Miami Neighborhood Restaurant Market Analysis

Alejandro Villasmil

February 23, 2020

1. Introduction

1.1 Background

Miami, FL is considered one of the cultural epicenters of the United States. Immigrants from South American Countries move to Miami to pursue new dreams and open new businesses. Given this, there is vast opportunity for business growth in the city of Miami if one is able to pinpoint the right market and location to open a new business.

Choosing the right location, depending on the type of business being considered, can either make or break a business. It is essential to find a location that will guarantee a wide customer base while at the same time not be incorporating itself into an overly saturated part of town.

1.2 Problem

My client, a young restaurant entrepreneur, has come to me for help in assisting him in choosing the optimal location in Miami for his new upscale Peruvian grill restaurant, Los Pollos.

1.3 Interest

My client's interest in my recommendation is essential, as a poor decision in choosing a location could very well run him out of business in the coming years.

2. Data Acquisition and Cleaning

2.1 Data Sources

In order to carry out this analysis I will need data on each of the distinct neighborhoods in Miami. This includes longitude and latitude data, as well as data corresponding to the various venues in the vicinity of these several neighborhoods. I was able to scrape the list of Miami neighborhoods off the third table found on this link. Using these names, I was then able to utilize geopy's geolocator to extract the latitude and longitude coordinates of each neighborhood. This was able to produce the dataframe shown below.

	Neighborhood	Municipality	Latitude	Longitude
0	Andover	Miami Gardens	25.968425	-80.212826
1	Biscayne Gardens	Golden Glades	25.921400	-80.217098
2	Bunche Park	Miami Gardens	25.920649	-80.236993
3	Carol City	Miami Gardens	25.940649	-80.245604
4	Coral Way Village	Westchester	25.747305	-80.317700
5	Cutler	Palmetto Bay	25.601008	-80.335983
6	Dadeland	Kendall	25.689273	-80.314201
7	East Perrine	Palmetto Bay	25.608716	-80.338942
8	Green-Mar Acres	Kendall	25.670937	-80.342831
9	Hawley Heights	Kendall	25.671813	-80.354861
10	Howard	Kendall	25.647326	-80.334219
11	Lake Lucerne	Miami Gardens	25.965092	-80.241438
12	Lakes by the Bay	Cutler Bay	25.572329	-80.325331
13	Norwood (Norland)	Miami Gardens	25.952025	-80.224928
14	Saga Bay	Cutler Bay	25.580373	-80.324675
15	Scott Lake	Miami Gardens	25.941482	-80.231993
16	South Beach	Miami Beach	25.774429	-80.133241
17	West Kendall	Kendall	25.671813	-80.354861

Figure 1; Combined Neighborhood Names with Geolocator Coordinates

The data is now ready to feed into the Foursquare API to obtain the venue information in each neighborhood.

2.2 Data Preparation / Feature Selection

Fetching the coordinates of the neighborhoods ran smoothly except for the neighborhood West Kendall, Kendall. I realized the coordinates that were returned originally (43.335, -78.07) were outliers compared to the rest of the neighborhoods seen above in Figure 1. I subsequently searched it on a map and came to the realization that it was referring to a similarly names neighborhood in Canada. Therefore, I had to manually modify the search address for that specific instance and add the city Miami after the neighborhood name in order to get the correct results.

Following this, I initiated several explore requests using the Foursquare API to obtain venue data within a 1000-meter perimeter of each neighborhood's coordinates. The data was parsed through in order to obtain the venue categories associated with each neighborhood. Once obtained, these were then one-hot encoded in order to facilitate cluster modeling further down the line. The results of this step led to the following dataframe (first 5 rows shown).

	Neighborhood	Accessories Store	Airport Service	American Restaurant	Argentinian Restaurant	Art Gallery	Asian Restaurant	Auto Workshop	BBQ Joint	Bakery	...	Tapas Restaurant	Tex-Mex Restaurant	Trail	Train Station	Res
0	Andover	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
1	Biscayne Gardens	0	0	0	0	0	0	0	0	0	...	0	0	0	1	
2	Biscayne Gardens	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
3	Biscayne Gardens	0	1	0	0	0	0	0	0	0	...	0	0	0	0	
4	Biscayne Gardens	0	0	0	0	0	0	0	0	0	...	0	0	0	0	

Figure 2; One-hot encoded venue category features

In figure 2, we can see the features that will ultimately be fed into the machine learning algorithm chosen. This final training set consists of 268 samples with 109 features. The features in this case are the unique venue categories returned by the

Foursquare API and the samples correspond to each venue returned by the Foursquare API for all the neighborhoods that were originally scraped from the Wikipedia table. This data structure will allow the machine learning model to accurately find relationships between each of the neighborhoods and the distinct venue categories housed within each one.

