Capstone Project - IBM Data Science Professional Certificate on Coursera

The Battle of Neighborhoods: Final Report

Gabriella Mansur

August 19, 2019

Contents

1 Introduction

1.1 Business problem

The opening of a new business is a matter which requires several considerations - from finding finacing to working the logistics of hiring staff and buying new equipment. This project will focus on one of these aspects, namely how to choose the right location, in this case, for a new Italian restaurant in the city of Miami, with prices ranging from mid-high to high.

1.2 Strategy

The idea is to look at which neighborhoods 1. already have a strong Italian cuisine culture/background in Miami, 2. have a higher-than-average medium income, facilitating the introduction of a restaurant of the same kind but more refined. By studying the locations where current Italian restaurants thrive in this city, one can make an educated decision in order to maximize the chances of a successful outcome when opening a new one.

1.3 Groups of interest

This report and its findings aims at to aid its main stakeholders, amongst who we could point businessmen in the city of Miami and its vicinities, who are interested in expanding their current business or investing in a new one; Italian cuisine chefs; and the Italian community of Miami.

2 Data

The data used in this project come from:

- **Foursquare** - through Foursquare one can retrieve the venues in the city of Miami. With the Foursquare data, one can identify which are the most popular venues in Miami, and observe the popularity of Italian restaurants in each neighborhood and where they mainly thrive.
- **Wikipedia** - `https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Miami` - used to retrieve the list of neighborhoods in Miami in Miami-Dade County, Florida, United States, together with its coordinates. This data is used together with the Foursquare data to understand how Italian restaurants are distributed in the neighborhoods of Miami. The dataframe scraped from the Wikipedia page can be seen in Fig. 1.
- **Statistic Atlas** - `https://statisticalatlas.com/county-subdivision/Florida/Miami-Dade-County/Miami/Household-Income#figure/neighborhood/`

`median-household-income` - after using the combined Foursquare and Wikipedia data and narrowing down the area where the Italian restaurant will be opened, we will use the Statistic Atlas data to look at the median household income per neighborhood in Miami and narrow our aimed location even further. This is important since Italian restaurants can range from street food to high-end cuisine, thus the need to ensure that our mid-high to high priced restaurant will fit its location.

3 Methodology

3.1 Scraping data from Wikipedia

The investigation of a proper location for the opening of a new Italian restaurant in the city of Miami with mid-high to high prices starts off with identifying the Miami neighborhoods. This is done through scraping a Wikipedia page and putting its content into a dataframe, seen in Fig. 1.

```
In [60]: df = pd.read_html("https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Miami", header=0)[0]

In [61]: df.head()
```

Out[61]:

	Neighborhood	Demonym	Population2010	Population/Km²	Sub-neighborhoods	Coordinates
0	Allapattah	NaN	54289	4401	NaN	25.815-80.224
1	Arts & Entertainment District	NaN	11033	7948	NaN	25.799-80.190
2	Brickell	Brickellite	31759	14541	West Brickell	25.758-80.193
3	Buena Vista	NaN	9058	3540	Buena Vista East Historic District and Design ...	25.813-80.192
4	Coconut Grove	Grovite	20076	3091	Center Grove, Northeast Coconut Grove, Southwe...	25.712-80.257

Figure 1: Dataframe scraped from Wikipedia

3.2 Data preparation

As with any other dataset, we must get this one into a shape that fits our needs. The most important information are the neighborhood name and its coordinates. We can drop columns such as ”demonym” and ”sub-neighborhoods” since we are not interested in this information, and, for now, we can keep the population columns. The coordinates column needs to be split into latitude and longitude, and we can drop any rows which do not have the coordinates information.

3.3 Generating map of Miami

The coordinates of Miami can be fetched through the geopy package, which we will generate a map of Miami with the folium package. This type of visualization will help us see how the neighborhoods are distributed across Miami. The point coordinates of each neighborhood are also tagged in the map, as seen in Fig. 2.

