Data Science 5K Final Presentation Pitch

Using the 2016-2017 NBA player data from basketballreference.com, I would like to explore the intricacies of each position and its related statistics. During exploratory data analysis, I would use grouping functions to juxtapose statistics between position groups, i.e. perform comparative analysis on counting stats (such as points, rebounds, assists, and steals) and on advanced stats (like true shooting percentage, usage rate, player efficiency rating). These types of comparisons are extremely valuable, but also face different challenges. Basketball Reference designates positions to the player on their own, with five structured and set positions – Point Guard, Shooting Guard, Small Forward, Power Forward, and Center. While the insights generated through this exploratory data analysis may be insightful, the modern basketball game and the positions players take is much more fluid. Often players blend skills that a point guard and shooting guard may have, or work as a combination power forward/center. This leads me to the model I hope to build – a classifier that ingests a vector of player data to determine which modern NBA position is best suited for this player. Players like Kyrie Irving and Draymond Green have changed the way NBA coaches look at utilizing players and my hope is to quantify this in an algorithm that will determine their best fit. After building this model, I will expand it to assess interesting players entering the NBA – Lonzo Ball, Markelle Fultz, De’Aaron Fox – and see where their college stats would place them and if their traditional position matches the position their stats dictate.

In addition to the positional data analysis and the classification model I hope to develop, another path I may take is modeling over age or amount of years in the NBA. This provides an interesting conundrum, as I can use age and service time as categorical or discrete variables. If I were to use them as categorical variables, my analysis would be similar to the analysis outlined in my first pitch – a mixture of analyses on both counting and advanced statistics to compare players based on their age and their tenure in the league. This can generate incredibly insightful analysis, as these two factors may not be as correlated as one would hypothesize. A large piece that I would be interested in is understanding the gap between rookies who enter the league after one year of college (Karl-Anthony Towns, Andrew Wiggins, John Wall) with those who stayed four years (Malcolm Brogdon, Jimmy Butler, Damian Lillard). How soon do these players reach their so called “peak?” Is there an optimal amount of time in college to play which would allow you to reach your peak at the earliest possible time? While I understand there is a significant impact from survivorship bias in these questions, these are still answers I would like to explore. Pairing this with time trend analysis could potentially generate interesting insights into the careers of many NBA players.