

# **Similar district**

## **Coursera Capstone Project**

by SERGII GUZENKO

20/05/2020

# *Report content*

1. Introduction
2. Data selection
3. Methodology
4. Results
5. Discussion
6. Conclusion

# *1. Introduction*

## Problem summary

- family of three moves from Italy to Manhattan
- they are looking for the most suitable neighbourhood

## Main tasks:

- reduce an impact of the relocation on their life
- find neighbourhoods at Manhattan similar to Italian

# *1. Introduction*

## Interested Audience

Companies/persons in following questions:

- relocation options;
- investment solutions;
- urban problems;

## *2. Data selection*

### Data Requirements

- Geodata for current residence with venues established using Foursquare.
- Manhattan neighbourhoods with venues established via Foursquare
- List of subway metro stations
- List of schools (from junior to high)
- Registered crime number

## *2. Data selection*

### Source of the data

- NYC Open Data – schools, subway station, crime number
- Manhattan merged neighbourhoods from Coursera lab
- GitHub – data repository

### *3. Methodology*

This method helps to visualize a map with all the required data to provide enough information for the final choice. Comparison between the two neighbourhoods is based on its similarity.

#### Used tools:

- Folium – mapping results
- Foursquare – clustering
- Nominatim – coordinate location

### *3. Methodology*

Steps:

1. Find location of Current residence
2. Find venues around
3. Determinate top 10 venues type
4. Download Manhattan clusters from labs
5. Determinate similar cluster in Manhattan based on op 10 venues comparison



### *3. Methodology*

Steps:

6. Upload schools location
7. Upload registered crimes
8. Upload subway stations position
9. Determinate more suitable neighbourhoods
10. Visualize complete map

## *Current residence neighbourhood*

Venues around

Used Nominatim and Foursquare

name	categories
Osteria Antiche Sere	Piedmontese Restaurant
Brasserie de La Mer	French Restaurant
Vale un Perù	Peruvian Restaurant
Bar Torrefazione Ferrucci	Coffee Shop
Piola da Celso	Piedmontese Restaurant
Parco della Tesoriera	Park
Plin & Tajarin	Piedmontese Restaurant
Hamburgeria	Burger Joint
Wasabi	Japanese Restaurant
Teatro Astra	Theater

