Choosing location for Opening Restaurant in Miami

Sai Sumanth A.S.

Introduction:

Background

A major factor in the success of a restaurant is location and accessibility. Business can be driven by proximity to areas with a high volume of pedestrians, such as a thoroughfare, a movie theatre, or shopping center. Unless a restaurant has built up a local reputation or has extensive advertising, they can be dependent on walk-ins. A restaurant owner must not only pick the right location, but also appeal to the primary audience that they are serving. For example, if the restaurant is located near a suburban neighborhood, the owners should consider making it family oriented. Understanding the clientele is a crucial element of running a successful restaurant.

Problem and Interest

Miami is famous for its vacation spot. Its is also a multicultural place with lots of tourists and the busiest time is from mid-February through May. The goal of this analysis is to open a oriental cuisine restaurant in an ideal location where one can run successful business. We will go through each step of this project and address them separately.

Data acquisition and cleaning

The Data is collected from Wikipedia by web scraping using Beautiful Soup library. The data includes Neighborhoods, Population, sub-neighborhoods, co-ordinates. To explore and target recommended locations across different venues according to the presence venues such as restaurants, gym, cafe etc. We will access data through FourSquare API interface and arrange them as a dataframe for visualization. By merging data on Miami neighborhoods and data on venues from FourSquare API interface, we will be able to recommend Ideal Location.

After Cleaning the data, the acquired data looks like

	Neighborhoods	Total Population	Population(per Km)	Latitude	Longitude
0	Allapattah	54,289	4,401	25.815	-80.224
1	Arts & Entertainment District	11,033	7,948	25.799	-80.190
2	Brickell	31,759	14,541	25.758	-80.193
3	Buena Vista	9,058	3,540	25.813	-80.192
4	Coconut Grove	20,076	3,091	25.712	-80.257

Plotting the locations to get a better understanding of neighborhoods
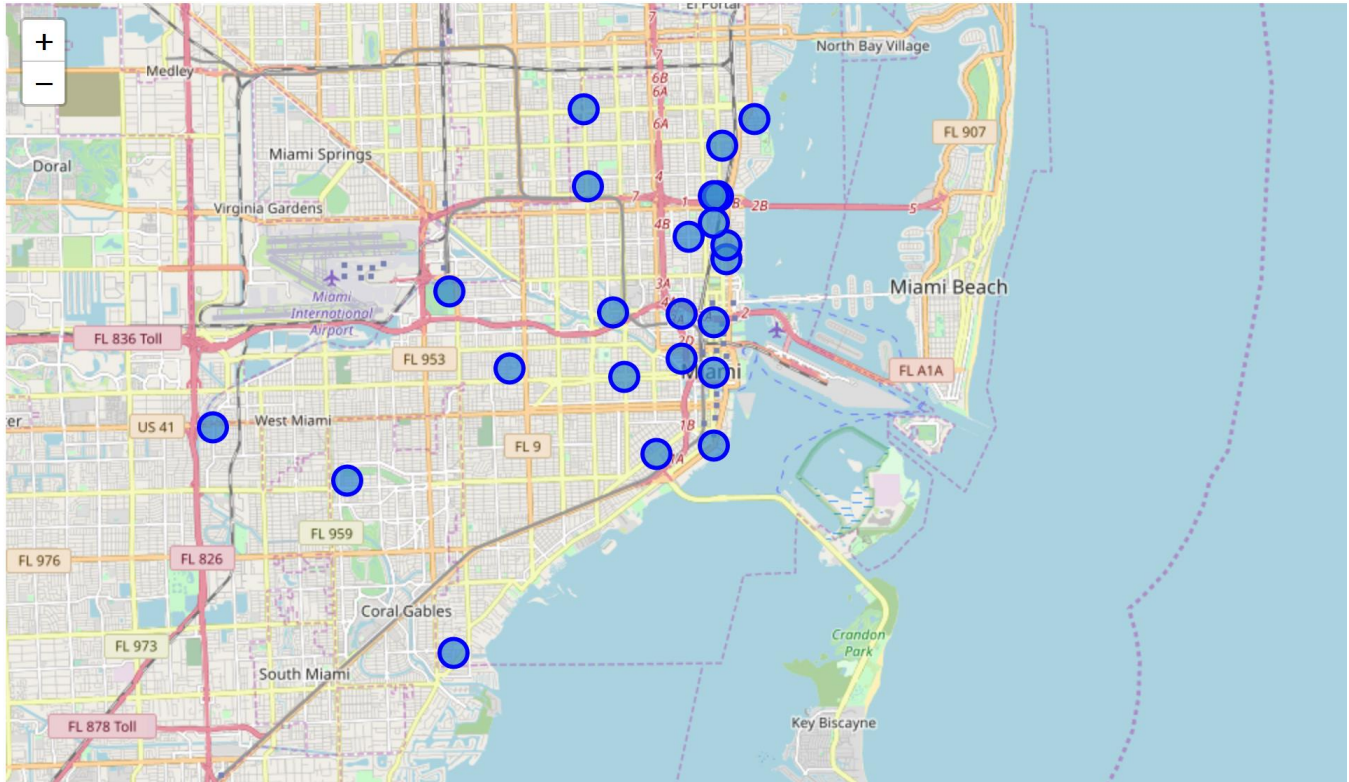

