Food Delivery in Manhattan

Selecting the appropriate location of a food-delivery office

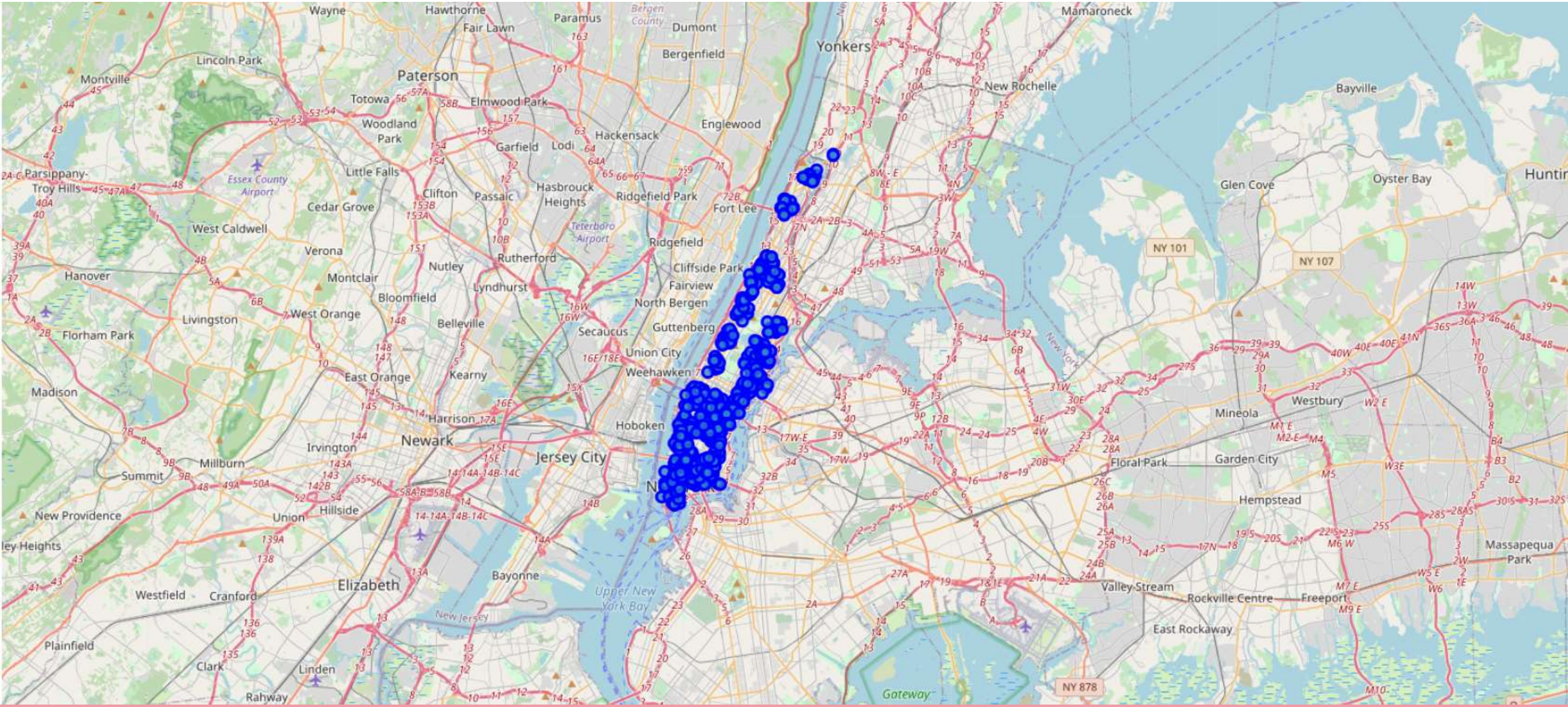
- Higher the number of restaurants in the proximity of the office, higher the revenues.
- Closer the restaurants to the office, lower the shipment costs.
- Higher the rating of the restaurants, higher the number of orders from potential customers.
- Higher rated restaurants have loyal customers, granting durable business relationships between venues and the food-delivery office.