## Map of venues around

## *Top 10 venues type in current neighbourhood*

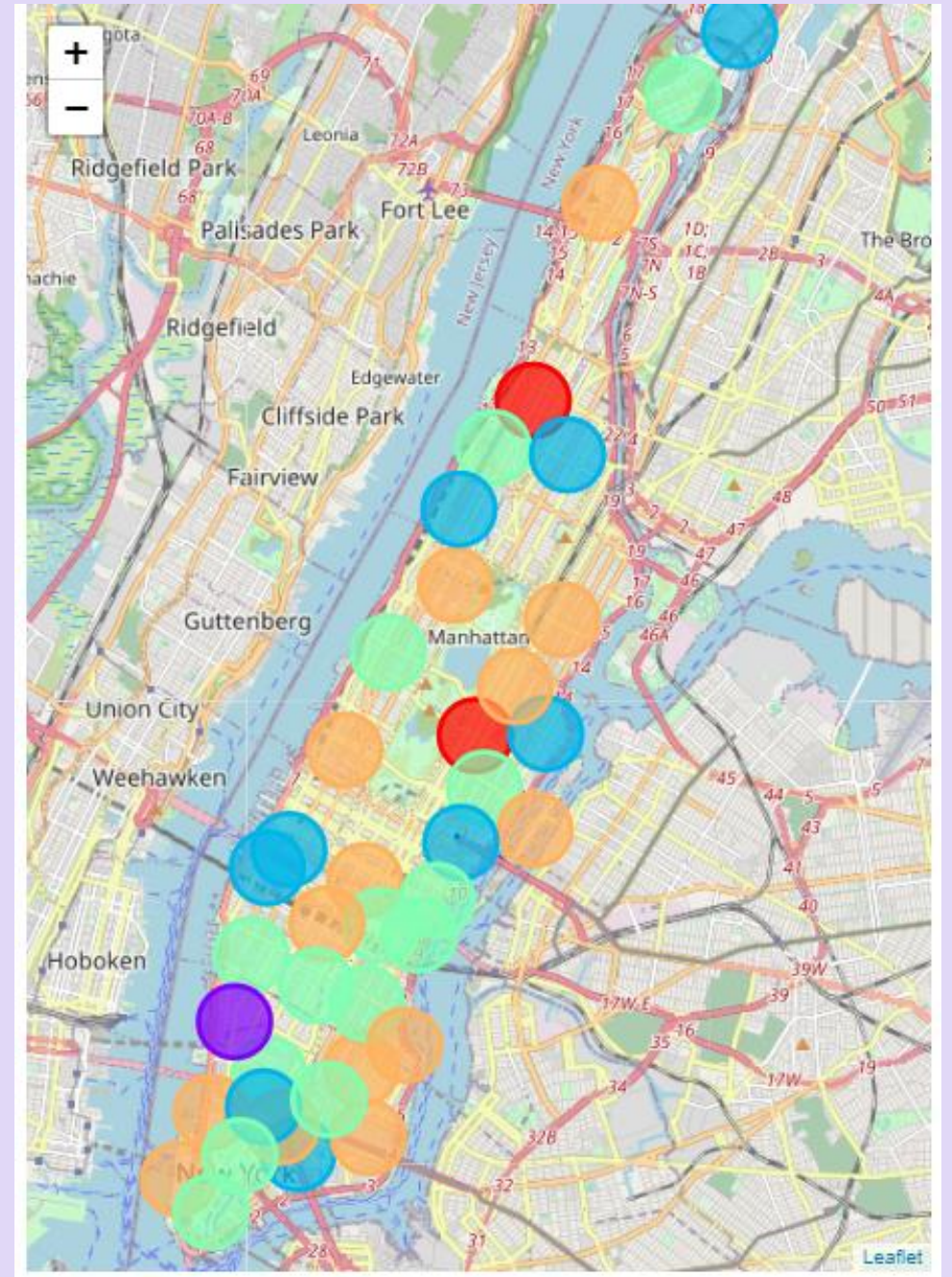
1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Pizza Place	Bus Station	Plaza	Chinese Restaurant	Japanese Restaurant	Piedmontese Restaurant	Italian Restaurant	Burger Joint	Café	Sandwich Place



# Manhattan clustered neighbourhoods

cluster # 5 (**orange**) resembles closer to the Italian place, therefore considered for future analysis.


Used Folium



## *Neighbourhoods in cluster #5*

Neighbourhoods are considered for future analysis.

Used Folium



	Neighborhood	Latitude	Longitude
2	Washington Heights	40.851903	-73.936900
7	East Harlem	40.792249	-73.944182
11	Roosevelt Island	40.762160	-73.949168
13	Lincoln Square	40.773529	-73.985338
15	Midtown	40.754691	-73.981669
19	East Village	40.727847	-73.982226
20	Lower East Side	40.717807	-73.980890
21	Tribeca	40.721522	-74.010683
22	Little Italy	40.719324	-73.997305
25	Manhattan Valley	40.797307	-73.964286
28	Battery Park City	40.711932	-74.016869
30	Carnegie Hill	40.782683	-73.953256
33	Midtown South	40.748510	-73.988713
37	Stuyvesant Town	40.731000	-73.974052