Obtaining and Exploring venues using FourSquare API

Now that we have our Neighborhood location, let's use Foursquare API to get venues in each neighborhood.

We're interested in venues in the 'food' category, but only those that are proper restaurants - coffee shops, pizza places, bakeries, etc. are not direct competitors so we don't care about those. So we will include in our list only venues that have 'restaurant' in the category name, and we'll make sure to detect and include all the subcategories of specific 'Italian restaurant' category.

After pulling the venues from foursquare API around neighborhoods, we will plot it over leaflet maps in folium library. As we can observe, Most of the restaurants are close to the beaches. This make sense as beaches attracts people across the land.
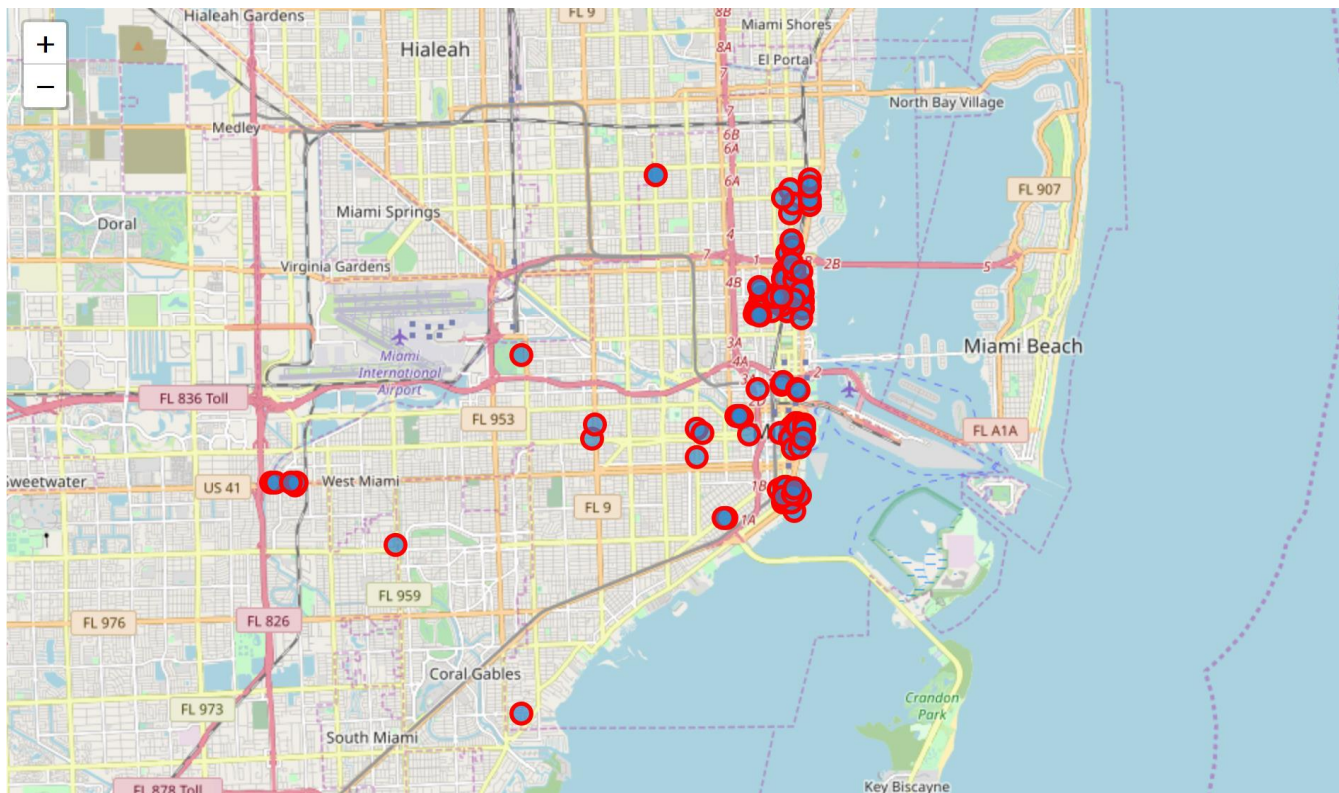
Data Visualization and Some Simple Statistical Analysis

