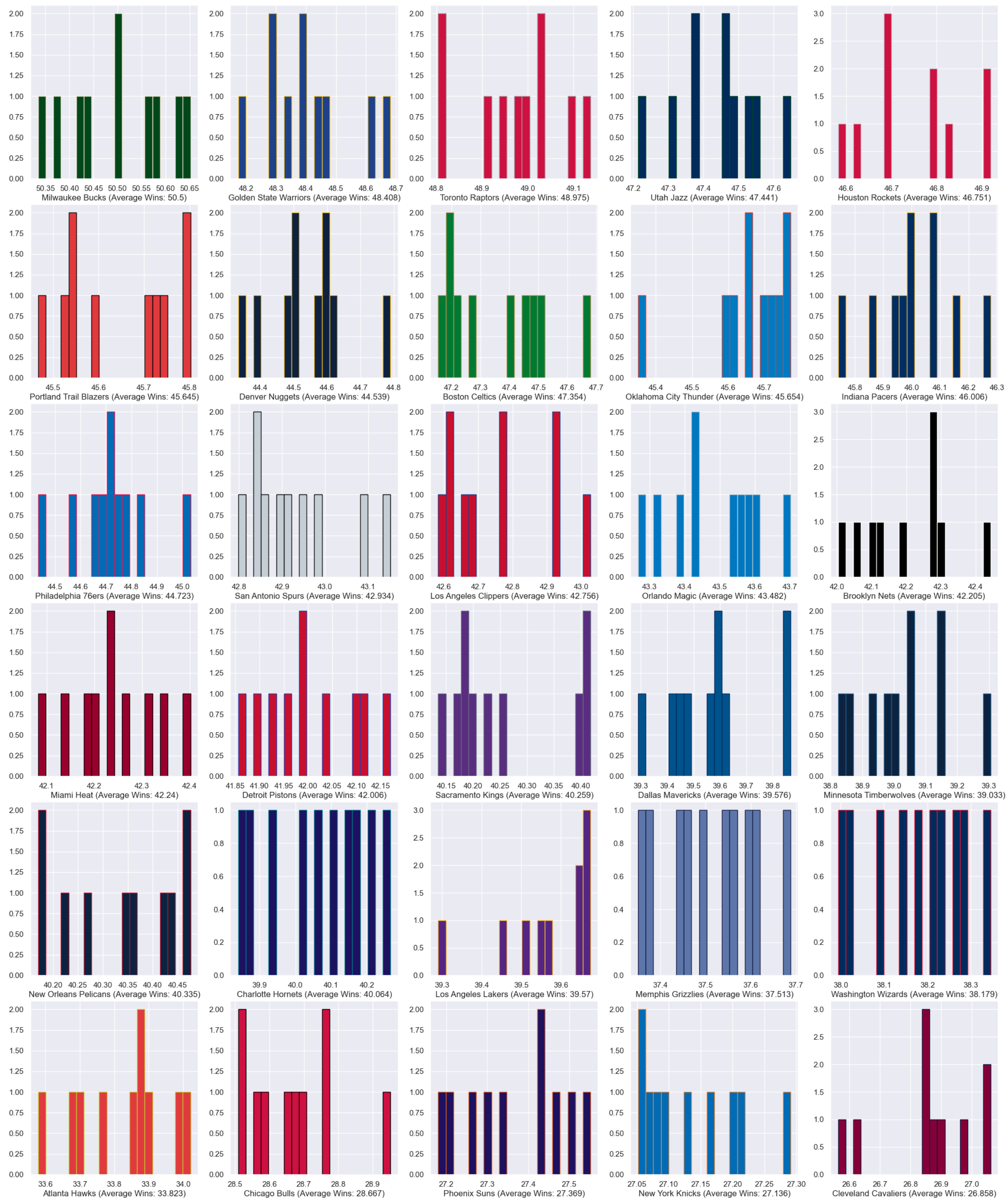
	2018-2019			Simulated					
	W	L	Pct	W	L	Pct	Δ W	95% Interval	(Max W, Min W)
Golden State Warriors	57	25	.695	48	36	.585	-9	(40, 57)	(64, 28)
Denver Nuggets	54	28	.659	45	37	.549	-9	(36, 53)	(61, 28)
Portland Trail Blazers	53	29	.646	46	36	.549	-7	(37, 54)	(61, 28)
Houston Rockets	53	29	.646	47	35	.573	-6	(38, 55)	(64, 29)
Utah Jazz	48	34	.585	47	35	.573	-1	(39, 56)	(65, 28)
Oklahoma City Thunder	49	33	.598	46	34	.561	-3	(37, 54)	(61, 28)
San Antonio Spurs	48	34	.585	43	39	.524	-5	(34, 51)	(57, 28)
Los Angeles Clippers	48	34	.585	43	39	.524	-5	(34, 51)	(61, 25)
Sacramento Kings	39	43	.476	40	42	.488	+1	(32, 49)	(56, 22)
Los Angeles Lakers	37	45	.451	42	40	.512	+5	(31, 48)	(55, 20)
Minnesota Timberwolves	36	46	.439	39	43	.476	+3	(30, 47)	(56, 21)
Memphis Grizzlies	33	49	.492	38	44	.463	+5	(25, 42)	(56, 21)
New Orleans Pelicans	33	49	.402	40	42	.488	+7	(32, 49)	(57, 25)
Dallas Mavericks	33	49	.402	40	42	.488	+7	(31, 48)	(56, 25)
Phoenix Suns	19	63	.232	27	55	.329	+8	(19, 36)	(47, 12)

