Final Assignment -Week 2

Segmenting Starbucks in New York City

1. Description of the problem

- I am looking for a new apartment in New York and my problem is that I am addicted to coffee…
- So I would like to find a suitable location where I can find the best coffee.
- My favorite coffee shop is **Starbucks**

- So **I would like to find and apartment that is very close to a Starbucks coffee** using a data science approach.

2. Data acquisition and cleaning

- It will be utilized the dataset from New York available in the following link.

https://geo.nyu.edu/catalog/nyu_2451_34572

- Once the dataset from New York is extracted and processed,
- The Starbucks coffee shops of New York will be displayed in a map and
- It will be displayed the location of the apartment to the nearest Starbucks coffee shop.

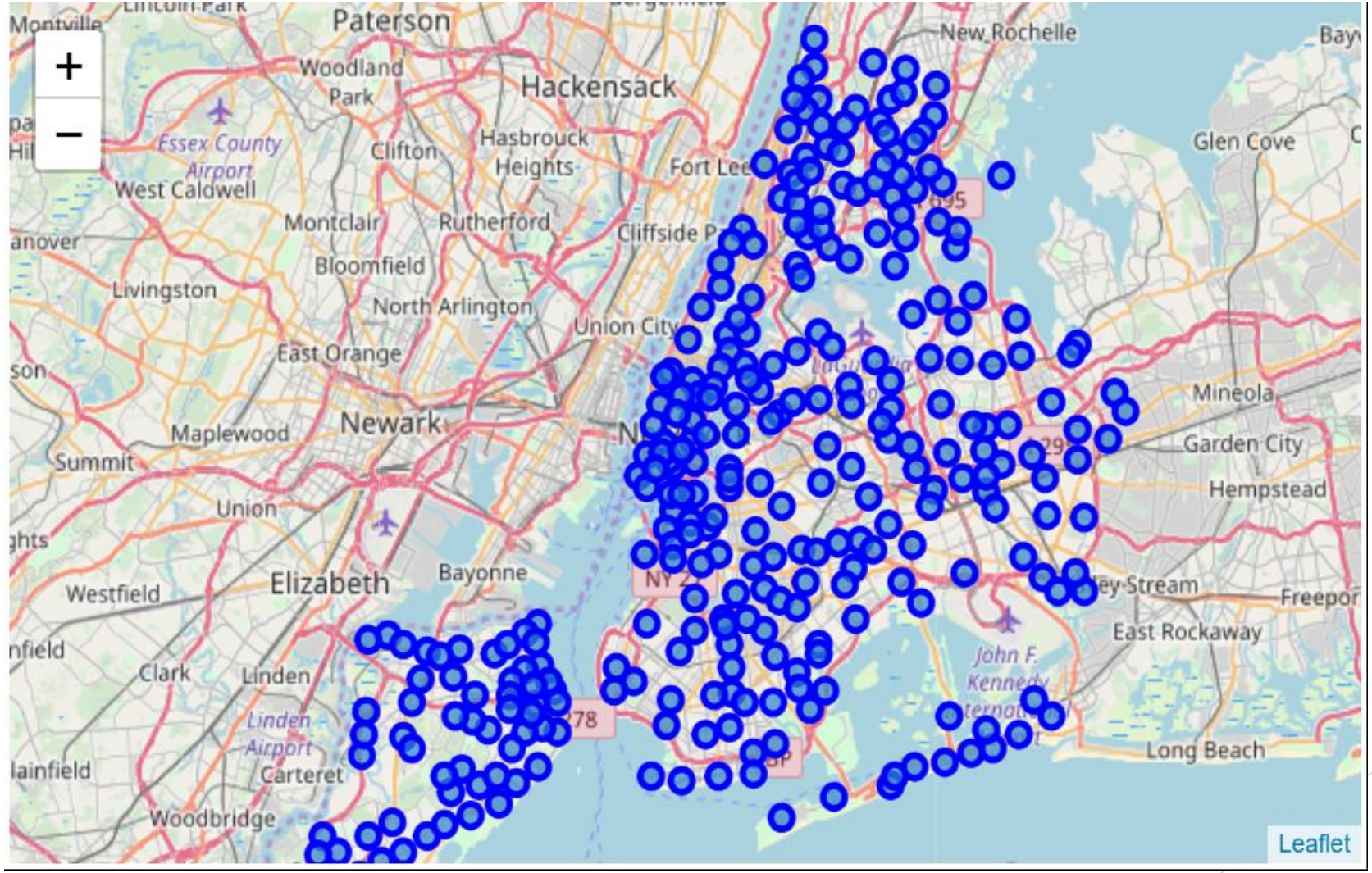
3. Methodology

- Initially the dataset will be shown in a map in order to confirm that the information is correct showing the location of all the neighborhoods of New York on a map.

- We will use the Foursquare API to explore neighborhoods in New York City and extract the relevant information about Starbucks coffee shops in each of them.

- We will get the most common venue categories in each neighborhood
- The we will filter the “Starbucks” venues”.
- Finally, the location of the apartments will be shown with the location of the Starbucks to identify the best location.

