

Clustering of NBA Team Cities and Performance Analysis

Luke Mohr
June 6, 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.

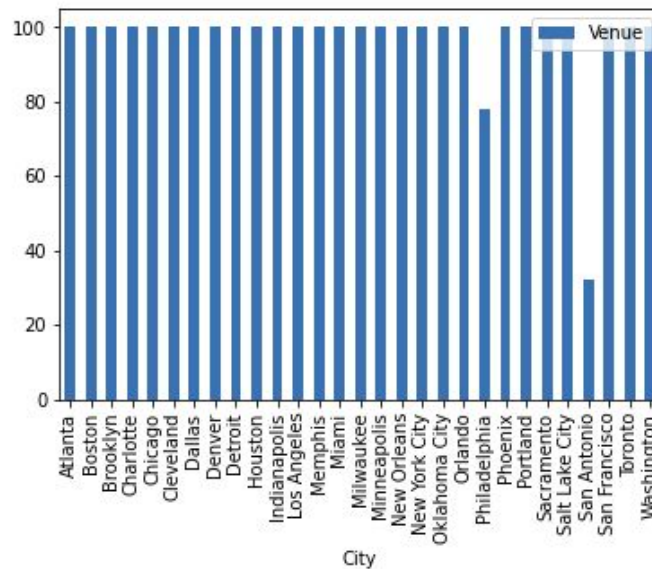
Methodology

Venue Data

To begin, I collected arena latitude and longitude values for each city and compiled these into a table.

	City	Latitude	Longitude
0	Miami	25.781638	-80.186974
1	Cleveland	41.496785	-81.688256
2	Sacramento	38.580456	-121.499660
3	Portland	45.532045	-122.666927
4	Orlando	28.539986	-81.383765

I then used the Foursquare API to retrieve information about venues within a 2km radius of each city's arena. (For simplicity, in this project we will later collect all games played at the New Jersey Nets as if they were played at the Brooklyn Nets' arena, Barclays Center.) This distance was chosen to be representative of the area around the arena without being too restrictive. Some exploratory analysis of the data revealed that San Antonio has a much lower venue count than all other cities:



A search on Google Maps shows that San Antonio's arena, AT&T Center, is less centrally located than many arenas, leading to this lower venue count. Initial experiments in clustering also revealed that this causes San Antonio to be a major outlier and often ended up in a cluster by itself. For this reason, San Antonio was precluded from the analysis that follows.

Venue information was grouped by each city according to venue type (bar, Italian restaurant, aquarium, etc.), and the ratio of each venue type to total number of venues is used as our base data set to compare cities.

	City	Accessories Store	American Restaurant	Antique Shop	Aquarium	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	...	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant	Warehouse Store
0	Atlanta	0.00	0.03	0.0	0.02	0.0	0.01	0.0	0.05	0.00	...	0.0	0.0	0.01	0.0
1	Boston	0.00	0.01	0.0	0.02	0.0	0.00	0.0	0.00	0.00	...	0.0	0.0	0.00	0.0
2	Brooklyn	0.00	0.01	0.0	0.00	0.0	0.00	0.0	0.00	0.00	...	0.0	0.0	0.00	0.0
3	Charlotte	0.00	0.03	0.0	0.00	0.0	0.00	0.0	0.01	0.01	...	0.0	0.0	0.00	0.0
4	Chicago	0.01	0.01	0.0	0.00	0.0	0.00	0.0	0.01	0.01	...	0.0	0.0	0.00	0.0

