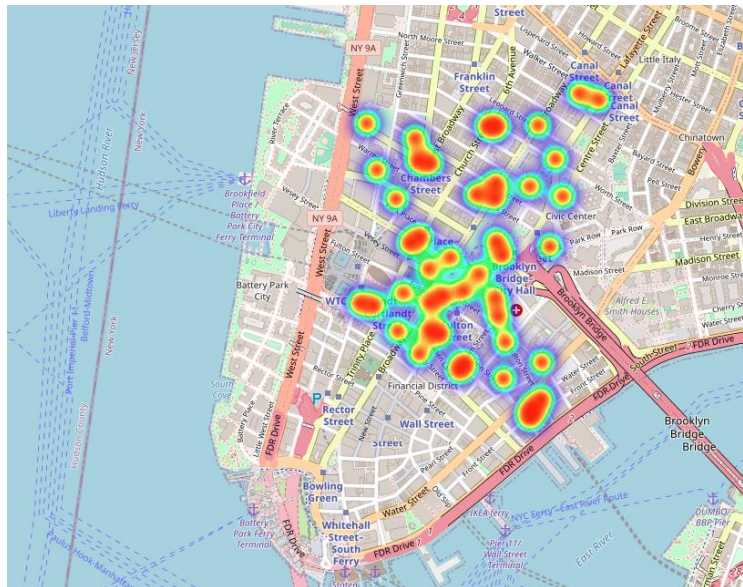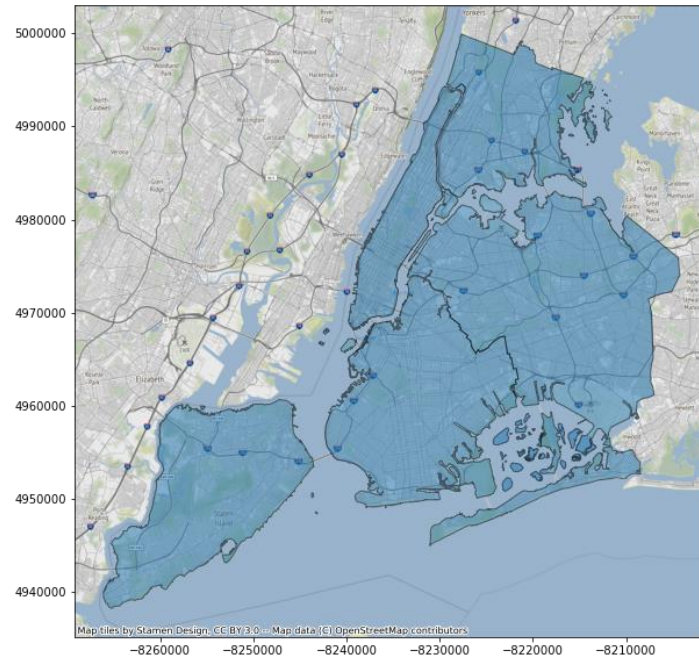# Peer-graded Assignment:

Capstone Project - The Battle of Neighborhoods

Report by Mohammed Burhanuddeen R

Github account [**@Metallurgist**]

Preface

Instructions for the Capstone Project Assignment:

Now that you have been equipped with the skills and the tools to use location data to explore a geographical location, over the course of two weeks, you will have the opportunity to be as creative as you want and come up with an idea to leverage the Foursquare location data to explore or compare neighborhoods or cities of your choice or to come up with a problem that you can use the Foursquare location data to solve. If you cannot think of an idea or a problem, here are some ideas to get you started:

In Module 3, we explored New York City and the city of Toronto and segmented and clustered their neighborhoods. Both cities are very diverse and are the financial capitals of their respective countries. One interesting idea would be to compare the neighborhoods of the two cities and determine how similar or dissimilar they are. Is New York City more like Toronto or Paris or some other multicultural city? I will leave it to you to refine this idea. In a city of your choice, if someone is looking to open a restaurant, where would you recommend that they open it? Similarly, if a contractor is trying to start their own business, where would you recommend that they setup their office? These are just a couple of many ideas and problems that can be solved using location data in addition to other datasets. No matter what you decide to do, make sure to provide sufficient justification of why you think what you want to do or solve is important and why would a client or a group of people be interested in your project.

Review criteria:
This capstone project will be graded by your peers. This capstone project is worth 70% of your total grade. The project will be completed over the course of 2 weeks. Week 1 submissions will be worth 30% whereas Week 2 submissions will be worth 40% of your total grade.

For Week 1, you will be required to submit the following:

1. A description of the problem and a discussion of the background. (15 marks)
2. A description of the data and how it will be used to solve the problem. (15 marks)

For the Week 2, the final deliverables of the project will be:

1. A link to your Notebook on your Github repository, showing your code. (15 marks)
2. A full report consisting of all of the following components (15 marks):
   - Introduction where you discuss the business problem and who would be interested in this project.
   - Data where you describe the data that will be used to solve the problem and the source of the data.
   - Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learning's were used and why.
   - Results section where you discuss the results.
   - Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.
   - Conclusion section where you conclude the report.
3. Your choice of a presentation or blogpost. (10 marks)

Topic: A new location to start Coffee venue in Manhattan, New York City

1. **A description of the problem and a discussion of the background.**

Client with well-established Coffee Enterprises outlets in Canada wants to try his luck in US, by opening a new Coffee chain outlet in the New York City, especially in the Manhattan Area. He is confident about his business skills, but is not very sure of where in Manhattan he wants to start this new business.

So, my task is to provide him an understanding about the business competition he will be facing and his fight for the Neighborhood.

2. **A description of the data and how it will be used to solve the problem.**

Using the Four Square location provider, the Cafe outlets around the NYC will be explored and analyzed. Usage of Folium maps, Clustering Analysis and KDE Heat Density Distribution plots to visualize how close the outlets around the NYC neighbor and provide possible location of interest to the client.

Introduction

New York City

Before to get right with the analysis lets understand about the New York City and its population.is the most populous city in the United States. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles (784 km2), New York is also the most densely populated major city in the United States. Located at the southern tip of the U.S. state of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass. With almost 20 million people in its metropolitan statistical area and approximately 23 million in its combined statistical area, it is one of the world's most populous megacities. New York City has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. Home to the headquarters of the United Nations, New York is also an important center for international diplomacy. **A record 62.8 million tourists visited New York City in 2017.**

S.No	Borough	Population (*2019 estimate*)	Population Density (*person/sq.mi*)
1	Manhattan	1,628,706	71,341
2	Brooklyn	2,559,903	36,147
3	Queens	2,253,858	20,767
4	Bronx	1,418,207	33,867
5	Staten Island	476,143	8,157

From the above data it is clear that the Manhattan borough has the highest Population Density, where it is seen that Manhattan being the most crowded area amongst the borough.

