



# Capstone Project - The battle of neighborhoods

## Business problem section

### Background and business problem

Pretend that you will travel to the US and you want to have a really good pizza while visiting there - just like the ones at home in Italy. The problem I aiming to solve is to analyse where we will find the best pizza in the US. However, let's say that you also would add some preferences with this pizza such as xx which will be covered in this analysis.

## Data section

The data that will be used for the analysis will be gathered via Foursquare API in order to collect information about the locations in the US. As a tourist, the highest probability is that you will visit the biggest cities in the US, and therefore the data will be limit to the top 5 biggest.

## Methodology

The main objective with the analysis is to identify the best pizza place, and therefore an API via Foursquare through the venues channel will be used to; firstly, get the highest density of pizza locations by using the near query to get the venus in the cities. Secondly, I will use the CategoryID to set it to show only pizza places since my search is limited to that specific target. This will furhter be run five times since the study will cover top 5 cities in the US.

Furthermore, to get an indicator of the density of pizza places, I have calculated a center coordinate of the venues to get the mean longitude and latitude values. After, I will calculate the mean of the Euclidean distans from each venue to collect the mean coordinates.

## Analysis

The first step was to identify the highest density of pizza locations in the cities: New York, Chicago, Boston, San Francisco and New Jersey. **We could see that;**