K-means clustering was used to sort these cities into three different groups. The first group was largely dominated by bars:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Labels
8	Detroit	American Restaurant	Farmers Market	Concert Hall	Lounge	Park	Coffee Shop	Hotel	Diner	Bar	Dessert Shop	0
9	Houston	Hotel	Bar	Vietnamese Restaurant	Cocktail Bar	New American Restaurant	Cajun / Creole Restaurant	Burger Joint	Baseball Stadium	BBQ Joint	Park	0
10	Indianapolis	Bar	Hotel	Steakhouse	Mexican Restaurant	Pizza Place	Brewery	Grocery Store	History Museum	New American Restaurant	Burger Joint	0
11	Los Angeles	Hotel	Coffee Shop	Bar	Burger Joint	Sushi Restaurant	Theater	Gym	Latin American Restaurant	Steakhouse	New American Restaurant	0
12	Memphis	Bar	Park	American Restaurant	Café	Southern / Soul Food Restaurant	Burger Joint	Music Venue	Hotel	Museum	Coffee Shop	0
14	Milwaukee	Bar	Hotel	Steakhouse	Coffee Shop	Pub	Brewery	Theater	Spa	American Restaurant	German Restaurant	0
18	Oklahoma City	American Restaurant	Brewery	Bar	Hotel	Pizza Place	Breakfast Spot	Coffee Shop	Sandwich Place	Pub	Restaurant	0
19	Orlando	American Restaurant	Bar	Hotel	Lounge	Coffee Shop	Brewery	Cocktail Bar	Theater	French Restaurant	Beer Bar	0
21	Phoenix	Coffee Shop	Art Gallery	Lounge	Pizza Place	Bar	Music Venue	Hotel	Mexican Restaurant	Cocktail Bar	Basketball Stadium	0

The second group had a high percentage of coffee shops:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Labels
2	Brooklyn	Cocktail Bar	Italian Restaurant	Grocery Store	Coffee Shop	Bar	Bookstore	New American Restaurant	Ice Cream Shop	Wine Shop	Pizza Place	1
4	Chicago	Coffee Shop	Restaurant	Pizza Place	Sandwich Place	Deli / Bodega	Café	Yoga Studio	Brewery	Park	New American Restaurant	1
7	Denver	Brewery	American Restaurant	Hotel	Coffee Shop	Italian Restaurant	Theme Park Ride / Attraction	Park	Pizza Place	Mexican Restaurant	Seafood Restaurant	1
15	Minneapolis	Brewery	Theater	Italian Restaurant	Coffee Shop	Farmers Market	New American Restaurant	Park	Hotel	Scandinavian Restaurant	American Restaurant	1
20	Philadelphia	Sandwich Place	Italian Restaurant	Pizza Place	Lounge	Sports Bar	Food Truck	Baseball Field	Bar	Pharmacy	Park	1
22	Portland	Coffee Shop	Hotel	Brewery	Park	Bookstore	Spa	Bakery	Donut Shop	Sporting Goods Shop	Lounge	1
26	San Francisco	Coffee Shop	Café	Brewery	Park	Art Gallery	Gym / Fitness Center	Gym	Wine Bar	New American Restaurant	Pizza Place	1
27	Toronto	Coffee Shop	Park	Yoga Studio	Gym	Hotel	Art Gallery	Plaza	Italian Restaurant	Café	Liquor Store	1

The third group largely features hotels:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Labels
0	Atlanta	Hotel	Art Gallery	Park	American Restaurant	Theater	Juice Bar	Coffee Shop	Mediterranean Restaurant	Burger Joint	Sandwich Place	2
1	Boston	Park	Historic Site	Italian Restaurant	Bakery	Hotel	Seafood Restaurant	Gastropub	Mexican Restaurant	Grocery Store	Sandwich Place	2
3	Charlotte	Hotel	Park	Italian Restaurant	Theater	American Restaurant	Brewery	Deli / Bodega	Performing Arts Venue	Coffee Shop	Pizza Place	2
5	Cleveland	Hotel	Coffee Shop	Theater	American Restaurant	Steakhouse	Lounge	Italian Restaurant	Park	Brewery	Deli / Bodega	2
6	Dallas	Hotel	Steakhouse	Coffee Shop	Italian Restaurant	American Restaurant	Restaurant	Pizza Place	Plaza	Mexican Restaurant	Cocktail Bar	2
13	Miami	Hotel	American Restaurant	Seafood Restaurant	Nightclub	Café	Park	Coffee Shop	Bar	Peruvian Restaurant	Japanese Restaurant	2
16	New Orleans	Hotel	Cajun / Creole Restaurant	Coffee Shop	Cocktail Bar	Seafood Restaurant	Hotel Bar	American Restaurant	Italian Restaurant	French Restaurant	Sandwich Place	2
17	New York City	Theater	Park	Hotel	Coffee Shop	Boxing Gym	Japanese Restaurant	Art Gallery	Mediterranean Restaurant	American Restaurant	Gym / Fitness Center	2
23	Sacramento	American Restaurant	Coffee Shop	Hotel	Seafood Restaurant	Pizza Place	Park	Dessert Shop	Café	Movie Theater	Sandwich Place	2
24	Salt Lake City	Hotel	Coffee Shop	Mexican Restaurant	Bakery	Park	Brewery	Burger Joint	Theater	Sandwich Place	American Restaurant	2
28	Washington	Hotel	Art Museum	History Museum	American Restaurant	Coffee Shop	Pizza Place	Plaza	Science Museum	Mediterranean Restaurant	Italian Restaurant	2

