# Site Recommendation for new Latin Restaurant

**José David Bonilla Placido**

**Junio 05, 2021**

## 1. Introduction

### 1.1. Background

**The Bronx is the only New York borough with a Hispanic majority**. At the 2010 Census, 53.5% of Bronx's population was of Hispanic, Latino, or Spanish origin (they may be of any race). At the 2009 American Community Survey, Puerto Ricans represented 23.2% of the borough's population, Mexicans made 5.2%. In 2005, more than 200,000 Dominicans called the Bronx home.

Restaurants and Culinary habits have always been one of the favorite ways to socialize and meet each other, have interesting conversation beside closing business contract. Knowing these habits, Restaurant income represent a good share of every country GDP (Gross domestic product).

### 1.2. Problem

My project it's a about determining **best restaurant location and segmenting** competitors in Bronx city to find **low Latin restaurant density** by means of data science **identify best location** for opening a new restaurant.

One the most complex decision for every restaurant it's finding a perfect location for introducing a new option for neighborhood and surrounding area at a selected city.

### 1.3. Interest

Restaurant groups, retail chains, and other enterprises can use this work and use POI data to help solve problems such as site selection, competitive analysis, and much more.

## 2. Data acquisition and cleaning

### 2.1. Data Sources

Basically, we used two data sources for this analysis. First the restaurant data used in this study was pulled from Foursquare's Places Database - a comprehensive dataset spanning 190+ countries and 50 territories. The data set is updated continuously and published daily as our world is constantly evolving. The Foursquare Places Database has a rich set of attributes you could use to

understand the world around you. This analysis only needed a few of them: the restaurant's location, the restaurant's categorization, and information telling us if the restaurants were part of a national chain.

Second for Bronx boroughs and its neighborhood we used

https://cf-courses-data.s3.us.cloud-object-torage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json

New York data Jason file from this address.

### 2.2. Data cleaning

In order to **segement the neighborhoods** and explore them, we will essentially need a dataset that contains the Bronx boroughs and the neighborhoods as well as the latitude and longitude coordinates of each neighborhood.

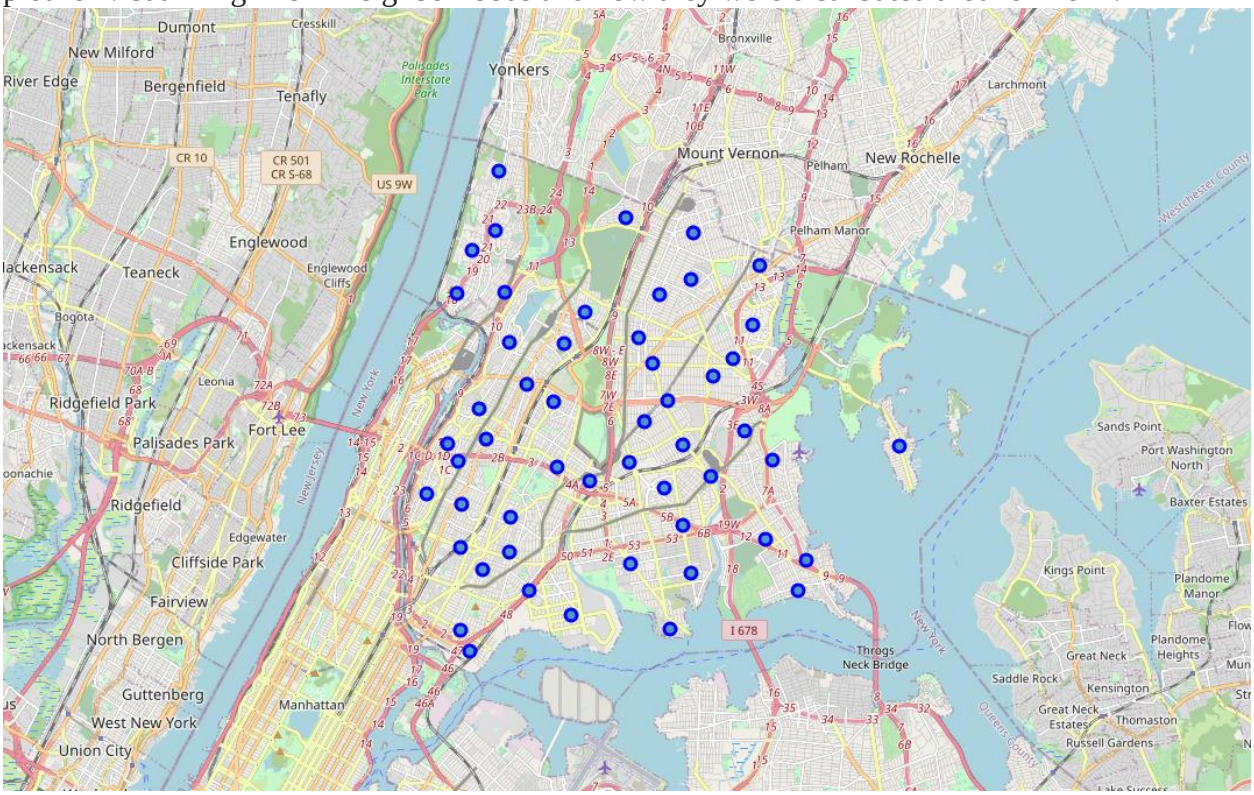
With data download into Jason file, we select four features for creating our first data frame

```
'Borough', 'Neighborhood', 'Latitude', 'Longitude'
```

