Brett Gordon

Capstone Project #2 - Project Proposal

My client, EBPN, is a growing basketball streaming service that has gained the rights to feature one National Basketball Association (NBA) game each night as their Premiere Game of the Night. The Premiere Game will cost most for the viewer to watch, with differing subscription options. This market, however, is continuously growing and competitive so if the chosen Premiere Games are often disappointing or boring, viewers are more likely to choose different options to watch NBA games. As a result, my client wants me to discover which NBA games will most likely have the most exciting results for each night during the season. This game will be chosen as the Premiere Game for the 2022-2023 NBA season.

EBPN understands that NBA viewers mostly enjoy certain aspects of a basketball game. For starters, viewers mostly want to watch close games that are decided by 5 points or less. They also want to watch games that have more points (220 total points or more) and less fouls as foul calls slow down the game and can bore viewers. Many viewers also prefer to watch the best players in the league (dependent on their RAPTOR score) as they tend to be more entertaining and interesting. Finally, teams that have better win percentages (typically above .500) tend to garner bigger audiences as they have higher stakes and tend to play better basketball than teams with lower win percentages.

We will be using data from two datasets related to this topic. The first (basketball.sqlite) is a Basketball Dataset that has information from over 60,000 NBA games and includes the scores, game summaries, officials, team information and the career statistics of over 4500 NBA players. The second dataset (historical_RAPTOR_by_player.csv) evaluates the quality of a players' capabilities over each season. All of this data can be used towards discovering what games the viewers would prefer to watch and enjoy.

We want to solve this problem primarily using the data in front of us instead of surveying fans since fans tend to have biases towards the teams they root for which would skew the data to unreliable results. We also want to limit ourselves to data from the year 2004 and more recently as the rules had some major changes then, affecting how the point totals were and the games were called by referees, plus most players that joined the league before 2004 have since retired.

We will solve this problem by using the data points above related to game scores, officiating and individual players. This data could expand as, for example, we might need to discover the tendencies of referees. To explain, some referees tend to call fouls more often than other referees and this would likely slow down the game. We might also need to determine if games are higher scoring due to my fouls being called or if the margin of victory is larger than 5 points. Finally, this information will be beneficial to EBPN but also the NBA as it could help them market games more easily and potentially work more exclusively with EBPN as a streaming service.