Data acquistion and cleaning:

- Json file of all New York neighborhoods: https://cocl.us/new_york_dataset
- API calls to FourSquare

From the data of New York, we selected the ones of Manhattan and created a database. Through API calls to FourSquare, data concerning all the venues in Manhattan were retrieved; then only restaurants were selected. A second cycle of API calls is used to collect the ratings of restaurants.

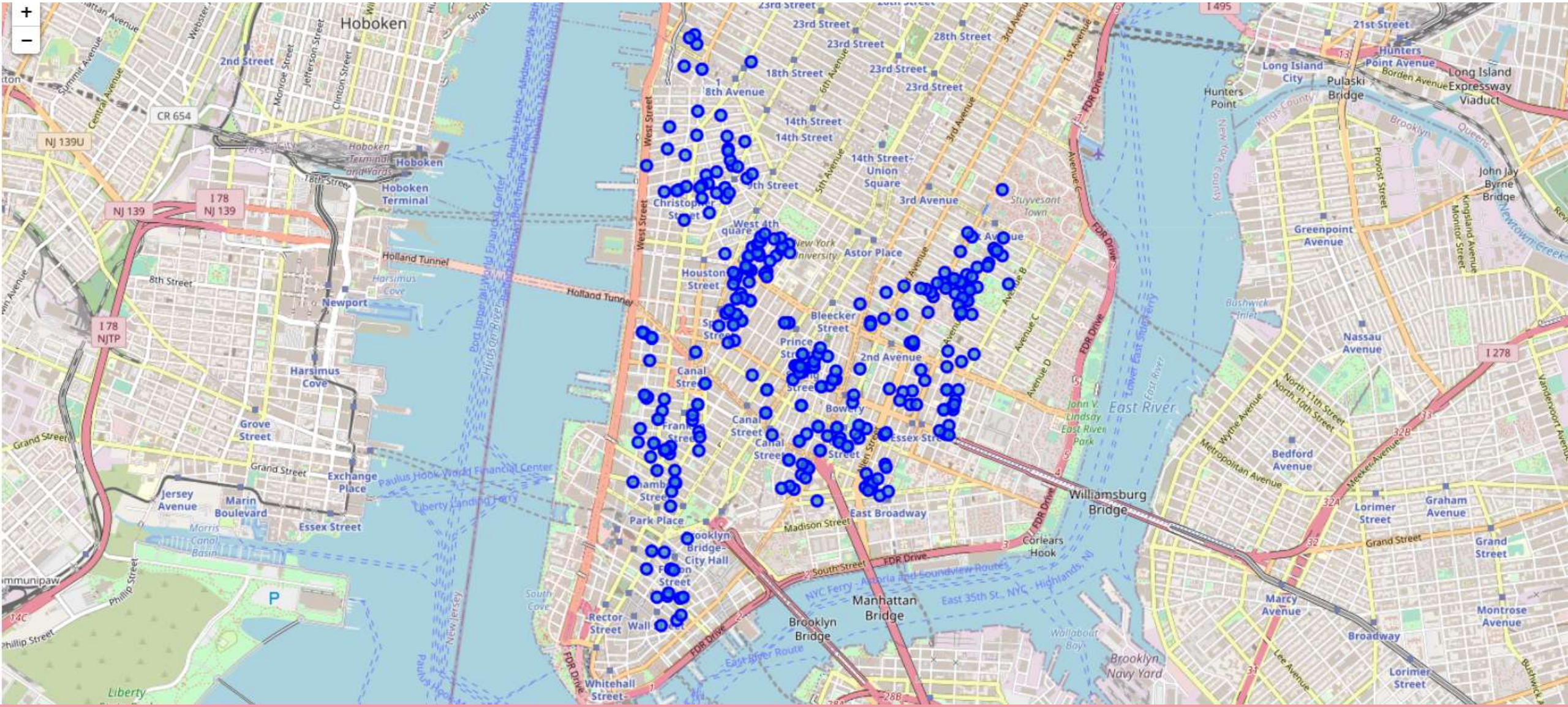
All restaurants in Manhattan:



Machine learning technique:

Clustering (an unsupervised machine learning technique) was chosen in order to find the area with the highest number of restaurants. Specifically, DBSCAN Clustering was used, which takes into account the density of points in order to form the clusters. The cluster containing the highest number of restaurants was then selected.