Therefore, we proceed to fill with looping methods from neighborhood data, but this one has information from New York complete, and our analysis was focused on Bronx borough. We took this data frame and slice to this and obtain its location data.

| | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|
| 0 | Bronx | Wakefield | 40.894705 | -73.847201 |
| 1 | Bronx | Co-op City | 40.874294 | -73.829939 |
| 2 | Bronx | Eastchester | 40.887556 | -73.827806 |
| 3 | Bronx | Fieldston | 40.895437 | -73.905643 |
| 4 | Bronx | Riverdale | 40.890834 | -73.912585 |

Using Google Maps geocoding API we have geographical coordinate of Bronx, we made first plot for visualizing Bronx neighborhoods and how they were distributed around Bronx.

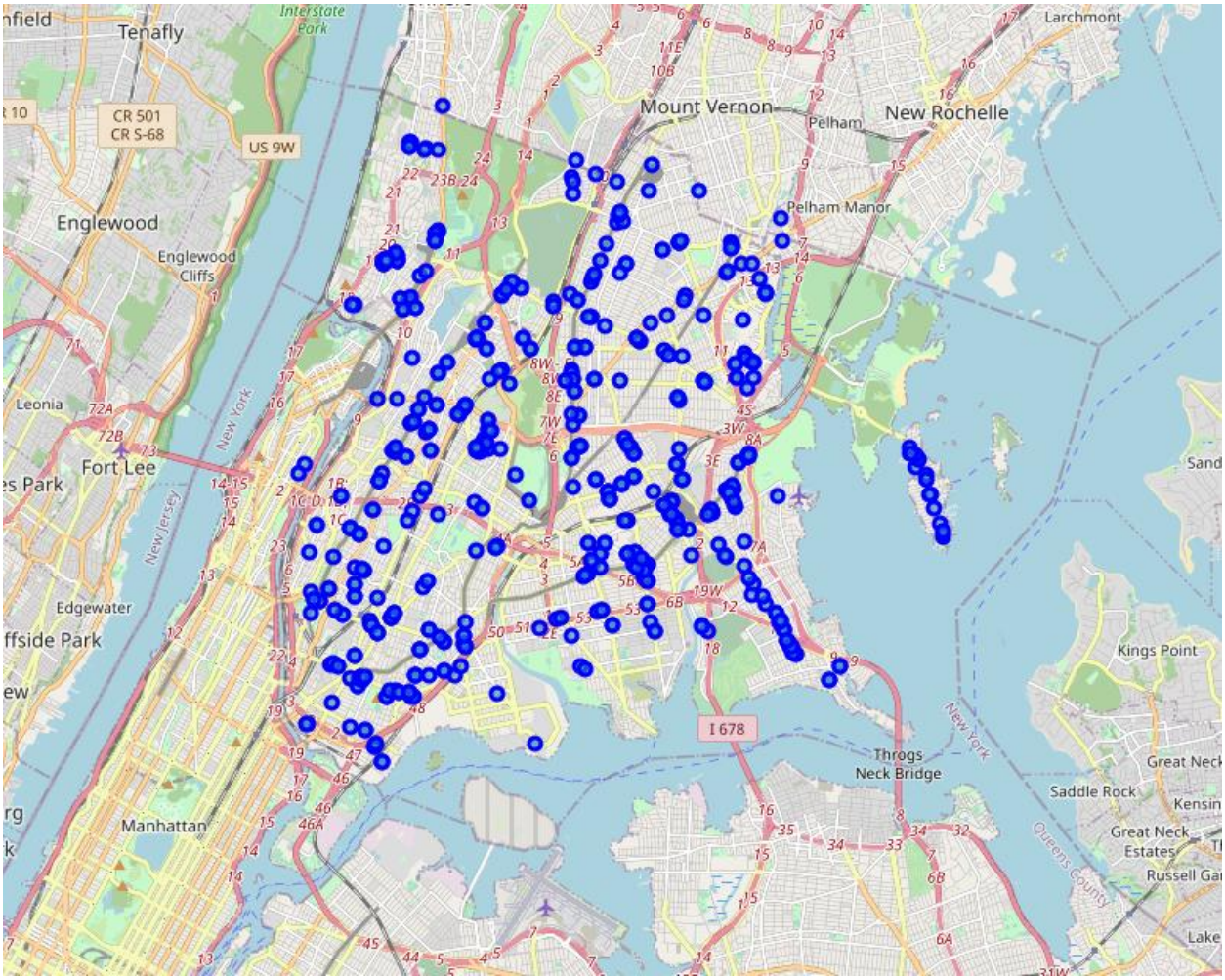



We this section complete, we start scraping venues and categories data using Four square API, and obtaining seven features (Neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venue, Venue latitude, Venue Longitude, Venue Categories. Picture below shows a sample.

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Wakefield | 40.894705 | -73.847201 | Lollipops Gelato | 40.894123 | -73.845892 | Dessert Shop |
| 1 | Wakefield | 40.894705 | -73.847201 | Ripe Kitchen & Bar | 40.898152 | -73.838875 | Caribbean Restaurant |
| 2 | Wakefield | 40.894705 | -73.847201 | Ali's Roti Shop | 40.894036 | -73.856935 | Caribbean Restaurant |
| 3 | Wakefield | 40.894705 | -73.847201 | Jackie's West Indian Bakery | 40.889283 | -73.843310 | Caribbean Restaurant |
