Data 319 Project

Link to Dataset: https://www.kaggle.com/datasets/mharvnek/nba-team-stats-00-to-18

Brief overview of your project idea and motivation

Our group's main goal and primary objective is to identify & determine which performance measures are associated with the success of NBA teams in the 2023-2024 season. Using methods such as Multidimensional Scaling (MDS) & K-means clustering will allow our team to explore the relationships between overall team performances. This will further allow our group to find patterns and recognize what makes a team successful amongst other teams in the NBA. Moreover, finding these patterns is meaningful in gaining insight into the driving factors that make a team successful in the NBA.

Description of Dataset

We received our dataset from Kaggle which acknowledged NBA.com for the data. It includes statistics from each team participating in the 2000-2024 basketball seasons. Thirty NBA teams were listed as participating in the 2023-2024 season. The dataset covers the teams' stats per season such as total wins, losses, free throw percentages, rebounds, assists, fouls, etc. Our dataset lists all teams and stats for each of the respective teams.

Research Question

What performance metrics are most strongly associated with the success of NBA teams in the 2023-24 season?

Multivariate Methods

In our analysis to determine what performance metrics are most strongly associated with NBA teams' success in the previous NBA season, we will use multidimensional Scaling (MDS) and K-means clustering methods.

We want to use MDS to visualize the level of similarity between teams on a 2-dimensional scale and understand what metrics lead to better team performance. We expect to see the best teams excel in some metrics, indicating that that metric may be important to an NBA team's success.

In addition to MDS, we also would like to use K-means clustering to group NBA teams into distinct performance categories based on their metrics. Whether K-means clustering reveals clusters with diverse success levels within each group or clusters that show clear success

differences compared to other clusters, both are important in understanding performance metrics that are most important to the success of NBA teams.

So, with the use of our multivariate methods (MDS and clustering), we should be able to come up with strong insights into understanding the performance metrics that are strongly associated with an NBA team's success.

Initial Observations

Some initial observations we have when looking at our datasets is that the data dates back to the 2000-2001 season. All data before the 2023-2024 season we want to avoid looking at because of how the game has changed over the years. Using the most recent data for our analysis will be most reliable. Our data is listed by totals per season. We are interested in looking at stats per game. To do this we will need to divide the totals by 82 games to get the averages of each stat per game.

Team members expected contribution statement

This project will be completed with equal effort from Mariah, Carter, and Daunte.