**Selecting a location in Miami for a New Restaurant**

Peter McGranaghan

March, 2019

**1. Introduction**

* 1. **Background**

One of the principal reasons why new restaurants fail is location e.g. see <https://www.cnbc.com/2016/01/20/heres-the-real-reason-why-most-restaurants-fail.html>

and <https://www.bostonglobe.com/magazine/2016/05/25/four-simple-reasons-that-good-restaurants-under/8hI9fAyYhFhycjajt6m46J/story.html>

It is not the only reason, as these articles point out, but it is a major reason and, therefore, it would make sense to research suitable locations before investing money to open a new restaurant. The geographic area that this report will focus on is the city of Miami, Florida.

* 1. **Problem**

Data that might contribute to determining location includes number of existing and planned restaurants, the cuisines they serve, neighborhood population size and density, average local income. This project aims to predict which of the city neighborhoods is the most suitable for introducing a new restaurant and which type of cuisine.

* 1. **Interest**

This type of investigation would be of interest to investors considering investing in a new restaurant in Miami since it would reduce the risk associated with at least one deciding factor i.e. location. Other factors would obviously need to be investigated e.g. zoning restrictions, proposed construction nearby, staff availability, etc. Those are not taken into consideration in this project.

**2. Data acquisition and cleaning**

**2.1 Data sources**

We will rely on the Foursquare location data for the neighborhoods of Miami for information on existing restaurants. For information on planned restaurants there is data available at <https://gis-mdc.opendata.arcgis.com/> from Miami-Dade county which includes the city of Miami. For neighborhood location data we will use data from <https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Miami> . This site also includes population size and density and average income data. Currently the site shows 25 neighborhoods but this may be reduced after Data Cleaning due to e.g. missing data (see below). For the Midtown neighborhood data source is <http://www.city-data.com/neighborhood/Midtown-Miami-FL.html> . Income data for Grapeland Heights neighborhood was obtained from <https://www.point2homes.com/US/Neighborhood/FL/Grapeland-Heights-Demographics.html> .

The BeautifulSoup tool <https://www.crummy.com/software/BeautifulSoup/bs4/doc/> will be used to scrape data from the required websites. An obvious problem with using neighborhoods as the basis of analysis is the variation in size and shape of the neighborhoods. This affects, for example, the radius size of the search and explore features of Foursquare. Instead of using a standard radius we will use the population density and size to estimate the radius.

**2.2 Data cleaning**

A preliminary review of the data shows some missing entries e.g. GPS coordinates for the Health District. This will probably be removed since this neighborhood would not be suitable for restaurant location due to zoning and other considerations. The Midtown neighborhood is missing population data. This neighborhood is relatively new and undergoing rapid development. It would be worth keeping in the analysis and finding the data from another source since it is potentially a good location for a new restaurant. The location coordinates data in the Wiki page will be split into Latitude and Longitude to make it easier to use both Folium for mapping and Foursquare for location data. Columns such as Demonym and sub-neighborhoods will be dropped since these don’t contribute to the decision process. The neighborhood of Virginia Key will most likely be dropped due to low population.

**2.3 Feature selection**

Features which will be considered are Density of Existing Restaurants in the neighborhood, Neighborhood Population, Neighborhood Population Density, Average Income.

**3. Methodology**

The aim of the study will be to find one or more neighborhoods of Miami city which will be a good choice for locating a new restaurant. Foursquare data is used to find areas which currently have a low density of existing restaurants and this will be combined with average income data to select likely candidate neighborhoods.

The initial list of 25 neighborhoods was reduced to 21 after removal of :

Health District - mainly hospitals, not zoned for restaurants

Virginia Key – low population

Allapattah – low income, low density of restaurants

Venetian Islands - high crime index due to proximity to Miami Beach.

The initial feature set from <https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Miami> was supplemented with income and crime data from other sources as detailed in Data Sources section above. The density index for existing restaurants was calculated as follows:

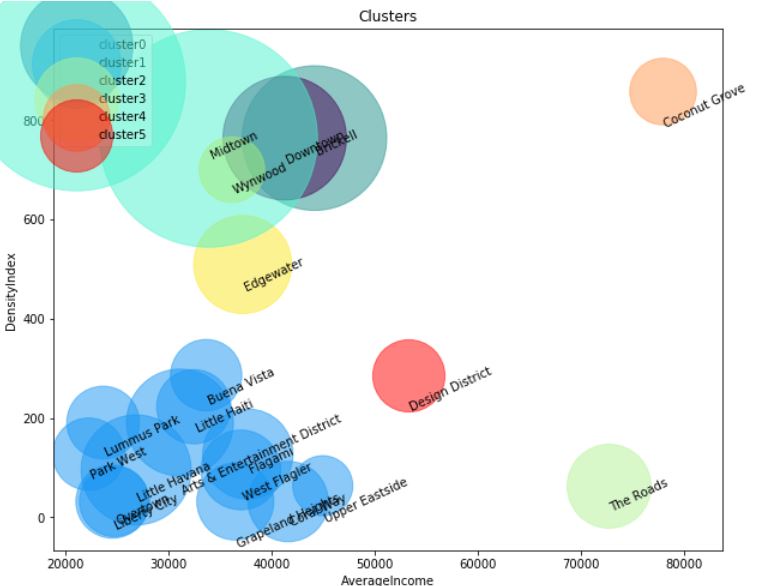
Density Index = (R \* T) / π, where T = number of restaurants reported by Foursquare for the neighborhood, R = circle radius passed to Foursquare in the explore query.

A radius of 100 was used.

After Foursquare returned the list of venues based on the GPS coordinate of each neighborhood a dataframe was created consisting of only the restaurants. This was then merged with neighborhood data. The density index for each neighborhood was then calculated and added to the merged dataframe.



Using the Scikit-learn package I then applied agglomerative clustering to the data to produce a cluster diagram showing the clusters based on average income and density of restaurants:



**4. Results**

The higher the density index the more restaurants were found by Foursquare in that neighborhood. The data suggests that The Roads neighborhood would be a good choice, followed by the Design District. These have relatively low density of restaurants and high income. The highest income was in the Coconut Grove area but the data suggests there is a high saturation of restaurants already in the area.

**5. Discussion**

The study was based on available public data but there are many sources for the same data and there are differences among the various websites. The differences are due to changing demographics and constant urban development and any decision based on the study should ensure that the latest and most accurate data is applied. This may change the choice of neighborhood closer to the date of decision. Also, the definition of neighborhood is very flexible. I found in collecting the data that even the neighborhood names vary over time for the same neighborhood e.g. The Arts and Entertainment District is also known as the Entertainment and Media neighborhood. A more stable parameter would probably be zip code (postal code).

**6. Conclusion**

Based on the data chosen for this project the Miami city neighborhoods of The Roads and Design District would be relatively good choices for locating a new restaurant.