New York have **283** pizza places

Chicago have **217** places

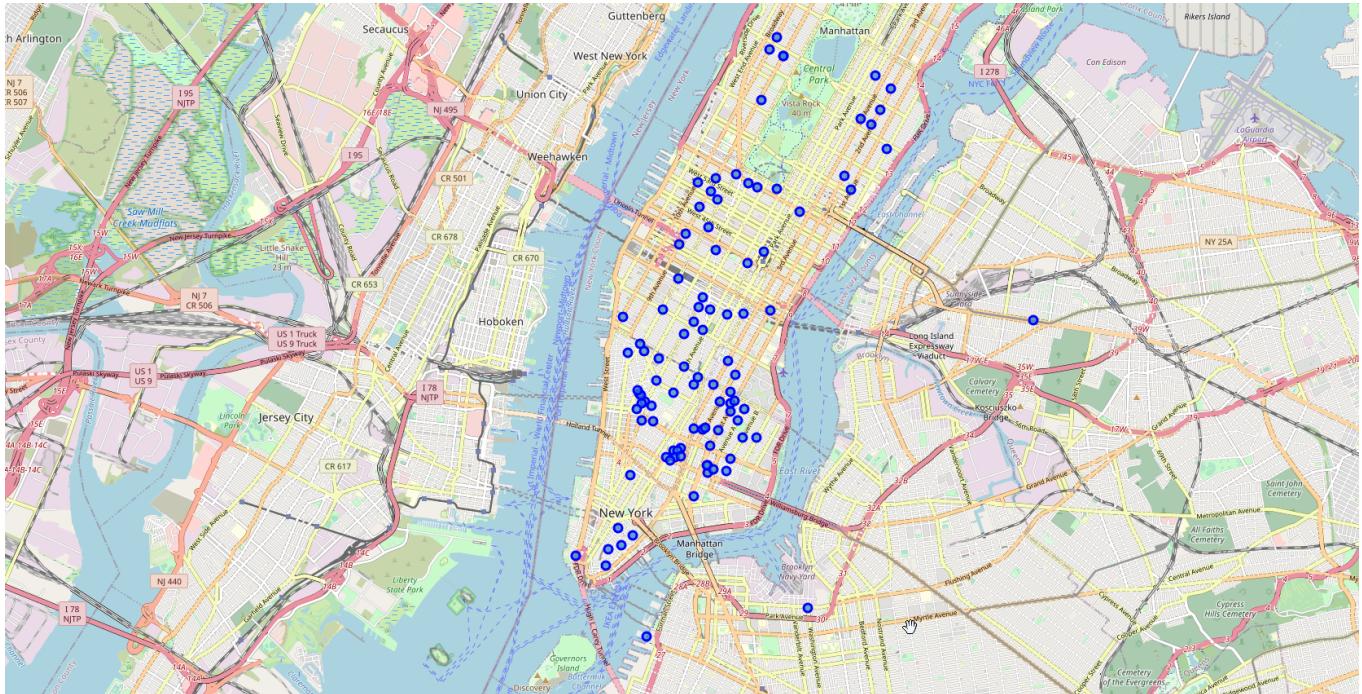
San Francisco have **169** places

Jersey City have **126** places

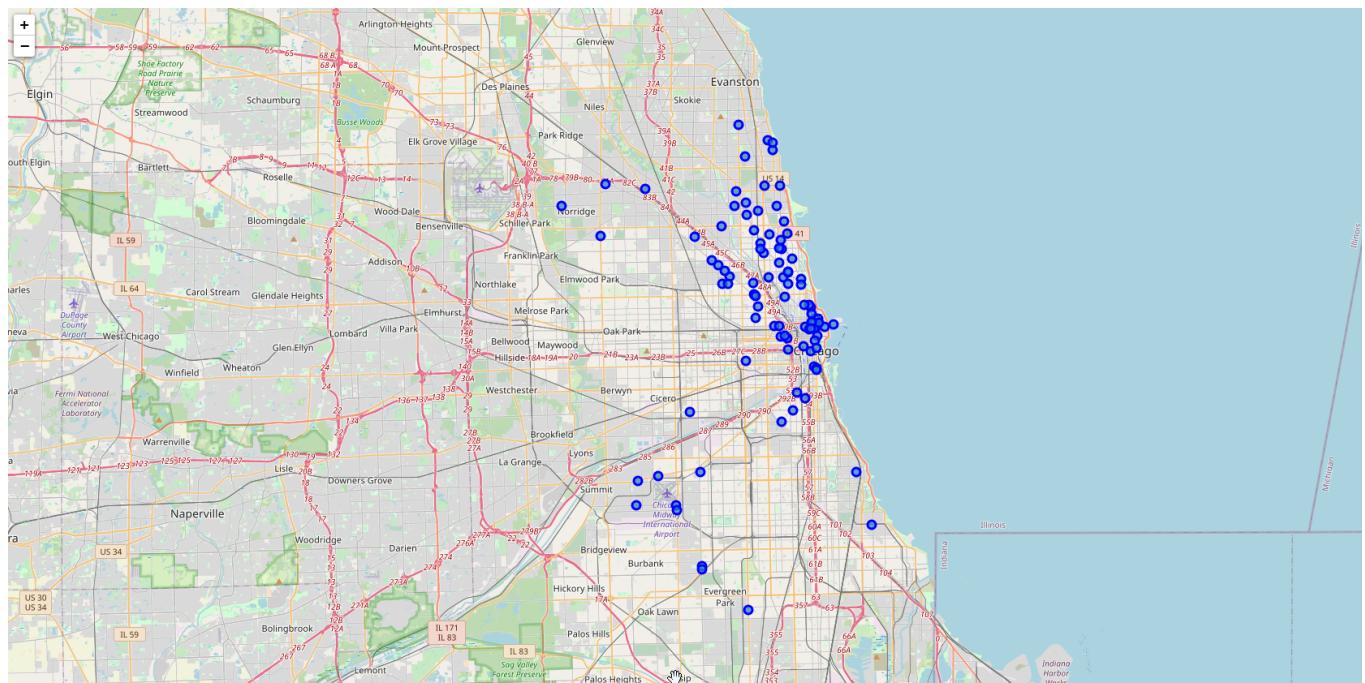
Boston have **184** places

Also, I did a check by visualising the actual places and here are the results:

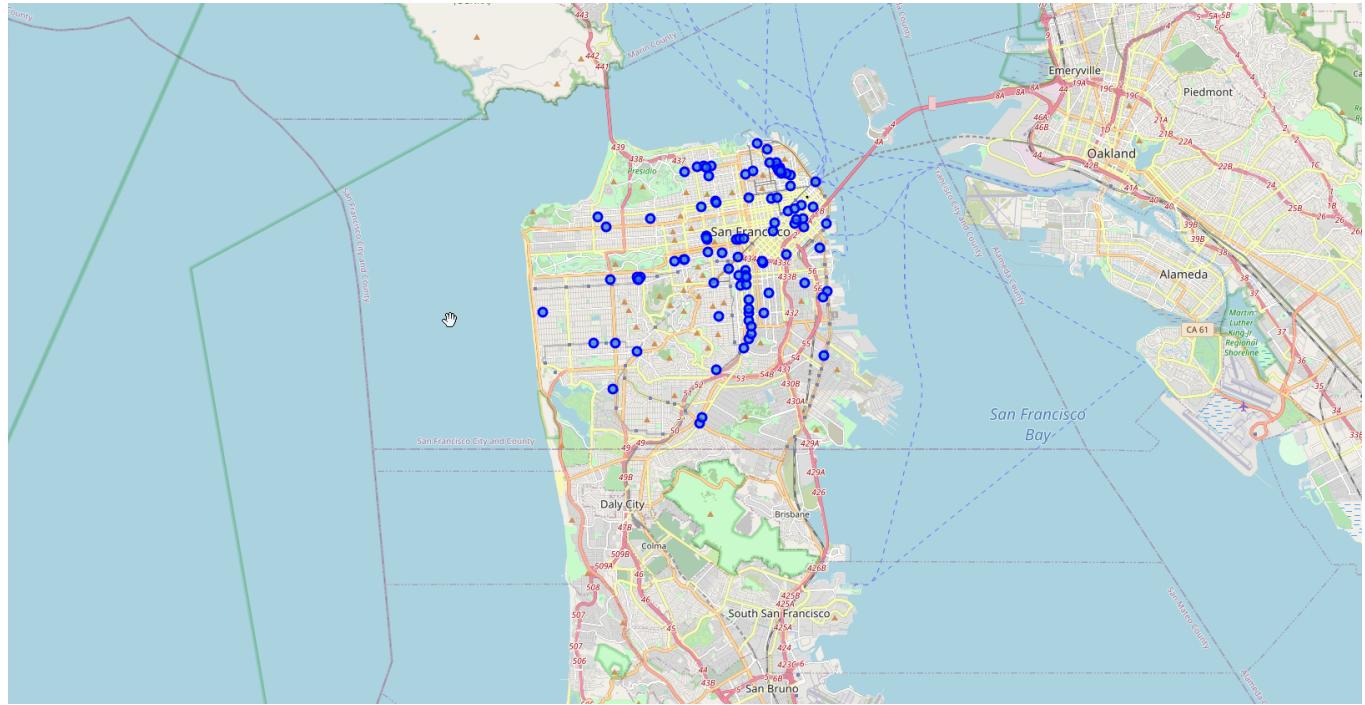
## New York



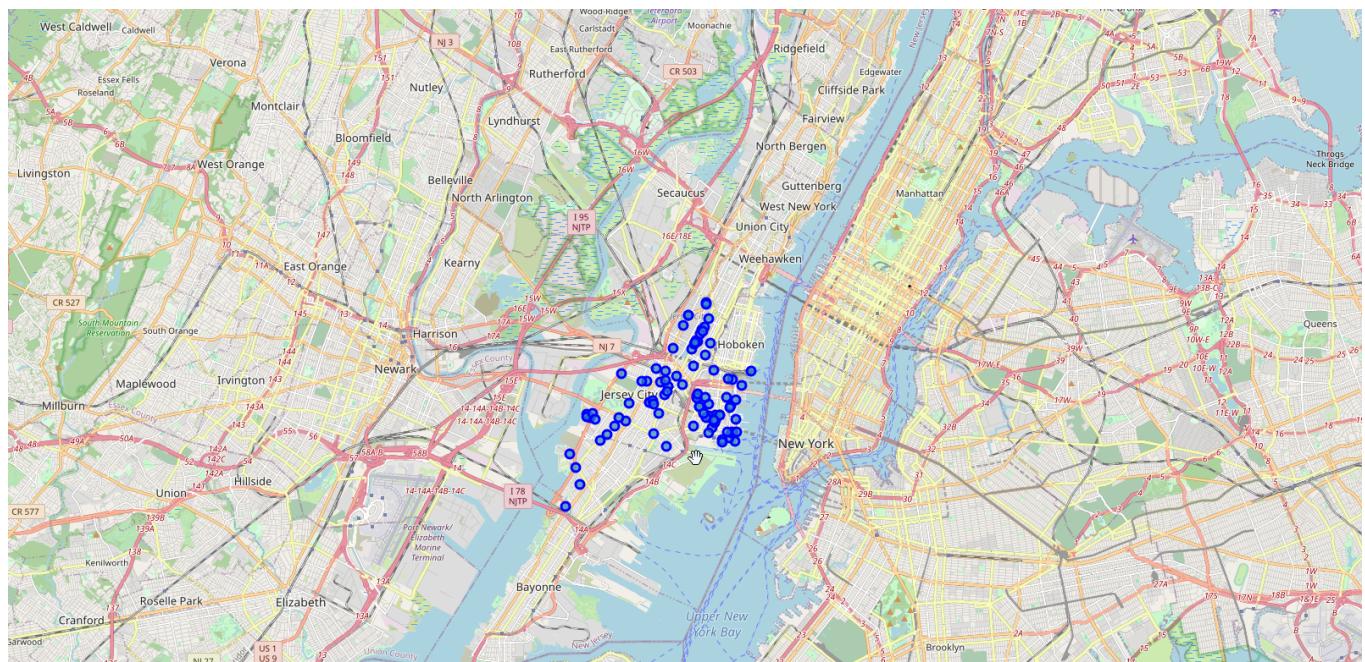
## Chicago



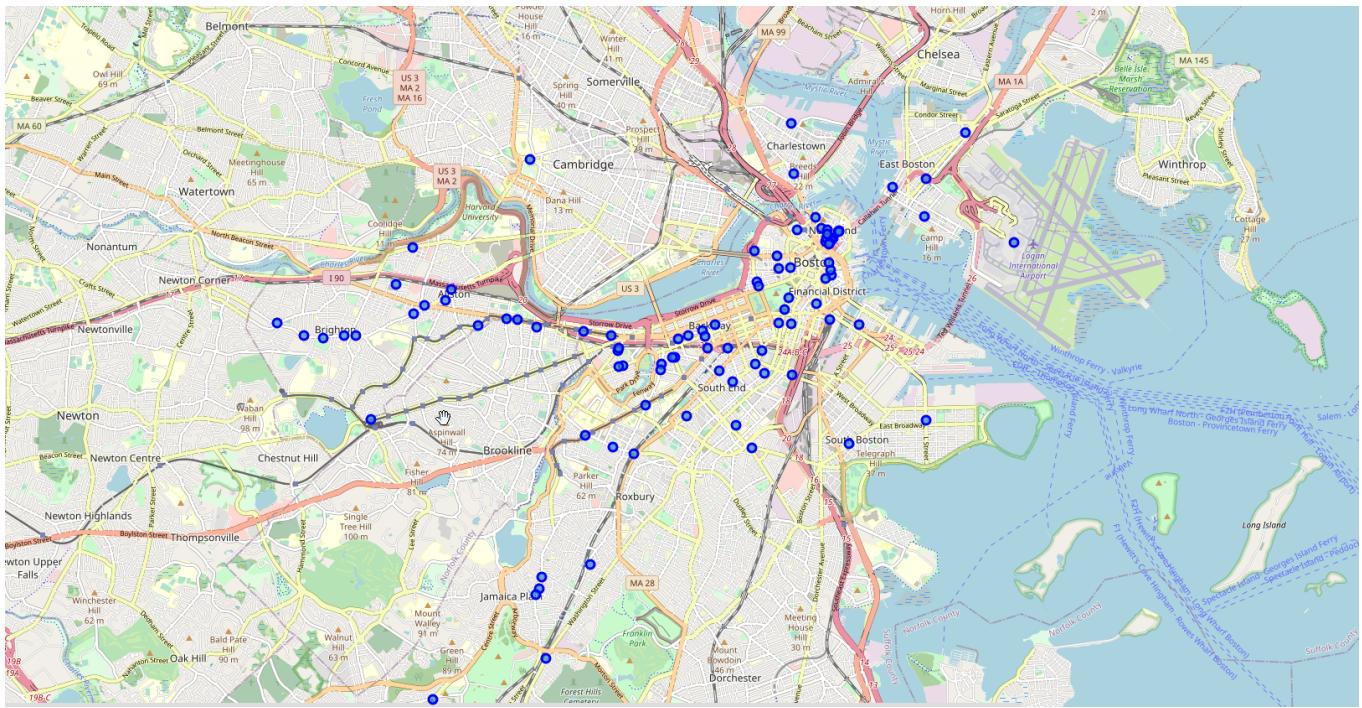
## San Francisco



## Jersey City



## Boston

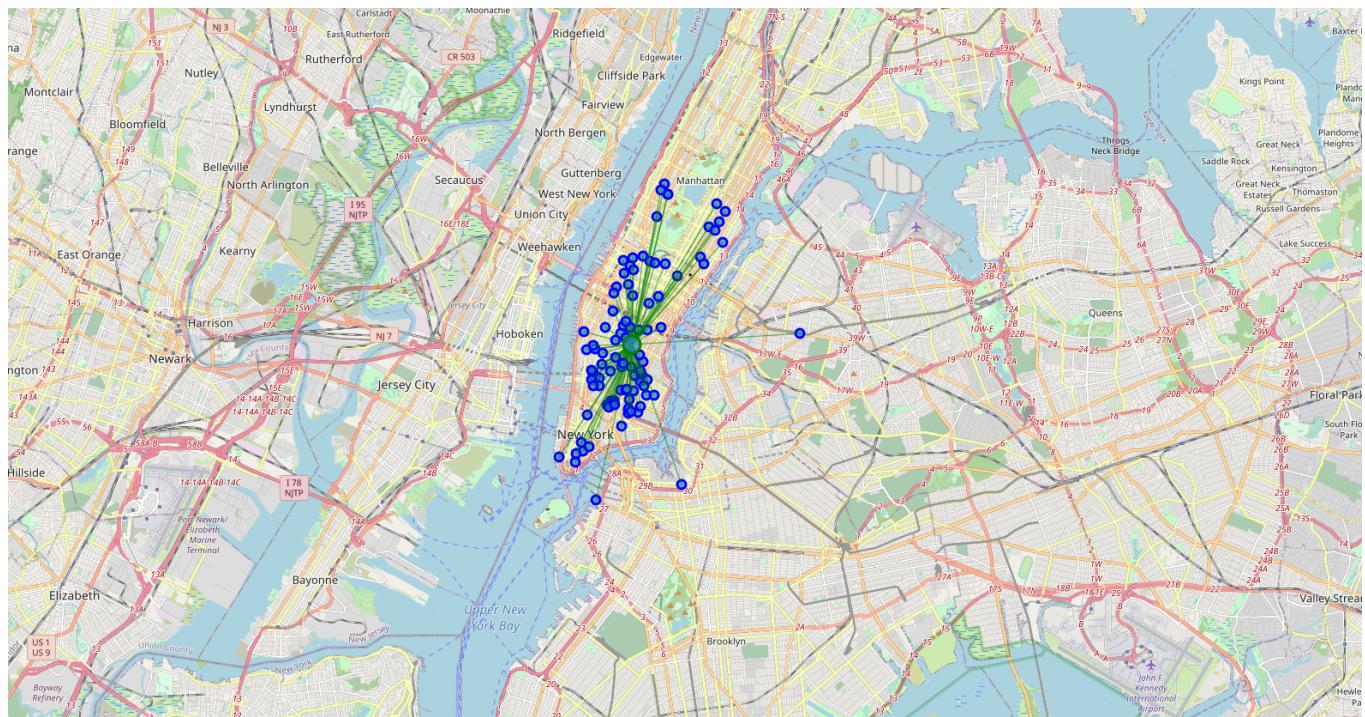


By looking at the maps I could clearly see that New York and Jersey City have the highest density when it comes to pizza places. However, this is not enough and required a furhter look at the concrete measurements of the density by using some basic statistics in order to first get the mean location of pizza places and next take the average of the distance to venues to the mean coordinates. By doing so I got got these values:

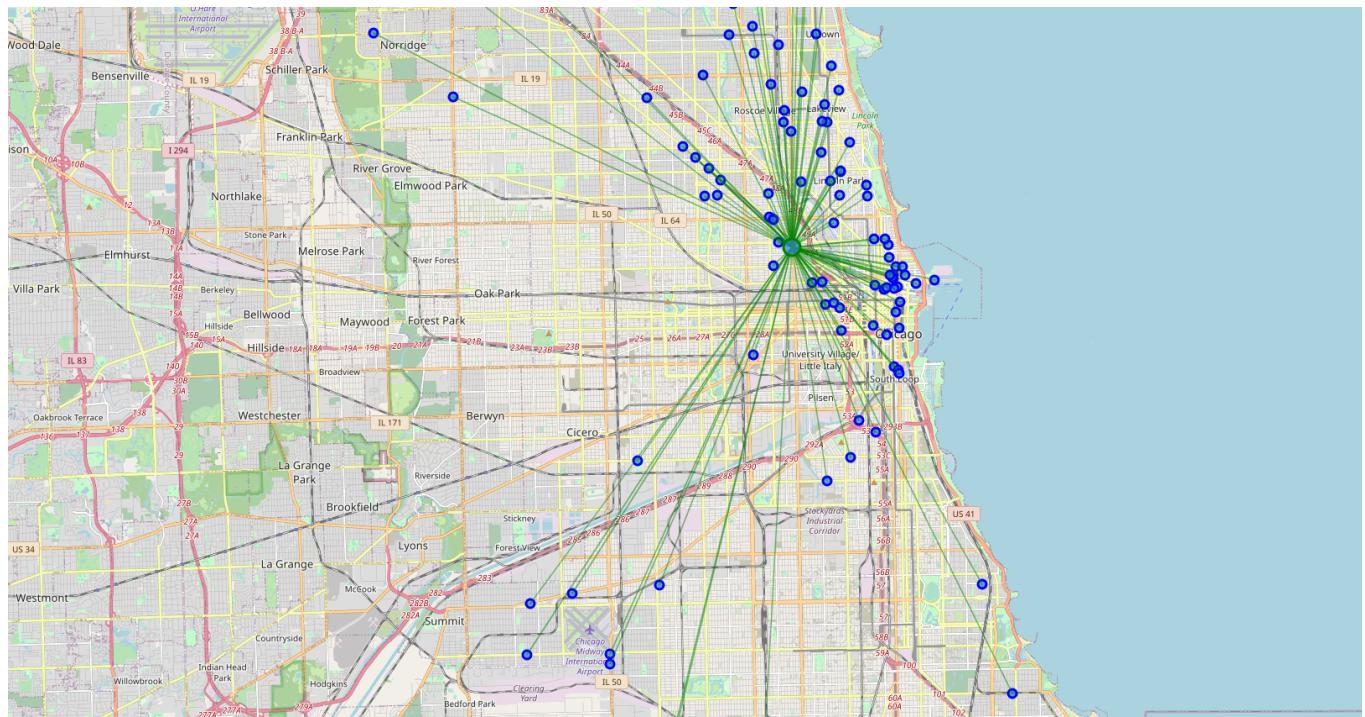
**New York, NY** Mean Distance from Mean coordinates 0.022591299517469826 **Chicago, IL** Mean Distance from Mean coordinates 0.06294178822671437 **San Francisco, CA** Mean Distance from Mean coordinates 0.02808016792198236 **Jersey City, NJ** Mean Distance from Mean coordinates 0.019473328512011254 **Boston, MA** Mean Distance from Mean coordinates 0.03490723805653442

Which actual looks like this:

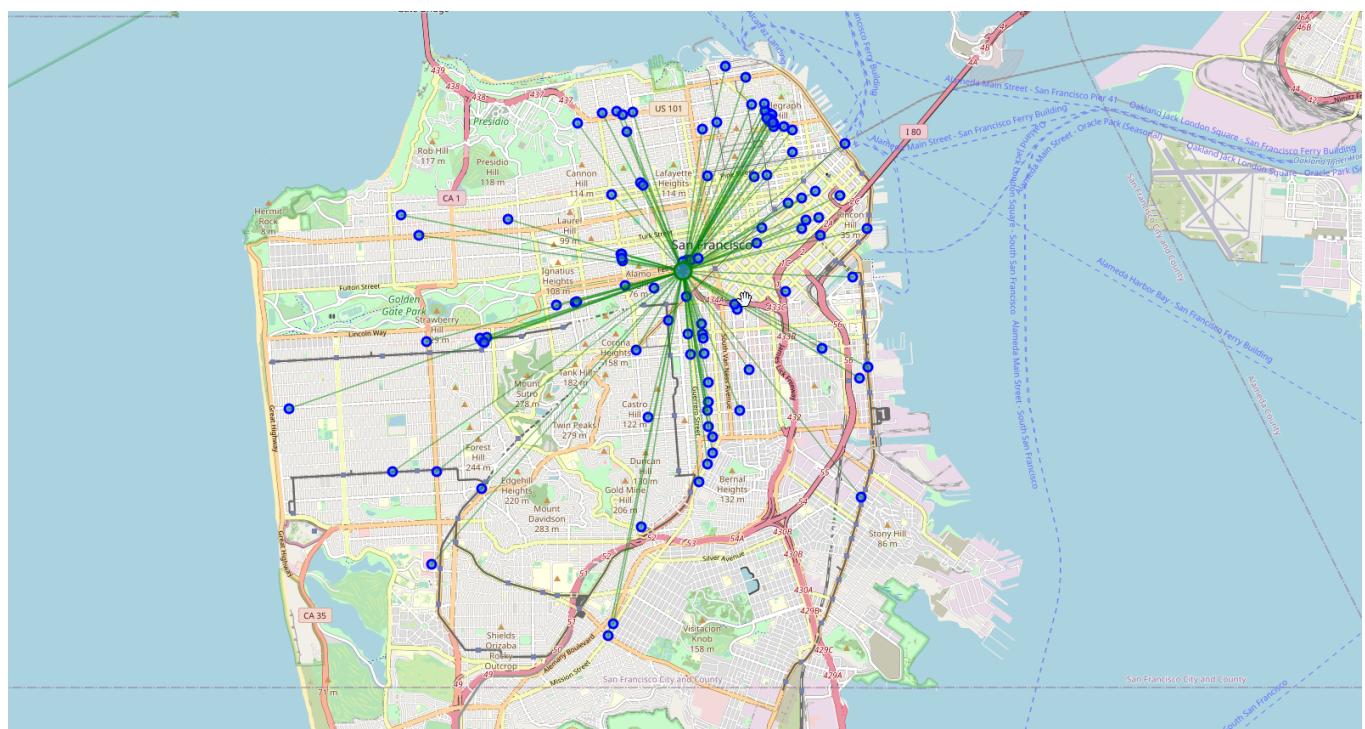
## New York



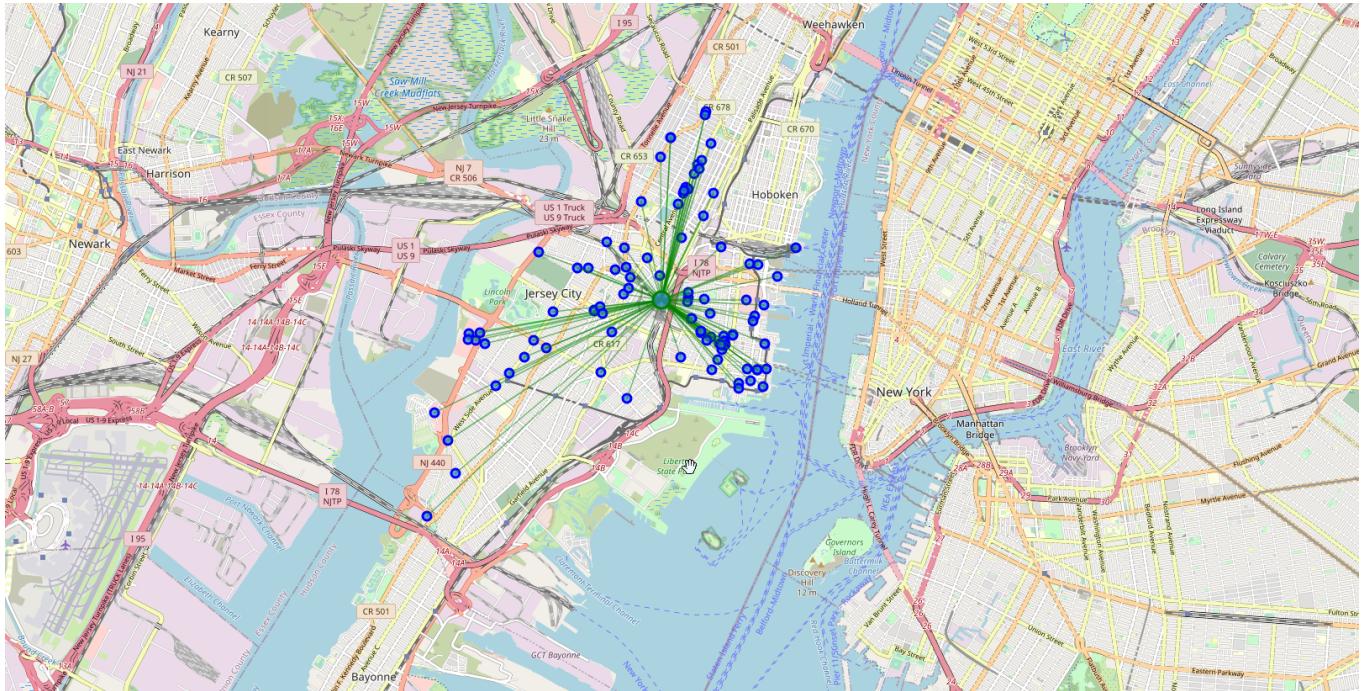
## Chicago



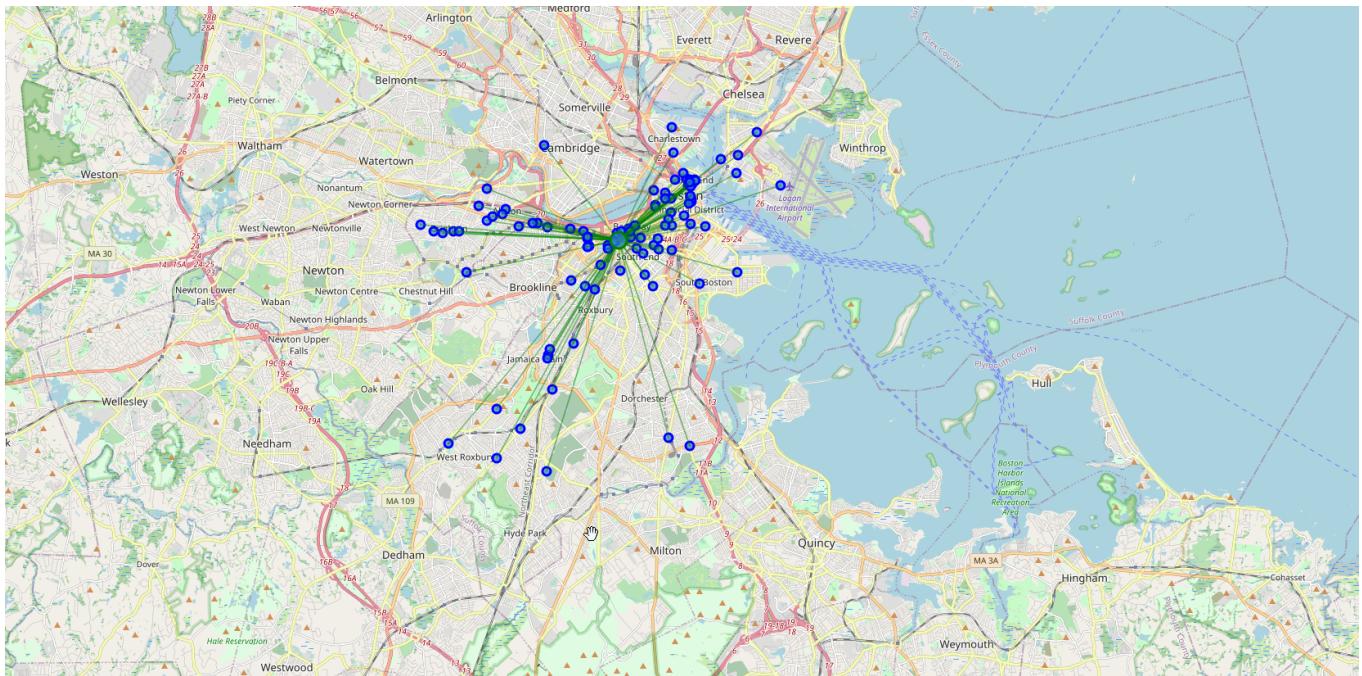
## San Francisco



## Jersey City



## Boston



By comparing the two cities, New York and Jersey City, I could see that the shortest mean distance to the mean coordinates is New York and therefore it is the best city where you can eat a really great pizza that meets all the pre-conditions. However, by visualising all maps we can identify one (1) pizza location that is really far away from the centre i.e. an outlier. If we would remove that city, what will the result then be? Well, it turned out that Jersey City had the lowest mean coordinates (0.019072529495376717) compared to the other cities so the ultimate recommendation for the tourist is **Jersey City!**

## Conclusion

This study concluded that Jersey City is the best city to have a pizza that meets the predefined conditions. However, it should be noted that if the outlier would not be identified then the recommendation would have been New York. This is something that is utterly important when analysing data that you have to consider if there are any noises e.g. outliers that can screw the results, and a recommendation for future analysis is that