# Finding the Best Place to Open a Pizza Shop

## Joshua Sebastian
April 28, 2020

## 1. Introduction

### 1.1 Background

Mario is moving to Queens, NY, from Italy. He grew up in and currently runs a top pizza shop in Naples. When deciding to move to the US he chose New York City largely because of its reputation for having the best pizza in the US. Being a family man, Mario wanted to move to the Borough best suited to raise a family and while that is subjective, he believes it to be Queens.

### 1.2 Problem

While Mario knows he is moving to Queens, he is not exactly sure which neighborhood in Queens he should set up his pizza shop. Queens is very large, densely populated area with many neighborhoods and since Mario has never been there, he needs help determining which neighborhood is the best fit for him. He wants to compete with the best and in area with healthy competition as he believes the cream always rises to the top.

### 1.3 Interest

Obviously, Mario has a vested interest in this info but any entrepreneur looking to open a pizza shop in Queens would find this report of interest and value.

## 2. Data acquisition and cleaning

### 2.1 Data Sources

There first data set required for the analysis is one containing info on the 81 neighborhoods located in Queens including their longitude and latitude coordinates. This dataset can be found at https://geo.nyu.edu/catalog/nyu_2451_34572.

The second data source required for this analysis is one that can provide up-to-date venue information for each of the 81 neighborhoods. For this project, the Foursquare.com API was used.