P.S : Studying the population dataset and analysis and visualizing with a choropleth map can be done, but since that is beyond the scope and far cry of the current assignment and problem statement. All the census data regarding the NYC is available in United States Census Bureau Department.

(https://www.census.gov/)

Here we are analyzing and visualizing data predominately with Four Square, that is under interest and judging criteria in this assignment.

PL-P2 NTA: Population Density by Neighborhood Tabulation Area*
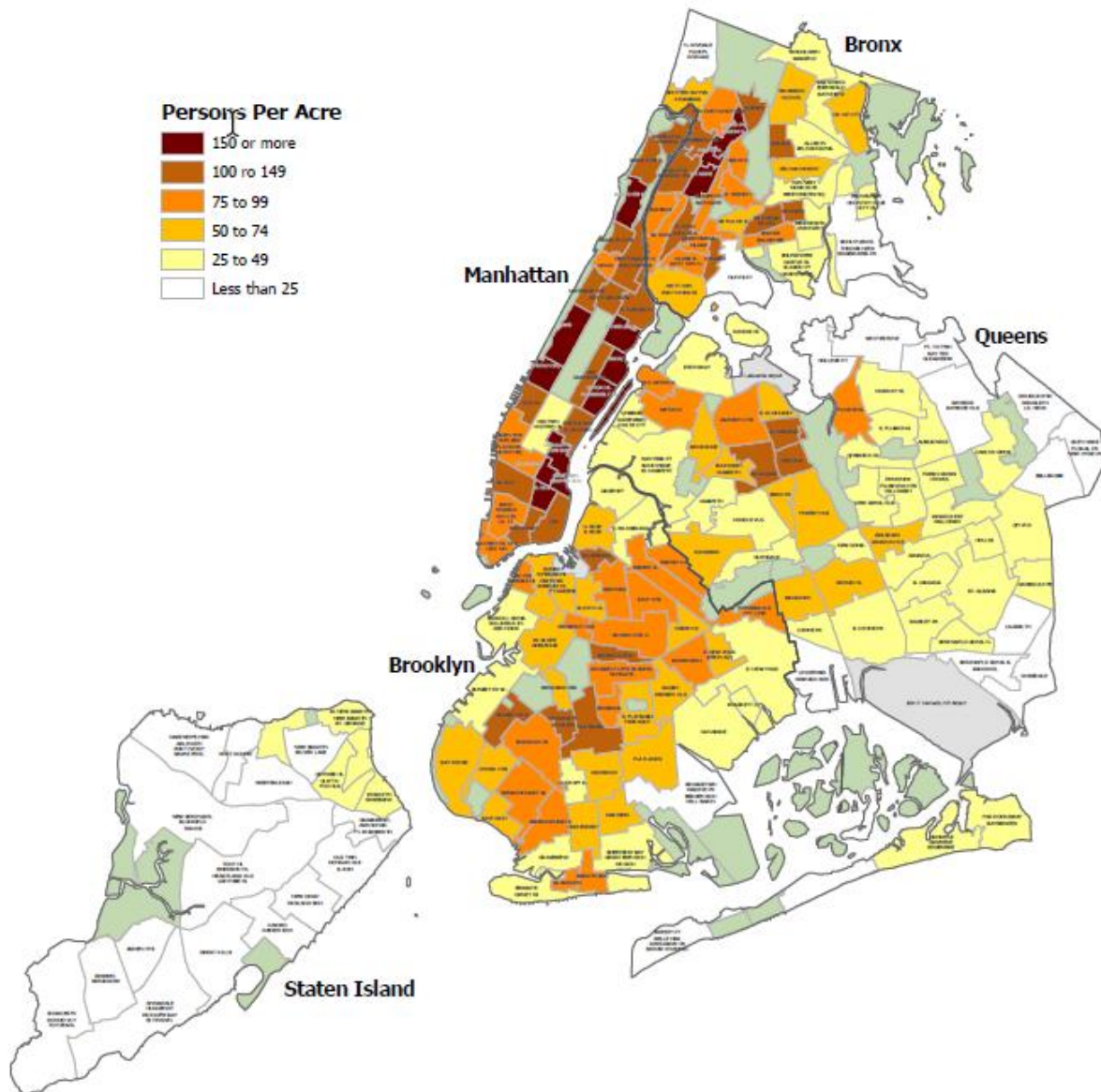## New York City, 2010

Figure1: Population Density of Neighborhood of NYC, 2010. Courtesy of NYC Planning Department.
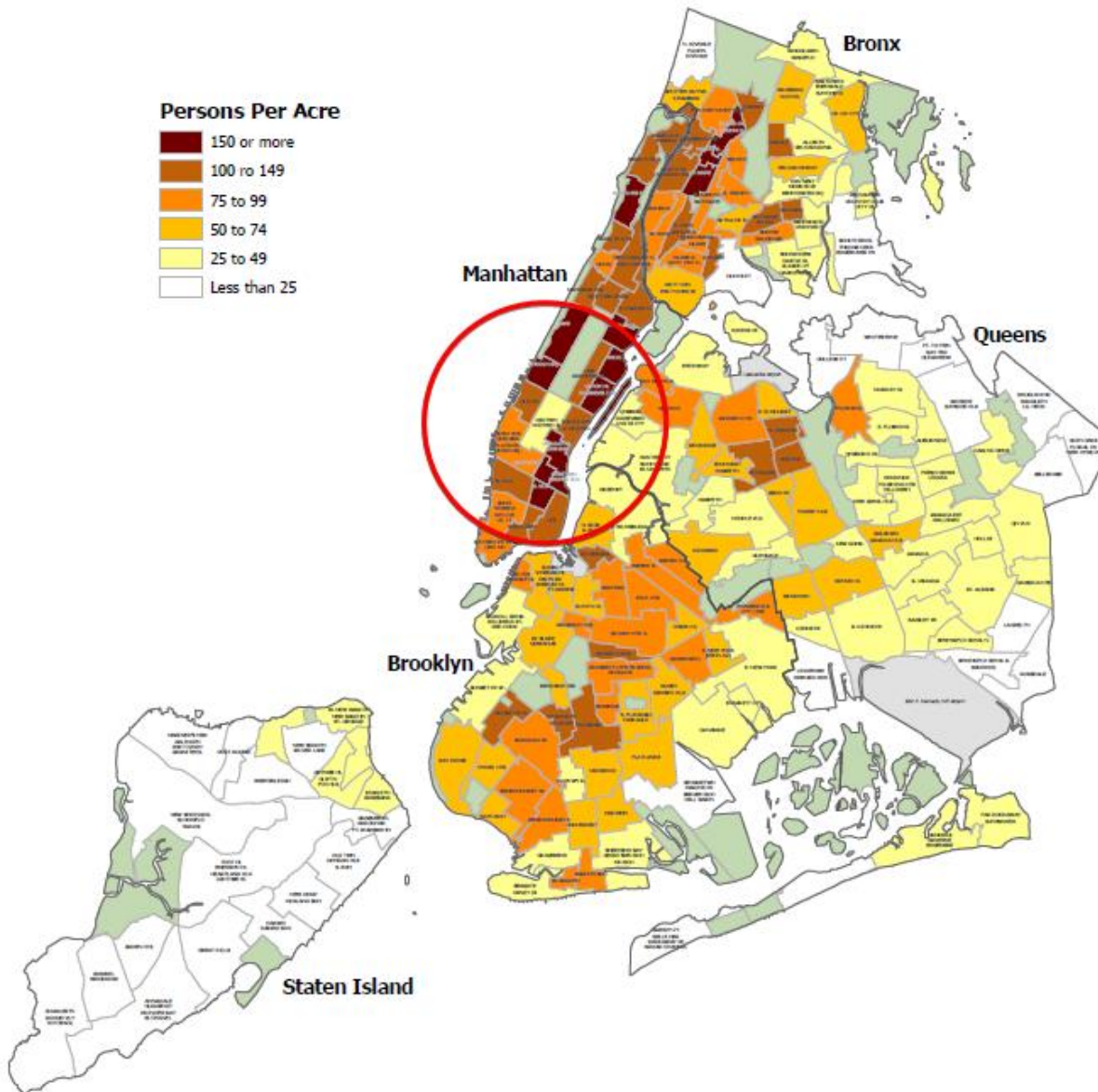
Figure2: Population Density of Neighborhood of NYC, 2010 with Manhattan Borough encircled. Courtesy of NYC Planning Department

