The Battle of Neighbourhoods

IBM Data Science Professional Certificate – Capstone Project

Opening a new Fitness Studio in the USA

- Sameer Phatak

(Jun2020)

Contents

Introduction

Background

The hypothetical client "ZZZ Fitness Studios" is a multi-national chain of fitness centres across the globe. However their presence in the USA is limited. The client has narrowed their search to five major cities in the USA and wishes to find the best amongst these to start a new Fitness studio. These cities are amongst the most populated and commercially amongst the highest ranking US cities. The client wants to make a decision that will be beneficial economically.

Business Problem

The client is unable to decide which city amongst New York (NY), Chicago (IL), San Francisco (CA), Jersey City (NJ), Boston (MA) of USA would be the best for opening a new Fitness Studio. The objective of this project is to analyse the current presence of Gyms/ Fitness Centres within these cities, to decide which city within US could be economically best for opening a new Fitness Studio. The client would like to leverage Machine Learning techniques to address this problem.

Thus simply put, applying Machine Learning techniques, can we recommend a city amongst the above five where "ZZZ Fitness Studios" could consider setting up a new Fitness studio?

Target Audience

The target audience are mainly those interested in opening a gym or a fitness studio in these cities. The methodology implemented in this project can also be a guide to analyse cities other business setups.

Analytical Approach

A visualization of the 5 cities with respect to the venue - Gyms/ Fitness centres backed by quantitative metrics would help arrive at a recommendation to the client.

Data Section

Data Requirements

Data required would be about locations of Gyms/ Fitness Centres in the five major cities in the USA: New York (NY), San Francisco (CA), Jersey City (NJ), Boston (MA) and Chicago (IL).

Data Collection, Understanding and Preparation

Use the FourSquare API to collect the location data for the five selected cities.

The Category ID for Gym/ Fitness centres is "*4bf58dd8d48988d175941735*"

The city codes are "New York, NY", 'Chicago, IL', 'San Francisco, CA', 'Jersey City, NJ', 'Boston, MA'
The URL is thus generated using the above parameters with the Foursquare API.
This data would help to identify the city where the Gym/ Fitness Centre could be setup based on the spread of this venue.

Methodology and Execution

The main target here is to asses which city would have the highest Fitness Studio density. The **Four Square API** was used through the venues channel. The near query was used to get venues in the cities. Also, the CategoryID was used to set it to show only Fitness Studio.

An Example of the requests:

`https://api.foursquare.com/v2/venues/explore?&client_id=&client_secret=&v=20180605&New York,NY&limit=100&categoryId=4bf58dd8d48988d1ca941735`

That "*4bf58dd8d48988d175941735*" is the Id of the Gym/Fitness Centre Category.

It should be taken into consideration that Foursquare limits to maximum of 100 venues per query.

This request was repeated for the five cities under study and their top 100 venues of the type Gym/Fitness Centre were considered. The name and coordinate data only from the result and plotted them on the map for visual inspection using **Folium**.