# Top 10 venues type in cluster #5

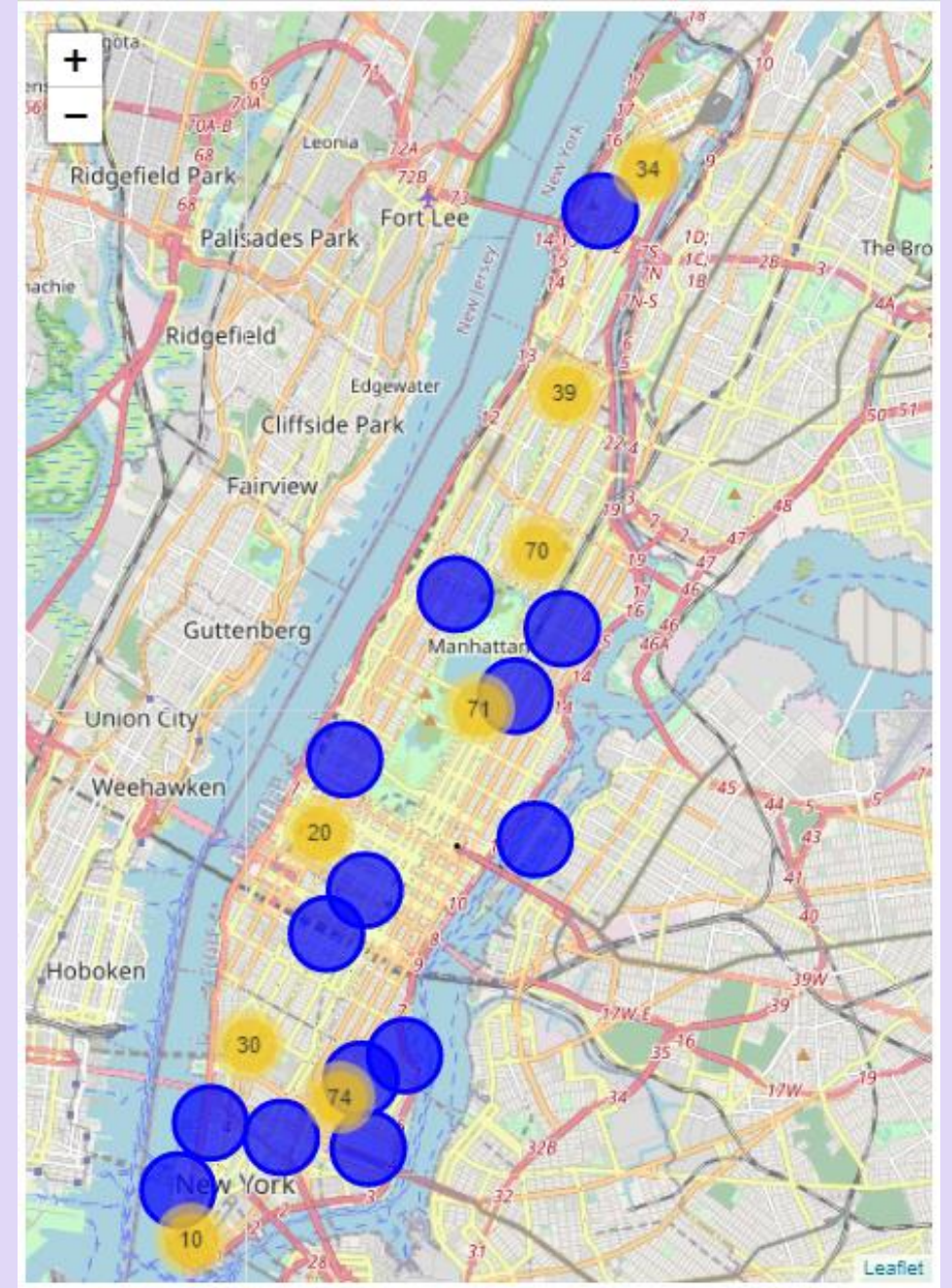
1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Café	Bakery	Mobile Phone Shop	Pizza Place	Sandwich Place	Park	Gym	Latin American Restaurant	Tapas Restaurant	Mexican Restaurant
Mexican Restaurant	Bakery	Latin American Restaurant	Deli / Bodega	Thai Restaurant	French Restaurant	Café	Taco Place	Street Art	Steakhouse
Coffee Shop	Sandwich Place	Park	Japanese Restaurant	Kosher Restaurant	Greek Restaurant	Baseball Field	Gym	Outdoors & Recreation	Dog Run
Theater	Gym / Fitness Center	Concert Hall	Plaza	Italian Restaurant	French Restaurant	Café	Opera House	Indie Movie Theater	Park
Hotel	Theater	Coffee Shop	Steakhouse	Food Truck	Cocktail Bar	Clothing Store	Spa	Bookstore	Sporting Goods Shop
Ice Cream Shop	Bar	Wine Bar	Mexican Restaurant	Cocktail Bar	Pizza Place	Coffee Shop	Chinese Restaurant	Speakeasy	Vegetarian / Vegan Restaurant
Chinese Restaurant	Coffee Shop	Café	Bakery	Latin American Restaurant	Park	Cocktail Bar	Japanese Restaurant	Pizza Place	Ramen Restaurant
American Restaurant	Italian Restaurant	Park	Spa	Café	Boutique	Wine Bar	Coffee Shop	Greek Restaurant	Gym