Methodology

 Outline Map with borderlines between its Burroughs of New York City visualizes with Geopandas to understand better its geo-location.

NYC is basically looks like an island city with 3 mass of lands surrounded by water and these geo masses are connected by bridges as a means of commute between them. It is very clear as to why Staten Island has the lowest population as it is very far away from other masses
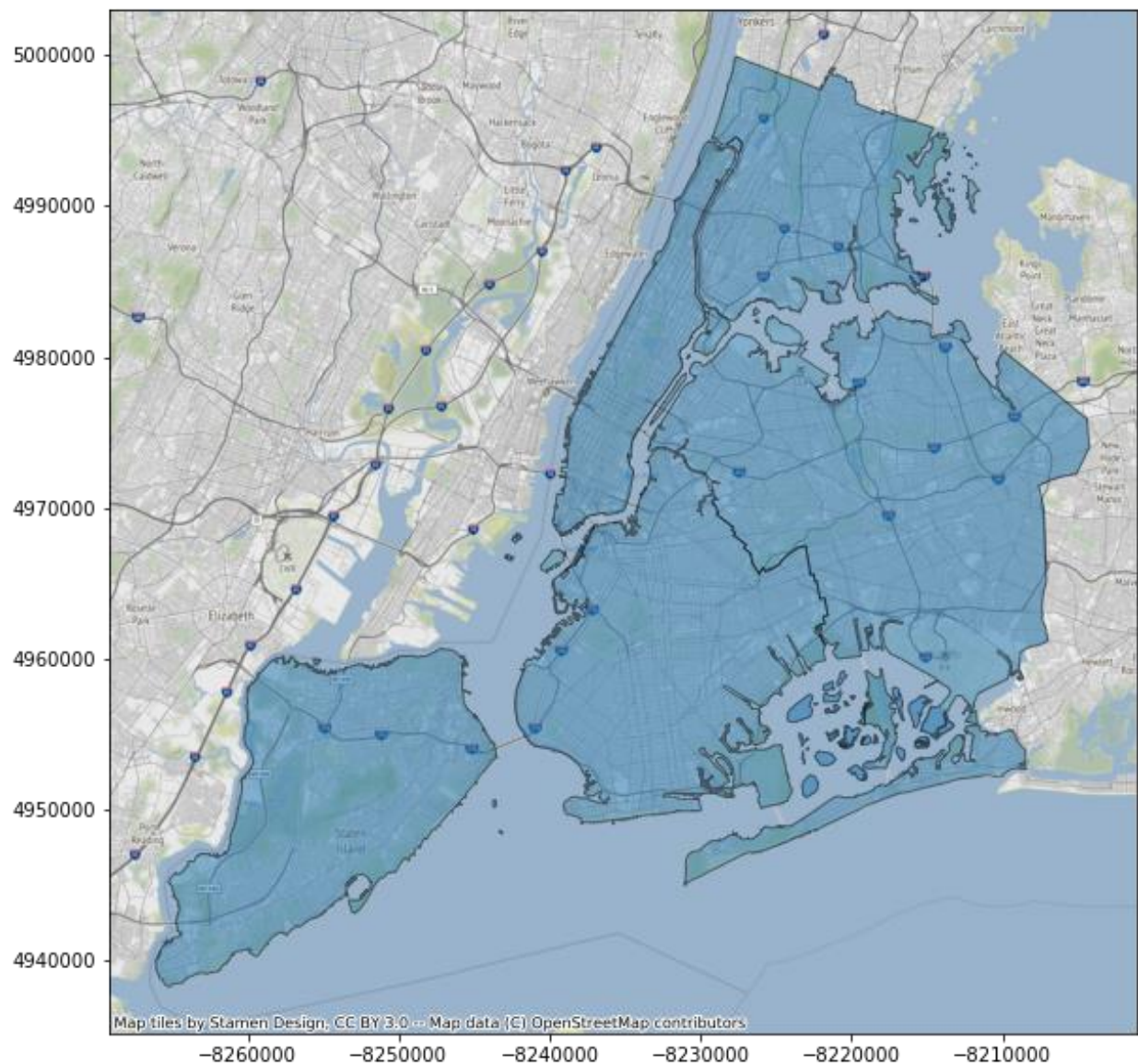

Figure3 : Outline map of New York City

Data Procurement

We need a location Data provider to procure our data and to create a Folium and Clustering Map. One such provider is **FourSquare**

FourSquare Developer Sandbox Account: Current account tier is Sandbox with

- 950 Regular Calls/Day
- 50 Premium Calls/Day
- 1 Photo per Venue
- 1 Tip per Venue
- 10 Limit
- 50 max return results (may impact workability for some cases)

The point to be noted is that the Four Square search results brings **up to max 50 results for a Sandbox account** to do this analysis, but user can always upgrade their account per needs anytime**.** But lets analyze in further stages if this limitation effect our analysis.