| 4 | Wakefield | 40.894705 | -73.847201 | Carvel Ice Cream | 40.890487 | -73.848568 | Ice Cream Shop |

We start slicing data from venues categories, for having ones corresponding to restaurant category. After this exercise we obtained 33 restaurant categories in order to analyze neighborhood similarity based on restaurant density.

We plot restaurant around Bronx and tried figure out the lowest restaurant zone, to go deeper into those neighborhoods.




## 3. Methodology

In this project we tried detecting areas of Bronx that have low restaurant density, particularly those with low number of **Latin and Caribbean Restaurant**.

In first step we have collected the required data: **location and type (category)** of every restaurant within Bronx borough. We have also identified **venues restaurants** (according to Foursquare categorization).

Second step in our analysis will be **segmenting areas based on restaurant density across different areas of Bronx** - we will use K-Means clustering Machine learning methods to identify borough cluster and focus our attention in those cluster with lowest Restaurant density.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with **lowest number of 1rst common venue of Caribbean restaurant**. We will present map of all such locations but also create clusters (using k-means

clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for later deeper analysis.

## 4. Analysis

### 4.1. Clustering neighborhood

We wanted to cluster neighborhood based on restaurant density, for this we use one hot coding for creating dummies variables with each restaurant categories 34 in total.