We will refer to these groups as Bar Cities, Coffee Cities, and Hotel Cities, respectively, throughout the rest of this report.

Basketball Record Data

We will examine how the type or “flavor” of a city affects the performance of a visiting team. To begin, we scraped <https://www.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. Each team's record at each other team's location is consolidated and stored. Home games are stored as if the team was playing against itself (as seen with the Miami Heat below).

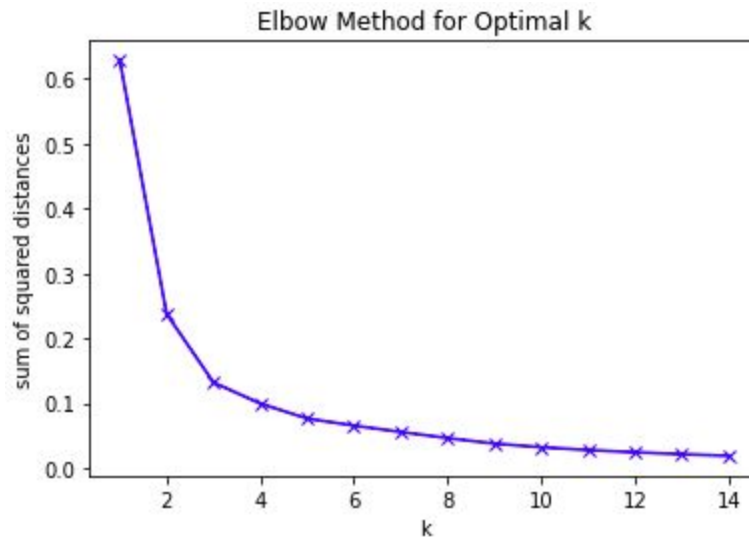
```
{
  'Miami Heat': {'wins': 303, 'losses': 143, 'ties': 0},
  'Cleveland Cavaliers': {'wins': 9, 'losses': 8, 'ties': 0},
  'Sacramento Kings': {'wins': 6, 'losses': 3, 'ties': 0},
  'Portland Trail Blazers': {'wins': 5, 'losses': 5, 'ties': 0},
  'Orlando Magic': {'wins': 12, 'losses': 8, 'ties': 0},
  'Philadelphia 76ers': {'wins': 13, 'losses': 9, 'ties': 0},
  'Atlanta Hawks': {'wins': 11, 'losses': 9, 'ties': 0},
  'Chicago Bulls': {'wins': 13, 'losses': 11, 'ties': 0},
  'Memphis Grizzlies': {'wins': 4, 'losses': 5, 'ties': 0},
  'Milwaukee Bucks': {'wins': 11, 'losses': 9, 'ties': 0},
  'Washington Wizards': {'wins': 12, 'losses': 8, 'ties': 0},
  'Phoenix Suns': {'wins': 8, 'losses': 1, 'ties': 0},
  'Los Angeles Clippers': {'wins': 4, 'losses': 6, 'ties': 0},
  'Utah Jazz': {'wins': 4, 'losses': 6, 'ties': 0},
  'Houston Rockets': {'wins': 4, 'losses': 5, 'ties': 0},
  'San Antonio Spurs': {'wins': 2, 'losses': 7, 'ties': 0},
  'Toronto Raptors': {'wins': 9, 'losses': 13, 'ties': 0},
  'Dallas Mavericks': {'wins': 7, 'losses': 3, 'ties': 0},
  'Indiana Pacers': {'wins': 9, 'losses': 16, 'ties': 0},
}
```