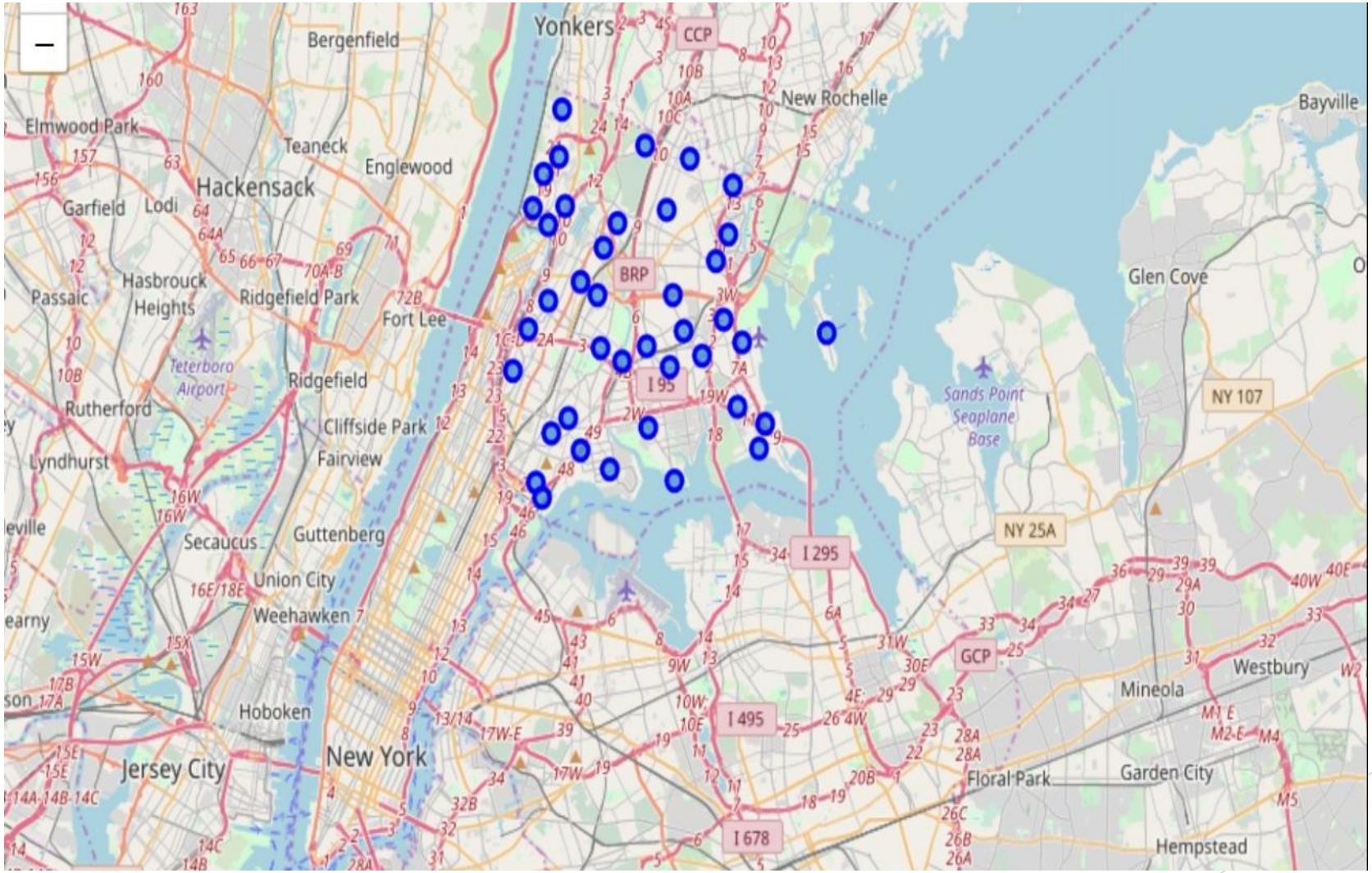
3.1 Map of New York with neighborhoods



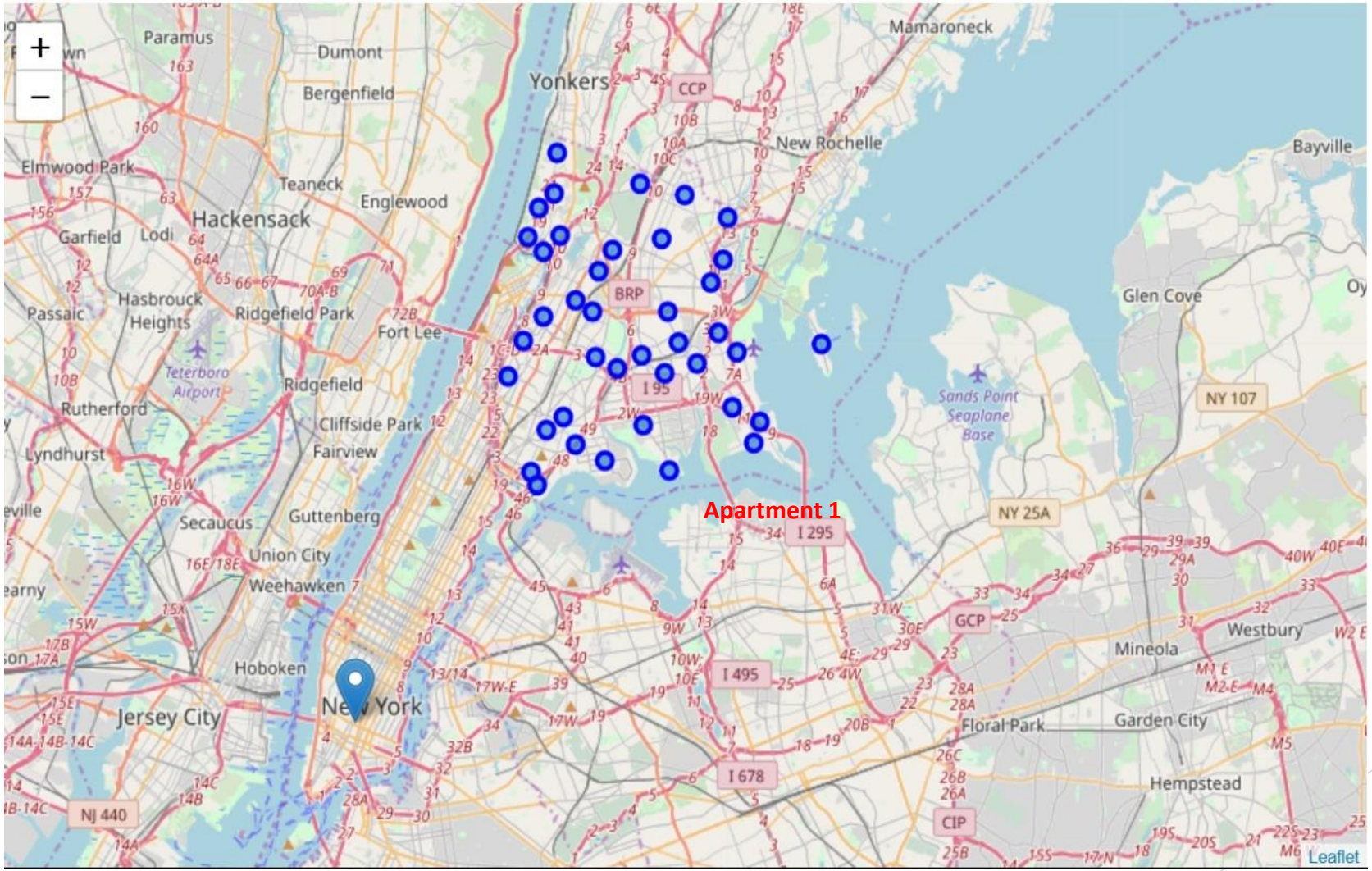
3.2 Dataset with all the Starbucks

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
131	Marble Hill	40.876551	-73.910660	Starbucks	40.877531	-73.905582	Coffee Shop
135	Marble Hill	40.876551	-73.910660	Starbucks	40.873755	-73.908613	Coffee Shop
832	Belmont	40.857277	-73.888452	Starbucks	40.860636	-73.890270	Coffee Shop
