



# **IBM Data Science Capstone Project**

by

Michael Margolis

June 10, 2020

## Table of Contents

Introduction.....	page 3
Data.....	page 4
Methodology.....	page 5
Results.....	page 6
Discussion.....	page 10
Conclusion.....	page 11

## **Introduction**

I am a musician and am in the process of saving money for a big vacation for when cities and venues open post-COVID-19. The following cities are on my bucket list:

- New York, NY
- San Francisco, CA
- Seattle, WA
- Boston, MA
- Chicago, IL

I will make my selection based on a combination of density and location of concert venues. The problem I want to solve is to analyze concert venue locations in the above cities to help inform a destination and find a convenient area to make a hotel reservation.

## Data

Foursquare application programming interface (API) will be used to research and collect concert venue information for the aforementioned locations.

- <https://developer.foursquare.com/docs/build-with-foursquare/categories>
- Foursquare venue (concert hall) location code: 5032792091d4c4b30a586d5c

The data set contains several important pieces of information including city name (e.g., New York, NY) and location of city center (e.g., latitude: 40.742185, longitude: -73.992602).

Using the Foursquare data, a map will be constructed for each of the cities showing each concert hall location and analyses conducted to inform a decision regarding where to go on vacation.

## **Methodology**

The methodology used is to examine the aforementioned Foursquare data for "Concert Halls" to do three things:

1. Calculate the total number of concert venues in each city.
2. Calculate the centroid (or mean coordinate) of said venues in each city.
3. Use this information to make a more informed decision regarding where to take a nice vacation post COVID-19.

The data will be examined to minimize influence of outlier data points (e.g., distant concert halls).

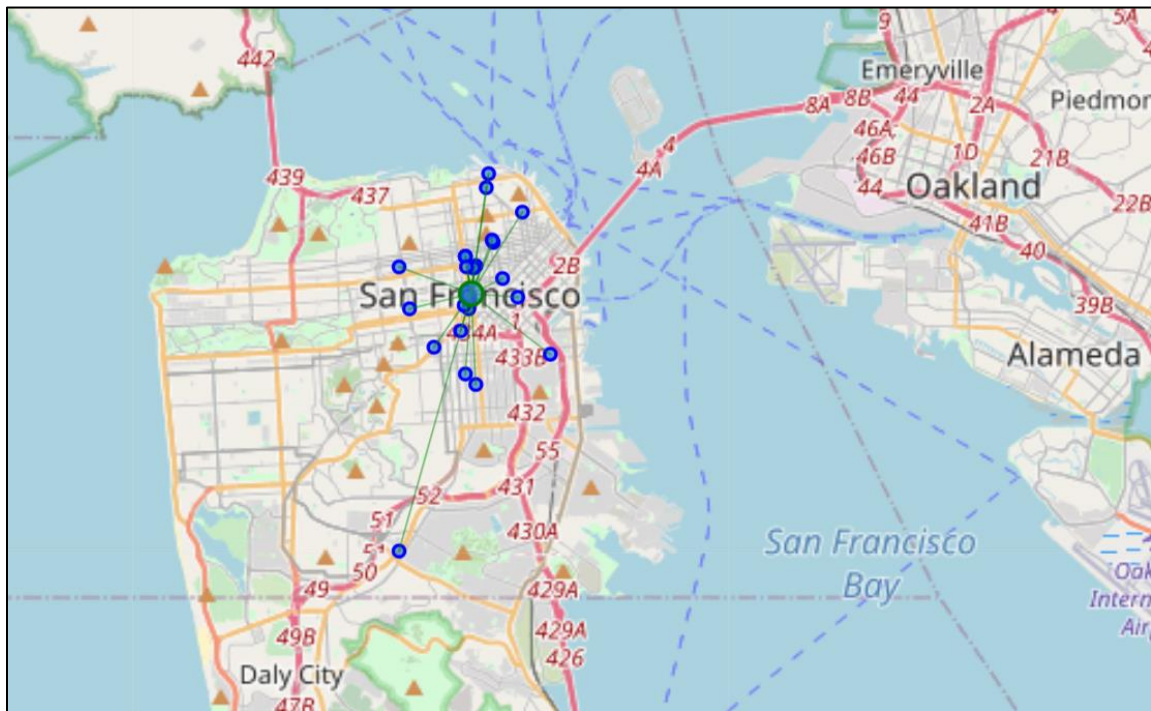
## Results

The table below lists for each city, the total number of concert halls, and the mean distance to the mean centroid/coordinate the respective cluster of said venues.