Visualizing New York with Folium:

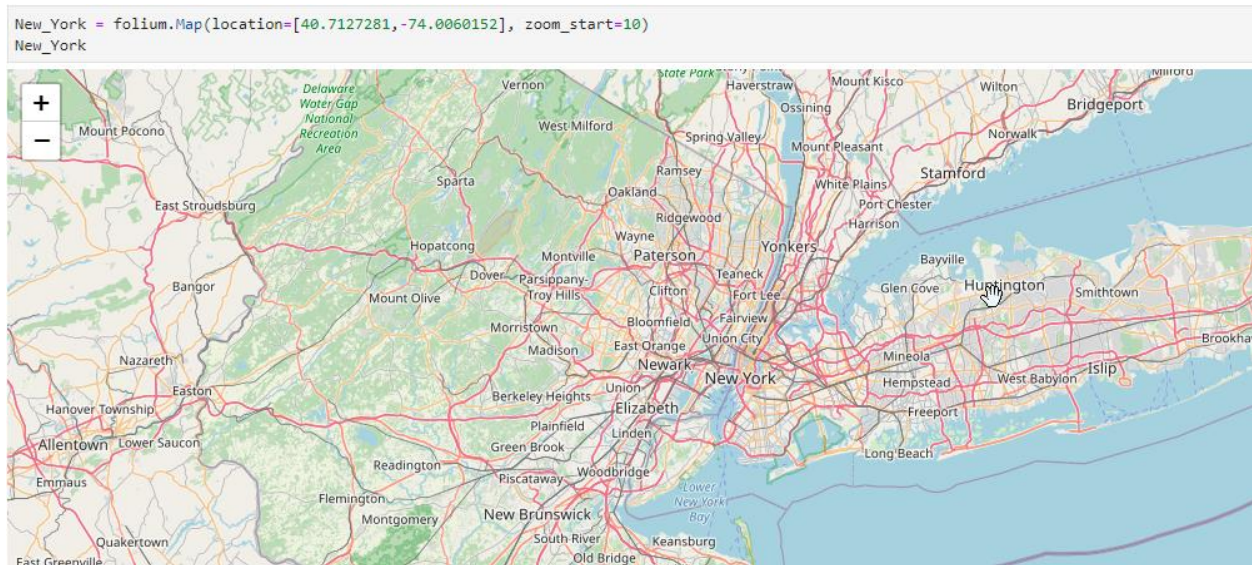Mapping basic NYC mapping with geo-coordinates to start with our basic visualization

Figure 4: Folium Map of New York City

Area of Interest (AOI):

With a starting geo point of **[40.7127281,-74.0060152]**, the search radius of 600 is set as our initial Area of Interest, keeping in mind that Four Square only return max 50 search results.

With Folium a circle can be drawn as an inbuilt function which is pretty good for pointing our area of interest and different shadings can also be done above the analysis which really gives a better visual understanding of the problem.

```
folium.Circle(
    radius=600,
    location=[40.7127281, -74.0060152],
    color='crimson',
    fill_color='yellow',
    fill=True,
    fill_opacity=0.4,
).add_to(m)
```

Figure 5: Plotting a circle in Folium

Let us visualize AOI-1 (also known as Customer's AOI) first to get an understanding of the land mass we are converging in our search query and in further analysis we are proposing our AOI i.e., AOI-2.
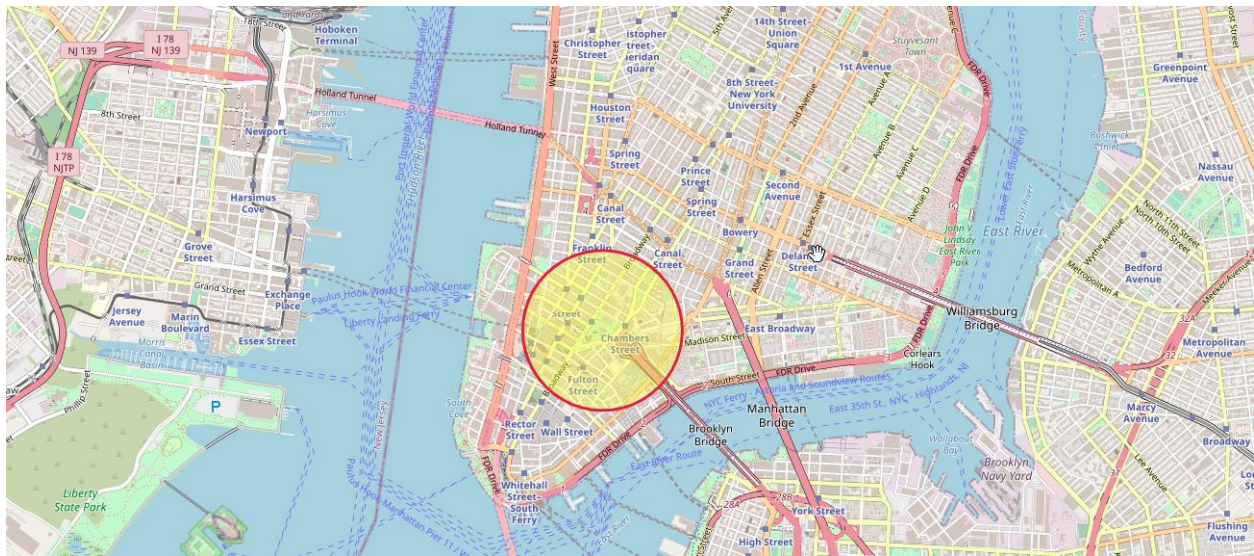

Figure 6: Customer Area of Interest (AOI-1)

Now let's start with the Coffee venues around the geo-coordinates [**40.7127281,-74.0060152**] and see if they fall within the Customer's AOI

```
search_query = 'Coffee'
radius = 600
print(search_query + ' .... OK!')

Coffee .... OK!
```

