**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).

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 Number of Existing Restaurants, Cuisines, Neighborhood Population, Neighborhood Population Density, Average Income. Other features may be added depending on further exploration of the data in week 2 of the assignment.