Table 3: Western Conference Results

3.2. Histograms for Simulated Wins by Team



3.3. Distribution of Mean Wins (10 × 1000 games)



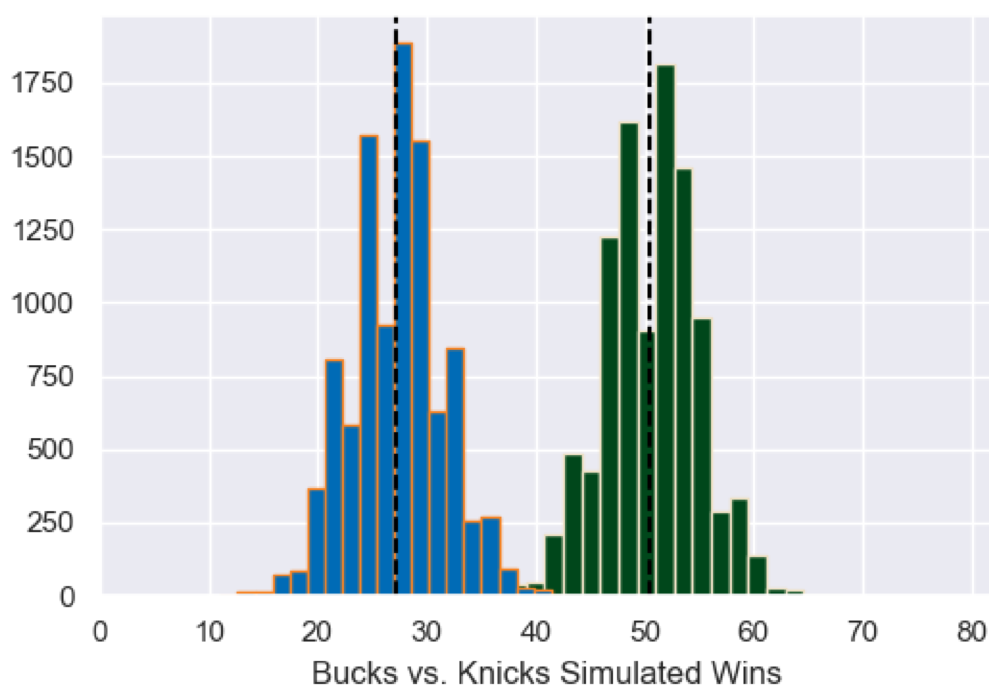
4. Analysis

4.1. Observations

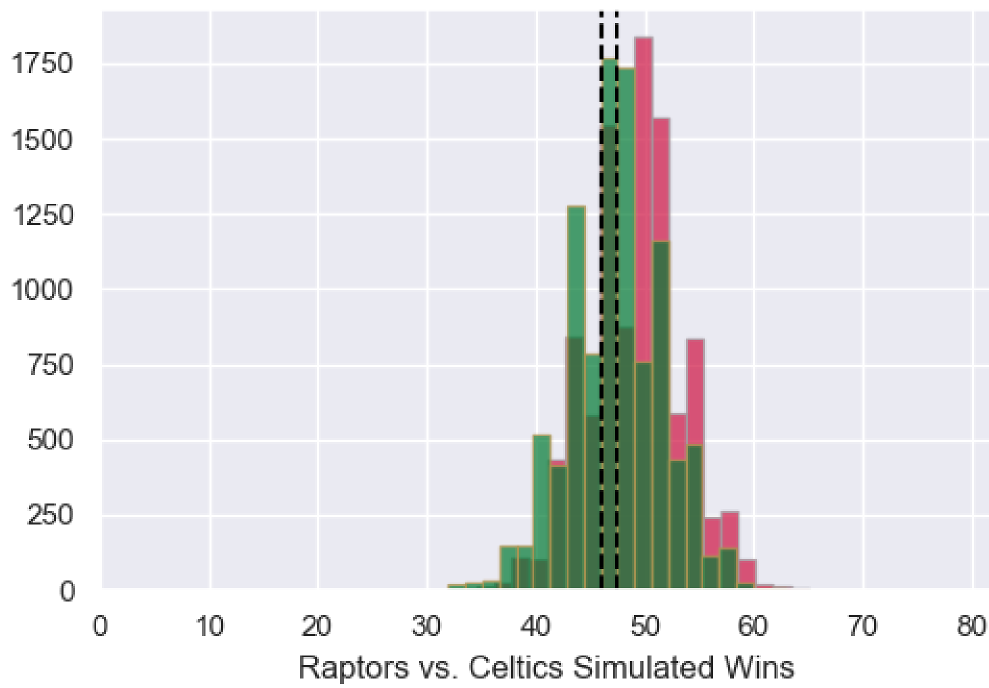
Immediately, we notice that the top teams (53 wins >) have their wins decreased by a relatively large amount, while the bottom teams (< 30 wins) have their wins increased. More importantly, the wide spread of wins and losses is replaced by very closely grouped teams, especially at the playoff level (top 8 teams in each conference). In the East, 6 games separate the top seeded Bucks and 5th seeded Pacers, while only 5 games separate the top seeded Warriors and the 8th seeded Los Angeles Clippers. The difference in results for the East and the West is not surprising, given that the west has historically been the much more competitive and balanced conference; this was reflected in both our simulation and the actual season.