1382	Brighton Beach	40.576825	-73.965094	Starbucks	40.577841	-73.961204	Coffee Shop
1896	Brooklyn Heights	40.695864	-73.993782	Starbucks	40.692469	-73.990971	Coffee Shop
2478	Bath Beach	40.599519	-73.998752	Starbucks	40.595227	-74.000017	Coffee Shop
3048	Georgetown	40.623845	-73.916075	Starbucks	40.625874	-73.917460	Coffee Shop
3547	Washington Heights	40.851903	-73.936900	Starbucks	40.850961	-73.938330	Coffee Shop
3937	Upper East Side	40.775639	-73.960508	Starbucks	40.773533	-73.959810	Coffee Shop
4045	Yorkville	40.775930	-73.947118	Starbucks	40.772356	-73.949984	Coffee Shop
4158	Roosevelt Island	40.762160	-73.949168	Starbucks	40.764980	-73.953124	Coffee Shop
4160	Roosevelt Island	40.762160	-73.949168	Starbucks	40.759360	-73.953153	Coffee Shop
4371	Lincoln Square	40.773529	-73.985338	Starbucks	40.771392	-73.982424	Coffee Shop
5442	Manhattan Valley	40.797207	-73.964286	Starbucks	40.795260	-73.965590	Coffee Shop