3. Methodology

3.1 Exploratory Data Analysis

After fetching the venue data using the Foursquare API, the saturation of venues in each neighborhood was visualized by counting the number of venues in each neighborhood.

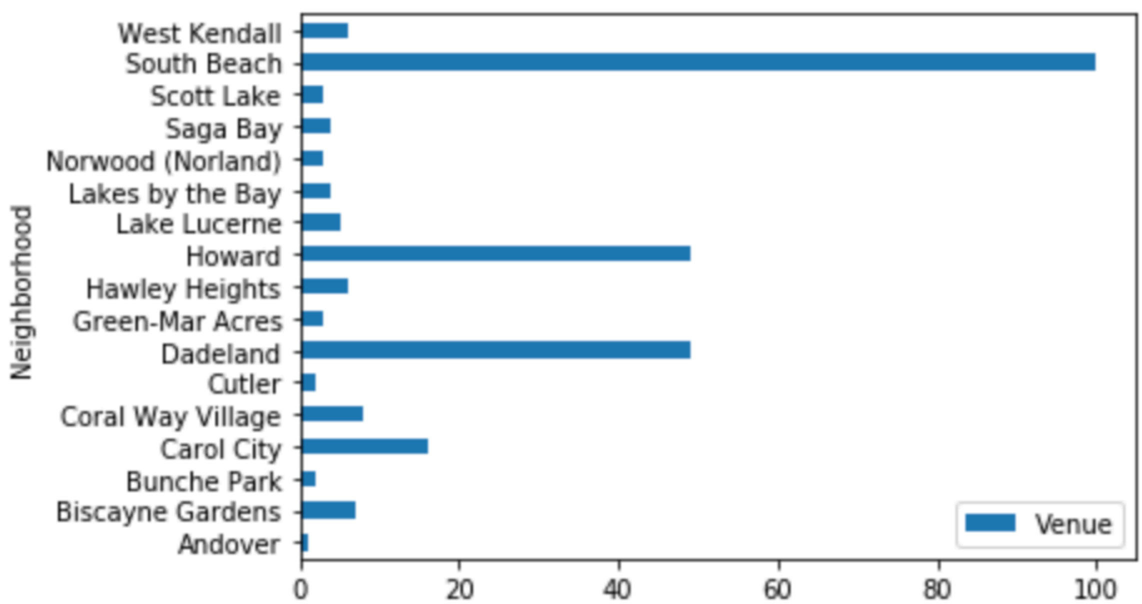

Figure 3; Number of Venues in Each Neighborhood

From this initial visualization, it seems like Carol City, Dadeland, Howard, and South Beach could all be potential contenders for the restaurant as they all have significant amount of foot traffic and business. However, South Beach could potentially be slightly over saturated.

Next, visualization of the most prevalent venue categories in each neighborhood was executed.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Dadeland	Clothing Store	Department Store	Coffee Shop	Furniture / Home Store	American Restaurant	Indian Restaurant	Italian Restaurant	Japanese Restaurant	Jewelry Store	Kids Store
Green-Mar Acres	Yoga Studio	Playground	Art Gallery	French Restaurant	Cosmetics Shop	Cuban Restaurant	Dance Studio	Deli / Bodega	Department Store	Discount Store
Hawley Heights	Pool	Burger Joint	Golf Course	Café	Bed & Breakfast	Soccer Field	Convenience Store	Cosmetics Shop	Cuban Restaurant	Dance Studio
Howard	Grocery Store	Restaurant	Hardware Store	Sporting Goods Shop	Pharmacy	Italian Restaurant	Fast Food Restaurant	Gym / Fitness Center	Furniture / Home Store	Burger Joint
Lake Lucerne	Restaurant	Casino	Yoga Studio	Coffee Shop	Convenience Store	Cosmetics Shop	Cuban Restaurant	Dance Studio	Deli / Bodega	Department Store
Lakes by the Bay	Housing Development	American Restaurant	Trail	Gym	Food Truck	Convenience Store	Cosmetics Shop	Cuban Restaurant	Dance Studio	Deli / Bodega
Norwood (Norland)	Park	Auto Workshop	Food Truck	Yoga Studio	French Restaurant	Cosmetics Shop	Cuban Restaurant	Dance Studio	Deli / Bodega	Department Store
Saga Bay	Beach	Grocery Store	Concert Hall	Yoga Studio	French Restaurant	Cosmetics Shop	Cuban Restaurant	Dance Studio	Deli / Bodega	Department Store
Scott Lake	American Restaurant	Discount Store	Snack Place	Yoga Studio	Food Truck	Convenience Store	Cosmetics Shop	Cuban Restaurant	Dance Studio	Deli / Bodega
South Beach	Hotel	Clothing Store	Seafood Restaurant	Beach	Pharmacy	Italian Restaurant	Cuban Restaurant	Park	Coffee Shop	Pizza Place
West Kendall	Pool	Burger Joint	Golf Course	Café	Bed & Breakfast	Soccer Field	Convenience Store	Cosmetics Shop	Cuban Restaurant	Dance Studio