Teams whose names have changed in the past ten years (New Orleans Hornets and Pelicans, Charlotte Bobcats and Hornets) were consolidated into their most recent incarnation. Similarly, the New Jersey Nets' record was consolidated with that of the Brooklyn Nets. The San Antonio Spurs are not included in this analysis since they were not included in the city classification above.

This data was then grouped by the city clusters determined in the previous section. The win percentage of each team at each city type (Label 0 for Bar Cities, Label 1 for Coffee Cities, and Label 2 for Hotel Cities) was calculated and stored in a table. Home games are not included to avoid inflating win percentages for each team's city type.

	Label 0	Label 1	Label 2
Miami Heat	0.510638	0.558140	0.534161
Cleveland Cavaliers	0.430556	0.436508	0.417722
Sacramento Kings	0.295302	0.274336	0.300813
Portland Trail Blazers	0.407407	0.428571	0.452703
Orlando Magic	0.309735	0.371681	0.362637

To examine if there was alignment between grouping cities based on venue type and teams based on performance, k-means clustering was performed on teams based on performance at the three city types. An elbow method test was performed to determine the optimal value for k, and we chose k=3 based on our result.



Examining the means across each group based on win percentage, we see that this method may simply be clustering based on overall performance. (Note that going forward, “Record Labels” refers to the clustering based on team performance, and “City Labels” refers to the clustering based on venue types.)

```
Record Labels
0    0.440517
1    0.328722
2    0.541704
dtype: float64
```

We examine more carefully the groups based on records to see if there is any alignment between record and city types. First, we look at the teams with the highest records (Record Label 2):

	Label 0	Label 1	Label 2	Record Labels	City Labels
Miami Heat	0.510638	0.558140	0.534161	2	2.0
Houston Rockets	0.463768	0.503937	0.567568	2	0.0
Oklahoma City Thunder	0.500000	0.540323	0.596026	2	0.0
Golden State Warriors	0.480000	0.576923	0.668966	2	1.0

Below are the teams with the second highest records (Record Label 0):

	Label 0	Label 1	Label 2	Record Labels	City Labels
Cleveland Cavaliers	0.430556	0.436508	0.417722	0	2.0
Portland Trail Blazers	0.407407	0.428571	0.452703	0	1.0
Atlanta Hawks	0.391892	0.434783	0.457831	0	2.0
Chicago Bulls	0.485714	0.422680	0.462366	0	1.0
Memphis Grizzlies	0.412162	0.500000	0.421429	0	0.0
Milwaukee Bucks	0.344828	0.491667	0.366667	0	0.0
Los Angeles Clippers	0.500000	0.406504	0.524476	0	0.0
Utah Jazz	0.391304	0.446281	0.430894	0	2.0
Toronto Raptors	0.489362	0.465347	0.437158	0	1.0
Dallas Mavericks	0.440994	0.424779	0.516129	0	2.0
Indiana Pacers	0.478261	0.361345	0.410256	0	0.0
Denver Nuggets	0.410256	0.436893	0.400000	0	1.0
Boston Celtics	0.477941	0.472000	0.494505	0	2.0

Finally the teams with the lowest records (Record Label 1):