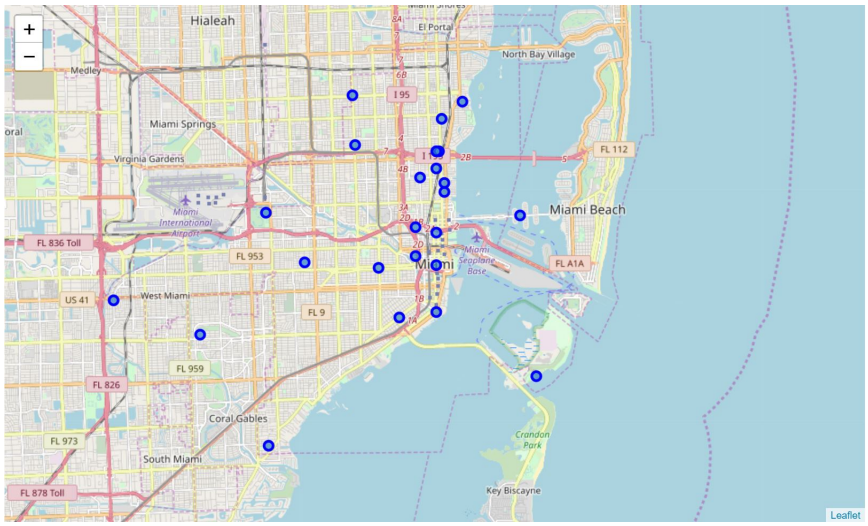

Figure 2: Map of miami with its neighborhoods

3.4 Retrieving Foursquare venues

With our Foursquare credentials and version, we can retrieve the venues of each neighborhood in Miami. We can sort these venues by how common they are in each neighborhood. Since we are interested in the Italian restaurants, we will look at which neighboorhoods have Italian restaurants as their top 3 most common venues.

4 Results and Discussion

4.1 Targeting existing Italian cuisine culture

As discussed in our business strategy, we begin by targeting neighborhoods with an existing Italian cuisine culture by locating those neighborhoods where Italian restaurants are most common.

Fig. 3 shows the fraction that Italian restaurants represents in each neighborhood. Brickell, Downtown and Midtown show the highest fraction: 10% of all venues in these places are Italian restaurants.

With that in hand, we can label the Miami neighborhoods as having Italian restaurants in their top 3 venues (purple marker) or not (red marker), as seen in Fig. 4.

