NBA DFS Capstone Milestone Report

Davy Gia Nguyen

The main goal of this capstone project is to look at two sets of NBA data and to come up with a model that will be useful in differentiating the stats that matter the most in predicting and projecting daily fantasy points for each player.

The task of wrangling and analyzing NBA data to aim at predicting daily fantasy sports is very close to my heart because it led me on a journey that opened my eyes to the extraordinary world of data science.

I was using Google Sheets as my main tool to analyze and construct DFS lineups when I first became interested in this hobby, but I soon realize after the first year that the highest-ranked DFS players were using much more sophisticated tools like R and Python. Further research led me to signing up for DataCamp and then I eventually decided to switch my entire career to Data Science and joining SpringBoard's program as I become convinced that Data Science is like a superpower. It can provide meaningful insights in countless industries, it can predict the future, and after listening to a data science podcast where one data scientist discussed how his team literally save lives using data science to predict and help patients with sepsis in emergency rooms.

Thus, when I was considering a few different ideas for my Springboard's capstone project, I ended up choosing to look at predicting NBA DFS because I can both hone my superpowers while analyzing the subject that I'm very interested in and have some domain expertise in.

I hope that the analysis and conclusions can provide some meaningful insights to the readers since I know the NBA league itself is the fastest growing sport and with on-demand services like Netflix being more popular, sports is becoming more and more attractice due to more lucrative TV contracts, and the DFS industry itself is just in its infancy with plenty of room for growth. Therefore, I hope that this phenomenon will lead to more and more daily fantasy players who might find this capstone useful.

My main hypothesis from playing DFS for two years is that performing clustering analysis to engineer a feature called DvC, defense vs clustering, will be very valuable in predicting future fantasy points performance for each player.

For my capstone project, I utilized two main sets of data and employed various data wrangling techniques to prepare them for analysis and for performing computations that will yield meaningful insights that will compose a model that can predict a player daily fantasy points. The first dataset is comprises of season stats for all active NBA player for the 2018 season acquired from stats.nba.com, while the second dataset was acquired from rotoguru.com.

In the first notebook, I ran a clustering operation to examine the differences the results might incur comparing to the conventional NBA positions. I learned that there are significant differences for all eight clusters, please see specific characteristics in the first notebook.

After clustering analysis Part 1 of the capstone, we will now move on to training and predicting fantasy points based on each player individual stats after we engineer a new feature known as DvC (defense vs cluster ratio) that we hypothesize to have high explanatory power and predictive ability for that player's performance on a certain matchup versus a particular team.