	Label 0	Label 1	Label 2	Record Labels	City Labels
Sacramento Kings	0.295302	0.274336	0.300813	1	2.0
Orlando Magic	0.309735	0.371681	0.362637	1	0.0
Philadelphia 76ers	0.323077	0.361905	0.303371	1	1.0
Washington Wizards	0.343284	0.398305	0.329193	1	2.0
Phoenix Suns	0.311111	0.360360	0.381944	1	0.0
Los Angeles Lakers	0.292857	0.403509	0.402685	1	0.0
Charlotte Hornets	0.333333	0.339806	0.282895	1	2.0
Detroit Pistons	0.252101	0.309735	0.345455	1	0.0
Minnesota Timberwolves	0.261438	0.252632	0.330935	1	1.0
New Orleans Pelicans	0.359477	0.330435	0.367521	1	2.0
New York Knicks	0.285714	0.345133	0.375796	1	2.0
Brooklyn Nets	0.307087	0.299065	0.329341	1	1.0

We can see that these groupings do indeed reflect overall team performance and do not appear to correlate to the city groupings in a meaningful way, meaning that a team does not necessarily perform better in a different city with a similar “flavor”.

Results

We will examine at which types of cities teams perform best or worst. This kind of data could be useful to understand if a team is underperforming at certain types of cities and if action could be taken to mitigate this.

The following are the teams that perform best at Bar Cities (Label 0):

	Label 0	Label 1	Label 2	Record Labels	City Labels
Chicago Bulls	0.485714	0.422680	0.462366	0	1.0
Toronto Raptors	0.489362	0.465347	0.437158	0	1.0
Indiana Pacers	0.478261	0.361345	0.410256	0	0.0

These teams perform best at Coffee Cities (Label 1):

	Label 0	Label 1	Label 2	Record Labels	City Labels
Miami Heat	0.510638	0.558140	0.534161	2	2.0
Cleveland Cavaliers	0.430556	0.436508	0.417722	0	2.0
Orlando Magic	0.309735	0.371681	0.362637	1	0.0
Philadelphia 76ers	0.323077	0.361905	0.303371	1	1.0
Memphis Grizzlies	0.412162	0.500000	0.421429	0	0.0
Milwaukee Bucks	0.344828	0.491667	0.366667	0	0.0
Washington Wizards	0.343284	0.398305	0.329193	1	2.0
Utah Jazz	0.391304	0.446281	0.430894	0	2.0
Los Angeles Lakers	0.292857	0.403509	0.402685	1	0.0
Charlotte Hornets	0.333333	0.339806	0.282895	1	2.0
Denver Nuggets	0.410256	0.436893	0.400000	0	1.0

These teams perform best at Hotel Cities (Label 2):

	Label 0	Label 1	Label 2	Record Labels	City Labels
Sacramento Kings	0.295302	0.274336	0.300813	1	2.0
Portland Trail Blazers	0.407407	0.428571	0.452703	0	1.0
Atlanta Hawks	0.391892	0.434783	0.457831	0	2.0
Phoenix Suns	0.311111	0.360360	0.381944	1	0.0
Los Angeles Clippers	0.500000	0.406504	0.524476	0	0.0
Houston Rockets	0.463768	0.503937	0.567568	2	0.0
Dallas Mavericks	0.440994	0.424779	0.516129	0	2.0
Oklahoma City Thunder	0.500000	0.540323	0.596026	2	0.0
Golden State Warriors	0.480000	0.576923	0.668966	2	1.0
Detroit Pistons	0.252101	0.309735	0.345455	1	0.0
Minnesota Timberwolves	0.261438	0.252632	0.330935	1	1.0
New Orleans Pelicans	0.359477	0.330435	0.367521	1	2.0
New York Knicks	0.285714	0.345133	0.375796	1	2.0
Boston Celtics	0.477941	0.472000	0.494505	0	2.0
Brooklyn Nets	0.307087	0.299065	0.329341	1	1.0

These cities perform worst at Bar Cities (Label 0):