Figure 7: Search Query around the geo-coordinates

Transforming the search query results from the Json file and creating a data frame, we begin to understand the venues around our geo-point

```python
# keep only columns that include venue name, and anything that is associated with location
filtered_columns = ['name', 'categories'] + [col for col in dataframe.columns if col.startswith('location.')] + ['id']
dataframe_filtered = dataframe.loc[:, filtered_columns]

# function that extracts the category of the venue
def get_category_type(row):
    try:
        categories_list = row['categories']
    except:
        categories_list = row['venue.categories']

    if len(categories_list) == 0:
        return None
    else:
        return categories_list[0]['name']

# filter the category for each row
dataframe_filtered['categories'] = dataframe_filtered.apply(get_category_type, axis=1)

# clean column names by keeping only last term
dataframe_filtered.columns = [column.split('.')[-1] for column in dataframe_filtered.columns]

dataframe_filtered
```

	name	categories	address	crossStreet	lat	lng	labeledLatLngs	distance	postalCode	cc	city	state	country	formattedAddress	neighborhood	id
0	Blue Spoon Coffee Co.	Coffee Shop	76 Chambers St	at Broadway	40.714428	-74.006859	[{'label': 'display', 'lat': 40.71442758460976,...	202	10007	US	New York	NY	United States	[76 Chambers St (at Broadway), New York, NY 10...	NaN	49c79540f964a520af571fe3
1	Proof Coffee Roaster	Coffee Shop	65 Nassau St	NaN	40.709476	-74.008546	[{'label': 'display', 'lat': 40.70947645413928,...	420	10038	US	New York	NY	United States	[65 Nassau St, New York, NY 10038, United States]	NaN	55f05a02498ea2939736f07a
2	For Five Coffee Roasters	Coffee Shop	1 Liberty Plz	Broadway	40.709554	-74.010576	[{'label': 'display', 'lat': 40.709554, 'lng':...	522	10006	US	New York	NY	United States	[1 Liberty Plz (Broadway), New York, NY 10006,...	NaN	5cf965947564f7002c79aa21
3	Coffee Cart	Coffee Shop	60 Centre St	Broadway and chambers	40.714389	-74.002791	[{'label': 'display', 'lat': 40.714389, 'lng':...	328	10007	US	New York	NY	United States	[60 Centre St (Broadway and chambers), New Yor...	NaN	4bffbb33daf9c9b65197faef
4	Hudson/Chambers Coffee Cart	Food Truck	Chambers St	at Hudson St	40.715638	-74.009613	[{'label': 'display', 'lat': 40.71563827489798,...	443	10007	US	New York	NY	United States	[Chambers St (at Hudson St), New York, NY 1000...	NaN	50a38463e4b0d1371722335f

Figure 8: First 5 venues from the search query

Below are our entire list competitors that we will be battling in AOI-1.. So we are working with this limitation, but let's examine if this dataset provides any confidence for further analysis

```
In [17]:  dataframe_filtered.name

Out[17]:  0                   Blue Spoon Coffee Co.
          1                For Five Coffee Roasters
          2                   Proof Coffee Roaster
          3                             Coffee Cart
          4             Hudson/Chambers Coffee Cart
          5          John & William St. Coffee Cart
          6                       Mary's Coffee Shop
          7                      Nick's Coffee Truck
          8          Periscope Coffee On John Street
          9                       Dante Coffee Shop
          10                 Coffee Room at the Lara
          11       Coffee Cart (Centre & Chambers)
          12          Coffee Cuisine of Church St
          13                            Coffee Truck
          14                 Coffee &  Donut Cart
          15                             coffee cart
          16                        sams coffee cart
          17                            Birch Coffee
          18                             Coffee Cart
          19                     Coffee & Croissants
          20                             Coffee Cart
          21                             Coffee Cart
          22                       Pride Coffee Club
          23           Lafayette/Worth Coffee Cart
          24               Jack's Stir Brew Coffee
          25                 Toby's Estate Coffee
          26                            I m a coffee
          27          National Coffee Assn of USA
          28                           Coffee Cento
          29                         Supreme Coffee
          30                            Sam's Coffee
          31                  Blue Spoon Coffee Co.
          32                          Coffee Studio
          33                             R&R Coffee
          34                            Think Coffee
          35                       Nobletree Coffee
          36                         Sawada Coffee
          37                     Blue Bottle Coffee
          38                         Rtisan Coffee
          39             MBJ Coffee Kiosk @ BMCC
          40              R.tisan Coffee Project
          41          Irving Farm Coffee Roasters
          42                               Starbucks
          43                               Starbucks
          44                               Starbucks
          45                       Squire's Diner
          46                               Starbucks
          47                               Starbucks
          48                       Hole In The Wall
          49                               Starbucks
```

Figure 9: List of Competitors

Analysis

Mapping the Coffee venues search results:

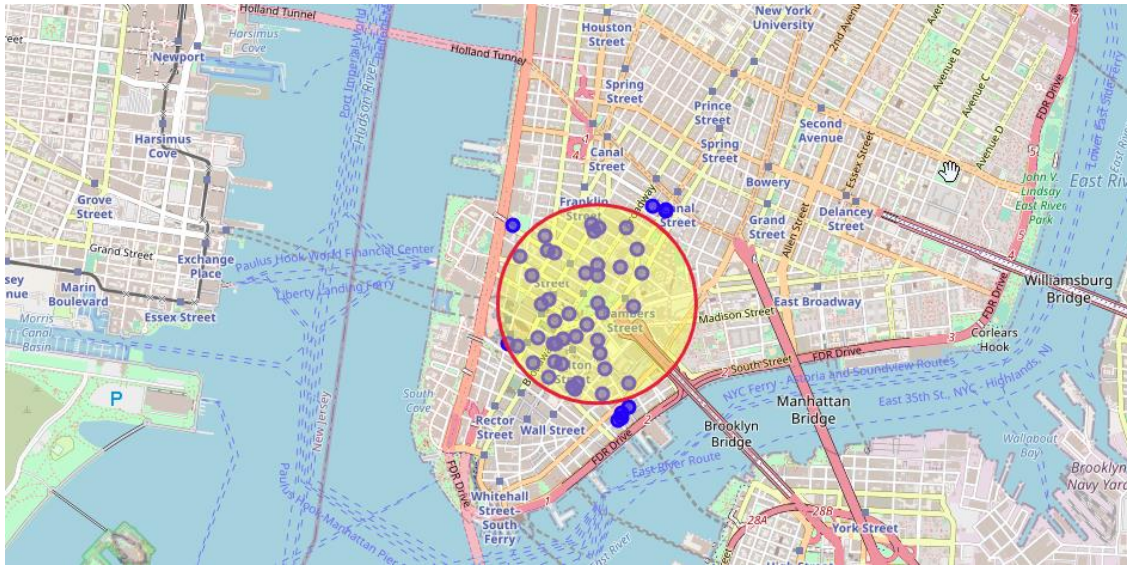Visualizing examining if 50 max search results, is good enough for further analysis

Figure 10: Mapping Coffee venues inside the AOI-1

All the venues fall within the AOI-1 with the radius of 600, with only 7 venues that are outside the radius. With this we can confirm that this **data is good** and is desirable for further analysis!!