Figure 4; Most Prevalent Venue Categories in Each Neighborhood

As can be seen in Figure 4 above, South Beach's leading venue are hotels followed closely by stores and restaurants. Dadeland's leading categories are mainly all comprised of stores and coffeeshops. This is insightful because it shows that just because a neighborhood is saturated with venues doesn't necessarily mean it's a negative factor for a new and upcoming restaurant. It all depends on the venue categories.

3.2 Machine Learning Algorithm

In this case, I chose to implement a k-means clustering algorithm in order to cluster the neighborhoods into independent clusters based on the distinct venue categories found in each one. The first step in implementing this model, would be to choose an appropriate value of k that will minimize the amount of inertia found in the resulting clusters. Inertia is used as a distortion metric in the model to gauge how internally coherent the clusters are. Essentially attempting to minimize the intra-cluster distances between samples while maximizing the inter-cluster distances.

The elbow method was implemented in order to determine what value of k would be the best fit for the model. The results can be seen below in Figure 5.

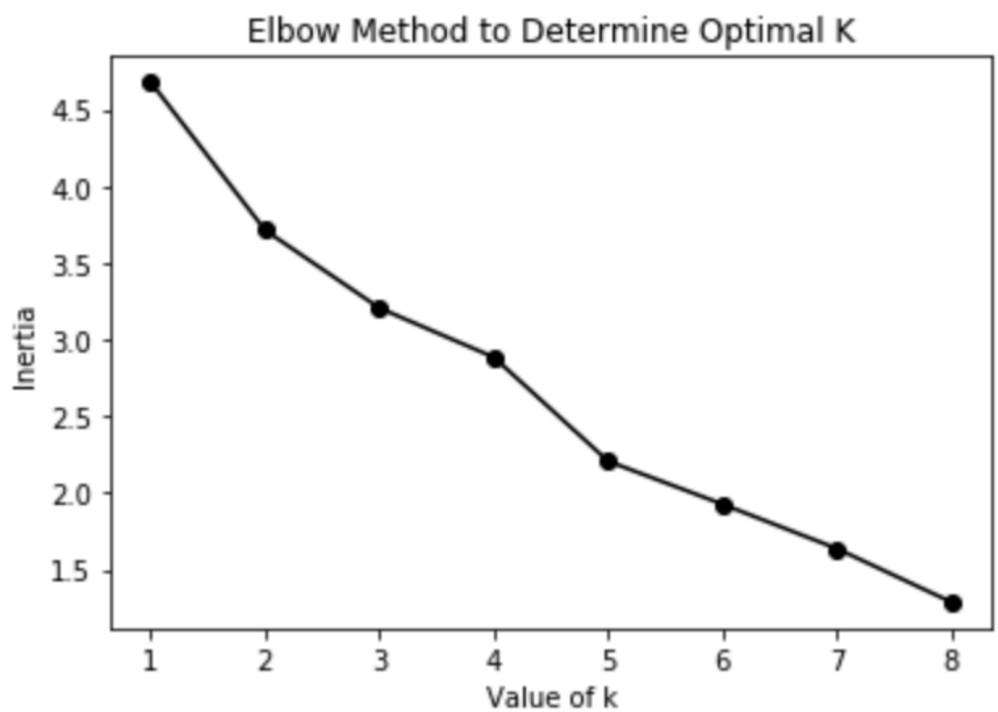

Figure 5; Elbow Method for Determining Optimal Value of k for k-means algorithm

Looking at the graph in Figure 5, the value of 5 was chosen for k as it resulted in smaller drops in inertia after increasing k thereafter. At this point all the parameters needed to fit the model and training data have been obtained and the model is ready to be fit.

4. Results

The k-means model, with 5 clusters as determined in Figure 5, was now fit with the one-hot encoded venue categories in shown in Figure 2. The result of this model was a cluster label (0-4) associated with each of the neighborhoods.