	Label 0	Label 1	Label 2	Record Labels	City Labels
Miami Heat	0.510638	0.558140	0.534161	2	2.0
Portland Trail Blazers	0.407407	0.428571	0.452703	0	1.0
Orlando Magic	0.309735	0.371681	0.362637	1	0.0
Atlanta Hawks	0.391892	0.434783	0.457831	0	2.0
Memphis Grizzlies	0.412162	0.500000	0.421429	0	0.0
Milwaukee Bucks	0.344828	0.491667	0.366667	0	0.0
Phoenix Suns	0.311111	0.360360	0.381944	1	0.0
Utah Jazz	0.391304	0.446281	0.430894	0	2.0
Houston Rockets	0.463768	0.503937	0.567568	2	0.0
Los Angeles Lakers	0.292857	0.403509	0.402685	1	0.0
Oklahoma City Thunder	0.500000	0.540323	0.596026	2	0.0
Golden State Warriors	0.480000	0.576923	0.668966	2	1.0
Detroit Pistons	0.252101	0.309735	0.345455	1	0.0
New York Knicks	0.285714	0.345133	0.375796	1	2.0

These cities perform worst at Coffee Cities (Label 1):

	Label 0	Label 1	Label 2	Record Labels	City Labels
Sacramento Kings	0.295302	0.274336	0.300813	1	2.0
Chicago Bulls	0.485714	0.422680	0.462366	0	1.0
Los Angeles Clippers	0.500000	0.406504	0.524476	0	0.0
Dallas Mavericks	0.440994	0.424779	0.516129	0	2.0
Indiana Pacers	0.478261	0.361345	0.410256	0	0.0
Minnesota Timberwolves	0.261438	0.252632	0.330935	1	1.0
New Orleans Pelicans	0.359477	0.330435	0.367521	1	2.0
Boston Celtics	0.477941	0.472000	0.494505	0	2.0
Brooklyn Nets	0.307087	0.299065	0.329341	1	1.0

These cities perform worst at Hotel Cities (Label 2):

	Label 0	Label 1	Label 2	Record Labels	City Labels
Cleveland Cavaliers	0.430556	0.436508	0.417722	0	2.0
Philadelphia 76ers	0.323077	0.361905	0.303371	1	1.0
Washington Wizards	0.343284	0.398305	0.329193	1	2.0
Toronto Raptors	0.489362	0.465347	0.437158	0	1.0
Charlotte Hornets	0.333333	0.339806	0.282895	1	2.0
Denver Nuggets	0.410256	0.436893	0.400000	0	1.0

Our results show one thing we may have expected, which is that, overall, most teams tend to perform worse at Bar Cities. A plurality of teams perform worst at these cities, and a very small minority perform best at them.

Interestingly, we also find that a plurality of teams perform best at Hotel Cities, while a minority perform worst at these cities.

Discussion

The information we found in our Results section is a good catalyst for further discussion. Note, for instance, that four teams (New York, Detroit, Phoenix, and Orlando) have a 5-6% lower win rate at Bar Cities compared to either of the other city types. The Warriors and Lakers, staggering, have 9% and 11% lower winrates, respectively, at Bar Cities than either of the

other city types. While we can't derive causal relationships directly from this data, it could be a driver for teams to evaluate internally whether players' activities when visiting opposing cities could be affecting team performance. This data could also be useful in determining spreads and odds, although a separate set of analyses should be performed to determine if city type is a significant variable in each team's performance.

Another possibility for future work is to examine the effect of home team performance on the win rate of visiting teams and whether this is skewing results for any city types. A weighting function could be implemented to normalize win values by reducing the value of a win against a home team with a poor record in a given year while raising the value of a win against a high-performing home team.

Conclusion

We've found that most NBA cities can be classified by three types: Bar Cities, Coffee Cities, and Hotel Cities. We saw that teams don't necessarily perform better in other cities of the same type. However, our results showed that, overall, teams seem to perform worse in cities with higher proportions of bars. This does answer our question of whether the "flavor" of a city may have an effect on team performance. Teams may want to perform an internal analysis to determine if there is a causal relationship between types of venues at visiting cities and the teams' performance, and whether there are any actions they can take to mitigate this.

Sources

- [Foursquare API](#)
- [Google Maps](#)
- [Basketball-Reference.com](#)