**Term Project**

**Darin Young**

**DSC 530**

**August 10, 2024**

I have been a fan of the NBA for as long as I can remember - I envy those that got to watch the legendary era of the 90's while Michael Jordan and the Bulls formed one of the greatest dynasties in sports. In my lifetime, I have seen many players start their career as rookies and end their careers - either due to age, injury or lack of ability. Typically speaking, the life cycle of an NBA player is around 10 years. However, some players exceed that mark, and in some cases significantly. In those cases, how does their performance stack up against their peers? For my term project, I decided to explore the question: Does a player’s age have an impact on their performance in the NBA? And if so, how much? My hypothesis is that, aside from some outliers, age will have a negative impact on performance.

With that said, I found that most statistics had very little, if any correlation with age. I compared age to three-point percentage, two-point percentage, free throw percentage, effective field goal percentage, points, minutes played, and turnovers, and all of them had an R2 of less than .15 and many of them had p-values of greater than .05 implying that their relationship was simply statistically insignificant. This shocked me and led me to think about reasons that age would not be related to a decrease in performance. After much thought and many recalibrations, I think I have concluded that there is simply a minimum level of skill/performance required to play in the NBA. Some players might fall under that level in their 20s, others might not fall under that level until their late 30s. Another way to think about it – all players decline with age (my original hypothesis). But, once they decline below NBA level, they simply don’t play anymore. So, players like Lebron James, Kareem Abdul Jabaar, and Dirk Nowitzki that were able to sustain their level of greatness late in their career are the exception and not the rule. In fact, I found that only ~5% of players play after they turn 35. But those that do play at that age have one thing in common: they are still able to maintain a level of play that allows them to compete at the highest level.

In the early stages of my analysis, I found that many players had three-point percentages of 0 or near 0. After some digging, I found that it was mostly due to players with a small number of attempts, so I excluded all players with less than 15 three-point attempts from my dataset. At the time, I felt that it would indirectly handle players that had very few minutes played. But, since then, I have realized that there were still many players in my dataset that rarely saw the court. Ultimately, I’m interested in players who are good enough to play for a relevant amount of time, so I think I missed on this part. I believe if I had also excluded players with less than 100 minutes (or so), it would’ve strengthened my analysis even further.

As for variables that may have helped, I probably should have used some number of non-scoring metrics. Blocks, steals, rebounds, and assists are all relevant to a player’s performance. I’m interested now to know how they might have correlated with age. My initial thought is that blocks would be most likely to negatively correlated as blocking requires athleticism that is lost with age.

As for challenges and assumptions, I think I was mostly ok. I think the only challenge I faced was while trying to use the ThinkStats2 functions – once I started using the industry standard libraries (numpy, matplotlib, seaborn, etc.) I didn’t have many issues. Generally, though, the biggest challenge I had was with understanding my results. I was sure that age would have a negative impact on performance, and it took me quite some time to process why that wouldn’t be the case. But isn’t that the point of EDA to begin with? Sometimes we make assumptions about our data that are just totally wrong and EDA reveals the truth.