Clustering of NBA Team Cities and Performance Analysis

Luke Mohr
May 29, 2020

Introduction

I recently heard a story on a basketball podcast about an underground poker ring in Los Angeles. The story had it that a contributing factor to a visiting NBA team's series of losses was the late nights spent gambling by some of its key players. Whether or not the story is true, it inspired a question: is there any correlation between the "flavor" of a city, i.e., the types of venues it has, and how well a team performs there? Will a team perform better in a city that is more similar to its home city?

To tackle this question, we can examine cities by venue type and cluster them to determine which cities are most similar to each other. We can then examine each team's win-loss record when playing in cities that fall within these clusters.

If we can find a correlation between a city's "flavor" and team performance, it would certainly be of interest to NBA teams and owners who could use this information to develop strategies to mitigate poorer performance in certain groups of cities. It would also be of use for scheduling purposes to ensure balance. It could also be a factor in developing odds in sports betting.

Data

Two main types of data are needed to perform this project: geolocation and venue information, and basketball win-loss records. Data was gathered as described below:

- Google was used to find basketball arena locations (latitude and longitude) for each team.
- Foursquare API was utilized to find the most common venues within a 2km radius of each arena.
- I scraped basketball-reference.com for the result of each game played in the 2009-2010 season up to the 2018-2019 season in the months of October-May.

The venue data found using Foursquare will be used to cluster NBA cities into groups to determine which are most alike. The NBA win-loss data will be used to determine each team's record at each city over the course of ten seasons. This will be consolidated to determine each team's performance against each city cluster group, and these records will be used to cluster NBA teams to investigate if there is a pattern of performance based on city type.