## *Manhattan schools and neighbourhoods of cluster #5*

7 of 14 neighbourhoods have a great schools choice, therefore considered for future analysis.

Used Folium





## *Sorted Neighbourhoods in cluster #5*

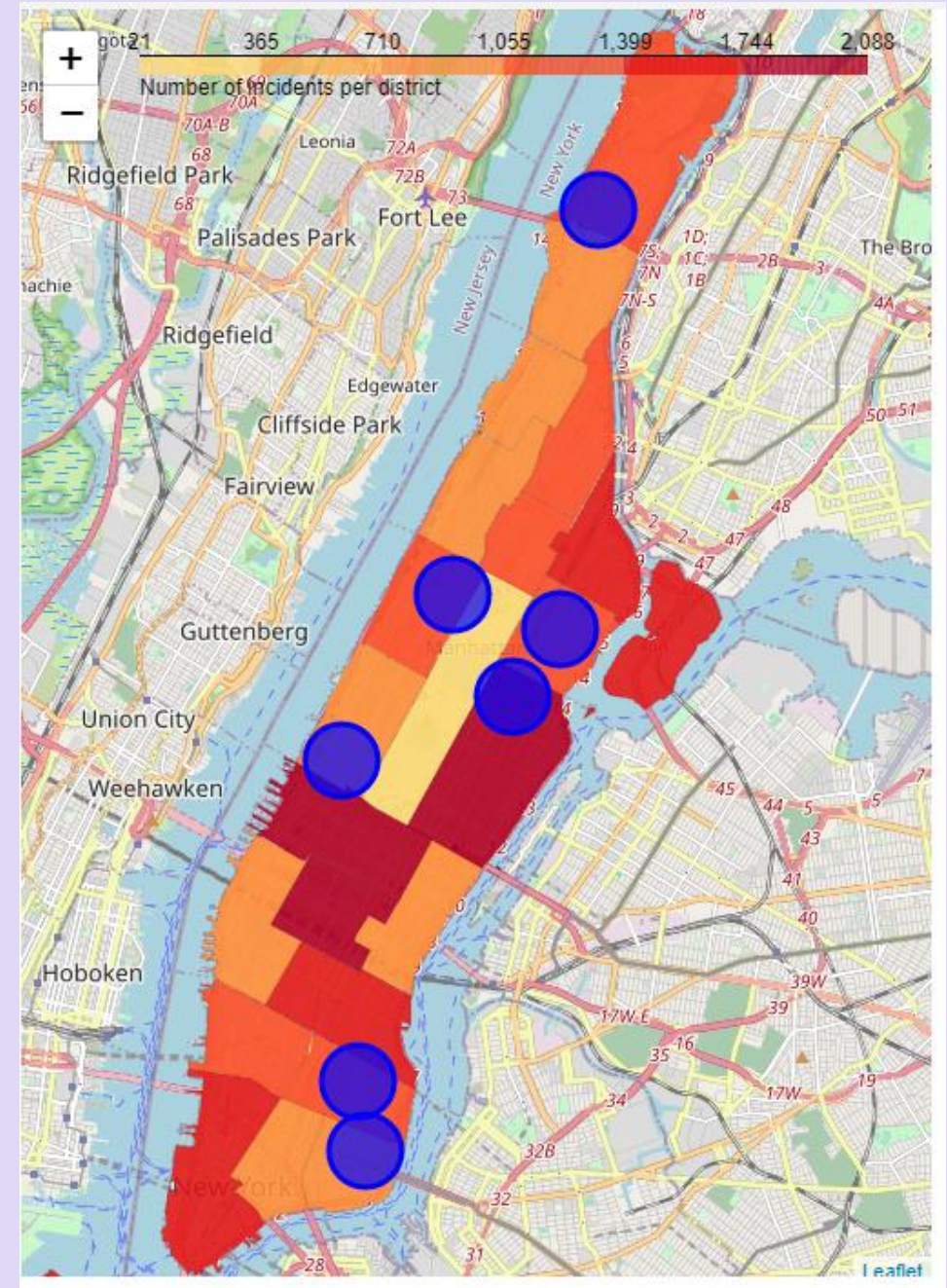
7 neighbourhoods with a great schools choice, therefore considered for future analysis.