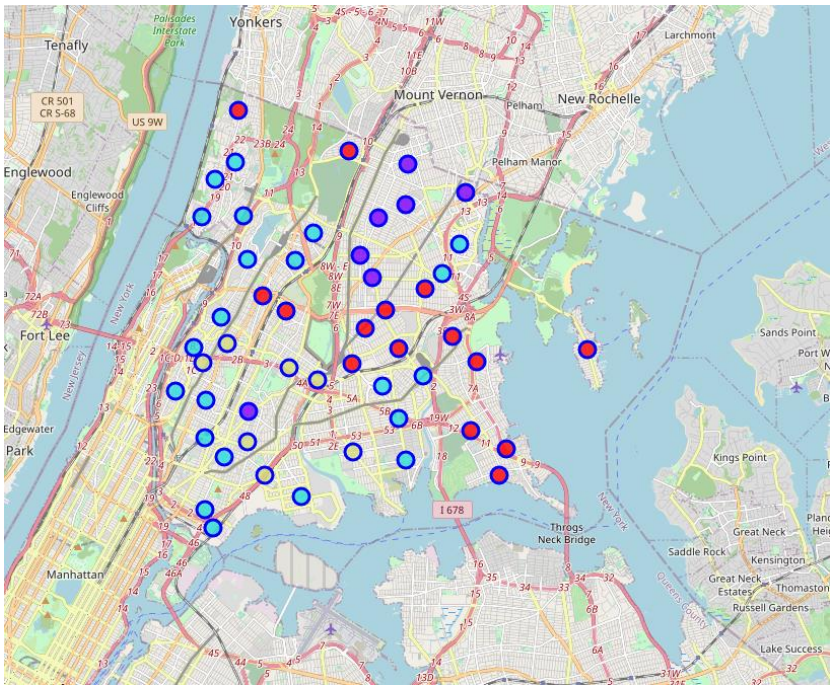
Now every neighborhood it's marked with every restaurant category, then we grouped rows by neighborhood and by taking the mean of the frequency of occurrence of each category.

But we want to analyze just top 10 most common venues for neighborhood, and we made a list of them and created a data frame to display top 10 venues for every neighborhood

### 4.2. Modeling

One the most used Machine Learning methods for clustering it's K-Means clustering for the ease of implementation and accuracy on this type of analysis.