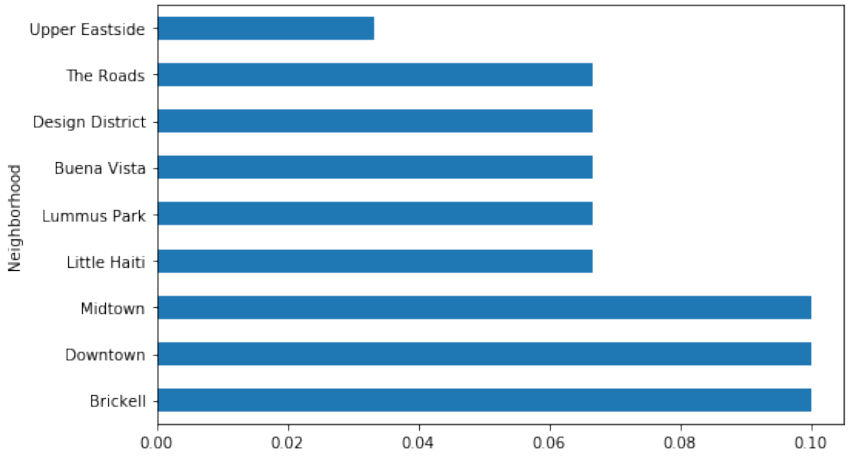

Figure 3: Percentage of Italian restaurants in each neighborhood

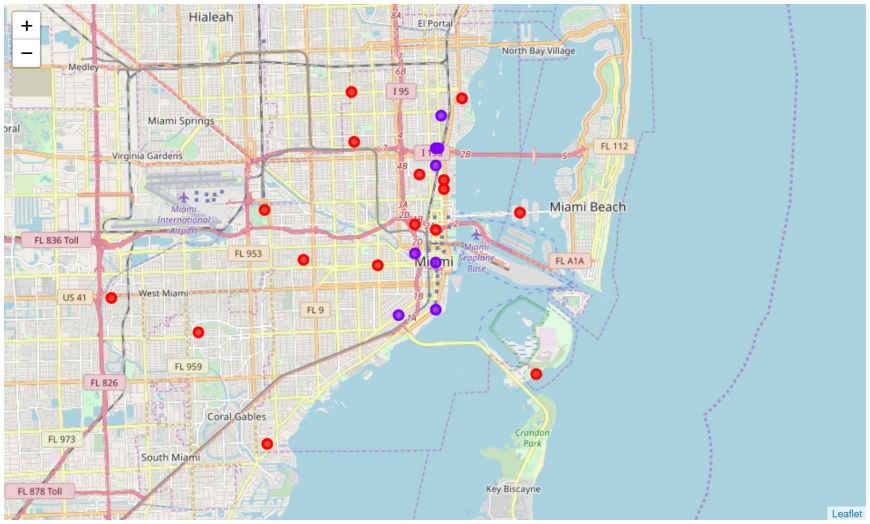

Figure 4: Map of Miami with differentiated markers

4.2 Targeting medium-high median household income

Now that we have narrowed down our potential locations based on existing Italian restaurants, we can explore the median household distribution in Miami so as to establish our restaurant in a location which median income matches our price range.

Since the only data available for the Miami distribution of household income per neighborhood is not retrievable per known methods, we will simply work with a snapshot of that data, as seen in Fig. 5. To align with our goals, we would like to look at those neighborhoods which median household income are above that of Miami, which are shown in green.

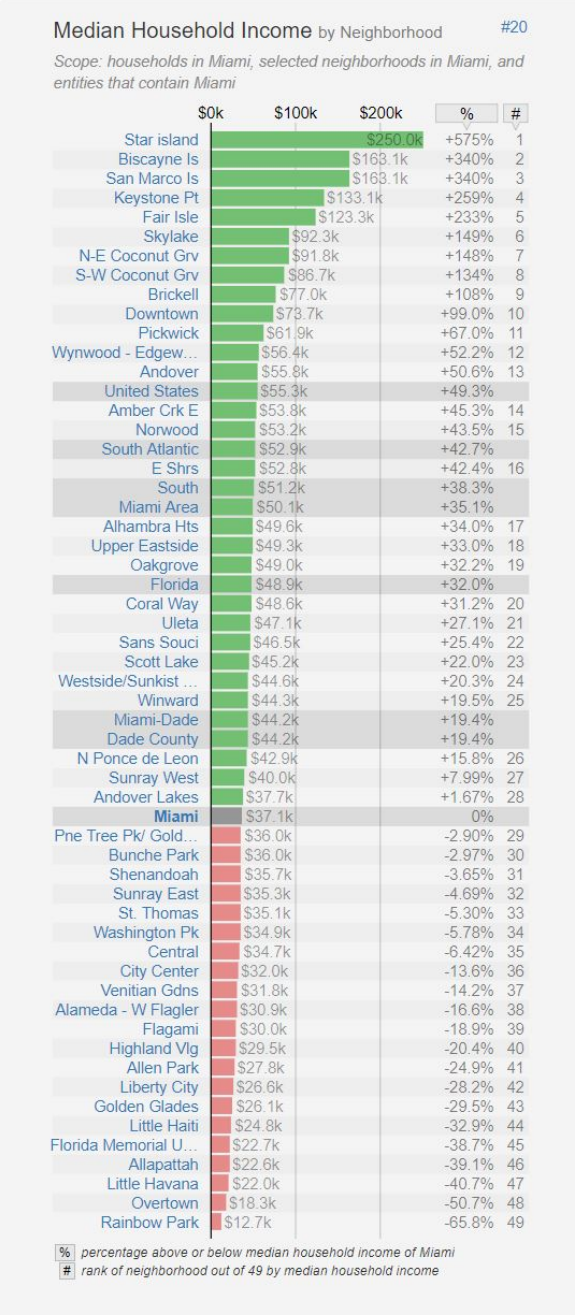

Figure 5: Miami median household income by neighborhood

5 Conclusion

Our conclusion of which location is the best to open a new Italian restaurant is the intersection of the neighborhoods selected in the previous section - i.e., neighborhoods that both have an established Italian restaurant culture and that have a higher median household income if compared to that of general Miami. Our best candidates are the neighboring neighborhoods of Brickell and Downtown.

This case study offers a simplistic approach in order to determine the location of a new Italian restaurant, which conclusion is based on the assumption that if a certain type of restaurant will thrive in a certain location, then a similar one will also do. Thus, it is recommended that, in order to achieve a well-fundamented conclusion, more aspects should be investigated, such as each neighborhood's dynamics, demographics and economics, which is out of the scope for this project.