	Neighborhood	Latitude	Longitude
20	Lower East Side	40.717807	-73.980890
19	East Village	40.727847	-73.982226
2	Washington Heights	40.851903	-73.936900
7	East Harlem	40.792249	-73.944182
25	Manhattan Valley	40.797307	-73.964286
30	Carnegie Hill	40.782683	-73.953256
13	Lincoln Square	40.773529	-73.985338

# *Manhattan registered crimes and sorted neighbourhoods of cluster #5*

4 of 7 neighbourhoods with the lowest crime numbers passing to next analysis.

Used Folium Choropleth map



## *Sorted Neighbourhoods in cluster #5*

4 neighbourhoods with the lowest crime numbers passing to next analysis.

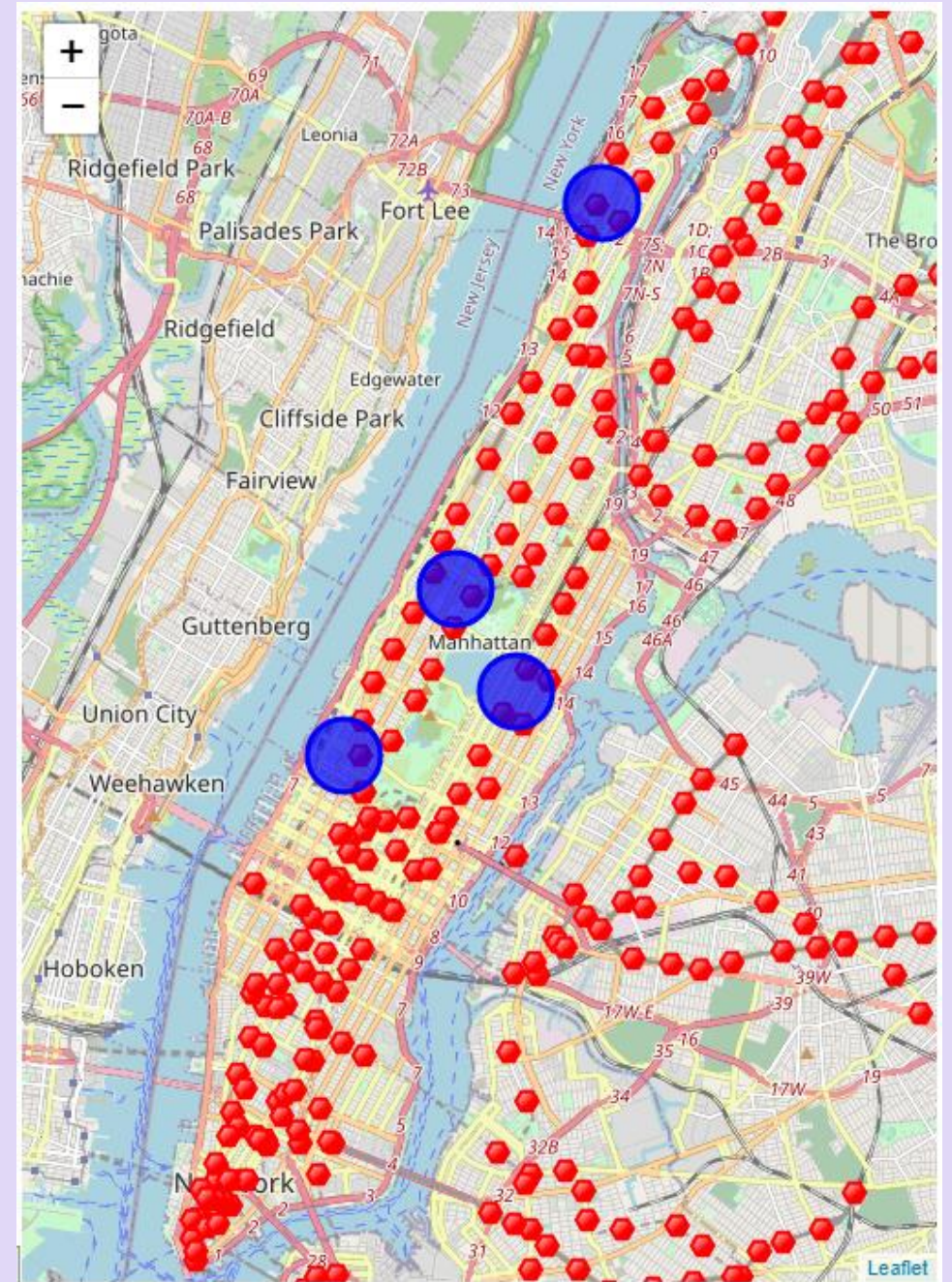
	Neighborhood	Latitude	Longitude
<b>2</b>	Washington Heights	40.851903	-73.936900
<b>25</b>	Manhattan Valley	40.797307	-73.964286
<b>30</b>	Carnegie Hill	40.782683	-73.953256
<b>13</b>	Lincoln Square	40.773529	-73.985338



# *Manhattan subway stations and sorted neighbourhoods of cluster #5*

4 neighbourhoods has a nice connection through subway, so will be considered for final analysis.

