

---

# Data

Now we could decide what kind of data we will need:

1. New York city neighborhoods data
2. The position data of all food, drink and grocery stores in Manhattan area

So we are going to use the Foursquare location data API to get the data we need and will use the centroid model of clustering analysis to find the centroid of each group in the data. The  $K$ -means clustering algorithm of this model works iteratively to assign each data point to one of  $K$  groups based on the features that are provided. Data points are clustered based on feature similarity. The results of this algorithm are the centroids of the  $K$  clusters, which can be used as our best-fit positions of delivery sites.

## 1. Neighborhood Candidates

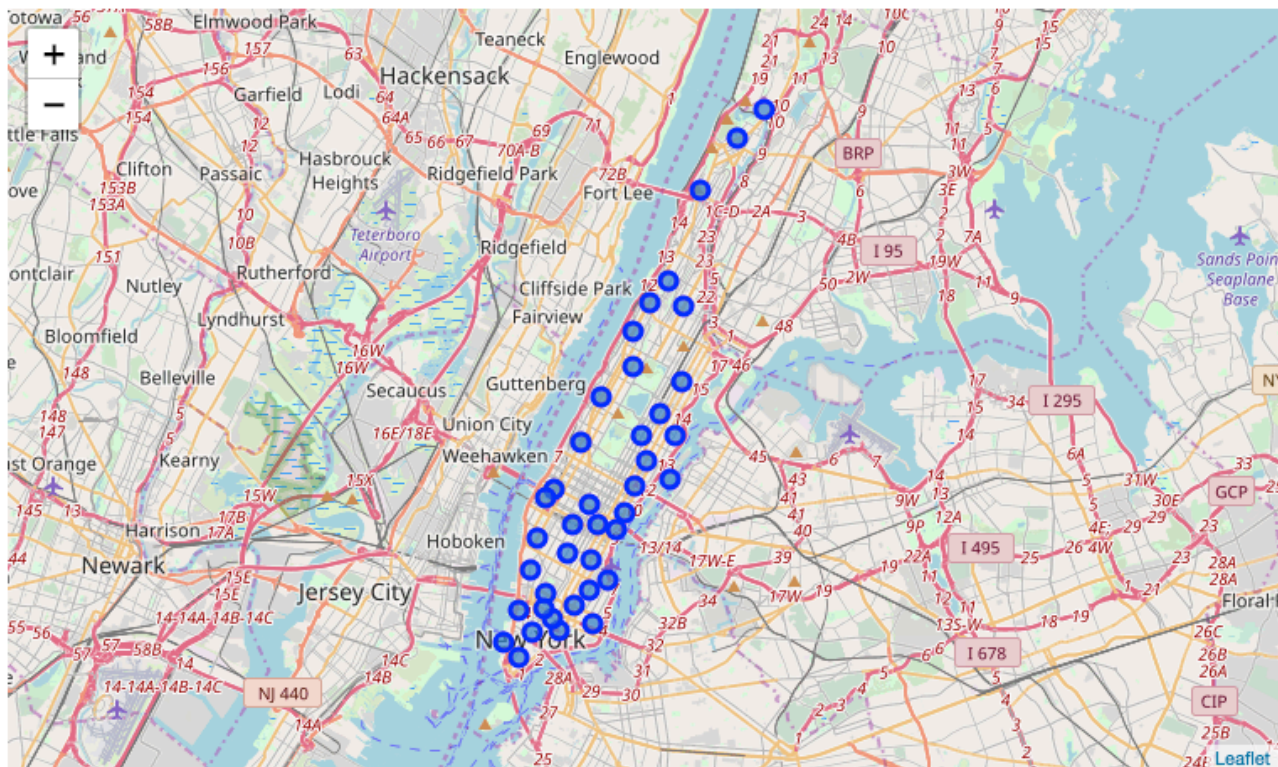
First of all, we need to find the New York City neighborhoods data. We download the data from [https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572). After some cleaning, we get a dataset like this:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Next, we slice the New York City data and create a new dataset of Manhattan data.

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Then we visualize the Manhattan with the 40 neighborhoods in it.



