

## DATA ESSAY IS310 Group 4

Justin Brown  
Alex Guzman  
Alan Kang  
Hayden Elmore  
Ahmad Sawwan

Our dataset represents the connection relationship between social media activity and on-court success during the NBA regular season. Basketball is one of the most globally popular sports and has remained relevant for centuries. As technology advanced over time and athletes began to improve, basketball as a whole began to shift. Throughout the years there have been various external factors many may speculate have been able to influence on-court play. In recent years, social media has been an added layer within the world of sports largely due to its massive influence on human tendencies, relationships, and perspectives.

In the 90's fans and players would have to wait for the newspaper and television shows to recap thoughts and stats while in today's modern era, this information is all at our fingertips. Twitter became a prominent platform after its creation in 2006 and the NBA quickly took action to become a part of this media platform to create a more unified marketing platform to promote basketball globally. In 2009 the NBA announced its partnership with Twitter and this created a new dynamic where social media would be implemented as a daily routine in many of these NBA athletes' lives.

Within media culture, as social media has been utilized more frequently, critics have become quick to point blame towards the prioritization of online activity versus additional focus to on-court play. This stems from an old school mindset where any attention drawn away from the basketball court was considered bad and unproductive. This relationship between old and new media has caused additional speculation and scrutiny on players in today's age. This motivated our group to analyze whether or not the activity of NBA players on Twitter would be reflective of positive success on the court.

Complexities we encountered included the wide range of social platforms that NBA players utilize, which can take away from how effectively we'll both gather and analyze this data. Due to Twitter delivering us the greatest quantity of data metrics to analyze frequency, this became the most ideal platform to dissect for us. Also, the range of players social media following throughout the NBA is extremely widespread which made us reset to create a strategy in order to understand who is influencing winning to the greatest extent.

While there are over 400 players in the NBA, only dozens have superstar presence online. When gathering data, we found that these superstar players were often incentivized to be active on social media because of their branding and marketing influence. In order to narrow down who was having the greatest influence on winning on a daily basis, we tracked the social media activity of the most frequent starting 5 lineups. This was an important decision because it allowed us to identify who the most consistent and depended on players were for the most and least successful teams in the league. Alternative approaches could have included gathering data from the 5 players who had the greatest point differential, greatest estimated win %, or even the most total minutes. These options didn't make the most sense because identifying who

plays together in the NBA is a huge part of knowing what you need for success. It's important to fill a well rounded team, and units of 5 are the most reflective of team chemistry. We decided to look at both the teams with the most and least amount of wins to identify a disparity between their social media activities.

To collect our data, we utilized Twitter's advanced search feature which gave us access to individual tweets from specific NBA players between our time period of choice. Initially, this feature gave us problems including being confused by fake impersonating accounts. On top of this, there were a few times when our advanced search delivered us no results until we entered the exact same search again. These were small issues we resolved but highlighted the detailed complexities that manually gathering data on Twitter can bring us. We voted to analyze data between the first day of the NBA season until the last. This period reflects the time when teams are traveling and playing consistently without large breaks or distractions. This specific time period also represents the exact dates that contribute to the regular season standings, which was a statistic we also took into factor.

In order to organize our labor efficiently and deliberately, we broke our data gathering into sections. Of our 5 person group, each individual was initially assigned 2 teams to gather social media data for the 2021-22 NBA season. This data collection included the social media posting frequency on Twitter for the 5 players who started the most together. Justin gathered data for the Celtics and Rockets, Haydn gathered data for the Celtics and Warriors, Alan gathered data for the Pacers and Pistons, Alex gathered data for the Suns and Magic and Ahmad gathered data for the Grizzlies and Heat. After inputting our data into a spreadsheet we color coded our values to highlight the main discrepancies in our data. Despite having a lot of valuable information, our group felt it would be more valuable if we had more information to compare this to. This prompted more data collection from two more teams in the 2022-23 season. Justin gathered data for the Bucks and Celtics, Haydn gathered data for the Nuggets and Cavaliers, Alan gathered data for the Rockets and 76ers, Alex gathered data for the Pistons and Spurs and Ahmad gathered data for the Trailblazers and Hornets. With this additional information, we now have more context to social media tendencies between seasons which will help us correlate results between the elite and below average teams. As for the main coding portion, it was hard to decide how each person would split up the work to make it equal for everyone. We decided the best approach to this would be for everyone to analyze, then get together and share our thoughts then compile the final version which took into account the group's ideas.

Reflecting on the project, our main focus was to dive deeper into the effects of social media activity during regular season play. The narratives from online platforms have begun an interesting dialogue regarding social media activity and leveraging these tendencies to become the scapegoat for negative on-court play. We encourage others to learn from our study in order to make greater correlations within sports media as well as increase confidence when promoting beneficial activities that can boost on-court play.

## **Dataset Structure:**

There are four main points in the dataset, the year, teams, players, and number of tweets. The year showcases which season of basketball matches we chose to analyze. From this, we pick the top 5 performing teams and the bottom 5 performing teams for each season. This is done by looking at the basketball reference website. From there we also collect the players that started the most for each team and compile them all into one dataset. After this step, we begin our data collection about the number of tweets from each player in each team. This data is then stored and worked on.

#### Potential Uses:

This dataset explores the relationship between social media activity and team performance. This dataset can have many potential uses for people interested in sports analytics, media studies, and digital psychological impact. Scholars can use this dataset to try and study whether social media engagements can have an effect on an athlete, sort of the opposite of what we did. This can also be beneficial to see if there are any psychological factors that play into it. If players have for example high social media presence show signs of tiredness or stress and there is a strong correlation to win rate or on-court efficiency, then scientists can try and make a connection and dive deeper into the matter. Another use would be to see how fan engagement with players can develop on and off the court. By analyzing players' performances and their next interactions with their fans, one can take a look at how the fans respond to the player and how the player can respond to the fans. These apply to a theoretical sort of approach to the usage of our data, meaning that with some changes to the dataset, it could be possible to achieve these things. The current limitation of our dataset is that it's kind of too small to make any concrete conclusions. Updating this dataset to include more entries with more players from the NBA would certainly achieve a level to make it possible. Another omission would be leaving out the text in the tweets. Leaving this out means our data is not verifiable, as people can't go find the tweets. Also, with text data, you can do more analysis on whether players' topics in tweets vary with winning and losing, something we thought about but were unable to get done.