Used Folium

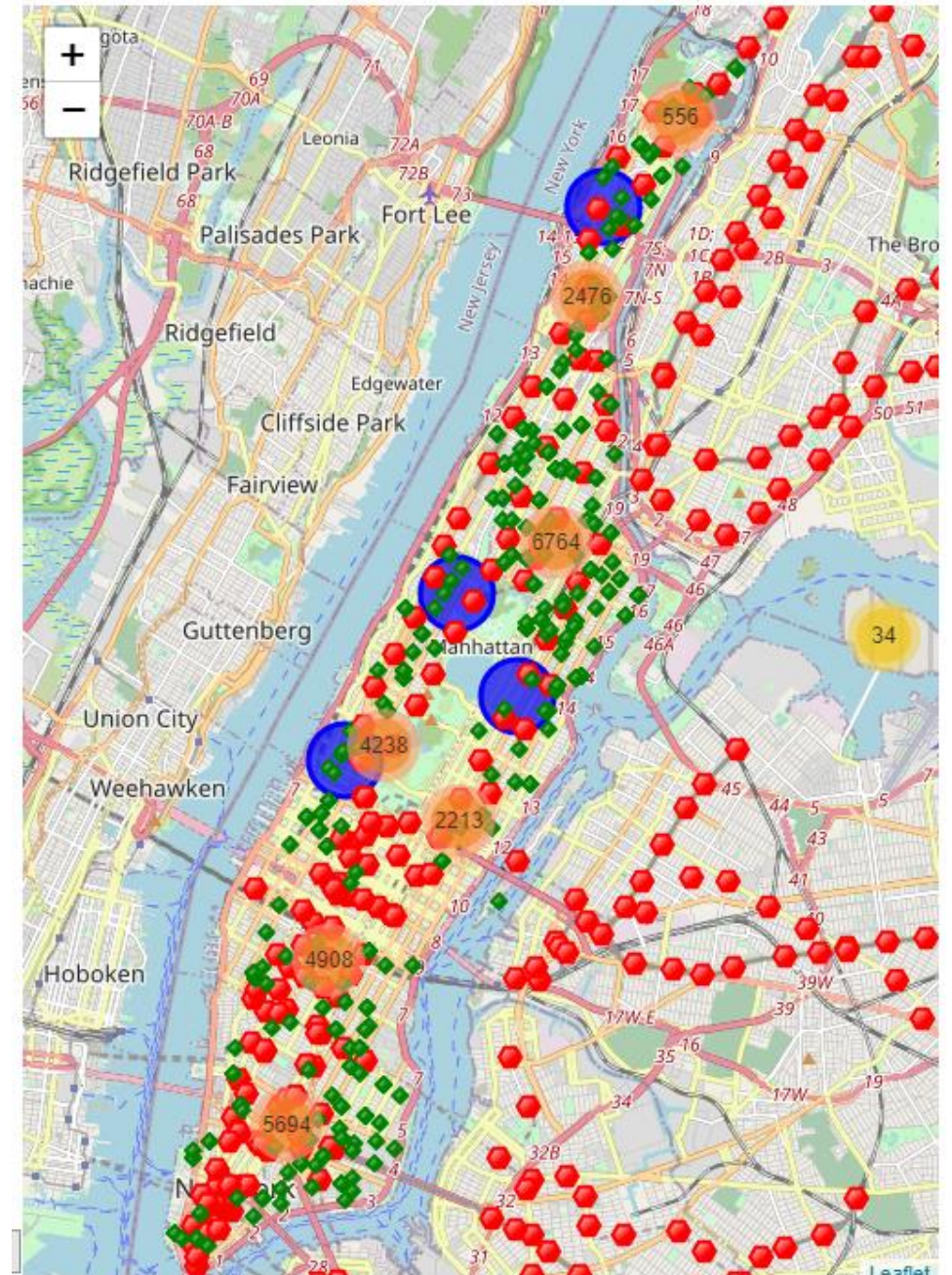




## 4. Results

Based on venues similar to Italian neighbourhood and key requirements we determined 4 more suitable districts, listed below:

- **Washington Heights**
- **Manhattan Valley**
- **Carnegie Hill**
- **Lincoln Square**



## *5. Discussion*

- Folium is a great tool for result presentation
  - various techniques of data visualization help understand better data set and obtained results
  - some further data analysis of the optimal neighbourhood may be required
- ✓ For future analysis house or rent price may be included

## 6. Conclusion

It is possible to compare various districts/cities to find answers on demand questions. In this work we use FourSquare and Folium to compare Italian city and Manhattan on short list of requirements.

During this project I applied several methodologies used during the course, such as data wrangling with pandas, basic data visualization, machine learning techniques.

For future projects, it could be considered to increase data amount available and applying other clustering algorithms such as DBSCAN.