

# **Capstone Project – The Battle of Neighborhoods (Week 2)**

## **Finding a Neighborhood in Cincinnati to Open a Latin American Restaurant**

Miguel Catalán

February 14, 2021

### **1. Introduction – Business Problem**

In Latin America some times the difficulties to find a good job or a job that pays enough to sustain a family properly is a big issue, for that reason many people use to migrate to the United States, regularly one single member is the one who looks to establish there and help their family. One common situation with this change of culture is the food, many of them look for places to eat the food they use to eat in their hometowns and that's not always easy to find.

This project has the goal to analyze Cincinnati, Ohio venues to establish which could be the best neighborhood to open a Latin American restaurant. To accomplish this objective we are going to analyze the status of each neighborhood regarding restaurants and Latin American restaurants, this will be made using the Foursquare API, to analyze the Cincinnati venues.

This project is aimed at entrepreneurs that are looking for opportunities to establish a new business in the food sector, so the result of this analysis will be data that helps them to choose the better option to start a new business.

### **2. Data**

The data required for this project will be obtained from different sources, which are listed below

#### **2.1 List of neighborhoods in Cincinnati**

This data will be extracted from [Wikipedia](#). From this list, we are going to use only the main neighborhoods, and discard the subclassification listed on that page. To get this data we are going to use BeautifulSoup to do the web scraping. An example of the data is listed next:

## District One [\[ edit \]](#)

### 1. Downtown

- [The Banks \(Riverfront\)](#)
- [Central Business District](#)
- [East Fourth Street District](#)
- [East Manufacturing & Warehouse District](#)
- [Fort Washington](#)
- [Lytle Park District](#)
- [Ninth Street District](#)
- [Race Street District](#)
- [West Fourth Street District](#)

### 2. Mount Adams

### 3. Over-the-Rhine

- [Brewery District](#)
- [Gateway Quarter](#)
- [Mohawk District](#)
- [Northern Liberties](#)
- [Schwartz's Point](#)
- [Sycamore-13th Street District](#)

### 4. Pendleton

### 5. Queensgate

### 6. West End

- [Betts-Longworth District](#)
- [Brighton](#)
- [City West \(Lincoln Court\)](#)

## 2.2 Latitude and Longitude of each neighborhood

LocationIQ API will be used to obtain the latitude and longitude of each neighborhood. This API is used as an alternative to the Google API. To get the data this API needs the name of the place to get the coordinates, in this case, will be the name of the neighborhood obtained in step 1. An example of a response from LocationIQ where we will extract the neighborhood coordinates is shown in the next image:

```
[{'place_id': '152023575', 'licence': 'https://locationiq.com/attribution', 'osm_type': 'way', 'osm_id': '274228529', 'boundingbox': ['39.1004091', '39.1012036', '-84.5196177', '-84.51932'], 'lat': '39.1006666', 'lon': '-84.51948', 'display_name': 'Cincinnati, Central Avenue, Betts-Longworth Historic District, Over-the-Rhine, Cincinnati, Hamilton County, Ohio, 45202, USA', 'class': 'tourism', 'type': 'artwork', 'importance': 0.11100000000000002, 'icon': 'https://locationiq.org/static/images/mapicons/tourist_art_gallery2.p.20.png'}]
```

With these data we are going to create a dataframe like the next:

	neighborhood	latitude	longitude
0	Downtown	39.1006666	-84.51948
1	Mount Adams	39.1075532	-84.4991114
2	Over-the-Rhine	39.1115038	-84.5143927
3	Pendleton	39.1104996	-84.5077559
4	Queensgate	39.1037497	-84.5377193

## 2.3 List of Venues

Foursquare API will be used to obtain the list of venues in each neighborhood. To get the list of venues in each neighborhood we are going to use the coordinates obtained from the LocationIQ API. With this data we are going to look for Latin American