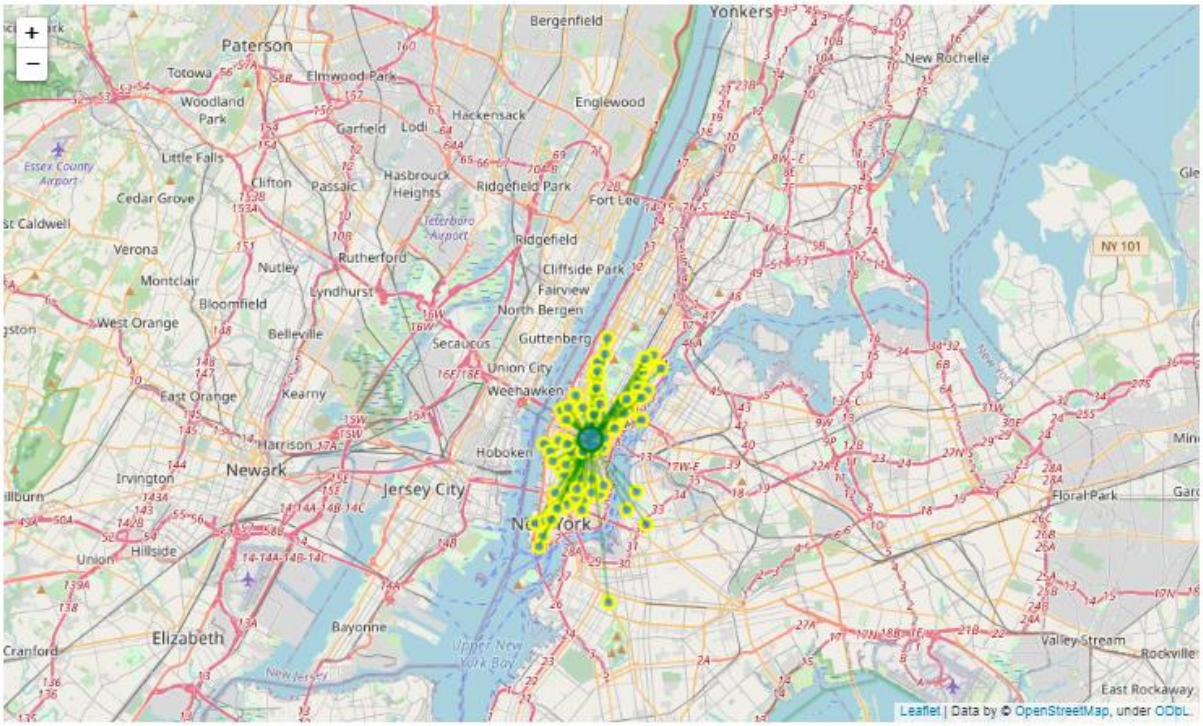
Next, to get an indicator of the density of Fitness Studios, a centre coordinate of the venues was calculated to get the mean longitude and latitude values. Then the mean of the Euclidean distance from each venue to the mean coordinates was calculated. That was the indicator used: **mean distance to the mean coordinate (MDMC)**.

Results

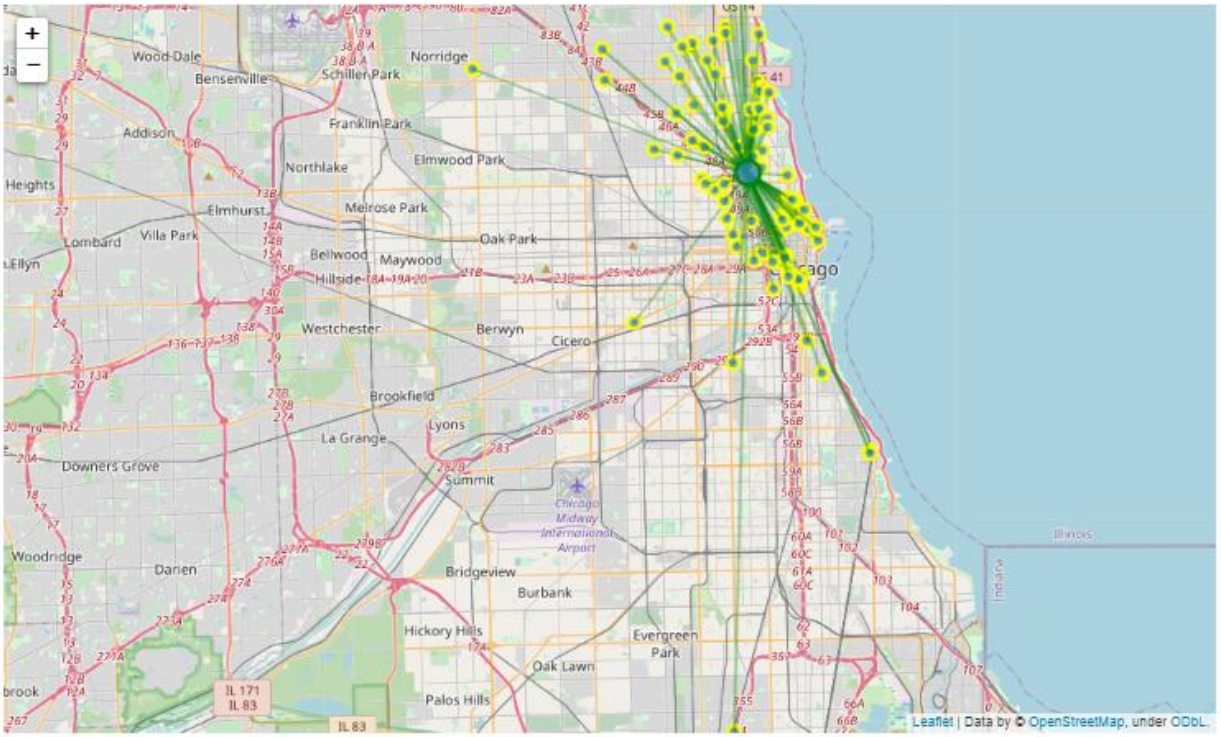
For the initial visual inspection it is observed that all cities have multiple Fitness Studios and often more than Foursquare would like to supply us.

