

# **IBM Data Science Capstone Project: Battle of the Neighborhoods**

## **Family-Owned Pizza Company Goes International**

### **Introduction/ Business Problem**

Our Client is one of the most popular family-owned pizza chains located in Italy. They want to bring their specialty pizza to the United States. They believe that their unique recipe and family techniques can separate them from other pizza restaurants to give patrons pizza they never tasted before. They want to examine various cities with high-density pizza venues to set up their first shop internationally.

### **Data Collection**

The most populated cities in the United States are New York, NY, San Francisco, CA, Jersey City, NJ, Boston, MA and Chicago, IL. I used the FourSquare API to collect data about the pizza venue densities in these major US cities to suggest a location to setup their first shop.

### **Methodology**

The main objective is to locate the city with the highest density of pizza venues for the client to set up shop. I used the Four Square API through the venues channel for the respective states mentioned above. I then used the near query to get venues in the cities, used the CategoryID to show only pizza places, and then narrowed it to the top 100 venues for each city being studied.

An example of my request code for one of the city is as follows:

[https://api.foursquare.com/v2/venues/explore?&client\\_id=&client\\_secret=&v=20180605&New York, NY&limit=100&categoryId=4bf58dd8d48988d1ca941735](https://api.foursquare.com/v2/venues/explore?&client_id=&client_secret=&v=20180605&New York, NY&limit=100&categoryId=4bf58dd8d48988d1ca941735)

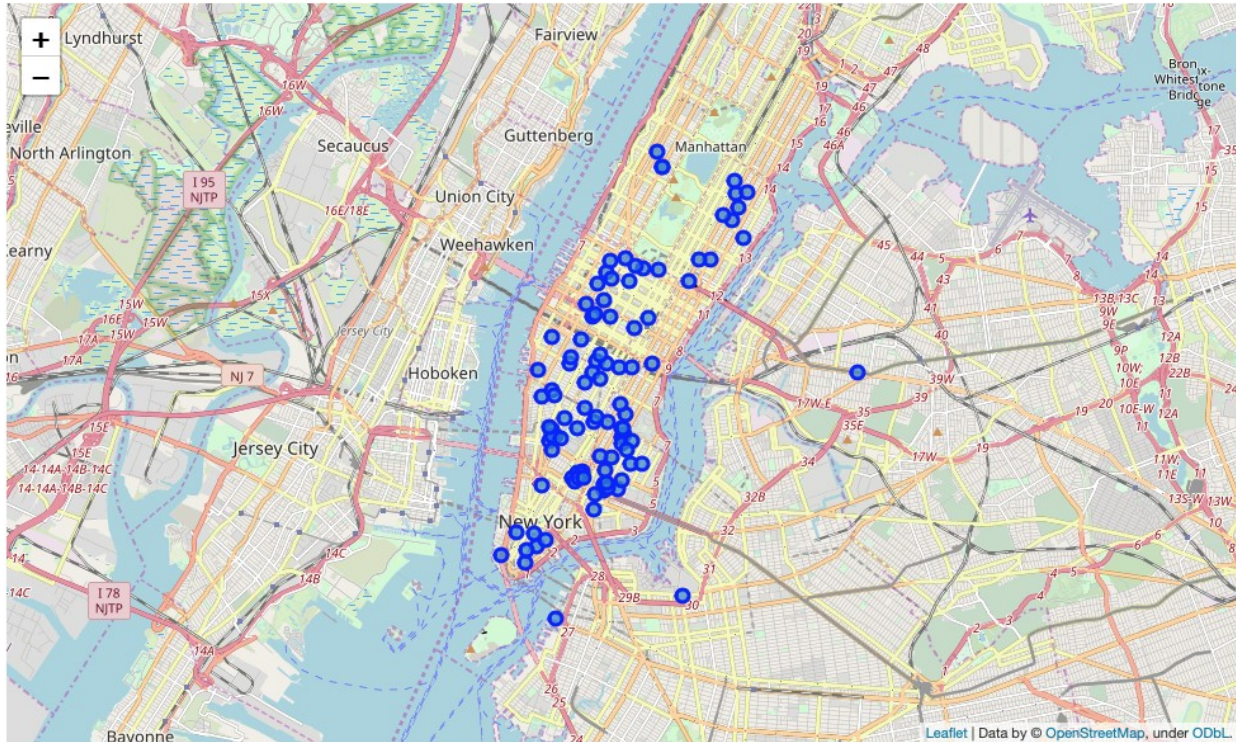
The 4bf58dd8d48988d1ca941735 is the ID of the category for pizza venues. Moreover, I repeated this request for each of the studied cities to save the name and coordinate data. From each result, I plotted them on the map for visual inspection.