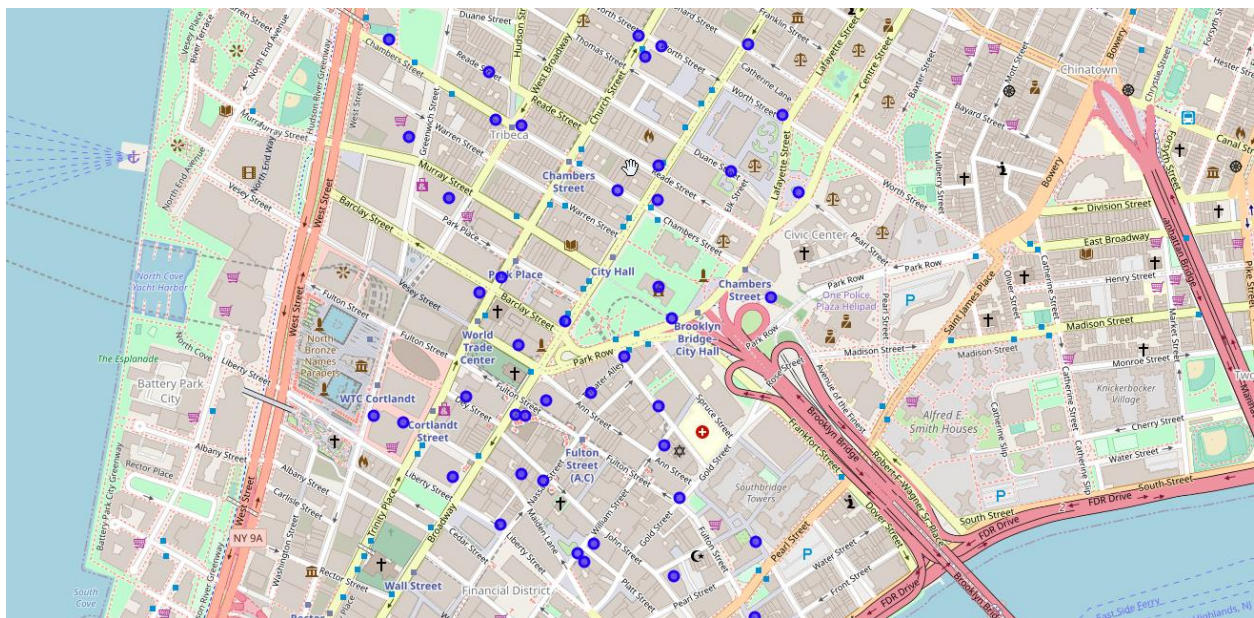Let's explore the venues further by zooming on the map with a level of 16

Figure 11: Coffee venues inside the AOI-1 with zoom level of 16

By visualizing, it's clear that venues are distributed almost evenly with more venues around Fulton and Park Street and there are no venues above the Brooklyn Bridge and venues only appear two blocks from the bridge

Clustering

With clustering we can club close venues together and yes venues are more crowded around Fulton, Park and Chambers street with maximum 15 venues around Fulton Street. On examining Fulton Street have two Churches, Synagogue, Supermarkets and other food restaurants which are the reason for this huge human footprint.

Figure 12: Clustering of Coffee venues

KDE Heat map

Plotting a KDE Heat map which reveals the Coffee venues density distribution around the area with heat signatures.
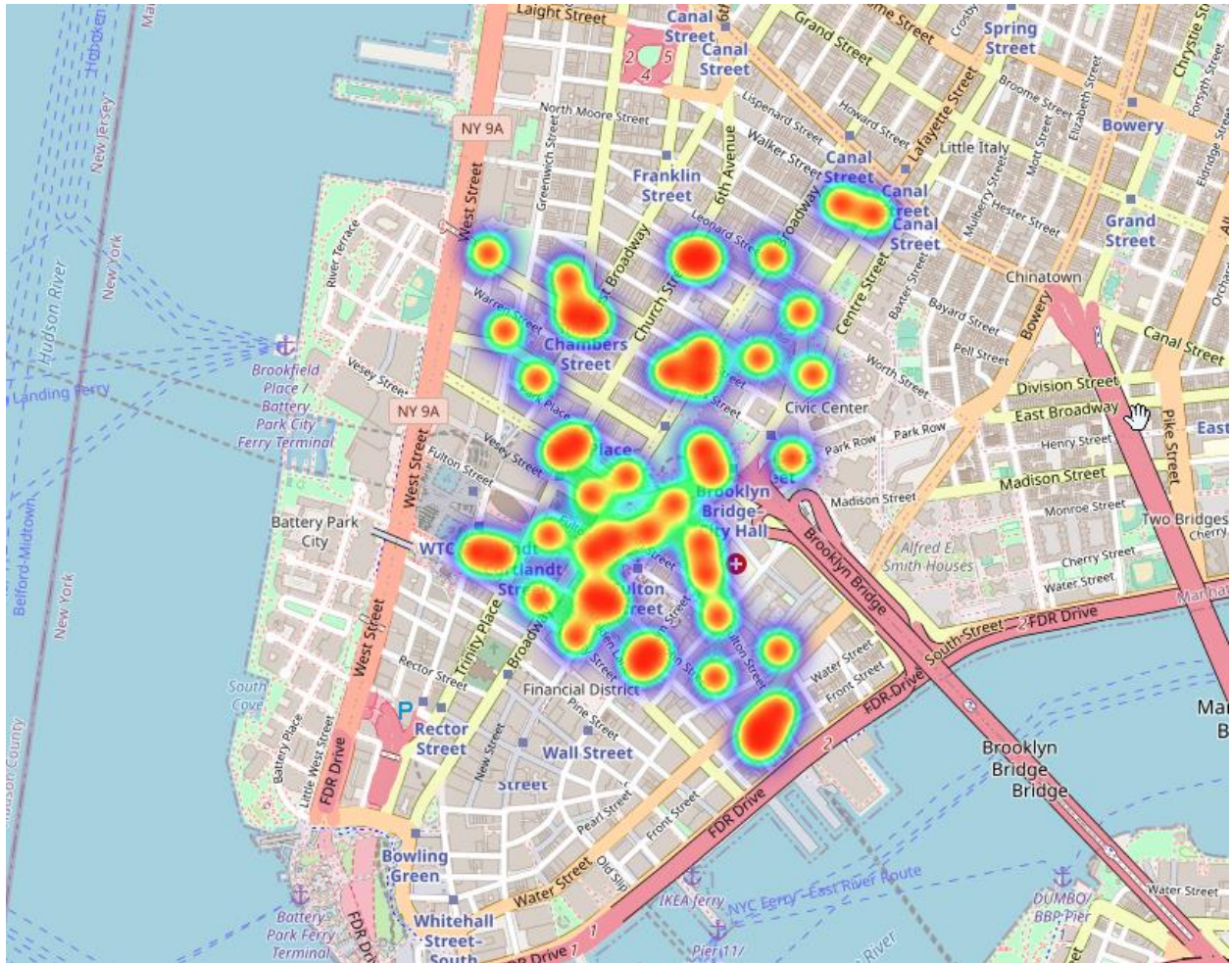