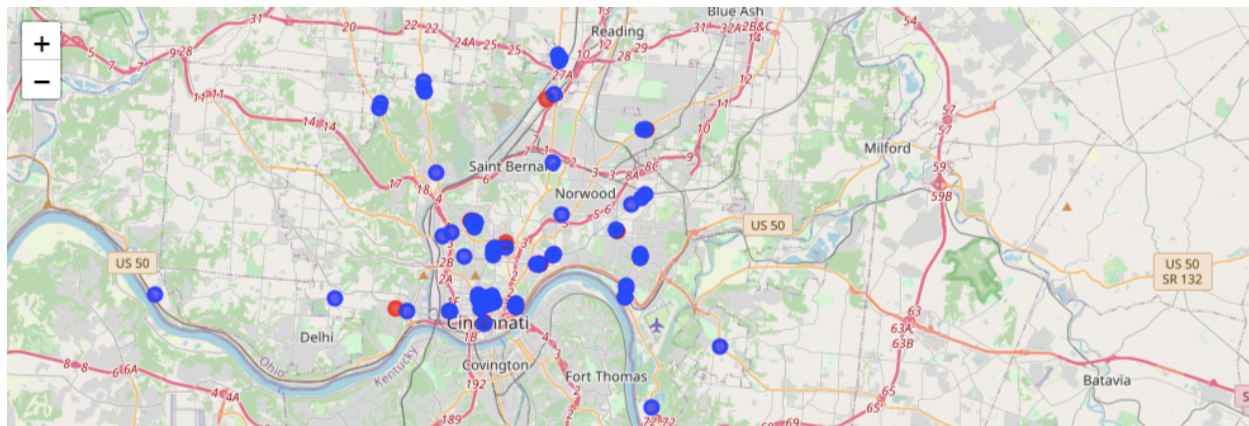
restaurants in the venues list and assign them to each neighborhood, to execute the analysis that we want to do. The list of venues will look like the next image:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Downtown	39.1006666	-84.51948	Maplewood Kitchen and Bar	39.101513	-84.515113	Breakfast Spot
1	Downtown	39.1006666	-84.51948	Mita's	39.101161	-84.514726	Latin American Restaurant
2	Downtown	39.1006666	-84.51948	Orchids at Palm Court	39.100626	-84.514335	New American Restaurant
3	Downtown	39.1006666	-84.51948	Hilton	39.100886	-84.514384	Hotel
4	Downtown	39.1006666	-84.51948	Abby Girl Sweets	39.101057	-84.514314	Cupcake Shop

### 3. Analysis

We are going to start creating a map that displays the physical distribution of these restaurants to get a first glimpse of what we are going to find in the end. For this, we are going to create a Folium map and create a mark for each of the restaurants.

The Latin American restaurants will be marked as red circles and the non-Latin American restaurants will be marked as blue circles.

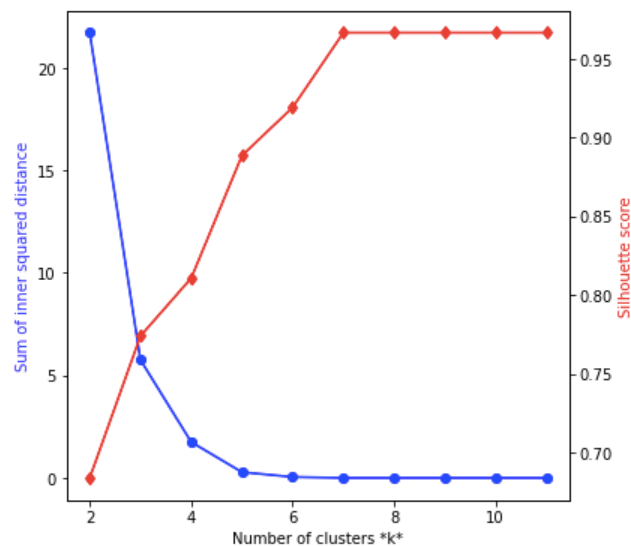


Now we are going to consolidate the data of the venues grouped by neighborhood, with the goal to find how many Latin American and non-Latin American restaurants exist in each neighborhood. To do this we are going to convert the Type variable values into variables, and do counting of how many restaurants of each type do we have in each neighborhood.

	Neighborhood	Latin American	Non Latin American
0	CUF	0	1
1	California	0	1
2	Camp Washington	0	2
3	Carthage	1	1
4	Clifton	2	6
5	College Hill	0	4
6	Columbia-Tusculum	0	5
7	Corryville	1	5
8	Downtown	1	1
9	East Price Hill	1	0
10	East Walnut Hills	0	3
11	Evanston	0	1
12	Hartwell	0	3
13	Hyde Park	1	2
14	Lower Price Hill	0	1
15	Mount Adams	1	3
16	Mount Airy	0	2
17	Mount Lookout	1	5
18	Mount Washington	0	1
19	Northside	0	1

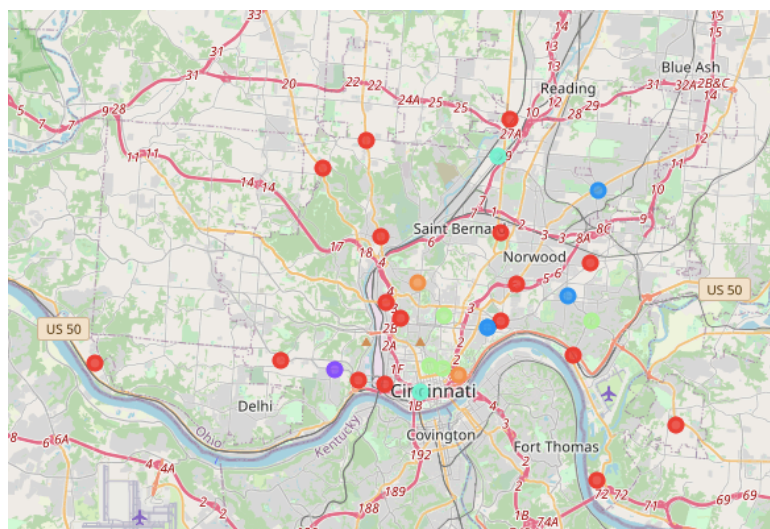
Now we are going to use K-means algorithm to cluster the neighborhoods based on the data obtained with the mean number of restaurants in each neighborhood. The first thing we are going to do is use the elbow method to get an optimum k number for the K-means algorithm.