Lastly, to get an indicator of the accuracy for the density of pizza venues, I had to calculate a center coordinate of the venues to get the mean longitude and latitude values. Afterwards, I calculated the mean of the Euclidean distance from each venue to the mean coordinates. This will be my indicator (mean distance to the mean coordinate).

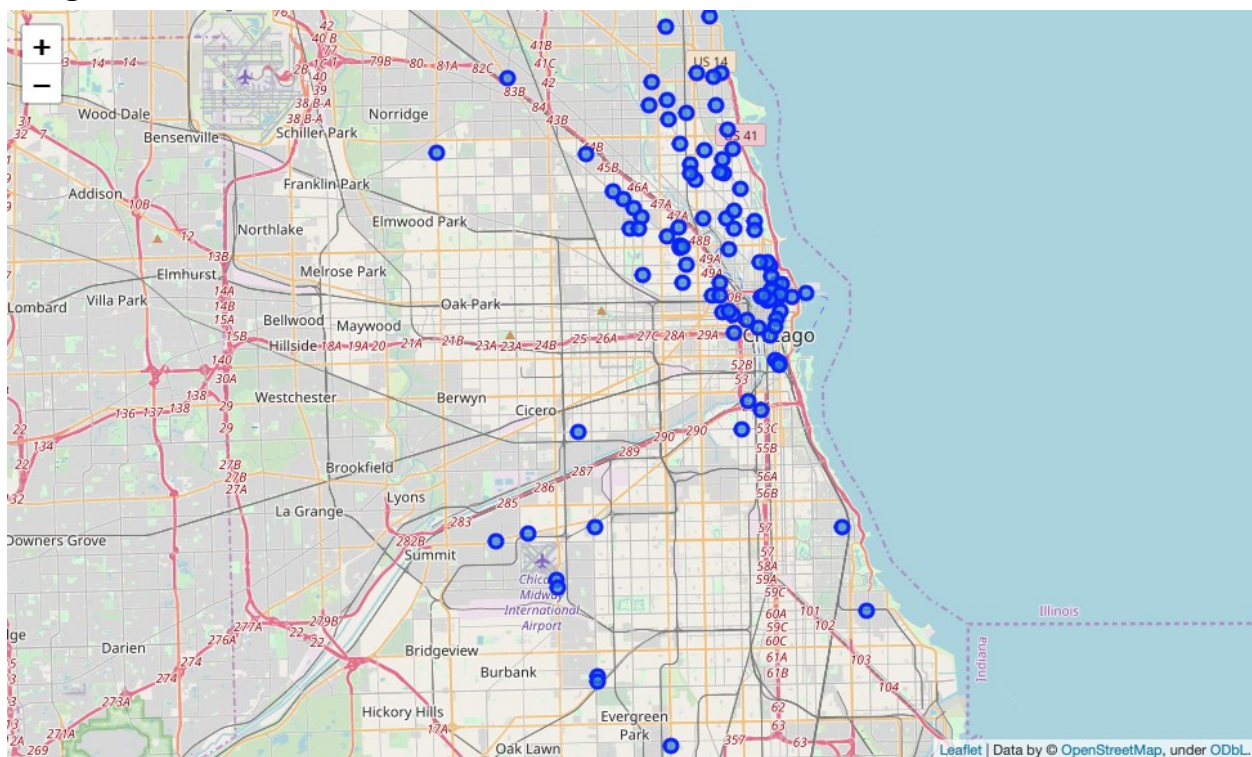
## Results

The following below are the initial clusters I created based of the Foursquare code for pizza venues of New York, NY, San Francisco, CA, Jersey City, NJ, Boston, MA and Chicago, IL:

### New York, NY

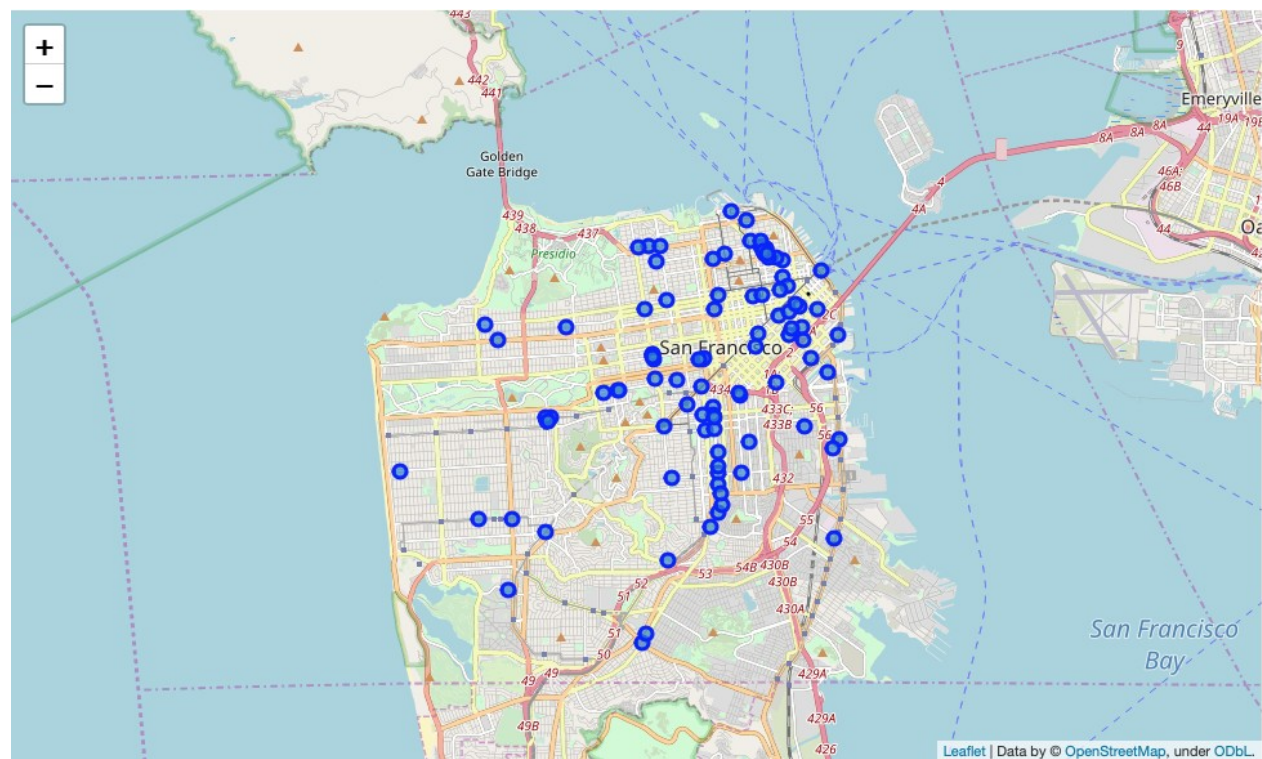


### Chicago, IL

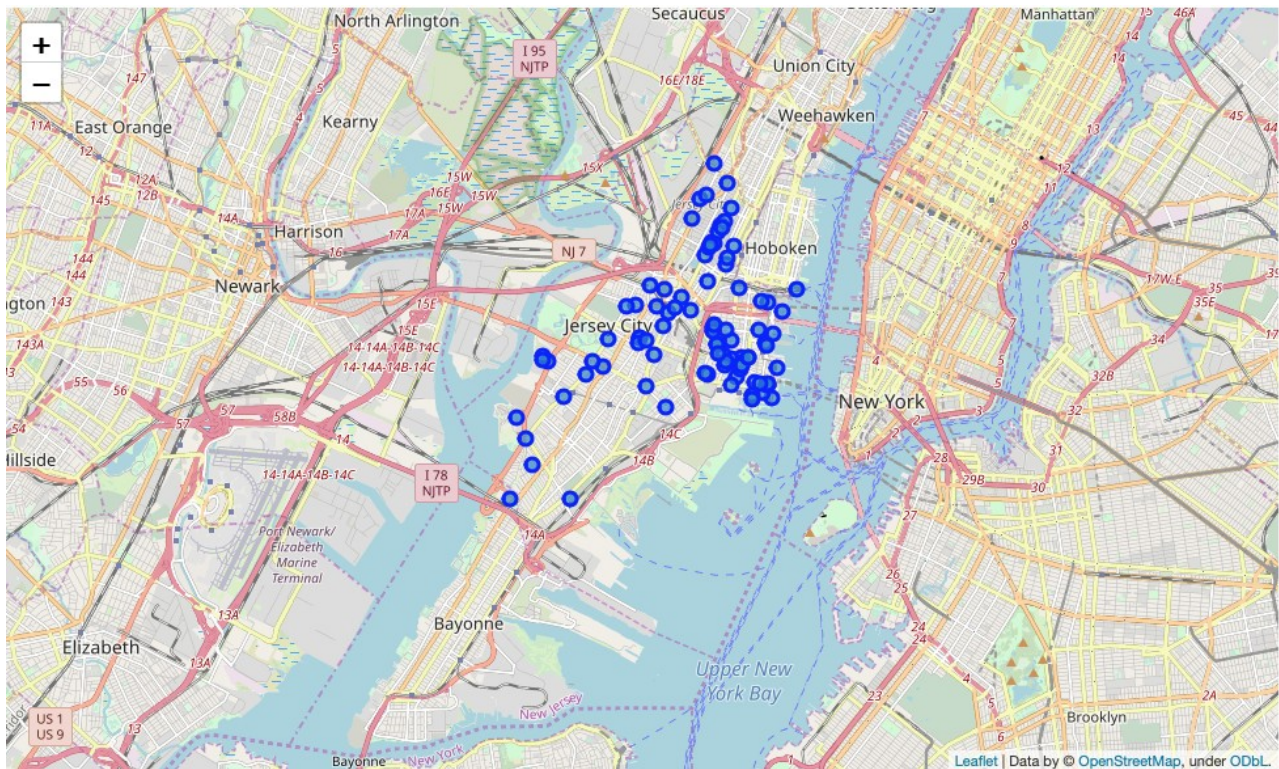




San Francisco, CA

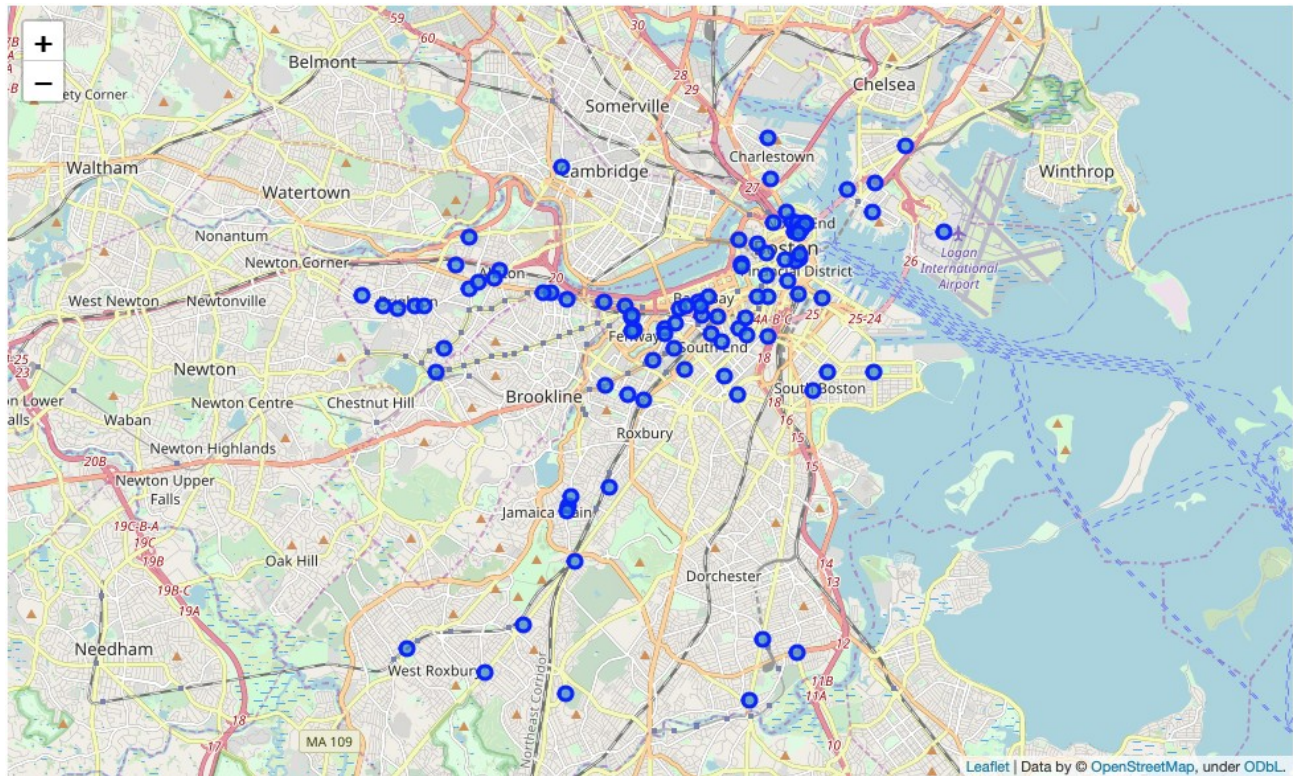


Jersey City, NJ





## Boston, MA



Visually, we observed that New York, NY