After Creating a Data-frame with the 10 Most Frequently Occuring Venue Category, we will use seaborn library for visualizations.
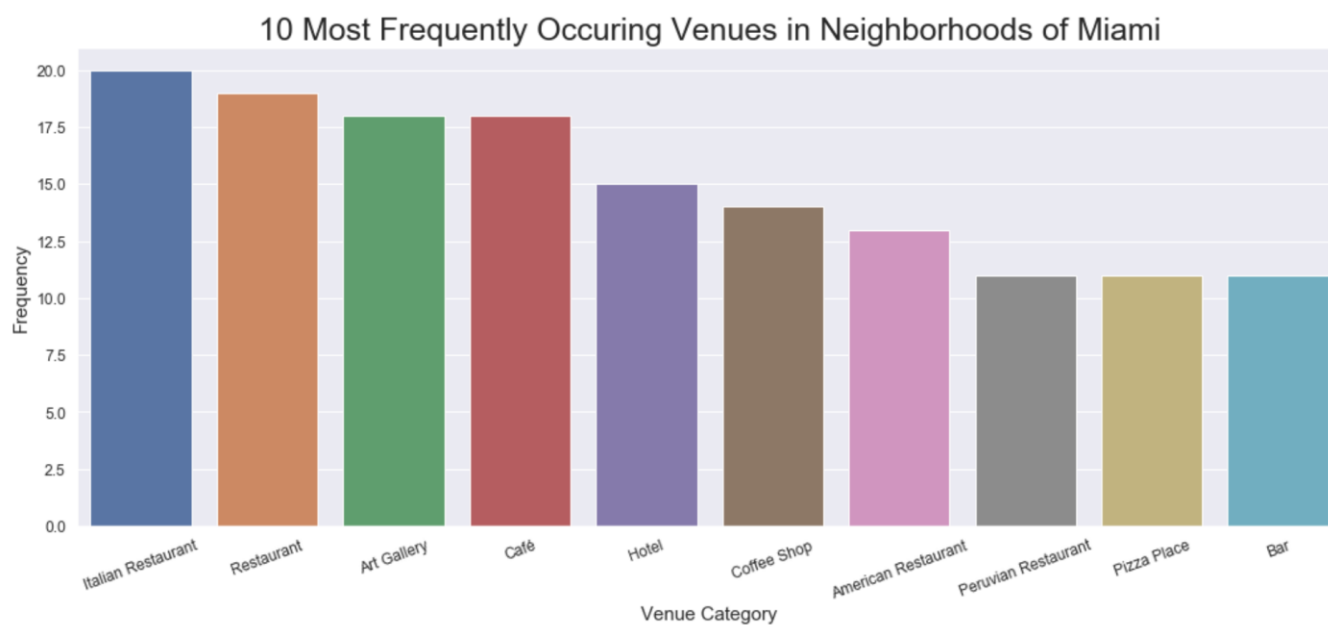
Now its a proven fact that Restaurants are most frequently visited venues and Italian restaurants are the number 1 most visited venue according to the data. Let's analyze the total number of Italian restaurants among neighborhoods.

