



Analysis of the 2018-2019 Season

by

Thi Quang  
Steven Green  
Kidist Gebremedhin

## Inspiration

With the NBA season just starting, we wanted to look back at the last normal season before Covid started, Season 2018. We wanted to identify which team performed the best based on the total score for the season, which are the most impacted players, and which stadium got the most wins. We wanted to do this because aside from being a huge NBA fan, I am also a huge fan of NBA stats. I know that there are some people that can find as much joy in analyzing NBA statistics as other people who watch basketball. I also thought that it would be a good challenge for my team to construct a data site with NBA stats on it, so that it can also be scaled up to show current NBA statistics as well. I also made my team aware that all NBA teams have analytics departments, and these are some of the datatypes and datasets that they look at so that they can further analyze their team's successes and failures. A link to these team departments can be found here.

<https://www.nbastuffer.com/analytics101/nba-teams-that-have-analytics-department/>

## Reasonings

We wanted to approach this like an NBA team would. The overall theme for this project was how would a team take these statistics to analyze their past season and how would they approach a new one in terms of strategy. We chose the Mavericks because it's the local team and some categories would be the easiest to quantify even for a novice. For example, a novice would know that Houston is a lot closer to Dallas than New York. Therefore, if the Mavericks had a game on the road against Houston, the team would need less rest than with a game in New York City. The first statistics we decided to analyze was which team with the most home wins and losses for teams and visualizing that. We wanted to include a map with all the teams with information about the team so that if we were to present this topic to top executives, we could show them where that team is geographically and what effect it would have on our team in terms of rest between practices and games. We could also visualize crowd data to strategize how to deal with crowd noise. Bigger stadiums may or may not have crowd noise. It would depend on the area, and how much they are winning at home and overall. We also wanted to show the points by city because that factors into the rest strategy of the Mavericks as well. If a team like Golden State for instance, scores the most and the Mavericks must play in San Francisco, we wanted to illustrate how these factors could contribute to Mavs' strategy. The final thing that we wanted to see was how much of an impact every NBA player. Using this we could determine things like which players to account for, which teams to account for which players to go after in the offseason and how that would affect our team. For player points per possession (PPP) and player impact estimate (PIE) we had to use formulas to find that information out and then clean our data to include that information. Links to those formulas can be found below.

<https://www.nba.com/.element/media/2.0/teamsites/thunder/statlab-en-october-1819-update.pdf>

<https://www.nbastuffer.com/analytics101/player-impact-estimate-pie/>

### Design Choices

For the design of these visualizations and the website overall we wanted to see which graph types showed our data the best. We chose to use the colors red, white, and blue, as the main theme colors for our dataset since those are the colors of the NBA. We chose to use bar graphs for the player wins and losses. We also chose bar graphs to show the players with the most impact, and the most points by city. We decided to use a bar graph with a filter to show the PPP. The reasoning behind it was that had we showed a graph without a filter it would have been too many players to show. We also wanted to show a map with all the team data on it so that it could easily be shown what team is at each location and how that affects the strategy. We used a satellite base for the map as well as a filter to show which conference a team plays in. We also used markers and color coded them to show which team had arenas with less than 18,500 seats (white), 18,500 to 20,500 seats (red), and then arenas with seating for over 20,500 (blue). Once again, this along with the other labels in the map show a particular strategy for dealing with that team. We decided to include markers, and, in that information, we wanted to show all these facts in that marker so that when a person clicked on the marker, they could see all the team information.

### Hypotheses/conclusions/stories in the data

Once we had our design choices down and the data was cleaned, we wanted to view our hypotheses based on the NBA results from the previous season. We wanted to approach this data with that view and how our team could predict the results of the next season. These hypotheses were that we believed that the Golden State Warriors would be the best team in the 2018 season. We chose this team because in the 2017 NBA season they did show that they were they best team by far. They swept their 2017 NBA Finals opponents; they had the third best record in the league that season and a large collection of talent. It seemed like they would be the team with the best record, the most points and the most wins. Also, since 2 of the last 4 league most valuable players (MVP) were on that team it was an easy choice. We also wanted to show the most impactful players and those would Kevin Durant and Steph Curry of the Warriors, Kyrie Irving of the Boston Celtics, Joel Embiid of the Philadelphia 76ers, Giannis Antetokounmpo of the Milwaukee Bucks, DeMar DeRozan of the San Antonio Spurs,

the 2016-2017 MVP James Harden, 4-time NBA MVP and 3x NBA champion LeBron James of the Cleveland Cavaliers, Russell Westbrook of the Oklahoma City Thunder, and Damian Lillard of the Portland Trail Blazers. Looking at the data, we found that our hypothesis was slightly incorrect. While the warriors did have the most points overall, they didn't have the most wins overall. The team with the most wins overall were the Toronto Raptors. They won the league championship over the Golden State Warriors. The Raptors also had the most home wins with 22. With player impact, we were shocked not to see any Warriors in the top 10 so we ran 11 names and even that eleventh player wasn't on the team we selected. They are James Harden (Rockets), LeBron James (Cavaliers), Anthony Davis (Pelicans), Kawhi Leonard (Raptors), Giannis Antetokounmpo (Bucks), Nikola Jokic (Nuggets), Russell Westbrook (Thunder), Paul George (Thunder), Karl-Anthony Towns (Timberwolves), Joel Embiid (76ers) and Nikola Vucevic (Magic). There are some stories that happened during the season of the dataset that are worth noting. One of these would be the last NBA season of Dirk Nowitzki. Dirk is a legendary player for the Dallas Mavericks and one of the best in the history of the NBA. It was also Dwyane Wade's last season, and he is a player for the Miami Heat among other teams. It was also the first season of Mavericks player Luka Doncic who would go on to win Rookie of the Year. It's also the last season that a very talented shooter named Klay Thompson played an NBA game. Klay unfortunately tore his ACL in the 2019 NBA Finals, then he suffered a torn Achilles just before the 2020 NBA season began. It's interesting to view the data from that perspective and find notable games with those stories in mind.

### Limitations/Future Work

If we had more time and available resources, some of the things that we could have included was that we could perform an analysis for all the seasons since 2018-2019. Visualizations could present the winning team for each season and the statistics for each of the players. We could then map a strategy based on the data to make predictions about each teams' habits and patterns to gain a strategic advantage over all other teams. For instance, we could make a graph of a player's impact over the course of his career up until that point. We could then use machine learning libraries such as scikit learn to predict the future of that player, or even a team over time. Such insight could be a powerful tool to explain player tendencies and use that against them. Then we could also use this for different rotations and different players. We could also use this data to analyze historical teams and players to figure out how teams from the past would compare against teams today. One of the biggest debates in sports is how would past players play against present players, such as LeBron James vs Michael Jordan or the 95-96 Chicago Bulls vs. 15-16 Golden State Warriors. We would also differentiate between regular and postseason. The reason that we would do this is that not all the players playing in the regular season make the playoffs and not all teams have the same players through the year. Players are traded or injured, and it would be a lot of work to see how that affects the impact of a player or even a team.