

# GELATO IN MANHATTAN

Coursera Data Science course capstone  
project

## **Business Problem**

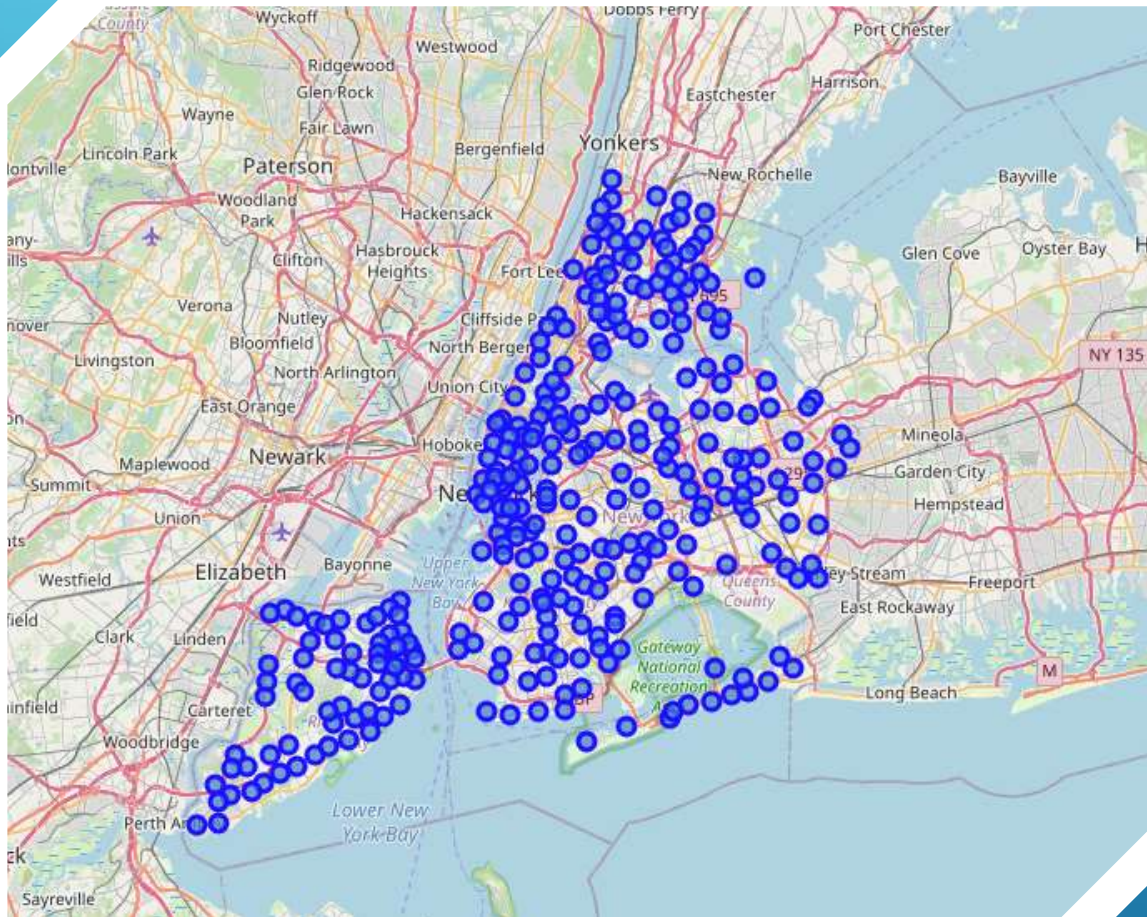
As a Gelato lover, I would like to open the ice-cream shop. Moreover, this is a high-margin business if done properly. With this work I would like to discover what neighbourhood in New York has a lesser number of such shops to open it there.

## **Instrumentation**

In this capstone project, I will leverage the FourSquare API and K-means algorithm to find the best place to open a Gelato shop. To visualize the result of my work I will use Folium module.

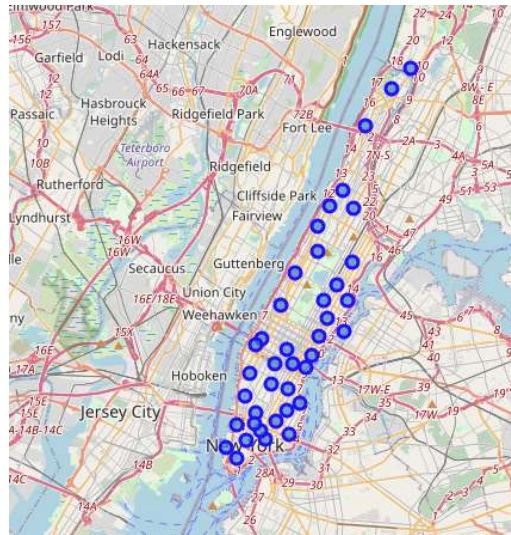
## **Target Audience**

Businessmen who are interested in high margin retail business in New York

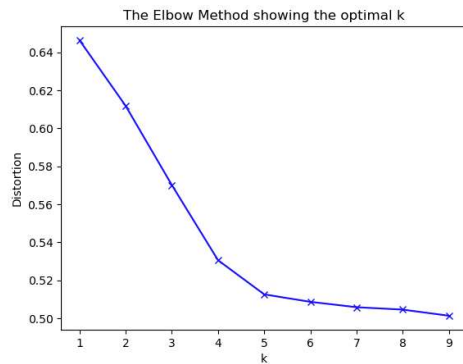


Here we load NYK  
neighbourhood data set and use  
Folium to mark them on the map