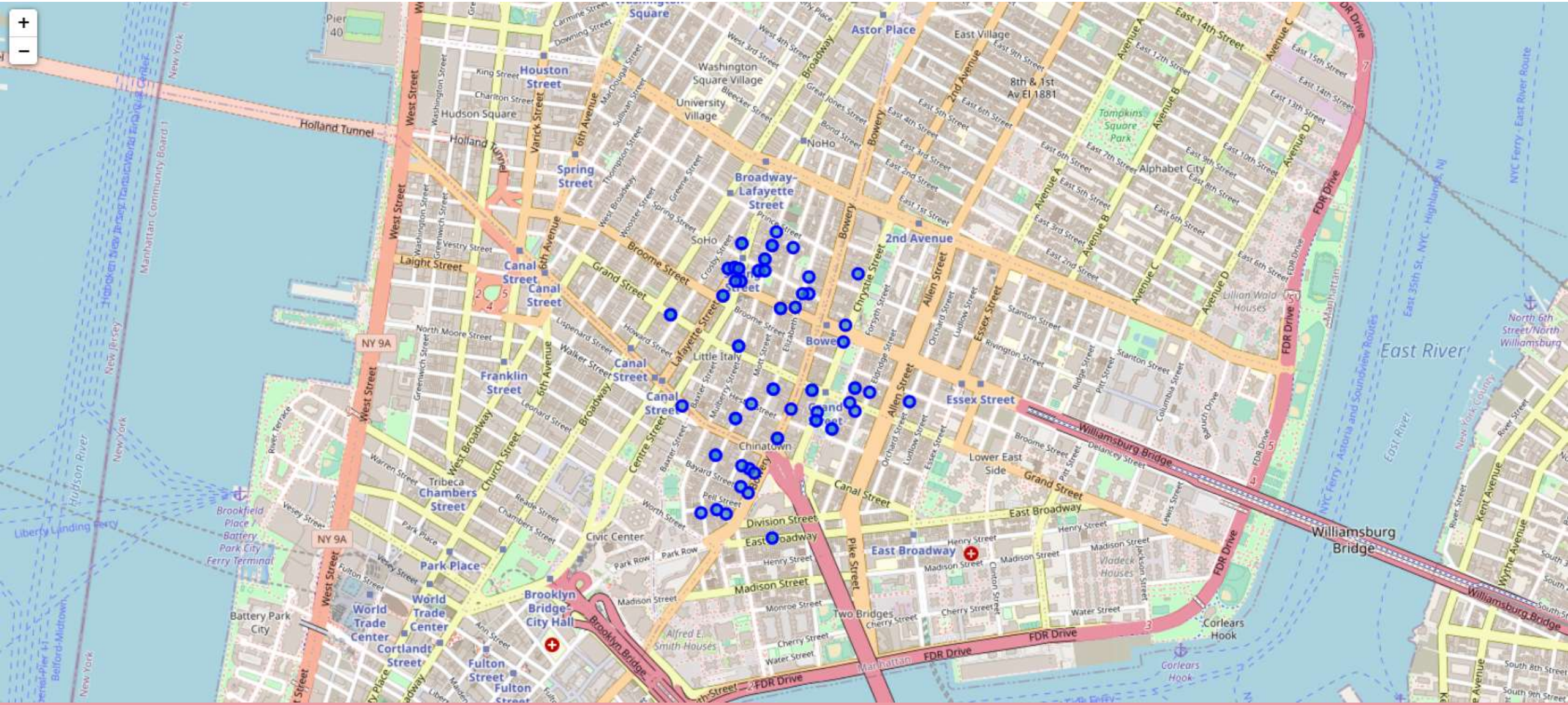
Biggest cluster of restaurants in Manhattan:



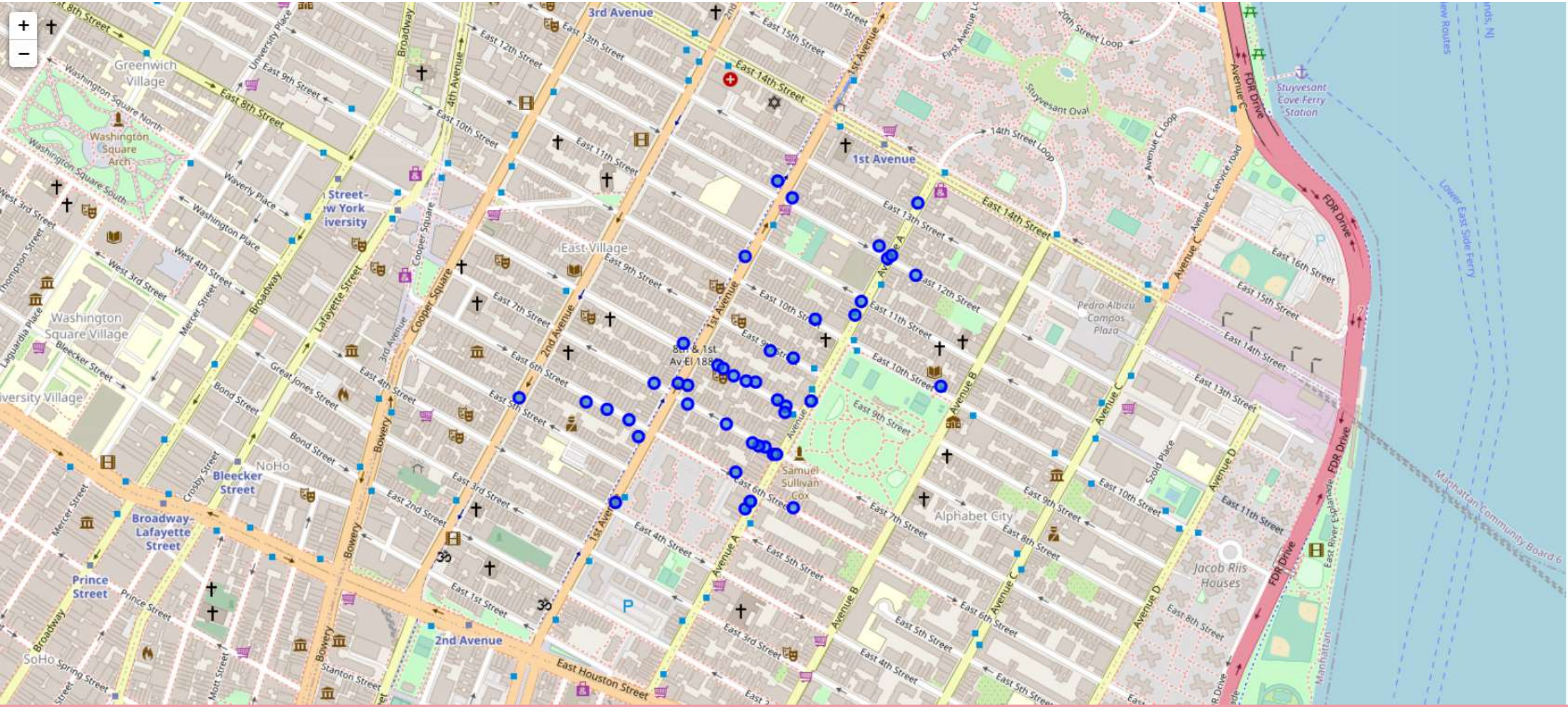
Creating the sub-clusters:

To reduce even more the area covered by the food-delivery office, sub-clusters of the previous cluster were created. We are now considering relatively small areas with high density of points, reducing the shipment costs.
Three clusters were selected based on the total number of points and their density.