and Chicago, IL seem to have the highest density of pizza venues for target locations to set up shop. However, we employed some statistics to get the mean location of the pizza venues to ascertain the validity of their actual density. Afterwards, we will take the average of the distances of venues to the mean coordinates to get a more concrete density. The results of this process can be seen below:

## New York, NY

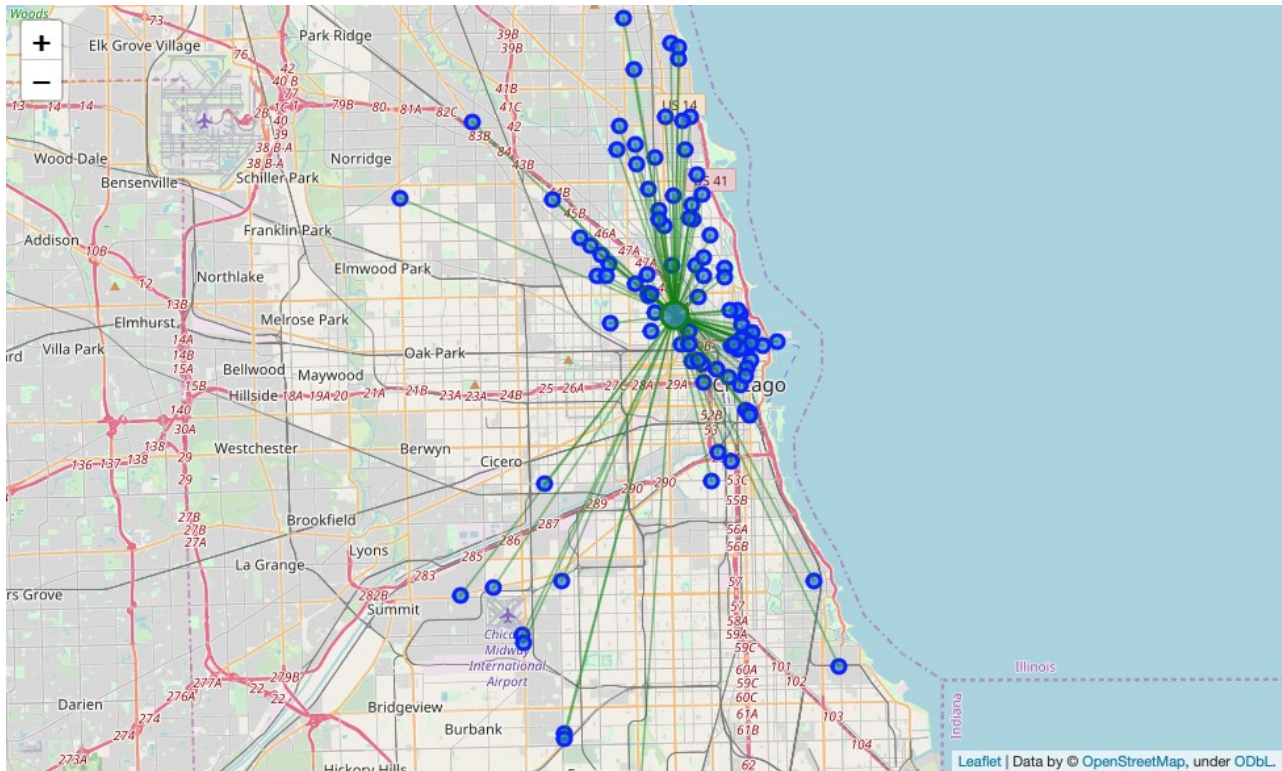
**Mean Distance from Mean Coordinates: 0.022656966339103884**





