[The-Battle-of-Neighborhoods](#)

PIZZA IN NEW YORK OR SAN FRANCISCO?

Data Science Capstone - IBM Data Science Professional Certificate on Coursera

By Kabwe C George

Contents

Introduction

We all love pizza! Right? Research has shown that pizza is one of the most popular foods in the world as ranked by [Taste atlas](#) and other researchers.

This project focuses on comparing pizza places in New York to San Francisco. Imagine your choice of which city to visit depends on which city has the best pizza. The decision is to choose a place with a high density of Pizza places as most of people love to change places where they eat pizza from until they find their best and favorite place. To solve this problem, an analysis of the Pizza stores' locations in New York and San Francisco will be conducted to find the best place for anyone who loves changing pizza places and anyone who wants to find the best place from .

Data section

The FourSquare API will be used to collect data about locations of Pizza stores in New York, NY and San Francisco, CA. These are two of the most populated US cities and they should contain some of the best Pizza places in the US.

The data obtained from FourSquare will have the following columns:

- Name of the pizza place
- The Address of the place for people to quickly locate it
- The Latitude and Longitude which will be used for mapping the places.

Methodology

The target here is to asses which city would have the highest Pizza store density. The Four Square API through the venues channel will be utilized together with the near query to get venues in the cities. Also, the CategoryID will be used to set it to show only Pizza Places. An Example of the requests:

```
https://api.foursquare.com/v2/venues/explore?&client_id=&client_secret=&v=20180605&Ne
w York, NY&limit=100&categoryId=4bf58dd8d48988d1ca941735
```
The 4bf58dd8d48988d1ca941735 is the Id of the Pizza Place Category. Also, Foursquare limits us to maximum of 100 venues per query.
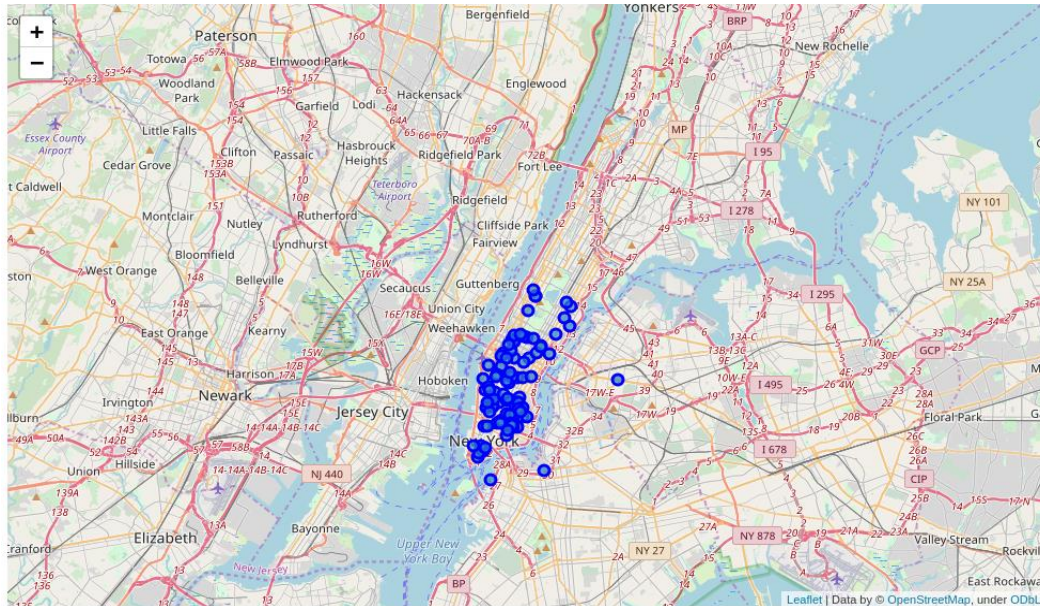
This request will be repeated for San Francisco to get its top 100 venues. The name and coordinate data will then get saved from the result and plotted on the map for visualization.