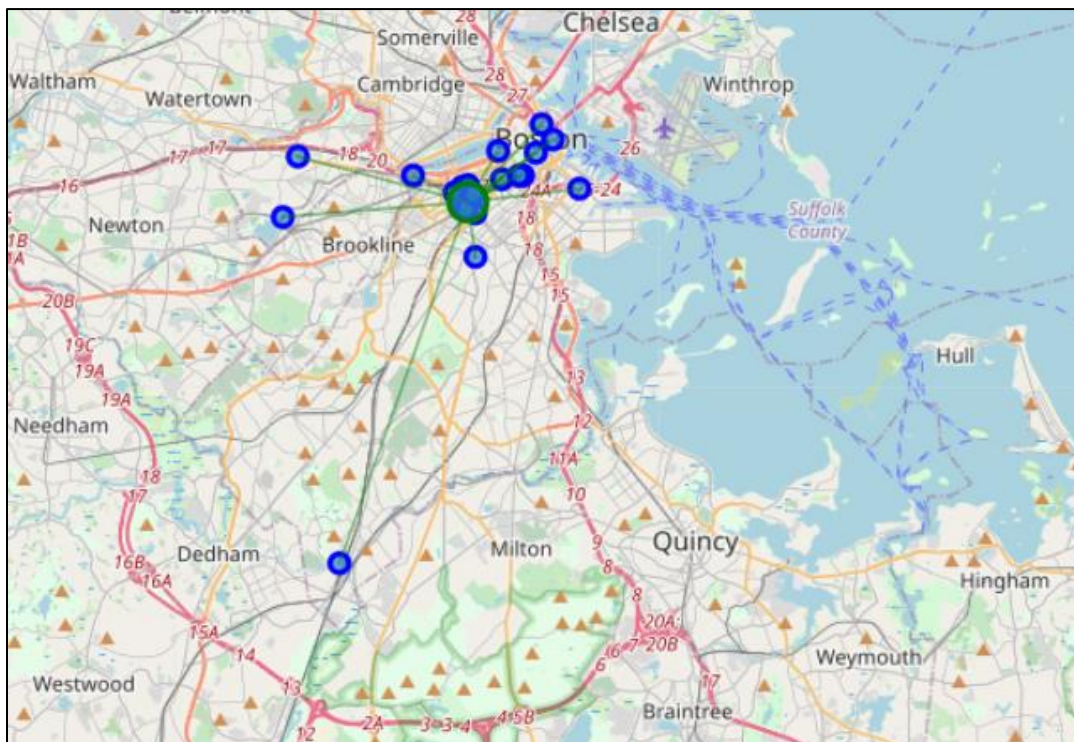
City	Num Venues	MDMC
San Francisco, CA	26	0.011
Boston, MA	25	0.014
Seattle, WA	17	0.035
New York, NY	80	0.037
Chicago, IL	45	0.049

Note: During the analysis, most of the cities had ~3 outlier data points. Therefore, the three most distant venues were removed from each data set as to not skew results.

Map of concert venues in San Francisco, CA:

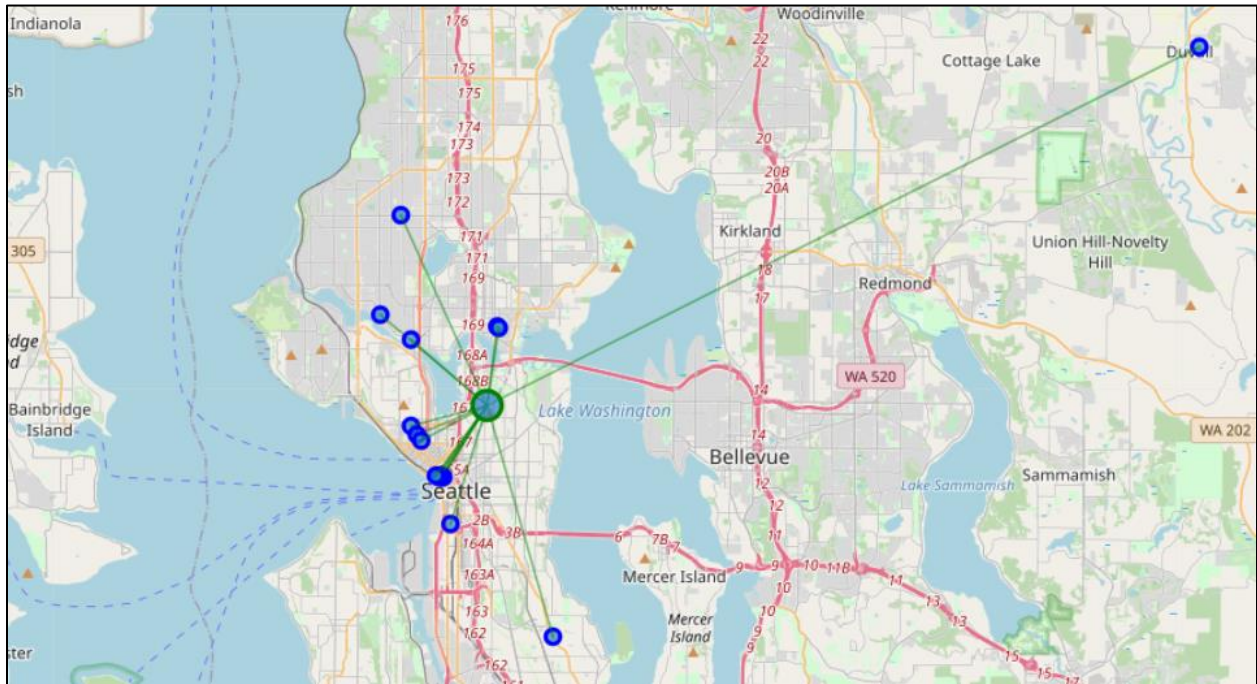


Map of concert venues in Boston, MA:

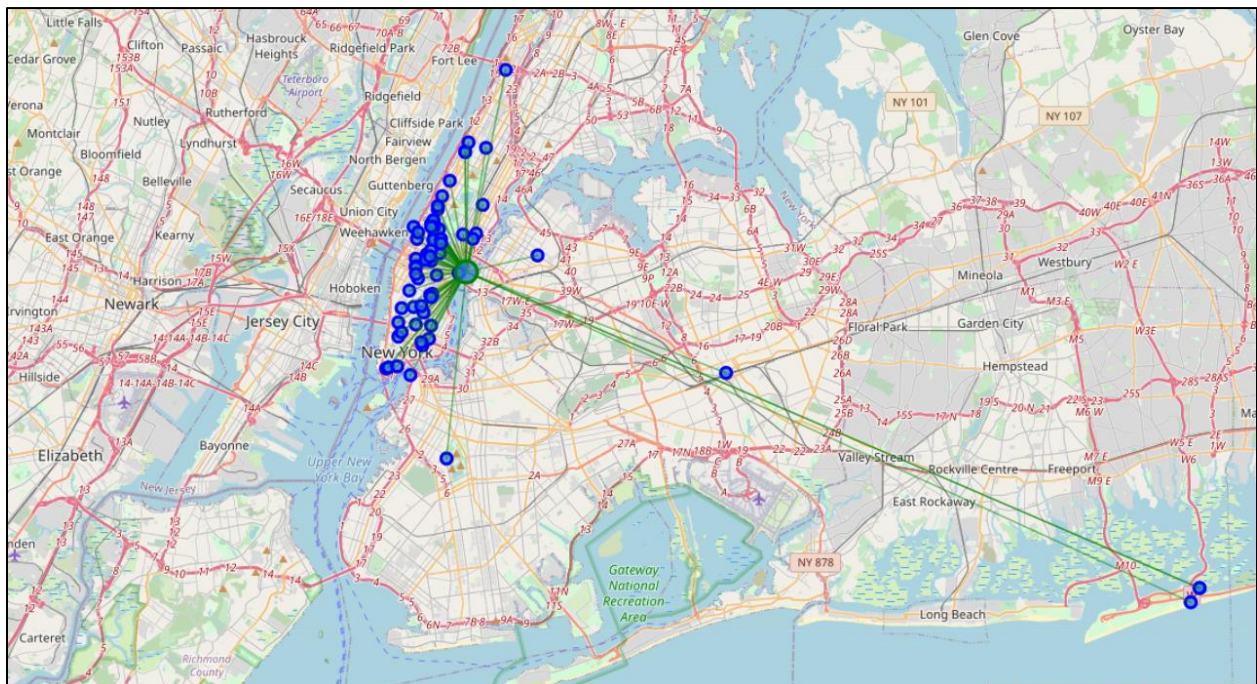




Map of concert venues in Seattle, WA:

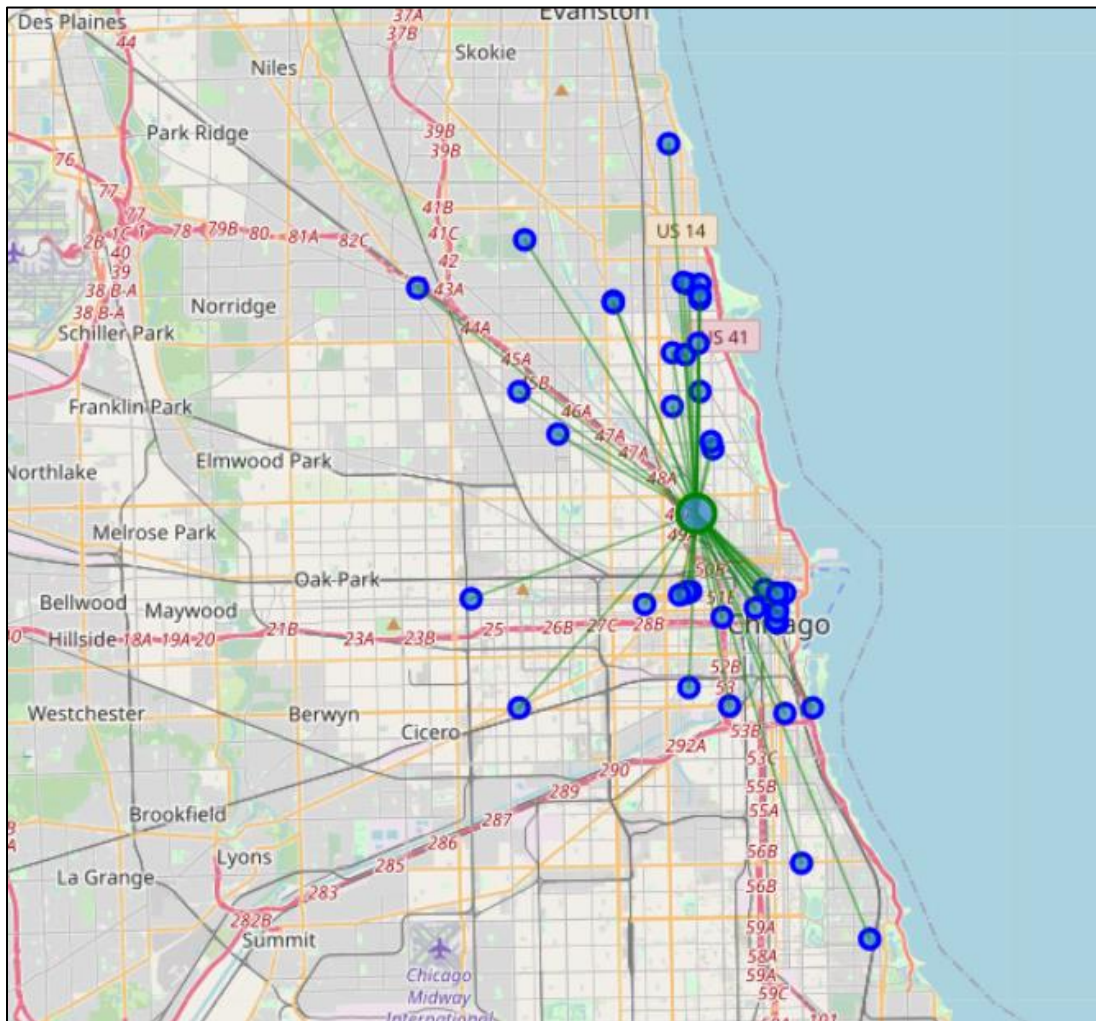


Map of concert venues in New York, NY:





## Map of concert venues in Chicago, IL



## **Discussion**

New York, NY clearly possesses the greatest quantity of concert halls relative to the other cities; nearly twice as much for each, but the mean distance to the mean centroid/coordinate (MDMC) is higher than other cities (0.037 km).

San Francisco, CA and Boston, MA each have ~25 venues with a similar MDMC of ~0.01 km.

All cities have several venues within a convenient distance from their respective concert hall MDMC.

## **Conclusion**

New York, NY appears to be the best choice for me to visit for my next vacation as this city offers nearly twice the number of concert halls relative to the other cities examined in this project.

While New York's MDMC is higher than three other cities (San Francisco, Boston and Seattle), it is only  $\sim 0.024$  km more, which is negligible.

Path forward: Keep an eye for when New York concert venues open post COVID-19 and when they do, make a reservation in a hotel vicinity the center of New York's concert venues.