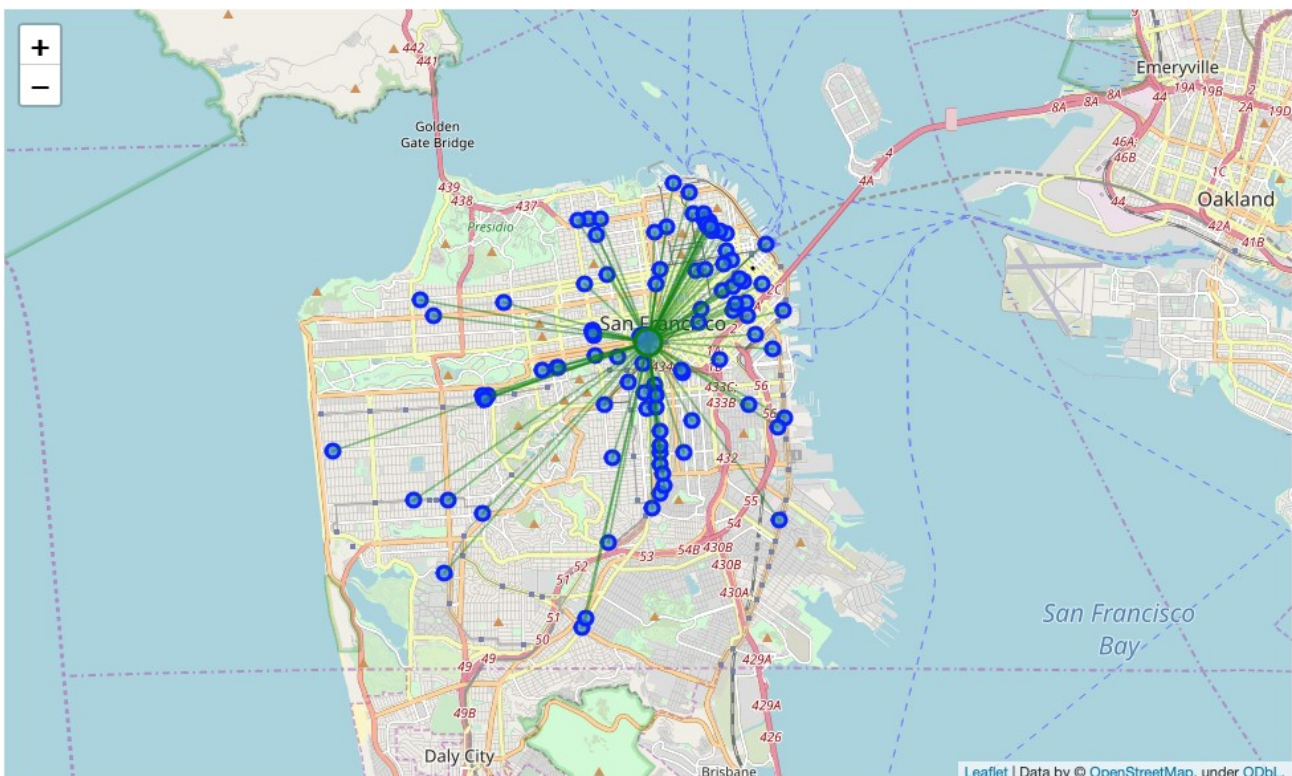
## Chicago, IL

Mean Distance from Mean Coordinates: 0.05723120018998896