Total no. of Restaurants among neighborhoods from all venues:
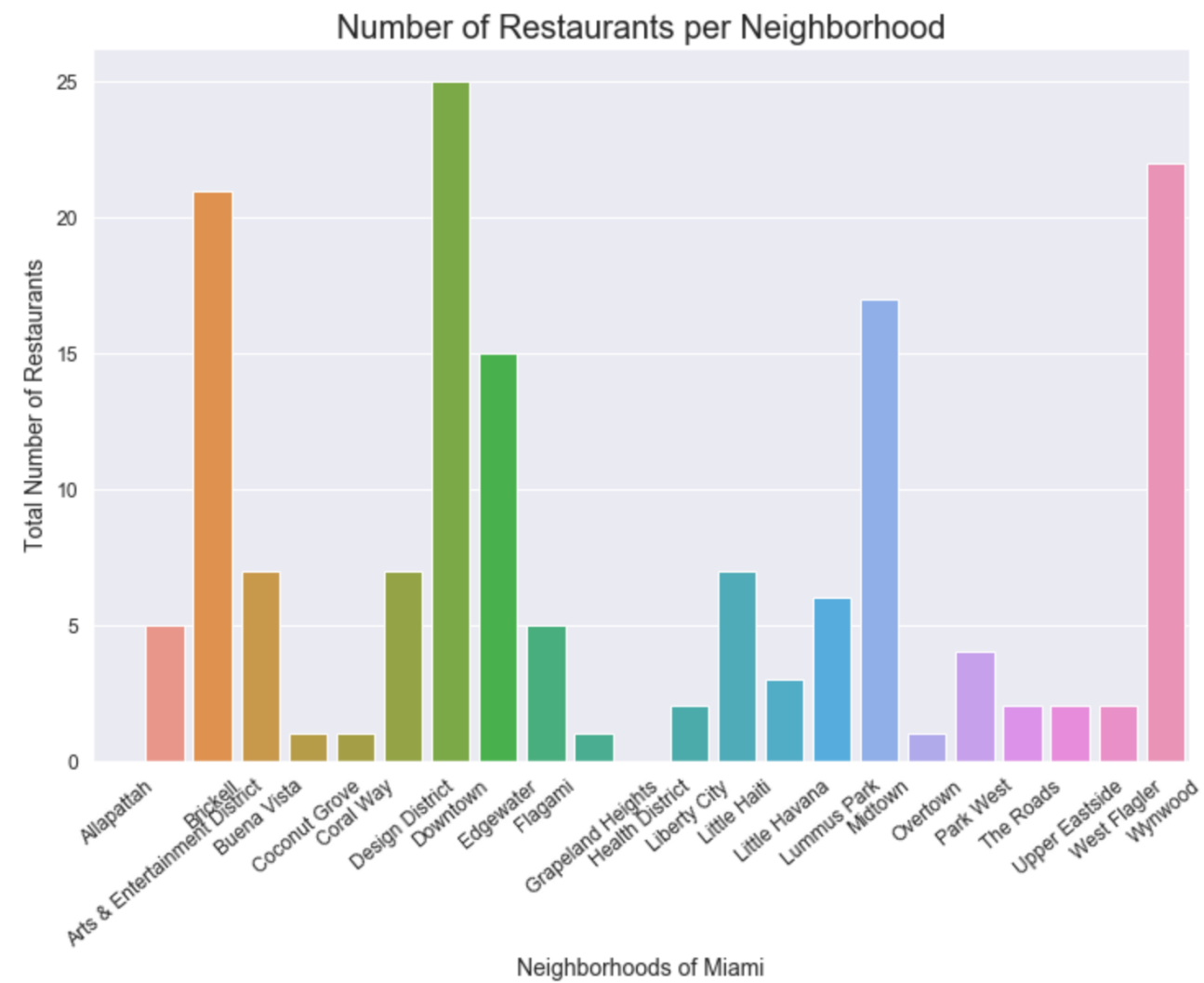
If one has decided to Open a restaurant, Above plot could be useful in choosing the proximity location.