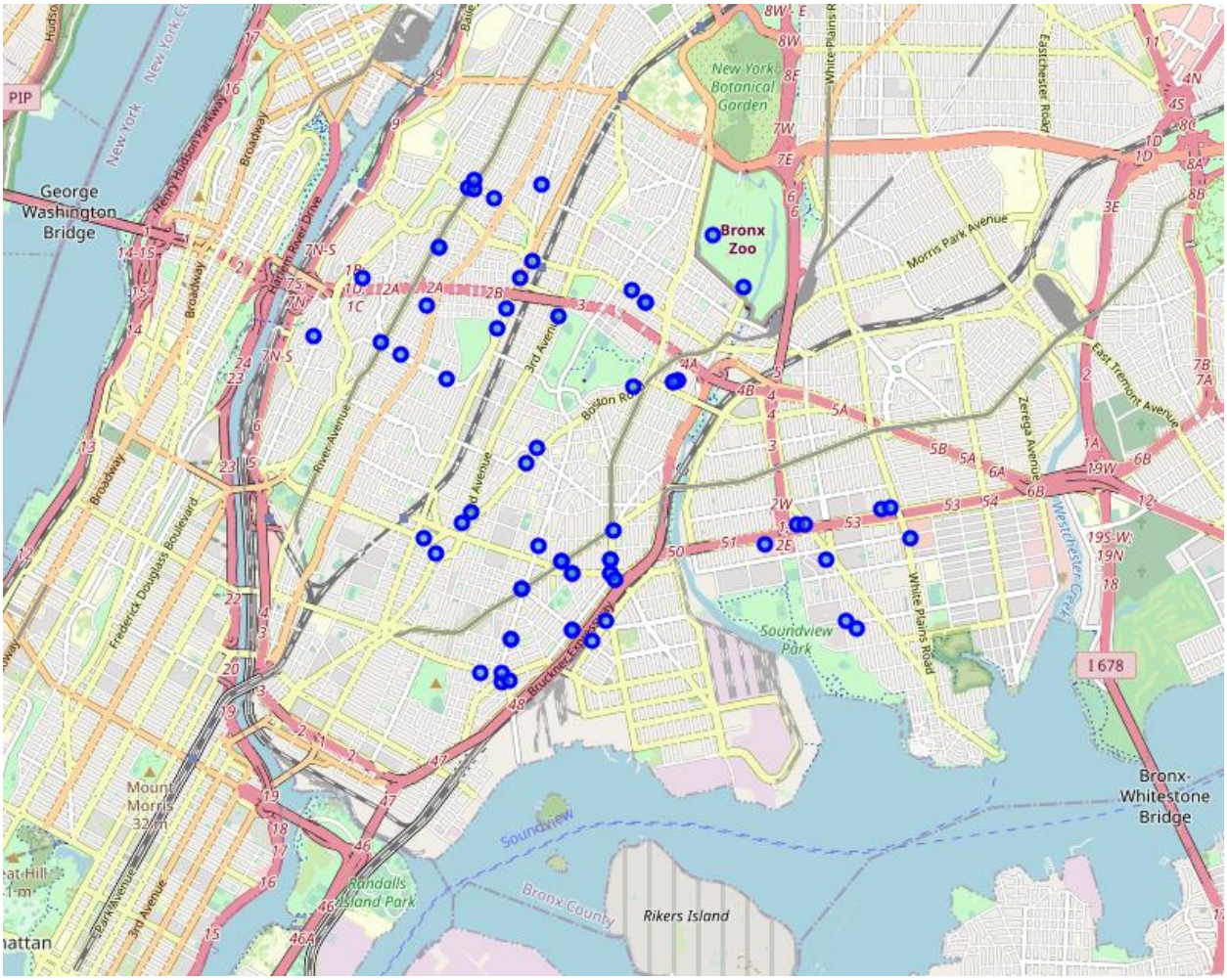
We ran k-means to cluster neighborhood into four cluster. After running model every neighborhood was marked with cluster number from 0 to 3, but we had to merged with previous data frame containing location data with this created a new data frame with one more feature (Cluster labels) , then we plot neighborhoods with cluster color coded and tried figure out cluster distribution around Bronx center.



We can figure out Cluster three has the largest Restaurant count, that represent our competitor. those neighborhoods are not the best site selection site just taking into consideration venues similarity.

**Cluster four** seems to be the best Neighborhood locations to be considered to open a new restaurant.

Sliced data frame from cluster number four and merged with Bronx venue restaurant data frame, later we plot to visualize using Folium package.




## 5. Results and discussion

Our analysis shows that although there is a great number of restaurants in Bronx 596 after slicing first data location, we find low restaurant density in the north region. Highest concentration of restaurants was detected south and and west from center, so we focused our attention to areas south, corresponding to boroughs **East Tremon, Mount Hope, Mount Eden, Morrisania, Longwood**. Other boroughs identified as potentially interesting was Soundview.

We first created Neighborhood clustering and segmenting them based on restaurants categories for exploring deeper our zone of interest. We divided in four groups of 10 most common type of

restaurant for making those with most Latin and Caribbean restaurants the zone of our interest. Result of all this is four cluster being cluster number one containing largest number of 1srt Most common venue of Caribbean restaurant, and cluster number four having the lowest 1srt Most common venue of Caribbean restaurant.

This cluster has the largest number of potential new restaurant locations based on number of 1srt Most common venue of Caribbean restaurant. Recommended zones should only be considered as starting point for more detailed analysis, like Latin population density. And what competitors offers.

### 5.1. Conclusion

Purpose of this project was to identify **Bronx borough low number restaurants** particularly Caribbean and Latin restaurants in order to help stakeholders in narrowing down the search for optimal location for a new Latin restaurant. By calculating restaurant density distribution from Foursquare data, we have first identified general boroughs that justify further analysis (East Tremon, Mount Hope, Mount Eden, Morrisania, Longwood), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.