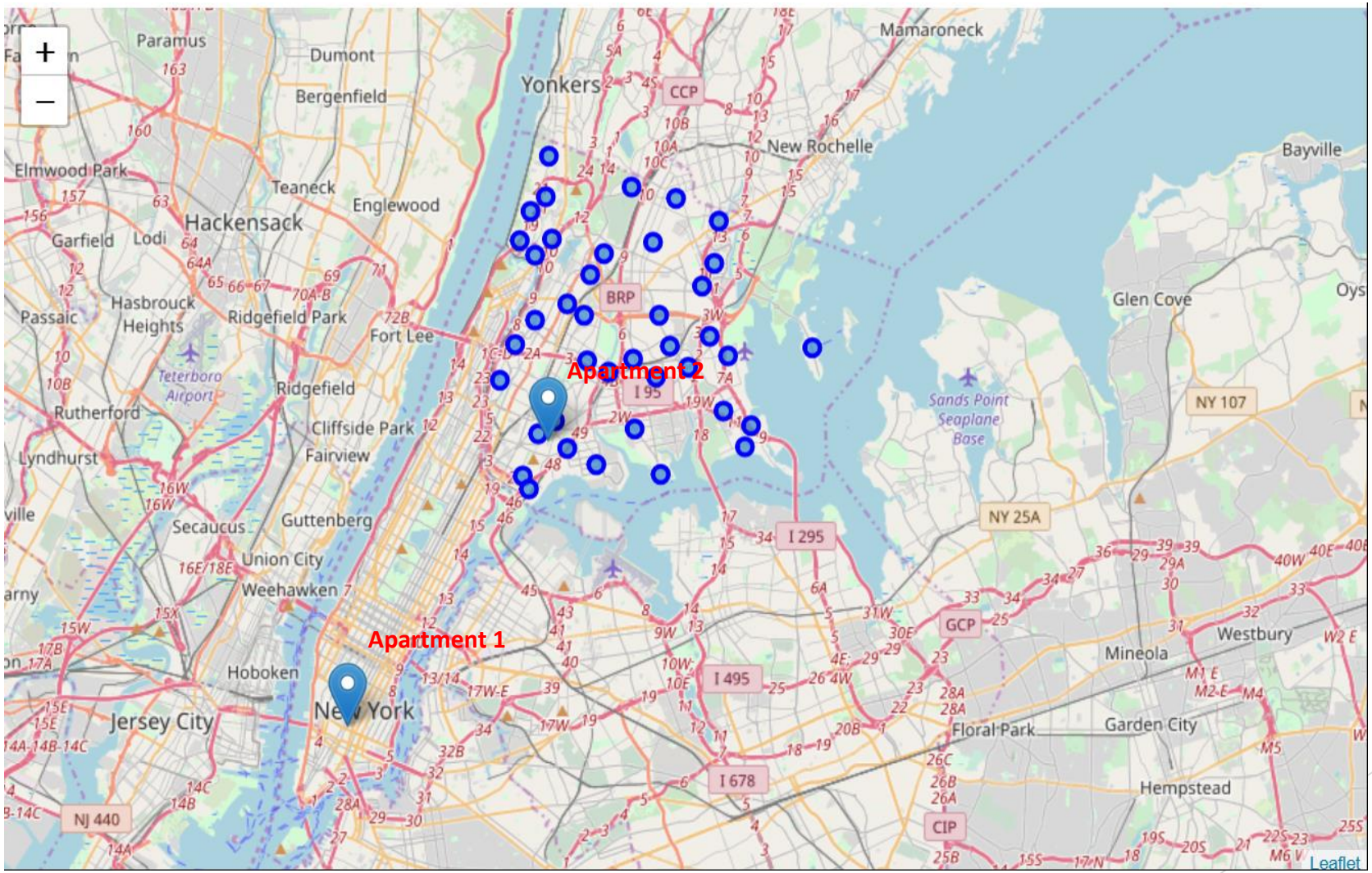
3.3 Map with all Starbucks coffees



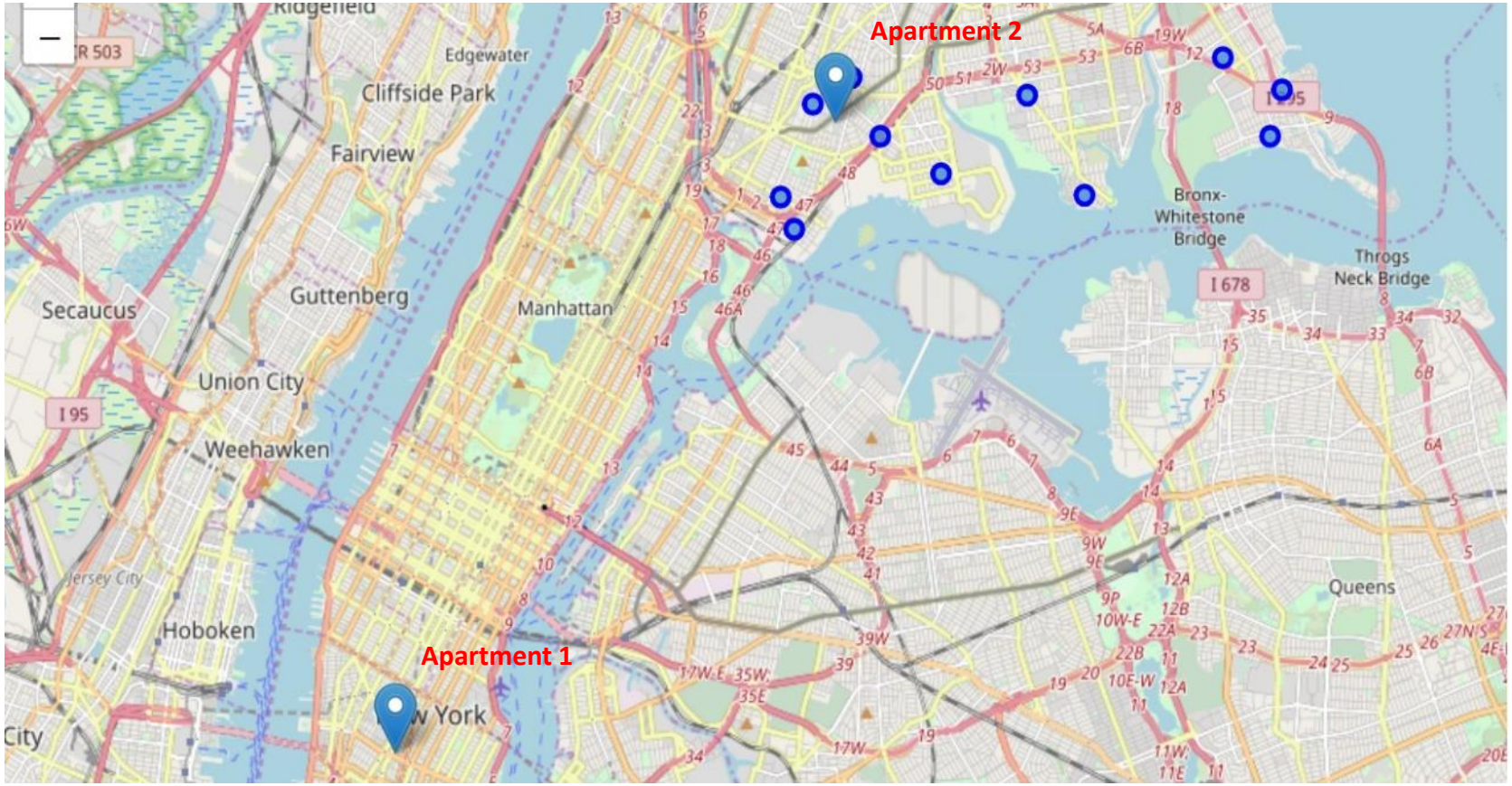
3.4 Location of our apartment 1



3.4 Location of our apartment 2



4. Results



- Looking to the map it is clear that the **Apartment 2** has much more Starbucks near it than **Apartment 1.**

5. Discussion

- The results are displayed showing the location of our apartment within the others places that we are interested, in this case, Starbucks coffee shop.
- The map provides a well first impression whether the selected location is adequate taking into account our initial premises.
- These initial requirements can be extended including any other type of venue that we may be interested in having close.

6. Conclusions

- This project has used two main applications of data science as it the management and visualization of big amount of data.
- In this case it has been managed all the venues in New York in order to find the best suitable location for our apartment.

- As a continuation of this project it could be extended with more automatic means to show the real distance between our location and the group of places that we are interested and even clustering these places in order to identify the neighborhoods more populated with this kind of venue.