Below are screenshots of the Geoplots generated with folium for each of the five cities and corresponding Mean Distance from Mean co-ordinates

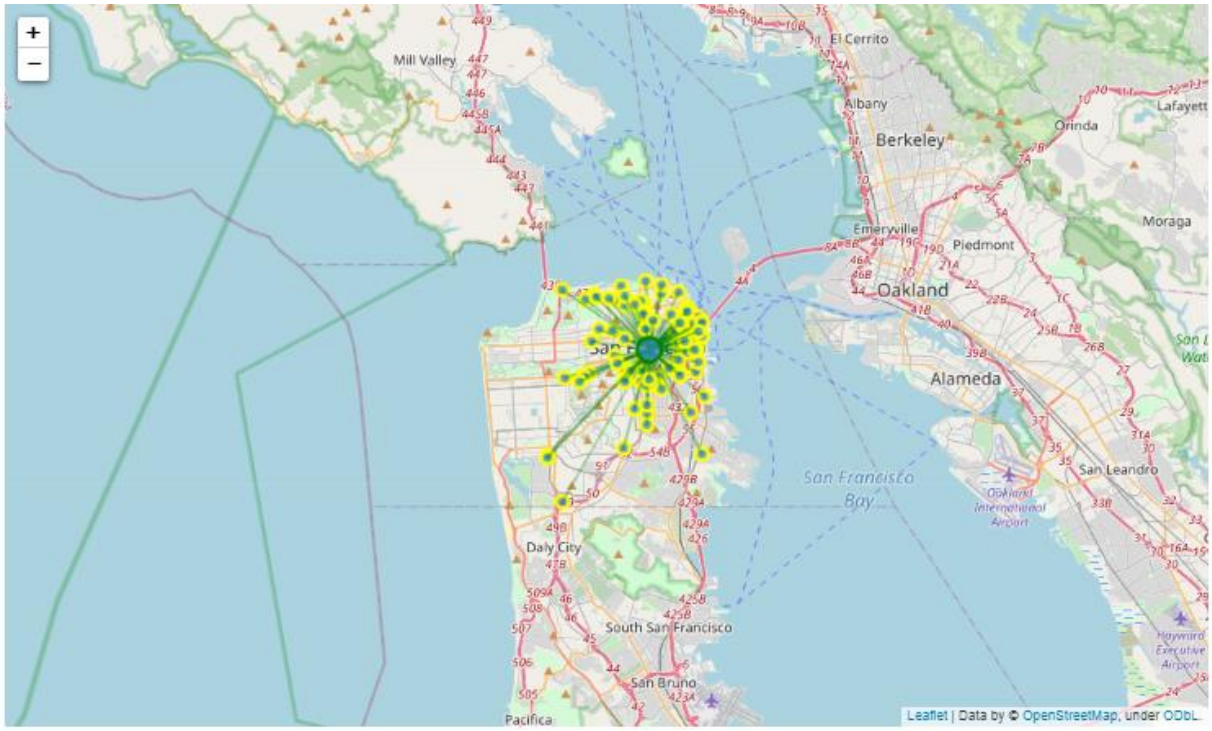
New York, NY
Mean Distance from Mean coordinates
0.02592323858922523