### 2.2 Data Cleaning

In the case of the neighborhood data. It was in json format and was pretty clean and complete to begin with. It did include data of all the neighborhoods for all 5 boroughs of New York City.
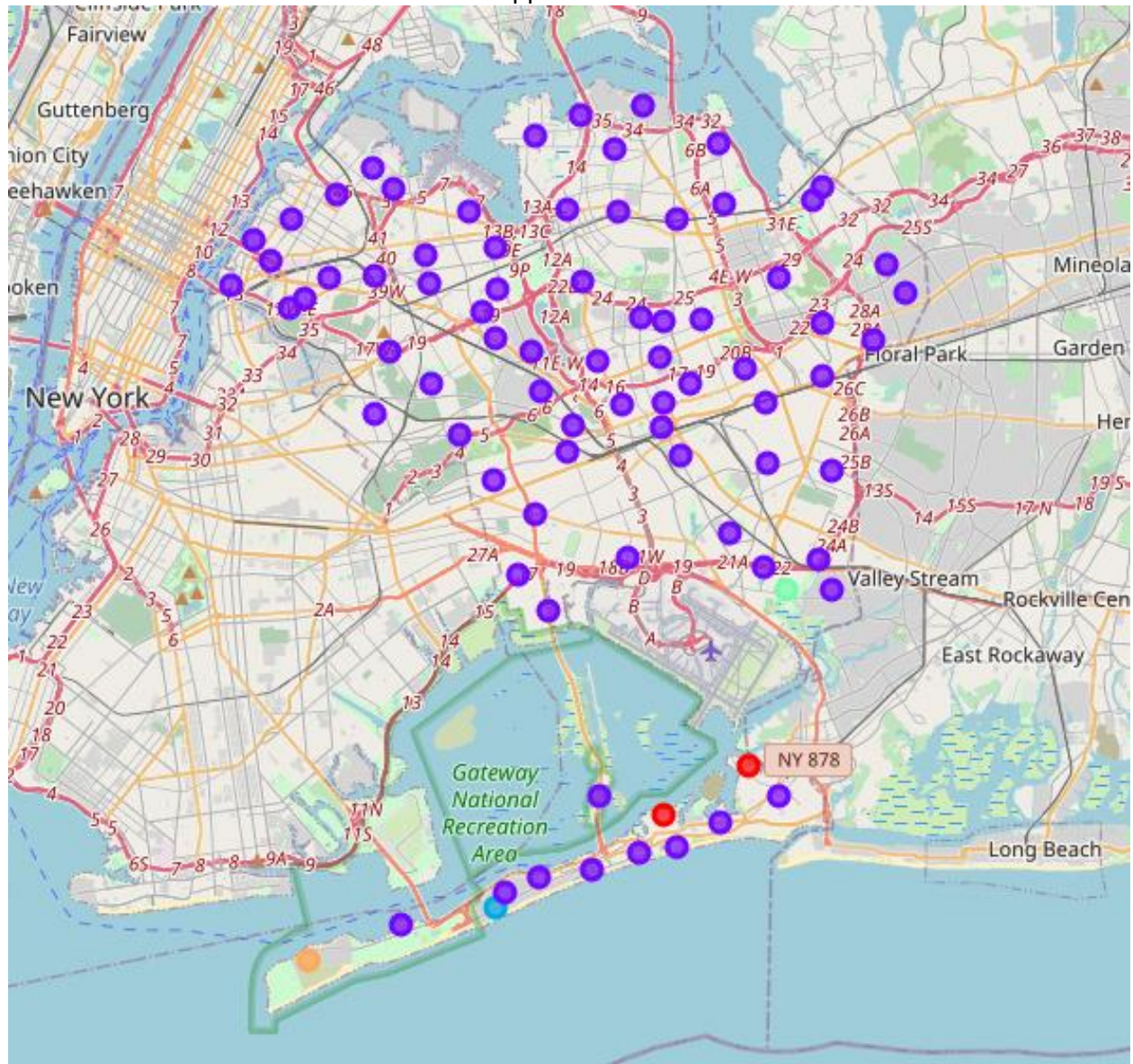
After mapping all the neighborhoods for reference, I filtered the data to only include the 81 neighborhoods of Queens.
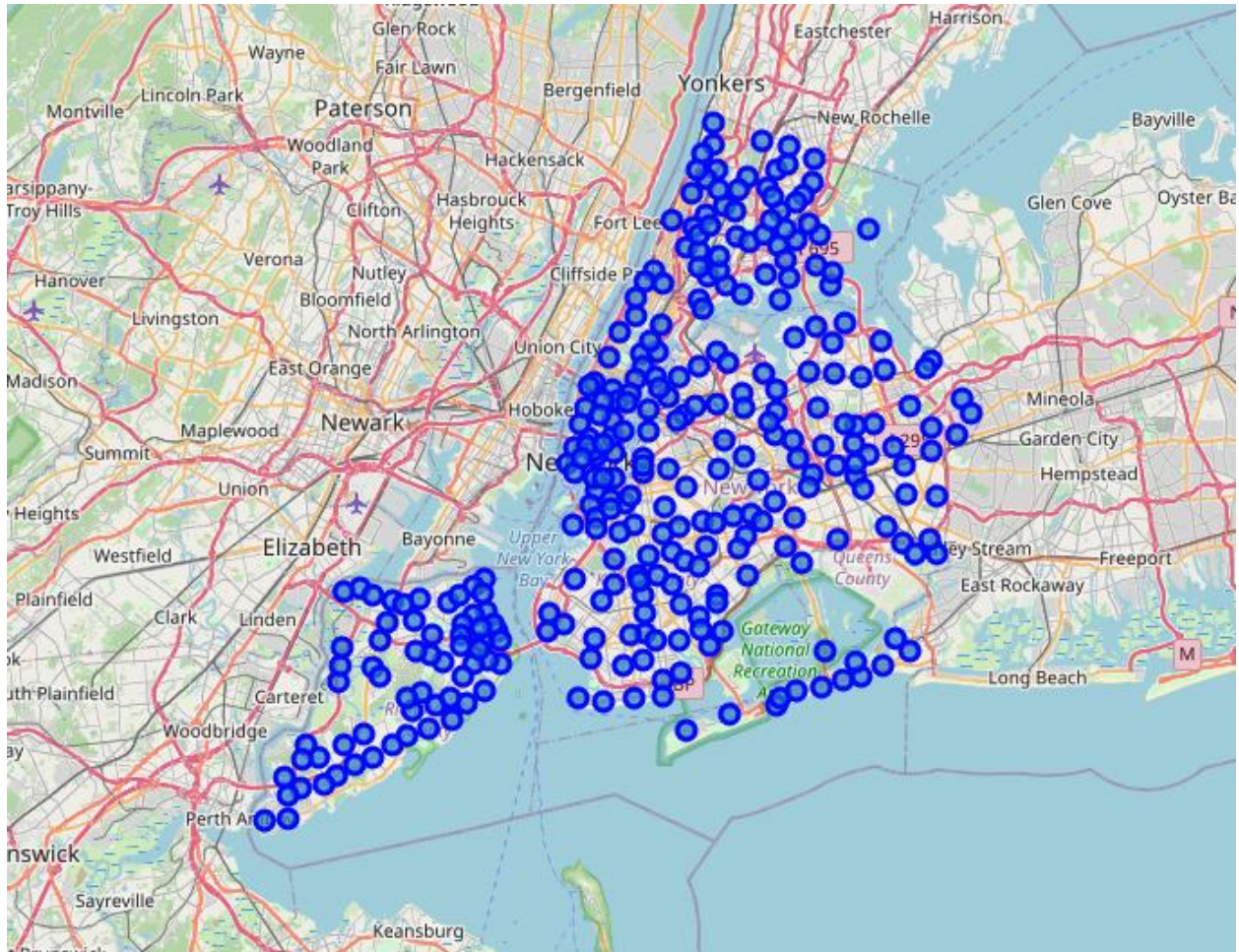


