Sejal Dua and Sook-Hee Evans
Megan Monroe
COMP152: Sports Analytics
9 February 2021

Web Scraping Lab: Bayesian Basketball

In a pre-game show before an NBA game, analysts typically go over "keys to the game," which are essentially strategic action items which, if performed well, will set the team up for success. In recent years, 3 point shooting has become an increasingly pivotal aspect of the game of basketball, yet the significance of 3 point shots and their effect on wins varies from team to team. When a team is successfully firing from beyond the arc in the first half, it tends to set the tone for the game. On the contrary, when they are in a shooting slump, it can create a landslide scoring margin between two teams, which ultimately influences the final outcome of the game. This phenomenon is interesting because the 3-PT field goal percentage of each team in the 2019-2020 NBA season ranged from 0.333 to 0.380, with the Atlanta Hawks registering in last place and the Utah Jazz leading the league in 3-PT shooting. However, the variance in this data is not significant enough to yield any actionable insights since it is unclear how these shooting streaks can lead to potential momentum swings which may influence game outcomes. This motivated us to investigate the probability of an arbitrary team--let's call it Team X--winning an NBA game given that they make more 3-PT shots in the first half, relative to their opponent.

We chose to examine these two events because if there is any effect on the posterior, it makes for great pre-game and mid-game insights for coaches: they should either adjust their defensive strategy by playing up on the 3-PT scorers or answer back with a similar high-scoring offensive strategy. The notion that event B may have an effect on event A, in this context, could also inform how sports betters go about predicting the winner of an NBA game using their own intuition about which team will make more 3s.

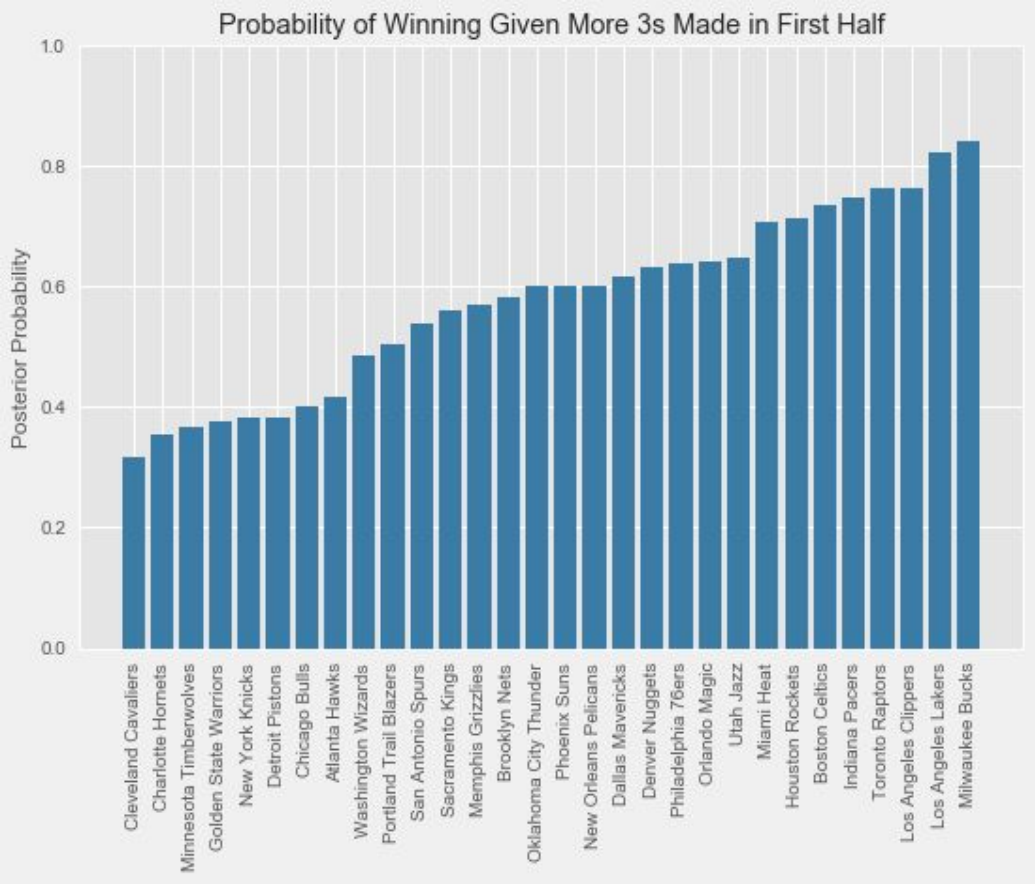
To perform our analysis, we scraped data from https://sports-reference.com. We initially chose the Utah Jazz as our team of focus due to their leading 3-PT field goal percentage in the 2019-2020 season. We scraped play-by-play data for each game under the "Schedule & Results" page for the Jazz in the particular season of interest. Using a regex expression to find "... makes 3-pt ...," we tallied up the number of made 3-PT shots by the Jazz and by the arbitrary opponent and also logged whether or not the Jazz won each game in question. After accumulating these numbers, we plugged our numbers into Bayes' Rule:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|\neg A)P(\neg A)}$$

where the prior, $P(A)$ = the probability of the Jazz winning a game, estimated by its win percentage and the likelihood, $P(B|A)$ = the number of games in which the Jazz made more 3s than its opponent in the first half, of the games with a winning outcome

We computed the posterior probability as 64.6%, which passes our sanity check because of the 78 games that the Jazz played in the shortened 2019-2020 season, they won 46 of them for a win percentage of 0.611. Of the 46 games they won, 35 of them were ones in which they out-scored their opponent from beyond the arc in the first half. However, some noise was introduced by the fact that of the 32 games they lost, they still made more 3s in 21 games. Perhaps the Utah Jazz consistently shoot well from the 3-PT line. We wanted

to dig a bit deeper, so we modularized our code and ran it on the 29 remaining NBA teams. Our posterior probabilities are plotted as a bar chart below:



We can initially observe that all but 6 teams have a greater than 50% chance of winning a game if they make more 3-PT shots in the first half, as compared to their opponent. Interestingly, all teams have a higher posterior probability than their win percentage (the prior), which indicates that event B has a *positive* effect on event A. If we take a closer look at the New Orleans Pelicans, which logged a win percentage of 0.417 for the 2019-2020 season, it is clear that their shooting in the first half has a strong effect on whether or not they win their games. Of the 30 games they won, they made more 3s than their opponent in 24 games while they shot relatively poorly from behind the arc for the majority of their losses. It appears that the data for the Cleveland Cavaliers yields such a low posterior probability because, well, the prior (0.292 win percentage) is low to begin with and there is no distinguishable signal which differentiates the Cavaliers' performance in the event that they shoot well from long range. The Golden State Warriors' wins don't have a clear separation based on event B, but the $P(B|\neg A)P(\neg A)$ term in the marginal makes it interesting because when the Warriors lose (49 games), it tends to also coincide with when they are not shooting well from 3-PT range (36 games or 73% of the time). If we were sports betters and Steph Curry was injured, we would definitely bet on the opponent to win! Easy money.