Figure 13: KDE Heat map

Creating a new AOI-2 based on Heat map

Plotting a new **AOI-2** circle which covers some packed Coffee venues based on the maximum density distribution. This new AOI is interesting as within this there are possible places to start a business.
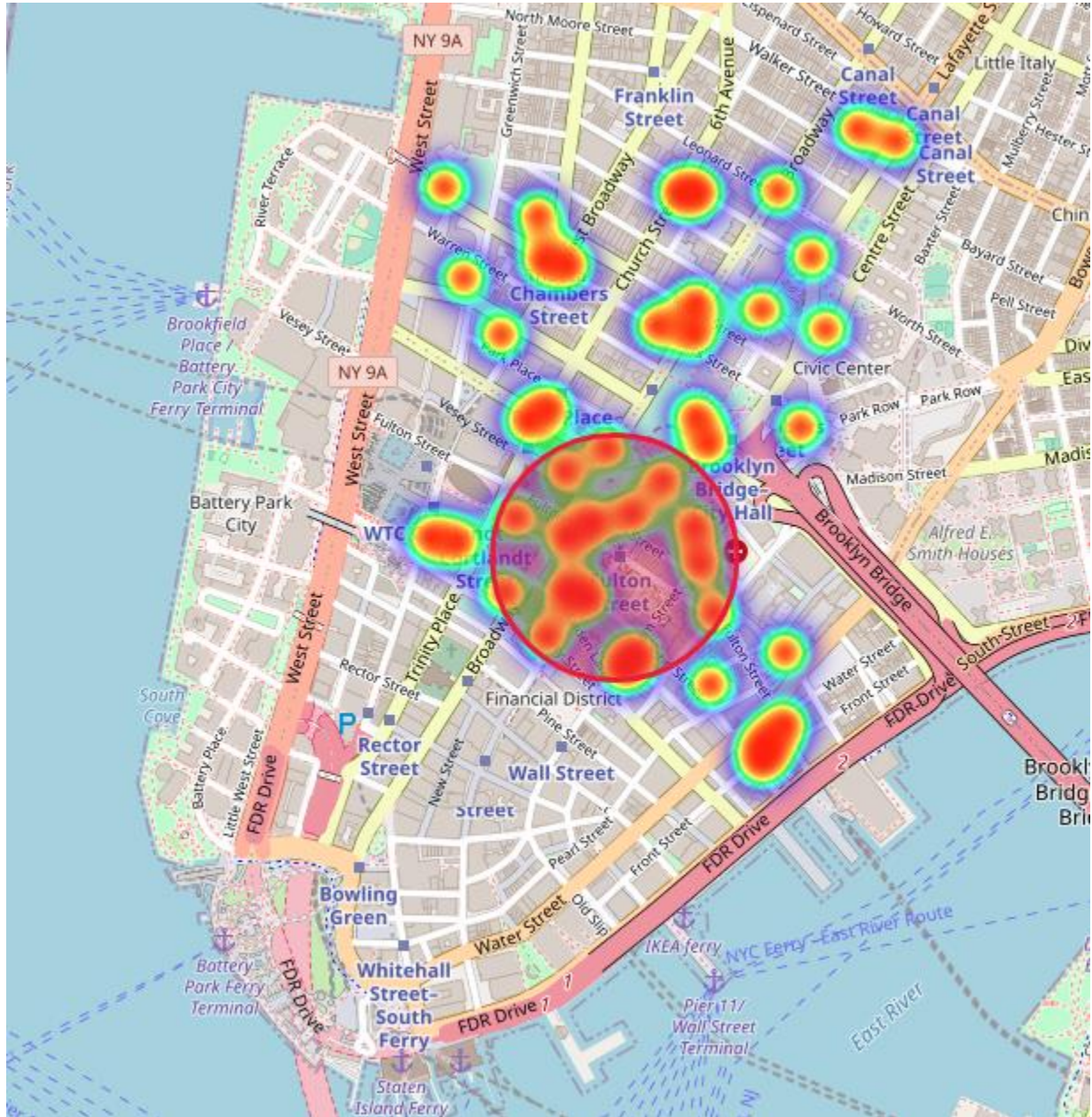

Figure 14: Plotting AOI-2 around dense coffee venues

Results

Superimposing both AOI's (AOI-1 and AOI-2):

Plotting both AOI's we can start to understand area locations with venues distribution, that AOI-2 is better than AOI-1.We have now narrowed down our key locations.
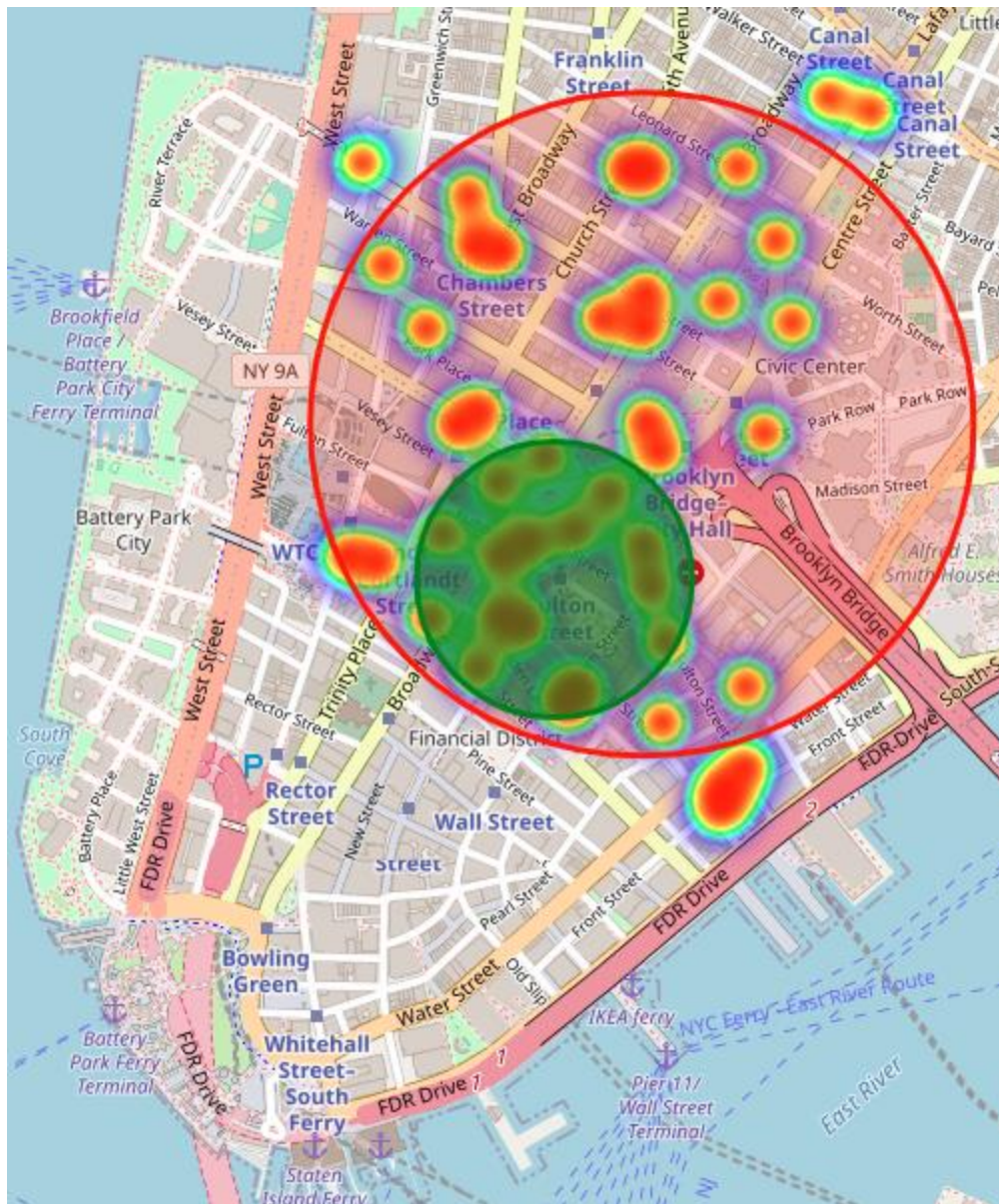

