Project 2 Regression Proposal:
Predicting Daily NBA Fantasy Points

Linda Ju

Domain

Motivation

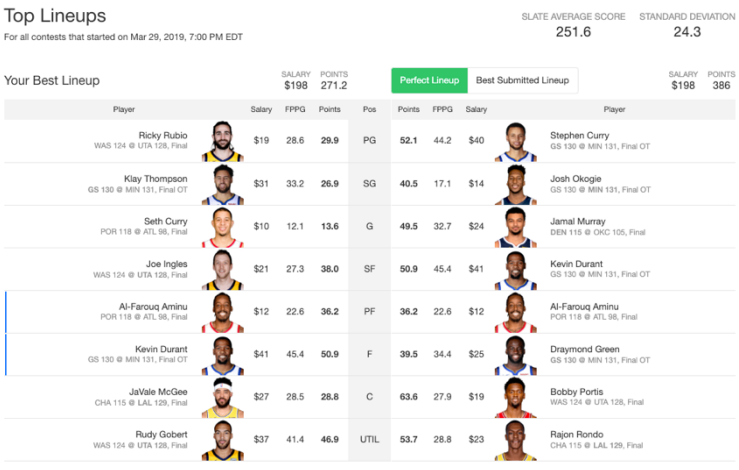
Daily Fantasy Sports (DFS) is a $7 billion industry with millions of users who spend thousands of dollars every day placing bets. I am one of those users. I play fantasy basketball and fantasy football, the former of which will be the subject of my second project. I will perform analysis to predict a basketball player's fantasy production which will hopefully make me a more successful DFS player.

What is daily fantasy basketball?

The objective of daily fantasy basketball is to create a team of NBA basketball players who score more fantasy points than your opponents' teams. Each player has a price associated with them to add to your fantasy team. Generally, the more productive players have higher price tags and vice versa. Each team manager is allotted a budget to spend on any player from any team playing that night in order to add them to their fantasy team.

As the night progresses and the NBA players rack up their stats, such as points, rebounds, assists, steals, blocks, and turnovers, your team's total fantasy points increase with positive production (for example points scored and number of rebounds) and decrease with negative production (for example, turnovers). If your players play well that night, they will score more fantasy points for your team. If your players play poorly, you'll have fewer fantasy points.

<u>Example</u>



The figure to the right is a screenshot of my daily fantasy results on March 29. The eight players on the left side of the screenshot from Ricky Rubio to Rudy Gobert are the players I chose for my team. I paid $198 out of my $200 budget to create my team. They scored 271.2 points. The eight players on the right from Stephen Curry to Rajon Rondo scored the highest points possible that night (386 points) within the $200 budget. The salary for each player are displayed, their average fantasy points per games is shown under 'FPPG' (Fantasy Points Per Game), and their actual points scored is provided under 'Points'. That night, my lineup came in 4th place compared to 19 other lineups in the contest and I won some money. The slate average score and standard deviation shown at the top right of the screenshot are the average and standard deviation of the scores in my contest. Finally, the middle column in gray is a player's position (point guard, shooting guard, center, etc.) Each team is

required to choose players that fill these positions. For example, you cannot have a team of all point guards.

Project Details

The goal of my project is to help me choose my daily fantasy team's players by projecting the fantasy points they will score that night. I will build a linear regression model based on past performance and other factors such as home vs. away game, number of rest days, and opponents' defensive ratings and pace.

Data
The initial dataset I will use is described below. As I work on my project, the actual final dataset may differ as I discard less relevant features and think of more relevant features.

Variable	**Type**	**Description**
Yahoo Fantasy Score	float	Number of fantasy points based on Yahoo's scoring methodology
Box Score Statistics	int/float	Individual box score statistics such as: minutes, usage, points, rebounds, assists, steals, blocks, turnovers, +/-
Home vs. Away	bool	True if home game, False if away game
Opponent Defensive Rating	Float	Estimate of number of points a team's opponent scores per 100 possessions
Opponent Pace	Float	Estimate of number of possessions a team has per game

Considerations
There are a couple of considerations that must be acknowledged before running the model and interpreting the results.

First is the position of players. There is a possibility that the results differ based on a player's position. However, to get started, I will model the data ignoring any references to position.

The second is game script, which evolves as the game is played. For example, let's say the Warriors are playing a lowly team such as the Suns. This matchup would seem great for a player like Stephen Curry. He's a top fantasy producer and he's playing a weak team. However, at halftime, the score is lopsided with Warriors up 82 to 47. Coach Kerr of the Warriors may decide to sit Stephen Curry for the rest of the game to keep him healthy as he's not needed for the rest of this game. This would be harmful to Curry's fantasy production. If he's not on the court, he won't be able to score fantasy points. It would be difficult to account for these situations in a linear regression model.