Total number of Italian restaurants among neighborhoods:

Total Italian Restaurants in Neighborhoods of Miami
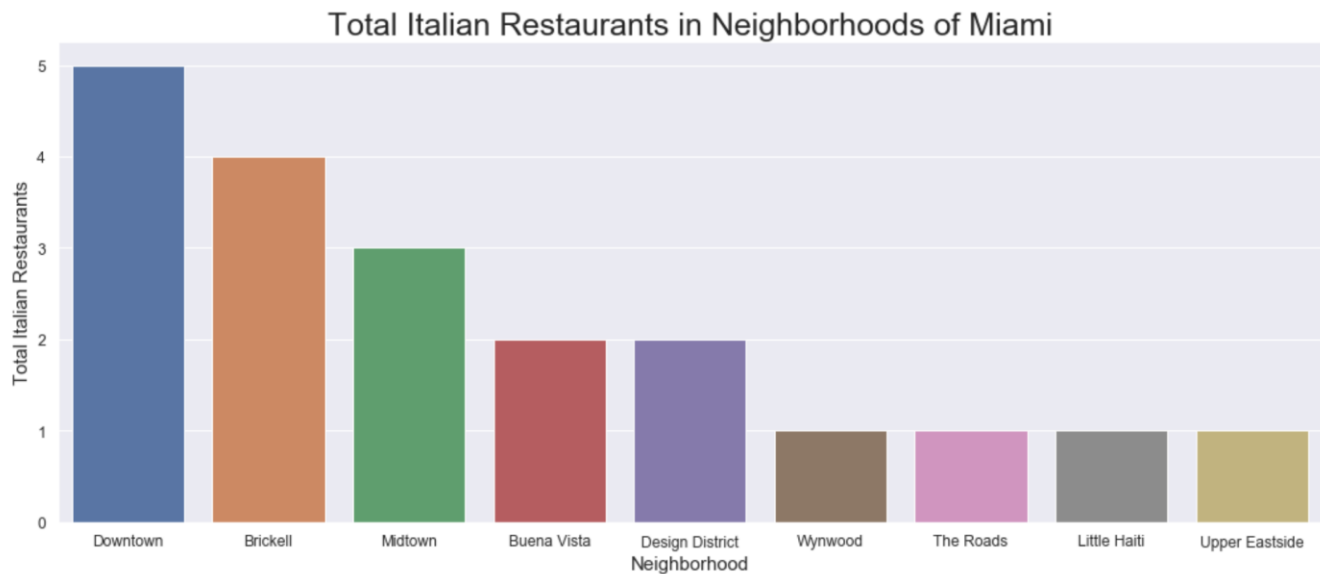
As we can infer, Downtown has most number of italian restaurants with a total of 5 followed by brickwell and midtown.

Modeling Using K-Means Clustering.

We have used silhouette method in obtaining the optimal k. The silhouette scores are

```
[0.29816592022856764,
 0.31443766971401155,
 0.2638595452638041,
 0.2757825332047923,
 0.24606758342000468,
 0.20433584023059218,
 0.19754501817814282,
 0.16670304365480537,
 0.061540562970160785,
 0.05497692083550228,
 0.11787262880502716,
 0.05021179896608412,
 0.12156146018781241,
 0.06825220400352448]
```