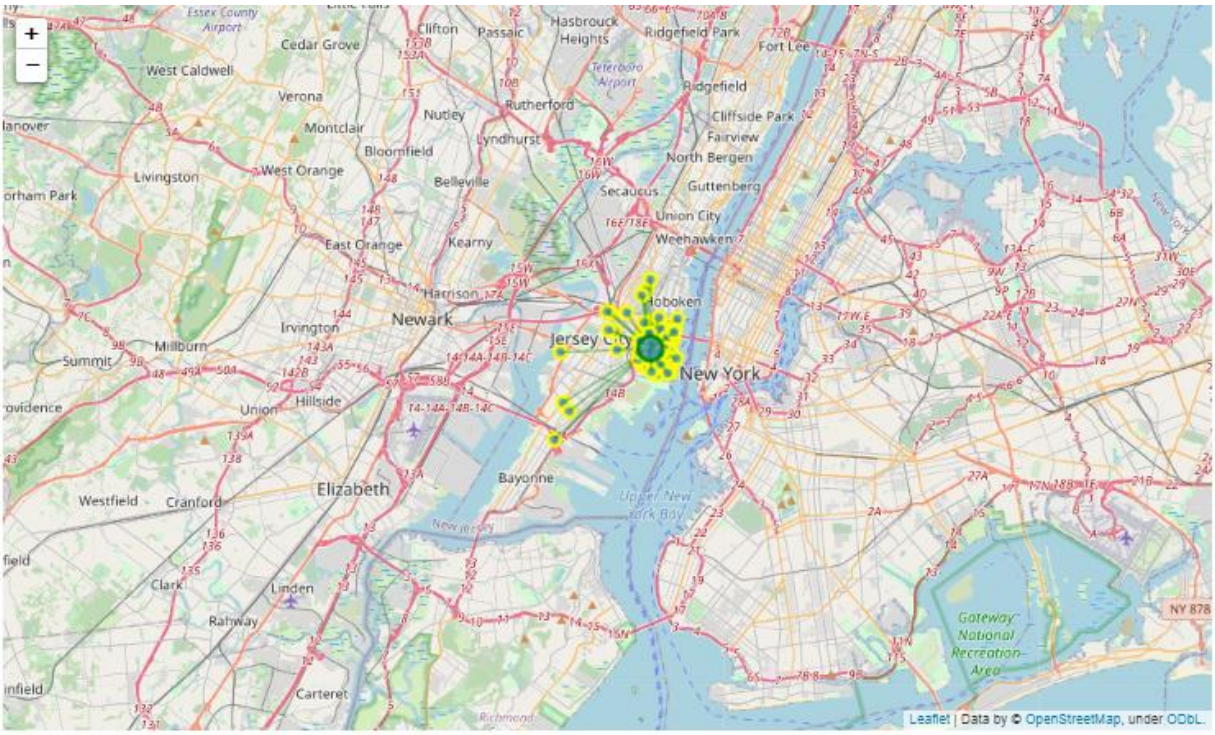
Chicago, IL
Mean Distance from Mean coordinates
0.0461274800911952