Next, to get an indicator of the density of Pizza Places, the center coordinate of the venues will be calculated to get the mean longitude and latitude values. Then the mean of the Euclidean distance from each venue to the mean coordinates will be calculated. This will be the indicator; mean distance to the mean coordinate.
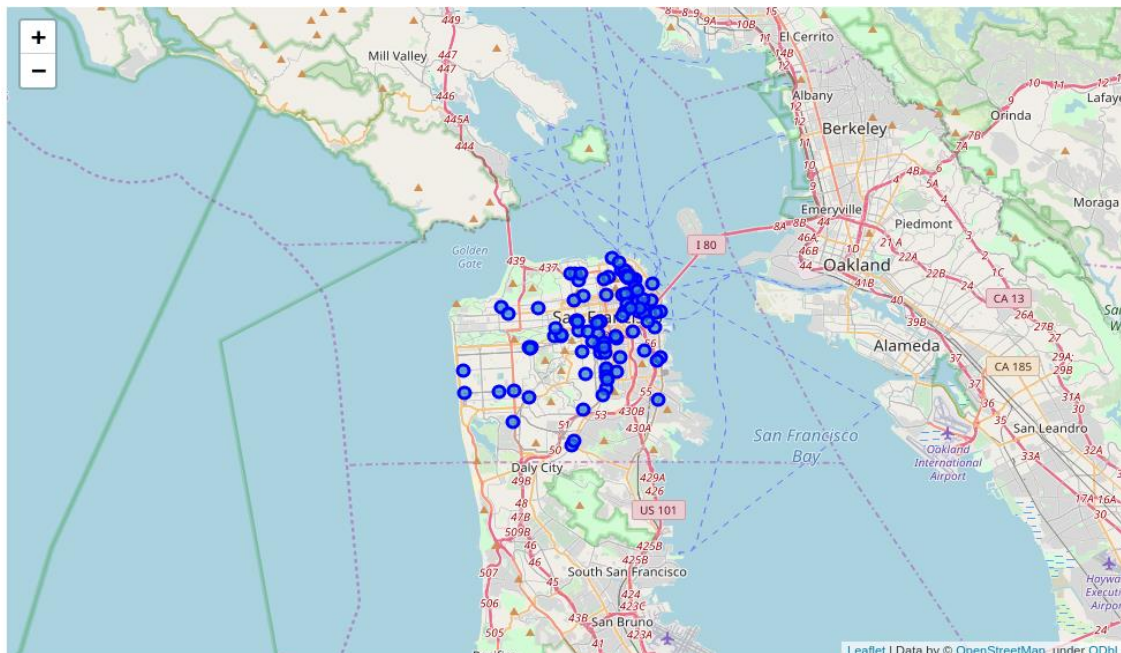
Results

The initial visualization showed that New York and San Francisco all have multiple pizza places and often more than Foursquare would like to supply us. The following are the pictures of the geoplot generated with folium:
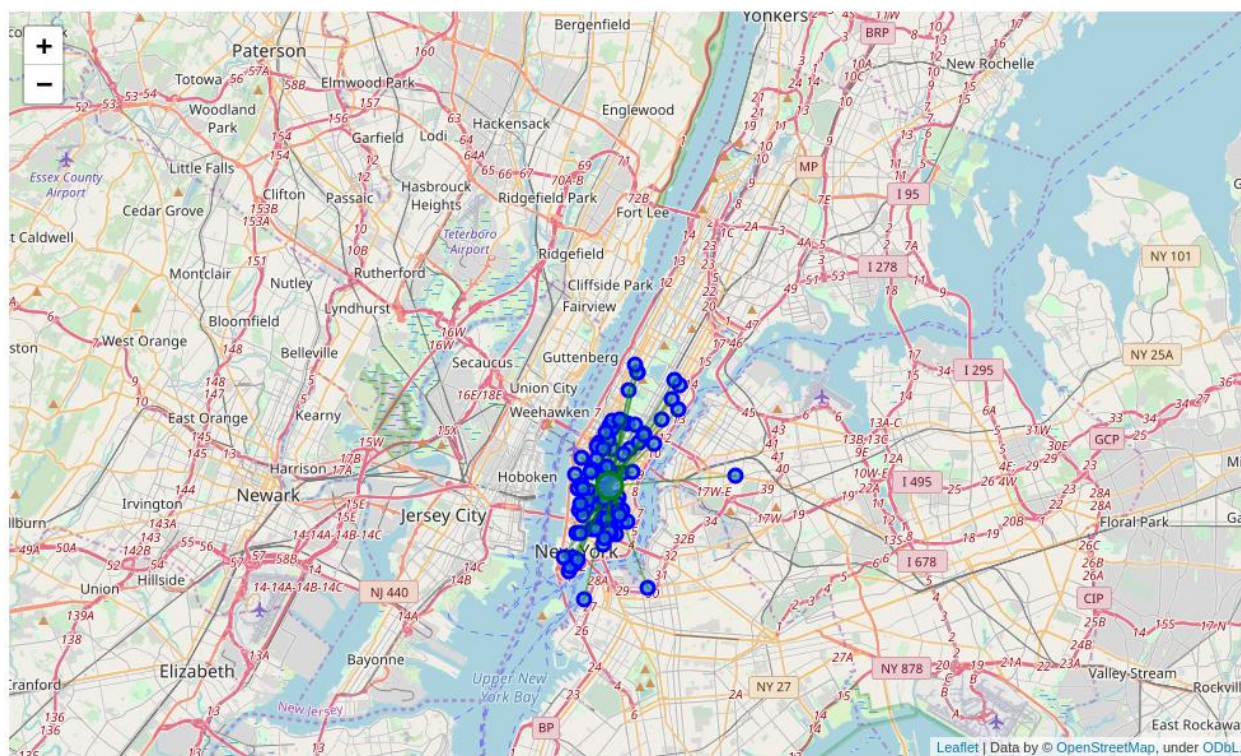
New York:

San Francisco:

The visualization shows that New York and San Francisco all have multiple pizza places.

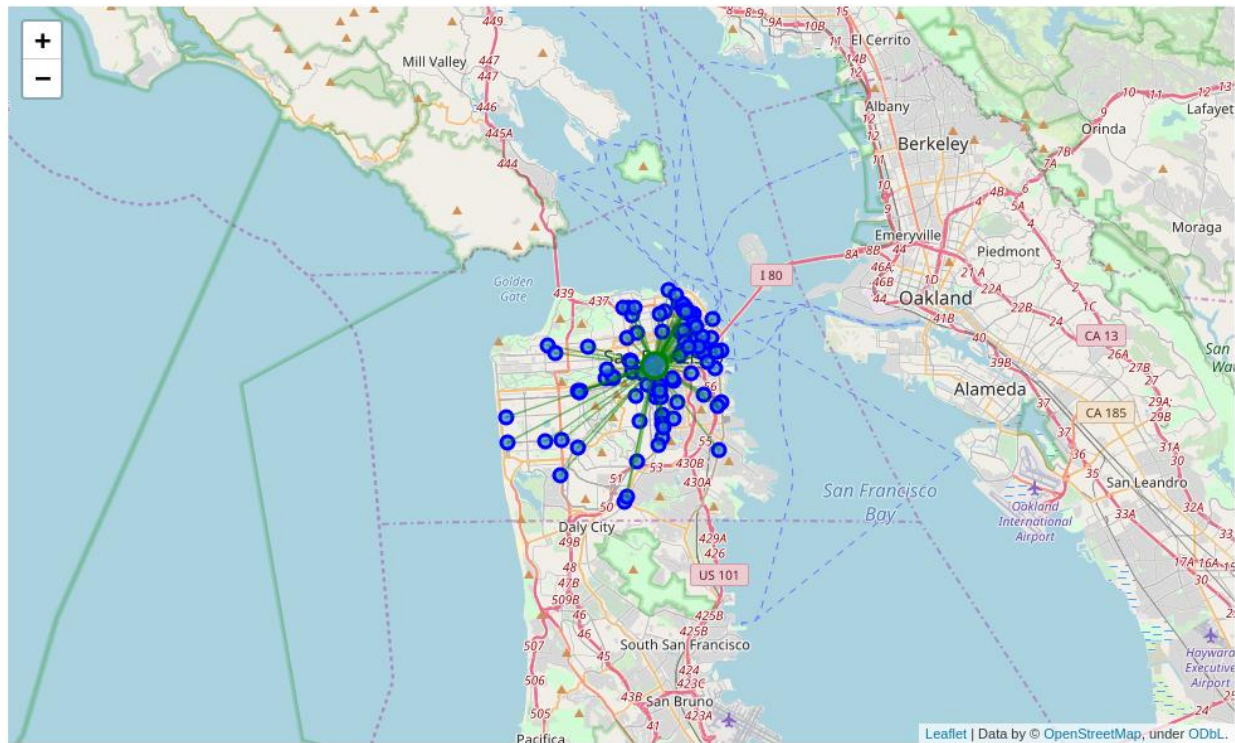| City	Number of Pizza Places
New York	271
San Francisco	167

Calculating the Mean coordinate and the mean distance to mean coordinate (MDMC).

The mean coordinate will be represented with a big green circle and distances with green lines.

New York:

San Francisco:

On average New York's top 100 Pizza places are 25% more dense than San Francisco

City	Mean Distance from centre
New York	0.02296898464975578
San Francisco	0.0288041694197 7326

The results show that New York ranks better than San Francisco.

Discussion:

It has been established that New York has its Pizza places more densely located as compared to San Francisco.

It can be therefore be concluded that New York is ranked as the best option in this comparison.

Conclusion:

An individual wishing to decide which city has pizza places close to each other can utilize these findings.

For future considerations, the pizza places can be ranked so and other information like availability of parking lots, children entertainment and other information can be gathered to better narrow down on which place one can choose