	Neighborhood	Municipality	Latitude	Longitude	Cluster Labels
0	Andover	Miami Gardens	25.968425	-80.212826	2
1	Biscayne Gardens	Golden Glades	25.921400	-80.217098	1
2	Bunche Park	Miami Gardens	25.920649	-80.236993	4
3	Carol City	Miami Gardens	25.940649	-80.245604	1
4	Coral Way Village	Westchester	25.747305	-80.317700	1
5	Cutler	Palmetto Bay	25.601008	-80.335983	0
6	Dadeland	Kendall	25.689273	-80.314201	1
7	Green-Mar Acres	Kendall	25.670937	-80.342831	1
8	Hawley Heights	Kendall	25.671813	-80.354861	1
9	Howard	Kendall	25.647326	-80.334219	1
10	Lake Lucerne	Miami Gardens	25.965092	-80.241438	3
11	Lakes by the Bay	Cutler Bay	25.572329	-80.325331	1

Figure 6; First 11 Neighborhoods with their respective clusters as predicted by the fitted k-means model

In order to better visualize these cluster labels a geographical map was generated with Folium showing each neighborhood and its color-coded cluster label.

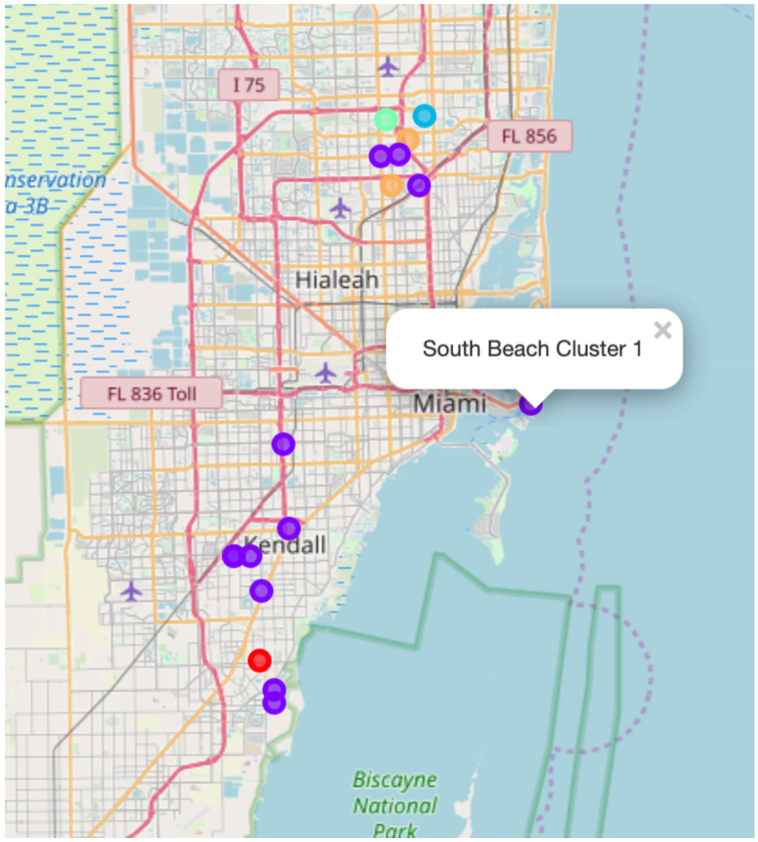

Figure 7; Geographical visualization of predicted neigborhood K-means cluster labels

Now that the clusters have been determined, it is essential that the discriminating venue categories for each cluster label be determined. This will help in deciding which neighborhood will be best suited for an up and coming upscale Peruvian Grill Restaurant. The neighborhoods and their most common venues for each cluster label are shown below.

<u>Cluster 1</u>

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Cutler	Grocery Store	Lighthouse	Yoga Studio	Food Truck	Convenience Store	Cosmetics Shop	Cuban Restaurant	Dance Studio	Deli / Bodega	Department Store