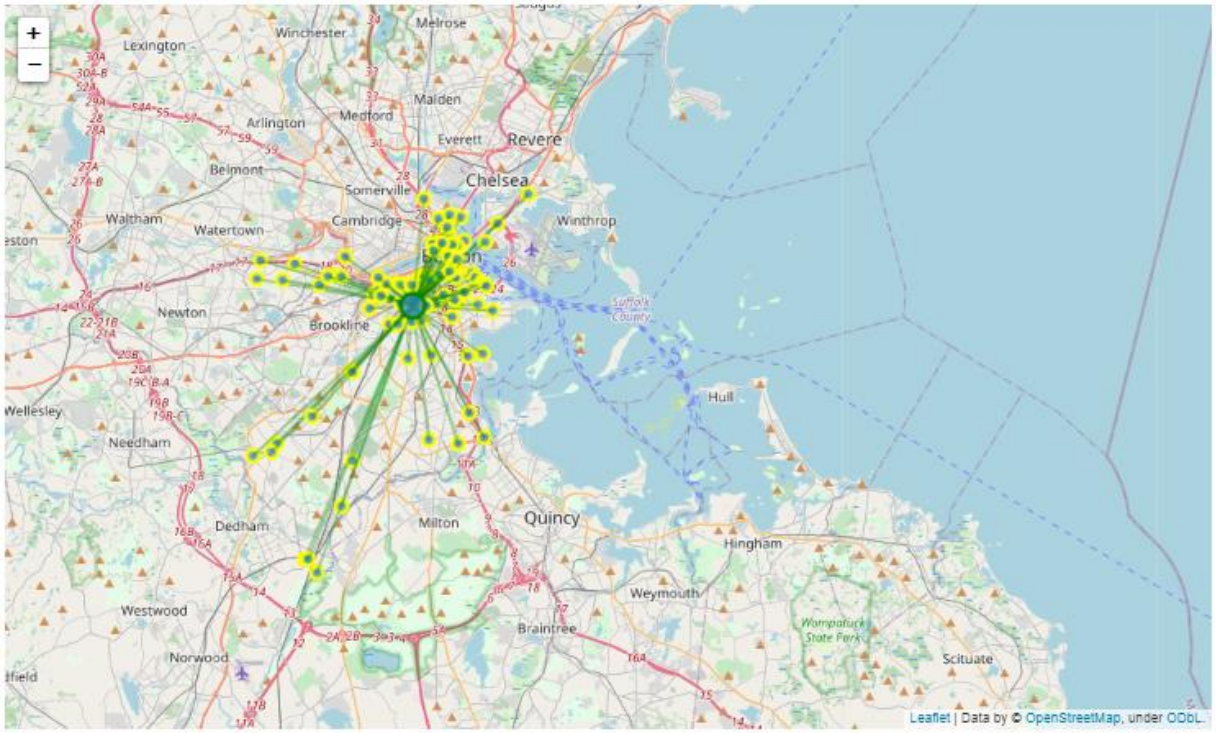
San Francisco, CA
Mean Distance from Mean coordinates
0.02555210180662777



Jersey City, NJ
Mean Distance from Mean coordinates
0.012395944321890763



```
Boston, MA
Mean Distance from Mean coordinates
0.036915033621342415
```



Discussion

There are a few figures that are really far resulting in a higher Mean Distance from Mean Coordinates (MDMC). However, they are not one-off figures. Moreover, such figures can be observed in each of the maps. Hence these would more or less impact the MDMC for all cities and removing them not have much impact on the final conclusion. Hence, I decided to retain them and not remove any.

With the current analysis, it is clear that Chicago would be the right choice to setup a new Fitness Centre.

One consideration to do further work on is to move the location of the Foursquare API query until we get all the Fitness Studios in each city and do the calculations again.

Conclusion and Recommendations

Chicago is the best city in the US to set up a new Fitness Centre for ZZZ Fitness Studios.

Highest MDMC for Chicago suggests that the gyms located within Chicago are more scattered as compared to other cities. A new fitness centre setup in proximity of workplace/ residence will benefit employees/ residents as they would not need to travel far. This way, ZZZ fitness would also gain economically by attracting more customers.