## San Francisco, CA

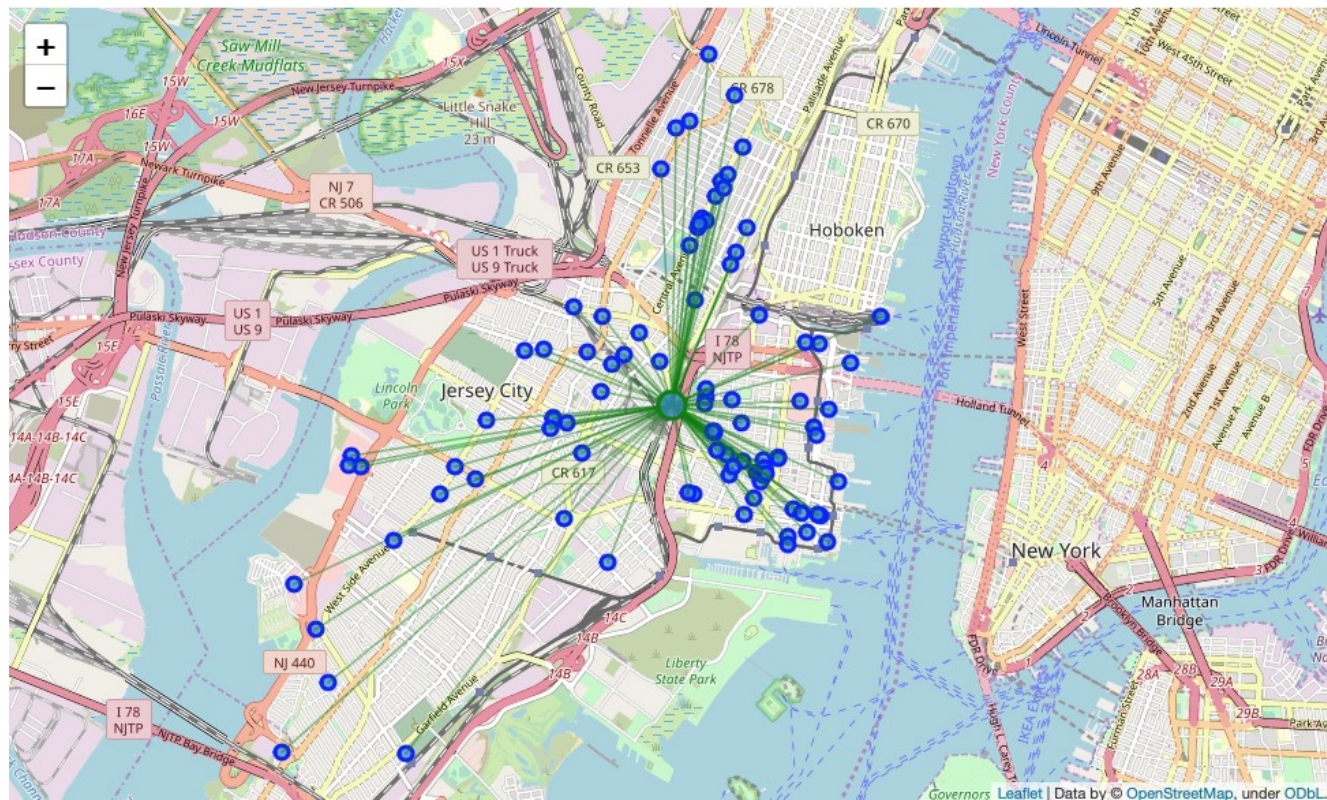
Mean Distance from Mean Coordinates: 0.02782241910063802