Cluster *2*

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Biscayne Gardens	Bus Station	Building	Airport Service	Home Service	Train Station	Yoga Studio	Cosmetics Shop	Cuban Restaurant	Dance Studio	Deli / Bodega
3	Carol City	Fast Food Restaurant	Department Store	Seafood Restaurant	Sandwich Place	Miscellaneous Shop	Strip Club	Bakery	Caribbean Restaurant	Trail	Grocery Store
4	Coral Way Village	Spa	Bakery	Cafeteria	Fast Food Restaurant	Market	Gas Station	Sushi Restaurant	Pizza Place	Greek Restaurant	Grocery Store
6	Dadeland	Clothing Store	Department Store	Coffee Shop	Furniture / Home Store	American Restaurant	Indian Restaurant	Italian Restaurant	Japanese Restaurant	Jewelry Store	Kids Store
7	Green-Mar Acres	Yoga Studio	Playground	Art Gallery	French Restaurant	Cosmetics Shop	Cuban Restaurant	Dance Studio	Deli / Bodega	Department Store	Discount Store
8	Hawley Heights	Pool	Burger Joint	Golf Course	Café	Bed & Breakfast	Soccer Field	Convenience Store	Cosmetics Shop	Cuban Restaurant	Dance Studio
9	Howard	Grocery Store	Restaurant	Hardware Store	Sporting Goods Shop	Pharmacy	Italian Restaurant	Fast Food Restaurant	Gym / Fitness Center	Furniture / Home Store	Burger Joint
11	Lakes by the Bay	Housing Development	American Restaurant	Trail	Gym	Food Truck	Convenience Store	Cosmetics Shop	Cuban Restaurant	Dance Studio	Deli / Bodega
13	Saga Bay	Beach	Grocery Store	Concert Hall	Yoga Studio	French Restaurant	Cosmetics Shop	Cuban Restaurant	Dance Studio	Deli / Bodega	Department Store
14	Scott Lake	American Restaurant	Discount Store	Snack Place	Yoga Studio	Food Truck	Convenience Store	Cosmetics Shop	Cuban Restaurant	Dance Studio	Deli / Bodega
15	South Beach	Hotel	Clothing Store	Seafood Restaurant	Beach	Pharmacy	Italian Restaurant	Cuban Restaurant	Park	Coffee Shop	Pizza Place
16	West Kendall	Pool	Burger Joint	Golf Course	Café	Bed & Breakfast	Soccer Field	Convenience Store	Cosmetics Shop	Cuban Restaurant	Dance Studio

Cluster 3

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Andover	Food	Yoga Studio	French Restaurant	Convenience Store	Cosmetics Shop	Cuban Restaurant	Dance Studio	Deli / Bodega	Department Store	Discount Store

Cluster 4

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
10	Lake Lucerne	Restaurant	Casino	Yoga Studio	Coffee Shop	Convenience Store	Cosmetics Shop	Cuban Restaurant	Dance Studio	Deli / Bodega	Department Store

Cluster 5

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Bunche Park	Salon / Barbershop	Park	Yoga Studio	Coffee Shop	Convenience Store	Cosmetics Shop	Cuban Restaurant	Dance Studio	Deli / Bodega	Department Store
12	Norwood (Norland)	Park	Auto Workshop	Food Truck	Yoga Studio	French Restaurant	Cosmetics Shop	Cuban Restaurant	Dance Studio	Deli / Bodega	Department Store

Figure 8; Most common venue categories for each cluster

5. Discussion

From the results presented above, each cluster's distinct features were able to be deduced from its unique venue categories as such:

- Cluster 1: Calm and tranquil, neighborhoods not associated with much foot traffic.
- Cluster 2: Touristy sections of Miami, associated with large foot traffic, shops, and restaurants.
- Cluster 3: Food centric neighborhoods.
- Cluster 4: Neighborhoods with an older population.
- Cluster 5: Neighborhoods with a large focus on the outdoors and urban development.

With these clusters and the data presented in the figures above, there is now enough information to provide a location recommendation to my client for his new restaurant. From the descriptions derived for each cluster, it is evident that one of the neighborhoods within cluster 2 would be the most suitable, as it would drive the most business and foot traffic for the new business.

Looking further into cluster 2, specifically using Figure 2, we can see that South Beach would probably not be the best choice for his first restaurant as it is heavily saturated with other competition and it would potentially not allow him to obtain an initial customer market that easily. From the looks of it, either Howard or Dadeland would be suitable locations as they are popular destinations but are not overly saturated.

Looking further into these two neighborhoods, specifically using figure 8 above, it is evident that Dadeland is populated with numerous shops and stores while Howard's second most common category is Restaurants.

Given this information, it would be a wise choice to bring this new restaurant to the Dadeland area as it will serve as a perfect complement to the plethora of stores present in the area. People will want a suitable place to eat while out shopping for the day.

6. Conclusion

In this study, I was able to assist my client in finding the most suitable location in Miami for his new upscale Peruvian Grill Restaurant, Los Pollos. I was able to scrape data off the internet regarding each of the neighborhoods in Miami, use geopy's geolocator to calculate the coordinates for each, and then obtain data about the different venues present in each neighborhood using the Foursquare API. I determined the best k-means clustering model to be used, using the elbow method, and then fit the model using the data obtained from the Foursquare API. Using the model's predictions, I was then able to make sense of the different clusters and recommend the neighborhood of Dadeland to my client for his restaurant.