The Foursquare data was very clean as well and had to be merged with the location data so that it could be analyzed and mapped.

## 2.3 Feature Selection

In order to segment and cluster the neighborhoods, the Foursquare API was used to download the venues and their types for each neighborhood. There was a total of 270 unique categories for the venues used in this analysis. The means of the frequencies were calculated for each category by neighborhood.

# 3. Methodology

## 3.1 Exploratory Data Analysis

In order to segment the different neighborhoods, k-means clustering was used with a k value of 5. Once the model was fit with the means of the frequencies of each venue category for each neighborhood the neighborhoods were segment into the following clusters.

1. Parks
2. Restaurants
3. Beaches
4. Delis/Bodegas
5. Monuments/Landmarks

The clusters were then color coded and mapped for visual reference.

## 3.2 Finding the Best Fit

Mario had stressed that he was the best pizza maker in Naples and soon would be the best in Queens and as such he wanted to be driven by competition. The neighborhoods that fell into the restaurant cluster were then analyzed to determine which had the most pizza places.

| Neighborhood | Pizza Place |
| --- | --- |
| Sunnyside | 5 |
| Sunnyside Gardens | 4 |
| Long Island City | 4 |
| Bellerose | 3 |
| Bayside | 3 |
| Ridgewood | 3 |
| Far Rockaway | 3 |
| Ozone Park | 3 |
| North Corona | 3 |
| Maspeth | 3 |

# 4. Results

As result of this analysis it was determined that Sunnyside, Sunnyside Gardens, and Long Island City had a high prevalence of restaurants in their neighborhoods and in particular of pizza places. There were seven other neighborhoods that were close behind.