Figure 15: Superimposing AOI-1 and AOI-2

NOTE: In AOI-1 we find many vacant or missing Coffee venues (in fact much less of food restaurants too), but in AOI-2 there are more venues available. There are other significant findings in AOI-2, venues such as burger joints; restaurants etc. were also available in plenty. This reveals that this AOI-2 has **more people footprint happening**, and this finding is **good for business!!**

Discussion

Which venue to choose for the business??

Based on the Four Square Data along with Clustering and Heat Map analysis, some possible locations are suggested. Also keeping in mind locations that should not be chosen for business.

- Those that are **in favour** for business are coloured in **green**. More available data and strong heat signature and larger footprint
- Those that are **not in favour** for business are coloured in **red**. Less available data with weak/no heat signature and no footprint
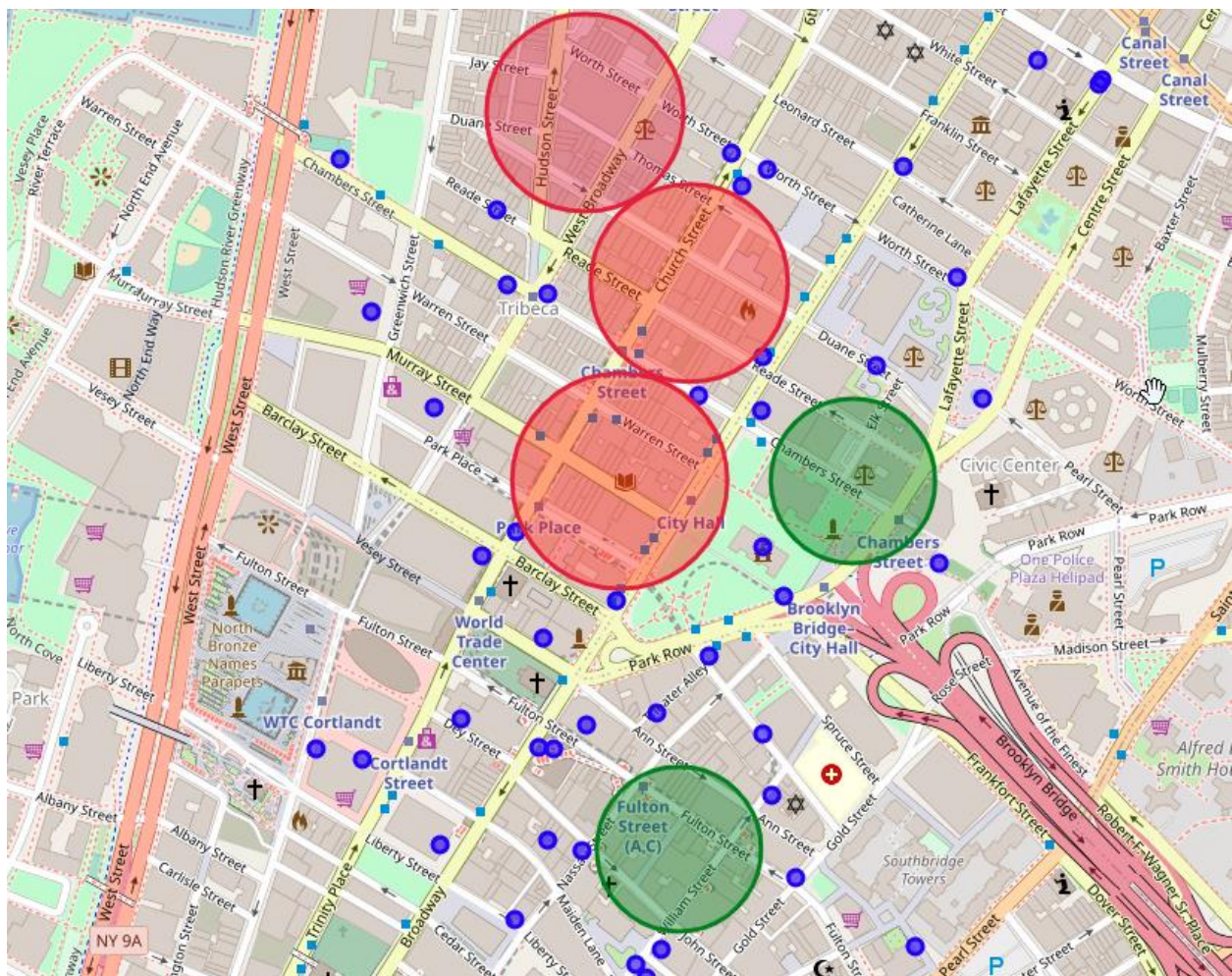

Figure 16: Possible venue locations in favour for business

Conclusion

The client can choose the proposed locations in **green circles** and also venues along the main roads connecting Park Row, Brooklyn and Chambers Street. This analysis excludes parameters such as venue cost, availability and other investment costs.

For Info.

Packages used: [pandas](), [numpy](), [folium](), [geopandas](), [geopy]()

Plugins used: [heatmap]()

Acknowledgements
Thanks to all fellow coders for sharing your knowledge and repository.

Thank you !!