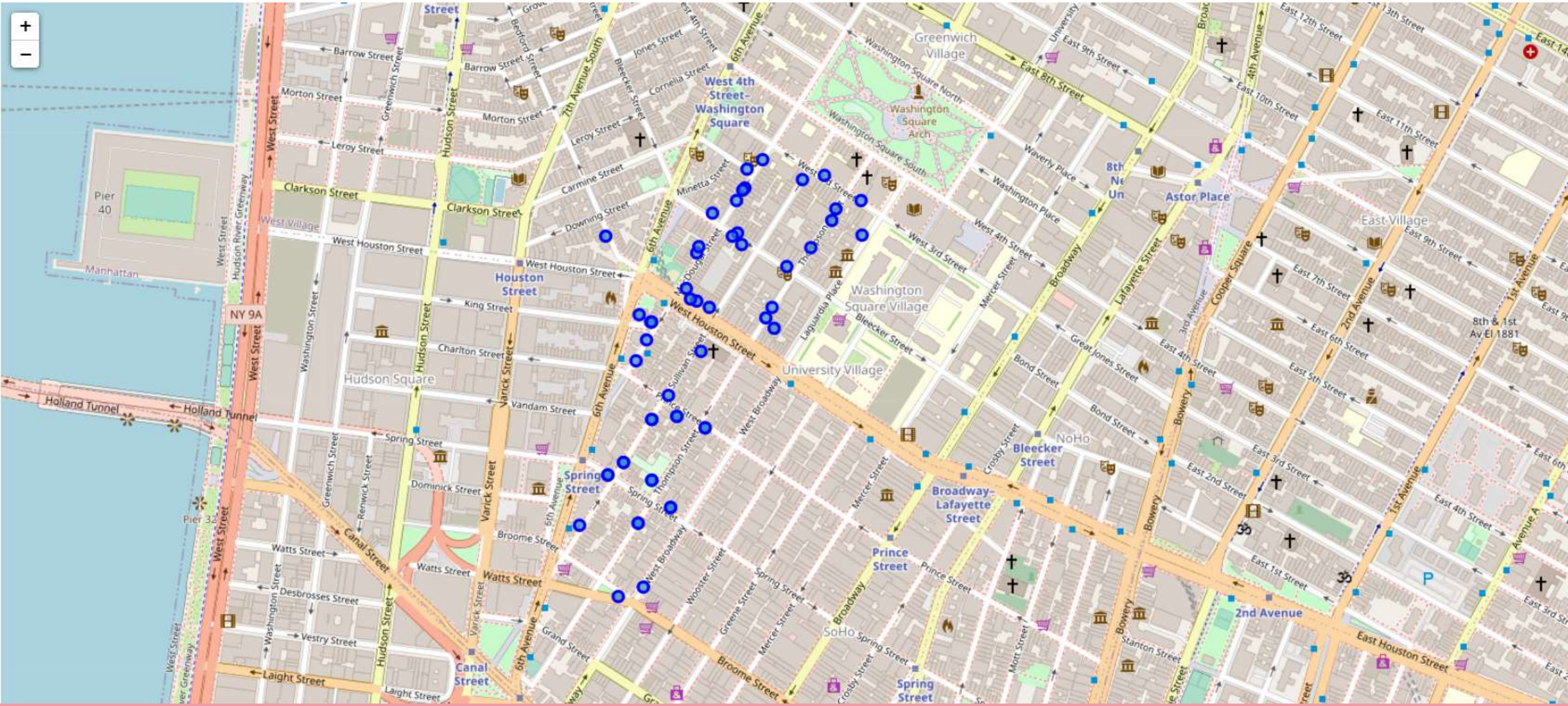
First sub-cluster:



Second sub-cluster:



Third sub-cluster:



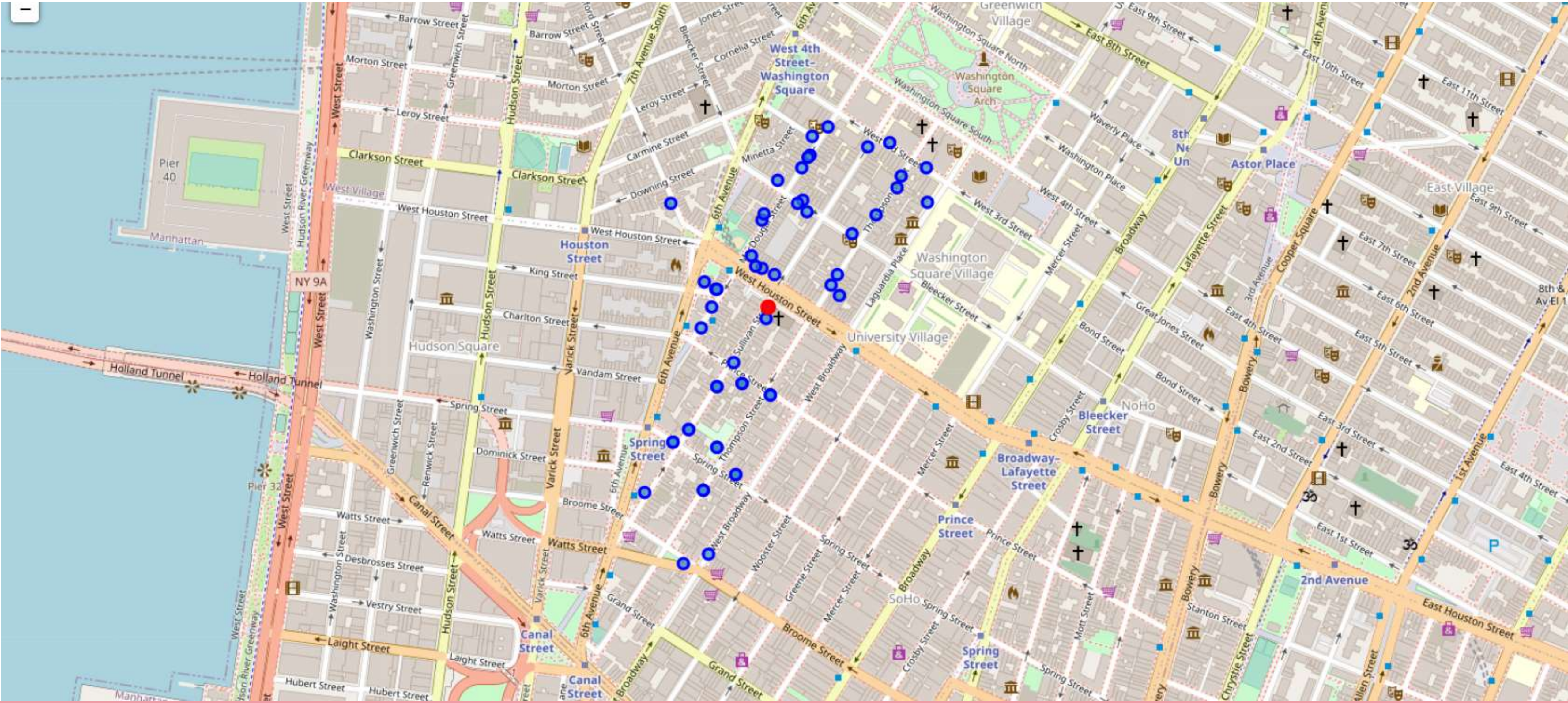
Average rating of the restaurants in each cluster:

As second criterion, the average rating of all the restaurants in each sub-cluster was considered:

- First sub-cluster avg rating: 8.60857142857143
- Second sub-cluster avg rating: 8.637777777777776
- Third sub-cluster avg rating: 8.752830188679246

The third sub-cluster is selected. The position of the office is the correspondig position of the centroid of the cluster, found by averaging the coordinates of all the restaurants in the sub-cluster.

Position of the food-delivery office (red point in figure):



Position of the food-delivery office in Google Maps:

Conclusion and future directions:

This project is a simple proof of concept. Many other factors can be taken into account in order to determine the best location of the food-delivery office, as number of reviews, number of daily customers, road traffic etc.
The number of information was limited by the daily quota of premium API calls to FourSquare and only restaurants information were retrieved. However, many other venues produce food that can be delivered (for ex. Ice-cream shops, bagel shops etc.).