The best value of k is 3. With that we build a K-Means model for clustering and obtained cluster labels. I have merged the cluster labels with the data for further cluster wise analysis.
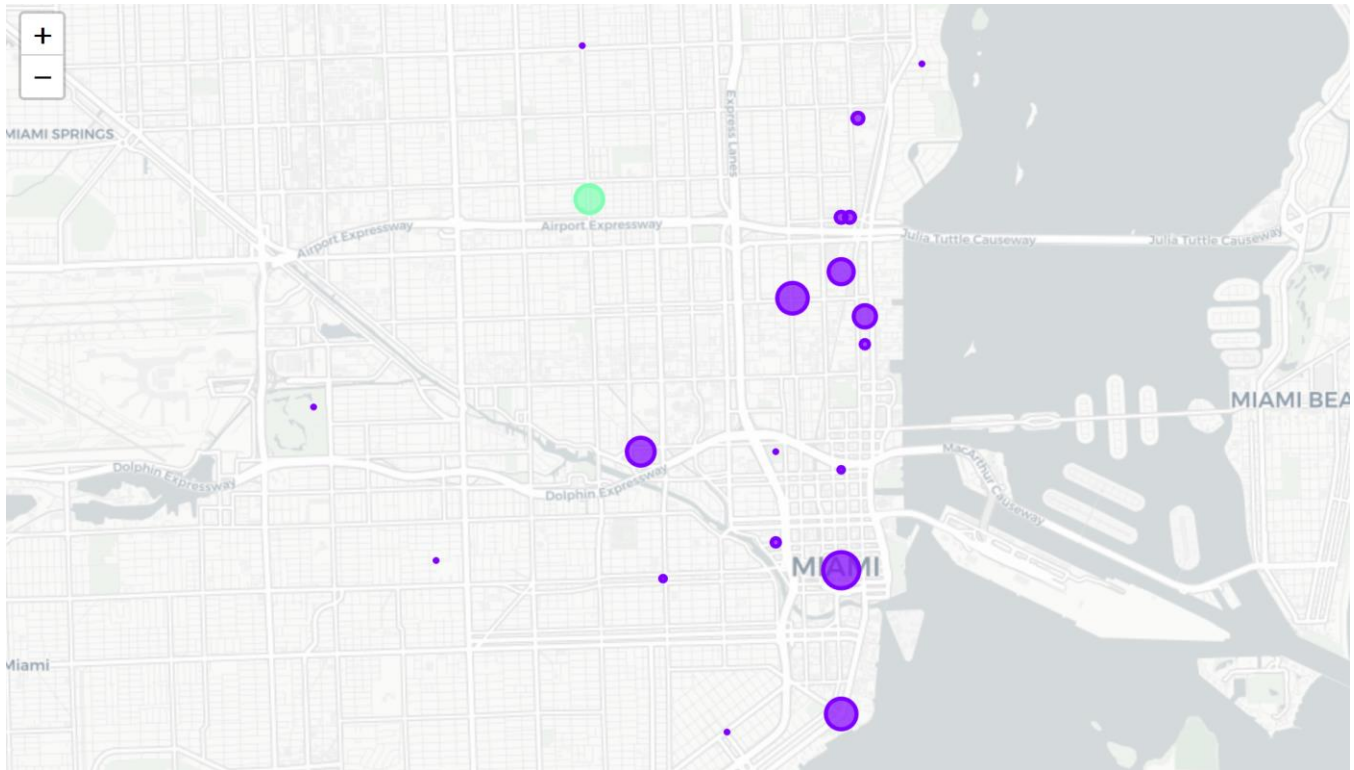
Neighborhoods of Miami segmented into 3 clusters based on the most common venues. The size of the circles represents number of restaurants as most common venues for each neighborhood.

Inference From these Results and related Conclusions.

We have successfully analyzed the neighborhoods of Miami and, as we started our analysis with pros and cons of opening a continental restaurant like italian restaurant in one of the busiest neighborhoods, the data exploration was mostly targeted on restaurants. I have collected the data from web resources like Wikipedia, python libraries like Geopy, and Foursquare API, to set up a very realistic data-analysis scenario. We have found out that

- continental cuisine restaurants top the charts of most common venues in the neighborhoods.
- Downtown, Midtown and brickell neighborhoods are dominated by restaurants with italian restaurants being most coommon venue whereas overtown and wynwood are dominated by bars, theatre, and cafe as most common venues.
- Downtown has maximum number of restaurants as the most common venue whereas has grapeheights area has the none.
- Since the clustering was based only on the most common venues of each district, Allapatah fall under one cluster and coconut grove fall under another cluster. since rest of the neighborhoods has most common venues, they fall under one cluster.

According to this analysis, Allapattah Neighborhood will provide least competition for an upcoming continental restaurant as stores & markets are the most common venue in this area and the frequency of

restaurants as the most common venue is very low compared to the remaining neighborhoods. The average Population per Km in Allapattah is denser compared to other neighborhoods and is close to beachs, downtown in miami. This Neighborhood could be a strategic location for continental restaurants. Few limitations and drawbacks of the analysis are

- The Model we build is completely relied on the data provided by FourSquare API. In actuality, There will be many more features or factors to estimate an ideaal location for a restaurant.
- This Assumptions we made is far from actuality. However, it certainly gives us some very important preliminary information on possibilities of opening restaurants around the neighborhoods of Miami.
- The will be more conclusive evidence if more factors/features proovided.

CONCLUSIONS:

- We got a wonderful exposure of how real life data-science projects look like.
- We have explored various methods and used libraries like Beautiful soup for web scraping and Foursquare API in obtaining neighborhoods of Miami and used Folium leaflet map to visualize neighborhoods.
- We have drawn few potential results using the data for an ideal location in opening a restaurant. Explored the areas of improvement.