---

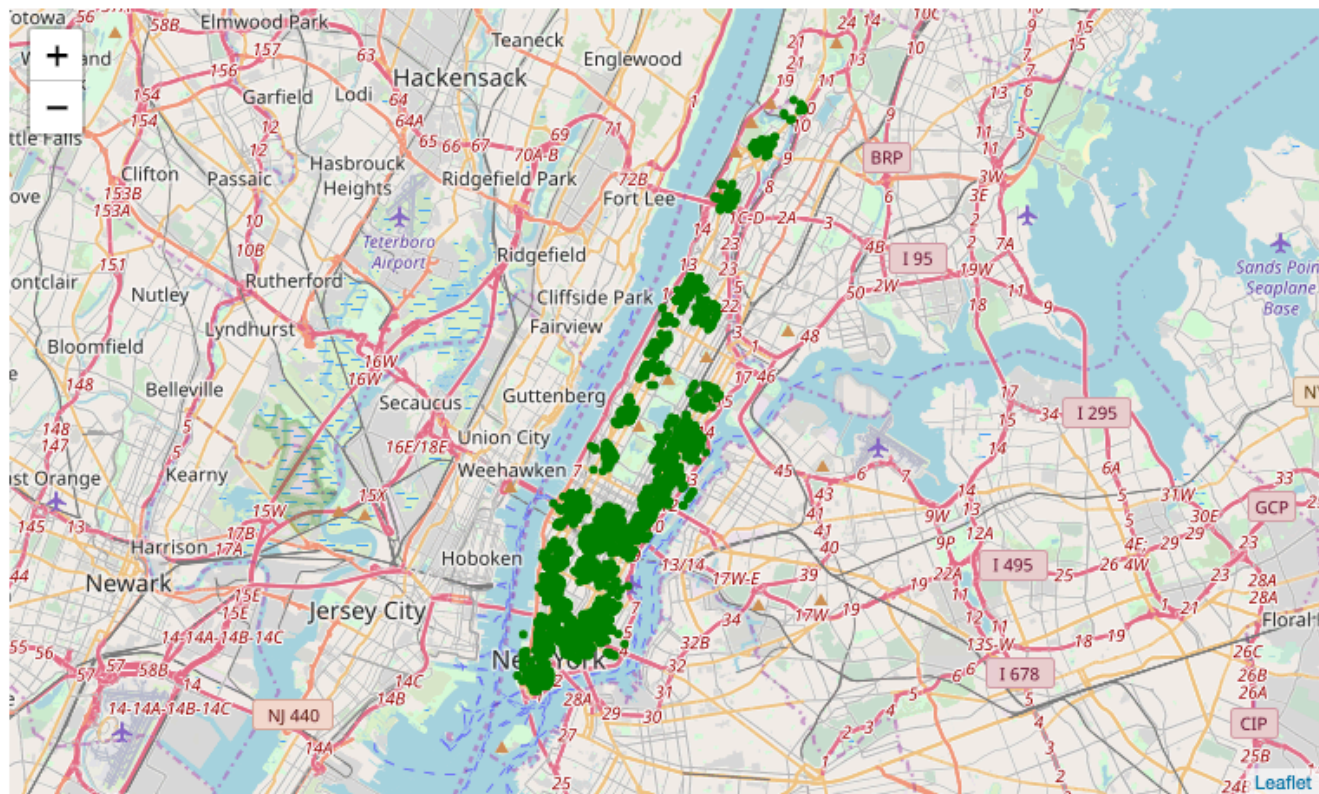
## 2. Foursquare

After have the neighborhoods candidates, we use Foursquare API to get information on food stores in Manhattan area in order to explore the stores and segment them.

	Venue	Longitude	Latitude
0	Arturo's	-73.910271	40.874412
1	Tibbett Diner	-73.908937	40.880404
2	Dunkin'	-73.906666	40.877136
3	Land & Sea Restaurant	-73.905873	40.877885
4	Parrilla Latina	-73.906073	40.877473
5	Subway Sandwiches	-73.909586	40.874667
6	Boston Market	-73.905412	40.877430
7	Auntie Anne's	-73.904947	40.877400
8	SUBWAY	-73.905385	40.878493
9	Subway	-73.905380	40.877720
10	Hernandez Grocery	-73.912591	40.875897
11	Terrace View Delicatessen	-73.912746	40.876476
12	Subway Sandwiches	-73.905308	40.878270
13	Pioneer Chinese Restaurant	-73.906086	40.879274
14	Kiki's	-73.992036	40.714476
15	Spicy Village	-73.993530	40.717010
16	Scarr's Pizza	-73.991649	40.715335
17	Cheeky Sandwiches	-73.991529	40.715699
18	Regina's Grocery	-73.991621	40.715453
19	The Fat Radish	-73.991950	40.715323

There are more than 2800 top-rated food stores in this dataset. We plot them on the Manhattan map.





Now we have all the top-rated food stores in Manhattan area. This concludes the data gathering phase. Now we're ready to use this data for analysis to produce the report on optimal locations for food delivery site!