Although the average wins changed, average position was still relatively consistent; top teams tended to remain at the top, while the bottom teams remained at the bottom. Out of 10,000 simulations of the season, the Bucks, the actual best team in the 2018-2019 season, had the best record in the league 2868 times; the Knicks, the worst team in the 2018-2019 season, finished last 3062 times. A further analysis of team positions within each conference might reveal "tiers", where teams of similar strength can be grouped together based on their range of records For example, based on the 95% confidence interval, we might state that the Magic, Pistons, Hornets, and Heat are in the same tier, as fringe playoff teams in the East (note that the 95% confidence interval is not the best measure; further data on team positions, which this project did not collect, is needed).

The Monte Carlo simulation also give us a way to compare teams with each other. For example, once again comparing the Bucks and Knicks, we can see that the histogram of their mean wins reveals a drastic difference in skill and performance:



Meanwhile, when comparing two closer teams in terms of performance, such as the Raptors (58-24) and Celtics (49-33), we see that they are relatively similar in average win distribution, suggesting that these two teams, despite having a 9 win gap during the regular season, are closer than the record may suggest. In fact, according to the project's methods of determining wins, the Raptors only have a 56% chance of defeating the Celtics at home. A possible interpretation from solely this information is that the Celtics and Bucks are evenly matched teams, and that the difference in their record might be explained by factors of luck, such as: strength of schedule, roster continuity (midseason trades, free agent signings, etc...), injuries, and other elements.



4.2. The Central Limit Theorem

When plotting histograms of the wins of each 10,000 seasons, we can see that the distribution is roughly normal, with only a small amount of teams having a slightly bimodal distribution (overall shape can still be described as "normal"). Across the simulations, the impact of variability and luck, which can be big factors during a single season, are reduced, resulting in win percentages growing closer towards .500.

Splitting the 10,000 simulations into 10 groups of 1000 seasons, we can see a clear convergence for each team's average wins (graphs shown in 3.3). This is consistent with the Central Limit Theorem, which states that random samples from any distribution will tend to have a normal distribution with a certain mean. It also offers a theoretical and conceptual explanation for the win percentages having a lesser spread and being centered around .500.

Note that this prediction method is not be the best for forecasting wins in a single season. It might make sense that better teams are better because they are able to maintain consistent levels of top play, while average and worse teams are unable to do so. This is a team qualitys lost in a large number of simulations, which will, more often than not, make predictions where team's are playing at their individual average skills.

5. Conclusion and Final Thoughts

Overall, the use of Monte Carlo methods confirms an element of luck in a single season. When using the method of Pythagorean Wins and Monte Carlo methods, the wins across all teams tend to be more centralized. Using the Central Limit Theorem, we can see that a team's wins will converge, but top teams tend to still remain at the top, while bottom teams tend to still finish at the bottom. Rather than using this method to predict a single season, this method is more suited to comparing the relative skill and strength levels of teams, with tiers forming.

A direct extension of this project is analyzing the proportion of place finishes in each conference (i.e. first, second, etc..). This would further expand on the current data, while also allowing us to see a more clear tier system. Another interesting application of this method is simulating the playoffs, where more closely matched teams play in best of 7 series, and only one team is crowned as the winner.

As for improvements and revisions for the idea of a season simulation, Pythagorean Wins only considers points scored and allowed, when there are many other metrics that can quantify and measure team performance and success. Additionally, we can also consider different models and approaches than the Monte Carlo Method, such as linear regression, random forests, etc...

6. Code

Python script used for simulation can be found here.