Capstone Project: A New Miami Coffee Shop

Introduction

Finding a good location for a new restaurant can be the difference between success and failure. In this project I will attempt to provide locations for a new coffee shop in Miami by using data about venue locations from Foursquare.

Data

The data used for this project will be venue data from Foursquare. Because there is a limit of 100 venues per API call, a grid will be used to create a series of API calls on overlapping 'neighborhoods' with a radius of 500 meters with centers spaced 500 meters apart. Coffee shops in this data will be identified and another data set will be created by pulling Foursquare data for 'neighborhoods' defined by a 500-meter radius surrounding each coffee shop. The Foursquare venue category will be used to create a new category to classify each location. This new feature will be used in the data analysis.

Methodology

The first data set (ds1) was all of the venue data from each 'neighborhood' defined by a 500 meter radius in an area covering approximately 10km by 10km centered on downtown Miami. This grid would create 441 potential neighborhoods, though some would not be relevant because they only contained water. Each neighborhood was an API call to the Foursquare data base to return all the venues in the neighborhood. In order to ensure all venue data for the area was gathered these neighborhoods will be fully overlapping with a distance of approximately 500 meters between the center of each neighborhood.

Ds1 contained many duplicate venues so the second data set (ds2) was created by removing duplicates from ds1. Then the third data set (ds3) was created by using all of the coffee shop locations in ds2. And ds4 was created by using each coffee shop location as the new center of a neighborhood and for each of these neighborhoods an API call to the Foursquare database was made to collect all of the venues in these coffee shop neighborhoods.

Each venue in ds1 and ds4 were assigned to a new category, Category2. Since the categories for each venue assigned by Foursquare are very numerous Category2 is used to create a smaller number of categories that seems more relevant to how each venue might help or hinder the inclusion of a new coffee shop in the neighborhood.

Finally, ds5 was created by combining ds1 (the original neighborhood data) with ds4 (the coffee shop neighborhood data). Ds5 was used for a kmeans analysis.

Several kmeans trials were run on ds5 using different numbers of clusters. The final analysis groups the neighborhoods into 10 clusters. The number of clusters was chosen by comparing the distribution of coffee shop neighborhoods to the distribution of the grid neighborhoods from ds1. The desired outcome would be to have the coffee shop neighborhoods concentrated in a small number of groups while having the grid neighborhoods more evenly distributed among all the groups so that groups could be chosen where the ratio of coffee shop neighborhoods to grid neighborhoods was significantly higher than in the other groups.

Clusters 0 and 8 were chosen as representing the best possible locations as their ratio was around 1 coffee shop neighborhood to 4 grid neighborhoods. There were a total of 124 grid neighborhoods in these two groups, so an additional step of processing was needed to further reduce the number of options for the new coffee shop.

A function was created to help determine the most desirable neighborhoods for a new coffee shop. The basis for the function was the assumption that a good neighborhood for a new coffee shop would tend to have more restaurants and shops but fewer coffee shops. The function added the number of restaurants and shops then divided that sum by one plus the number of coffee shops in the neighborhood.

This 'RankBy' function was used to filter the potential neighborhoods. A threshold of 10 was decided to create a set of locations to choose from that was small enough to be useful in filtering the choices yet large enough to provide several different options for the new coffee shop owner to choose an area that best fits their business model and brand.

Results

The final list of neighborhoods includes a count of the number of coffee shops, restaurants, and shops in each neighborhood. The RankBy function described above allows the new coffee shop owner to choose neighborhoods that already have one or more coffee shops but also have a very large pool of potential customers evinced by the large number of shops and restaurants or choose an area with no competition but also likely less neighborhood traffic. Since the neighborhoods in the former group would probably require the owner to pay more in rent than the latter, the amount of available capital to devote to the business as well as the brand and associated customer base for the new business are all things the owner will need to consider when choosing a new location.

A few broad areas contain most of the neighborhoods identified:
1. The area east of I-95 from the Financial District north to the Miami River.
2. Areas south and west of The Shops at Midtown Miami
3. The area along US41 from NW 27th Ave to I-95.
4. The area along FL 972 between SW 37th Ave and SW 17th Ave

Discussion

Future options for refining this analysis would include creating a more refined ranking function and incorporating traffic data or other data that ties area activity to time of day. An additional layer of analysis could be to include this additional data with a business forecast model in order to generate a map with the dollar value the business would be willing to pay in rent for a lease in a specific location. This could then be cross referenced with lease data to help a business quickly assess which real leases would fit the business's needs.

Conclusion

This project helped analyze location data to help identify the best locations for a new coffee shop in Miami. This specific analysis benefited from the fact that a coffee shop is a distinct category that does not compete directly with other restaurants. The analysis would need to be altered if a different target venue was used. However, this analysis could easily be repeated coffee shops in different cities.