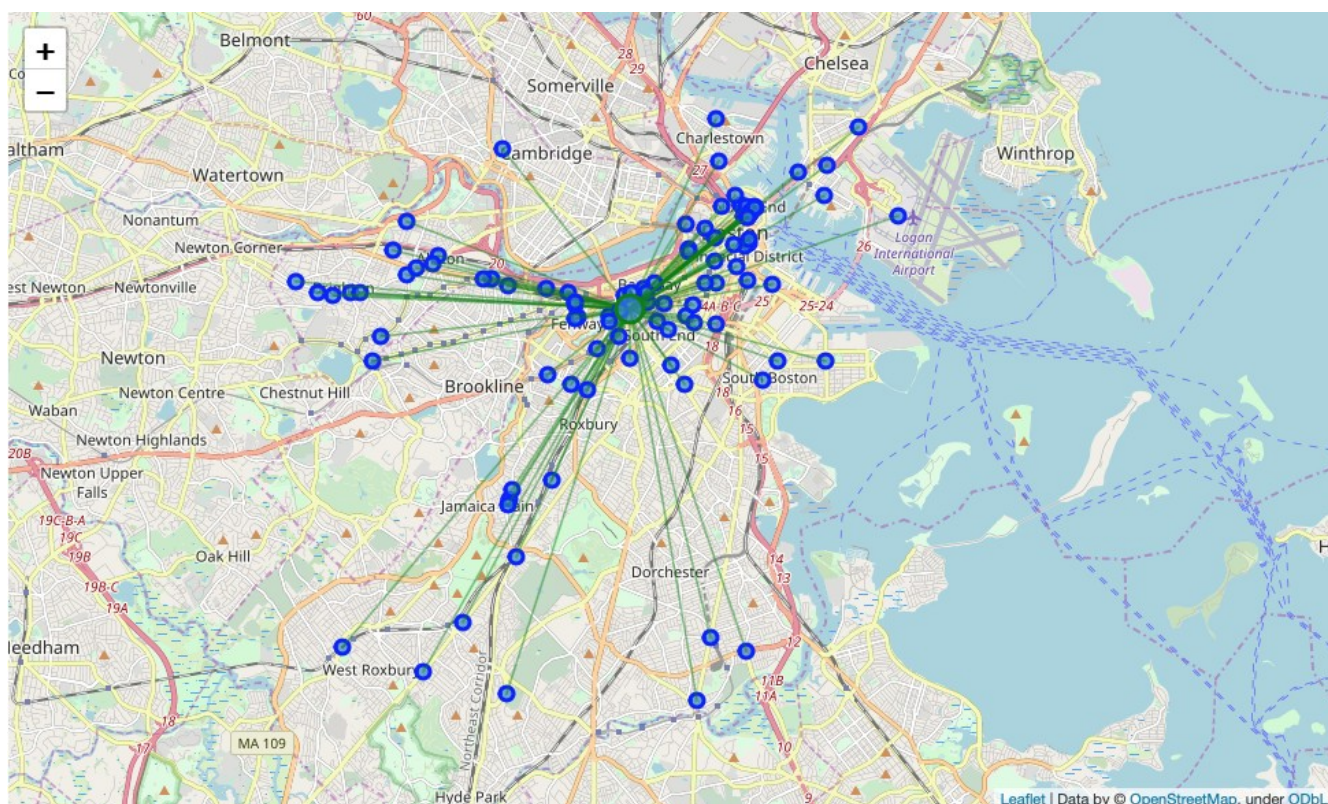
**Jersey City, NJ**  
**Mean Distance from Mean Coordinates: 0.018821184781042995**

**Mean Distance from Mean Coordinates: 0.018821184781042995**



**Boston, MA**  
**Mean Distance from Mean Coordinates: 0.03534053436469107**

**Mean Distance from Mean Coordinates: 0.03534053436469107**



## **Discussion**

Our initial observations suggested that either New York, NJ or Chicago, IL were going to be the optimal cities for our client to set up shop. However, after determining the mean coordinates of each city cluster with the Euclidean distance, our assumptions changed to indicate a denser cluster within Jersey City, NJ.

One thing to note is our data was limited to Foursquare only providing us with the top 100 venues for each city for analysis. Thus, our results may have differed if additional venues were provided.

## **Conclusion**

Interesting enough, although New York, NY and Chicago, IL appeared to be denser visually on the first clustering, the "Mean Distance from Mean Coordinates" performed on the second clusters show that Jersey City, NJ would be denser.

Based on these findings, I would suggest the client to set up their first International Italian Pizza Shop in Jersey City, NJ.