As result, we are going to obtain the list of clusters that should be assigned to each neighborhood assigned by the result of the K-means algorithm.



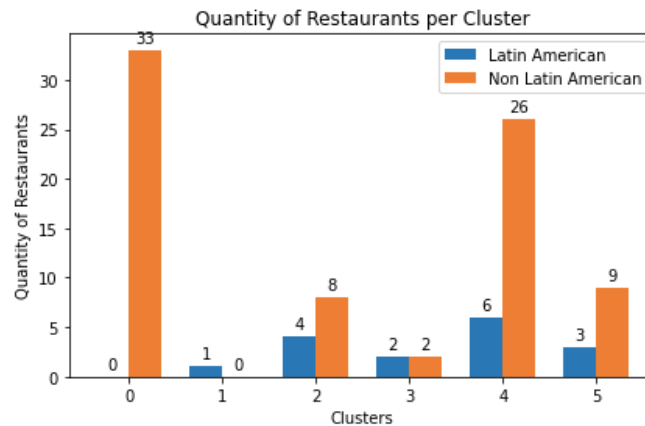
	Neighborhood	Latin American	Non Latin American	Cluster
0	CUF	0.000000	1.000000	0
1	California	0.000000	1.000000	0
2	Camp Washington	0.000000	1.000000	0
3	Carthage	0.500000	0.500000	3
4	Clifton	0.250000	0.750000	5
5	College Hill	0.000000	1.000000	0
6	Columbia-Tusculum	0.000000	1.000000	0
7	Corryville	0.166667	0.833333	4
8	Downtown	0.500000	0.500000	3
9	East Price Hill	1.000000	0.000000	1
10	East Walnut Hills	0.000000	1.000000	0
11	Evanston	0.000000	1.000000	0
12	Hartwell	0.000000	1.000000	0
13	Hyde Park	0.333333	0.666667	2
14	Lower Price Hill	0.000000	1.000000	0
15	Mount Adams	0.250000	0.750000	5
16	Mount Airy	0.000000	1.000000	0
17	Mount Lookout	0.166667	0.833333	4
18	Mount Washington	0.000000	1.000000	0
19	Northside	0.000000	1.000000	0
20	Oakley	0.000000	1.000000	0

Now we can create a Folium map to display the neighborhoods clustered, and give us an idea of the distribution and classification of each neighborhood.



Now using the previous dataframe we can create a summary of how many restaurants Latin American and non-Latin American exists in each cluster.

Cluster	Latin American (Mean)	Non Latin American (Mean)	Latin American (Count)	Non Latin American (Count)	
0	0	0.000000	17.000000	0	33
1	1	1.000000	0.000000	1	0
2	2	1.333333	2.666667	4	8
3	3	1.000000	1.000000	2	2
4	4	0.733333	3.266667	6	26
5	5	0.500000	1.500000	3	9



In this chart, we can see that a good option could be to use cluster 1 or 3 since this cluster has the less quantity of restaurants which the minority are Latin American restaurants. Let's list the neighborhoods that belong to each one of the clusters.

From the previous analysis, we can say that good options for opening a Latin American restaurant are the neighborhoods East Price Hill, Carthage, or Downton which are the neighborhoods that belong to clusters 1 and 3. With this, we conclude our analysis.

## 4. Results and Discussion

The analysis we made was focused on identifying all restaurants that are 500 meters around the central coordinates of each of the neighborhoods. After getting this data we classify each of the restaurants in Latin American or non-Latin American. This was executed only for cities in Cincinnati, and we get a result that three neighborhoods are good options to open a Latin American restaurant.

We have to keep in mind that this project doesn't imply that the neighborhoods identified as options are in fact the best place to open a new restaurant, since we should have other things in mind, as the Latin American population in each of these neighborhoods or around them. Also have to keep in mind that we only analyzed the venues that were 500 meters around the central point of the neighborhood, so probably there were more restaurants that we could be analyzed. Another point to take into count is the price for rent for a restaurant in these neighborhoods. What we want to left clear is that the resulting locations are only starting points for more detailed analysis.

## 5. Conclusion

The purpose of this project was to identify neighborhoods with a low number of Latin American and non-Latin American restaurants in order to make visible to all the persons' interested possible places to open a Latin American restaurant. To do this we list all restaurants around 500 meters from the central location for each neighborhood, group these restaurants as Latin American and non-Latin American, and segment them for each cluster.

In the end, the person who will take the decision to open a new Latin American restaurant should have in mind other variables as rent, Latin American population near each neighborhood.