Now let's look closer at Manhattan area, and also load all the venues we have in its heghbourhoods into a dataframe



	name	categories	lat	lng
0	Harlem Cycle	Cycle Studio	40.817201	-73.942592
1	Belle Harlem	American Restaurant	40.816890	-73.942680
2	Harlem Hops	Beer Bar	40.814336	-73.944494
3	NYPL Schomburg Center for Research in Black Cu...	Library	40.814575	-73.940928
4	Yatenga	French Restaurant	40.814306	-73.944430
5	Shrine World Music Venue	Music Venue	40.814403	-73.944242
6	Harlem YMCA	Gym / Fitness Center	40.814790	-73.942910



	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	Shop Frequency
0	Manhattan	Marble Hill	40.876551	-73.910660	2	0.041667
1	Manhattan	Chinatown	40.715618	-73.994279	3	0.030000
2	Manhattan	Washington Heights	40.851903	-73.936900	0	0.011628
3	Manhattan	Inwood	40.867684	-73.921210	0	0.016949
4	Manhattan	Hamilton Heights	40.823604	-73.949688	1	0.000000
5	Manhattan	Manhattanville	40.816934	-73.957385	1	0.000000
6	Manhattan	Central Harlem	40.815976	-73.943211	1	0.000000

After that, as we are interested only in gelato, we filter out the rest of venues.

Next we use unsupervised machine learning algorithm K-means to cluster ice cream shops based on their number per each neighbourhood.

We also leverage elbow method to determine right number of K.



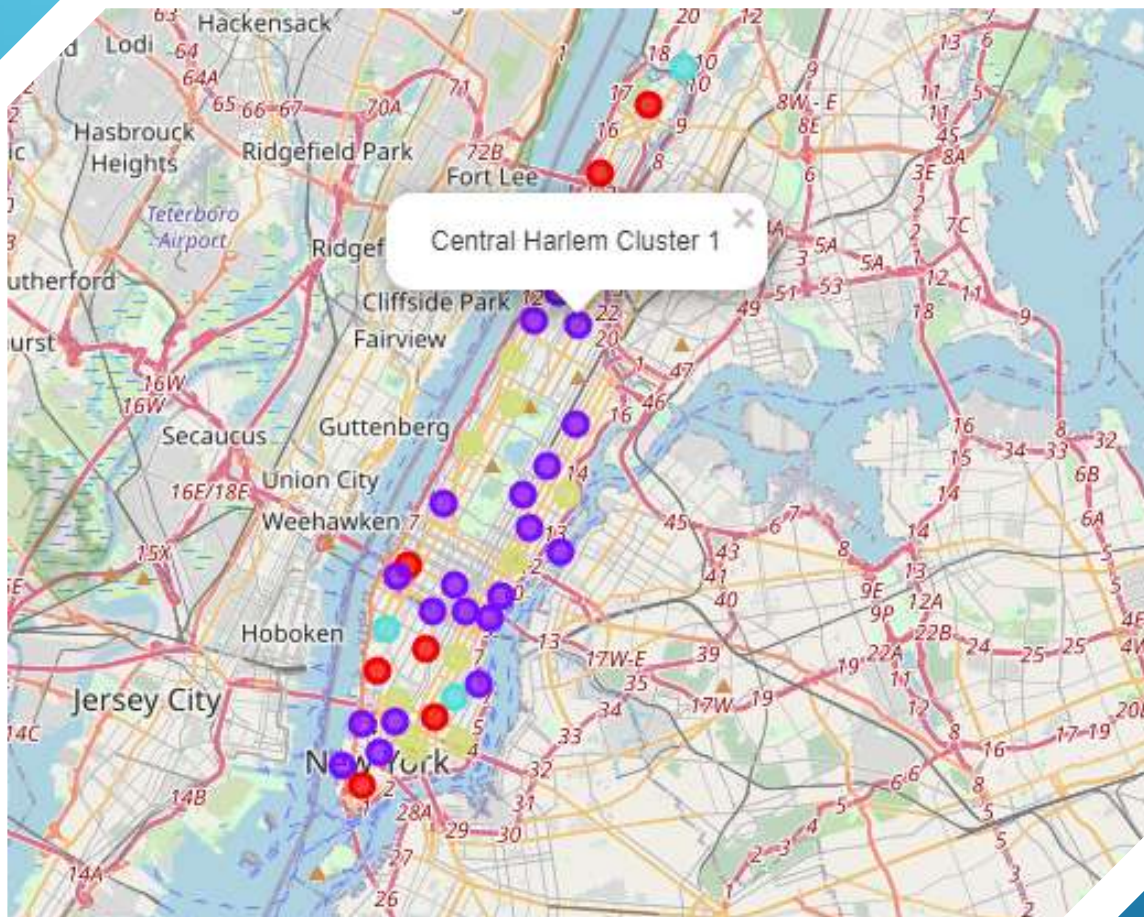
## Conclusion

New York is a very crowded place in terms of venues, including ice cream shops. But still there are vacant locations.

As we can see, 2nd cluster has no ice cream shops at all, and this is the area we should look at.

You can click on any dot on the map to understand to which cluster it belongs to.

In our case we should look at purple dots on the map.



Cluster Labels	Latitude	Longitude	Shop Frequency
0	0	40.769023 -73.978637	0.011225
1	1	40.760994 -73.974851	0.000000
2	2	40.782811 -73.965334	0.040556
3	3	40.754607 -73.976968	0.024222