# 5. Discussion

Once the top 10 neighborhoods for pizza places were identified, they were mapped and color coded based on the number of places in the neighborhood.
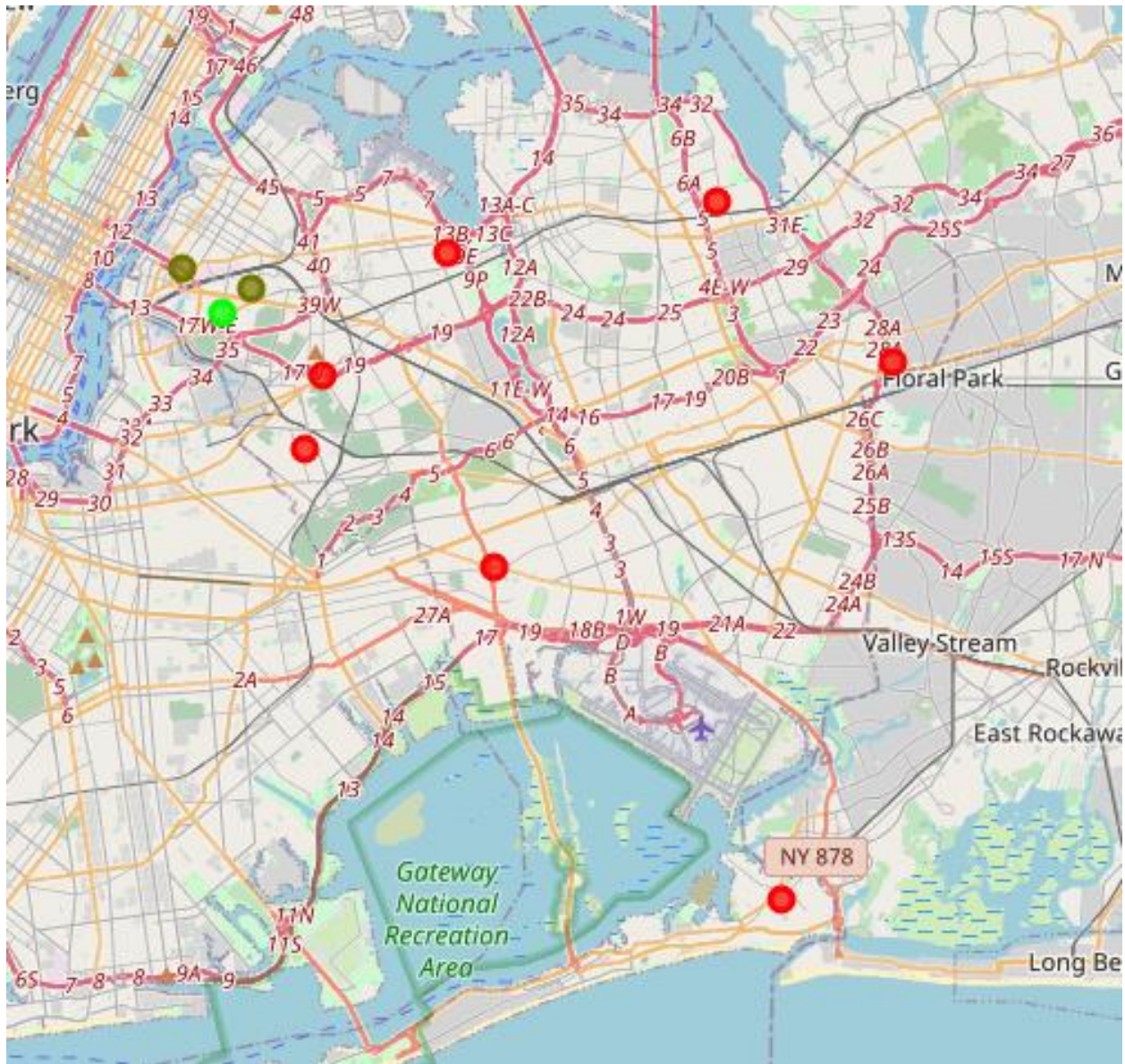
The top neighborhood, Sunnyside with five pizza shops was coded bright green, while the runner ups, Sunnyside Gardens and Long Island City with four pizza shops each were code dark green. The rest of the top 10 were coded red and each contained three pizza shops.

## 6. Conclusions

Based on Mario's criteria it looks like Sunnyside would be the best fit for him to embark on his pizza domination of New York City. As you can see in the final map, Sunnyside Gardens and Long Island City are both very close to Sunnyside and would also be suitable, giving Mario a bit more flexibility in the area room to select his final location based on personal tastes of the area.