Capstone Final Project
Where to open my new Bar

Problem

- Let's suppose that someone wants to open a new bar in a city (take New York for our example of city). The first reflex that comes is to find where there is a high and low density of Bars to look for a place where he can benifit from his new Bar. For exemple, the majority will choose to open a new Bar where there is no high density of Bars. In this report, we are going to use you will use the Foursquare API to explore neighborhoods in Toronto.

Data sources

- Using data in our previous lab, we're import it from https://cocl.us/new_york_dataset. The data we're going to explore is composed 5 boroughs and 306 neighborhoods.

Our first data set is composed of 4 columns :

- Borough
- Neighborhood
- Latitude
- Longitude

Data sources

	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
5	Bronx	Kingsbridge	40.881687	-73.902818
6	Manhattan	Marble Hill	40.876551	-73.910660
7	Bronx	Woodlawn	40.898273	-73.867315
8	Bronx	Norwood	40.877224	-73.879391
9	Bronx	Williamsbridge	40.881039	-73.857446

The data is composed 5 boroughs and 306 neighborhoods.

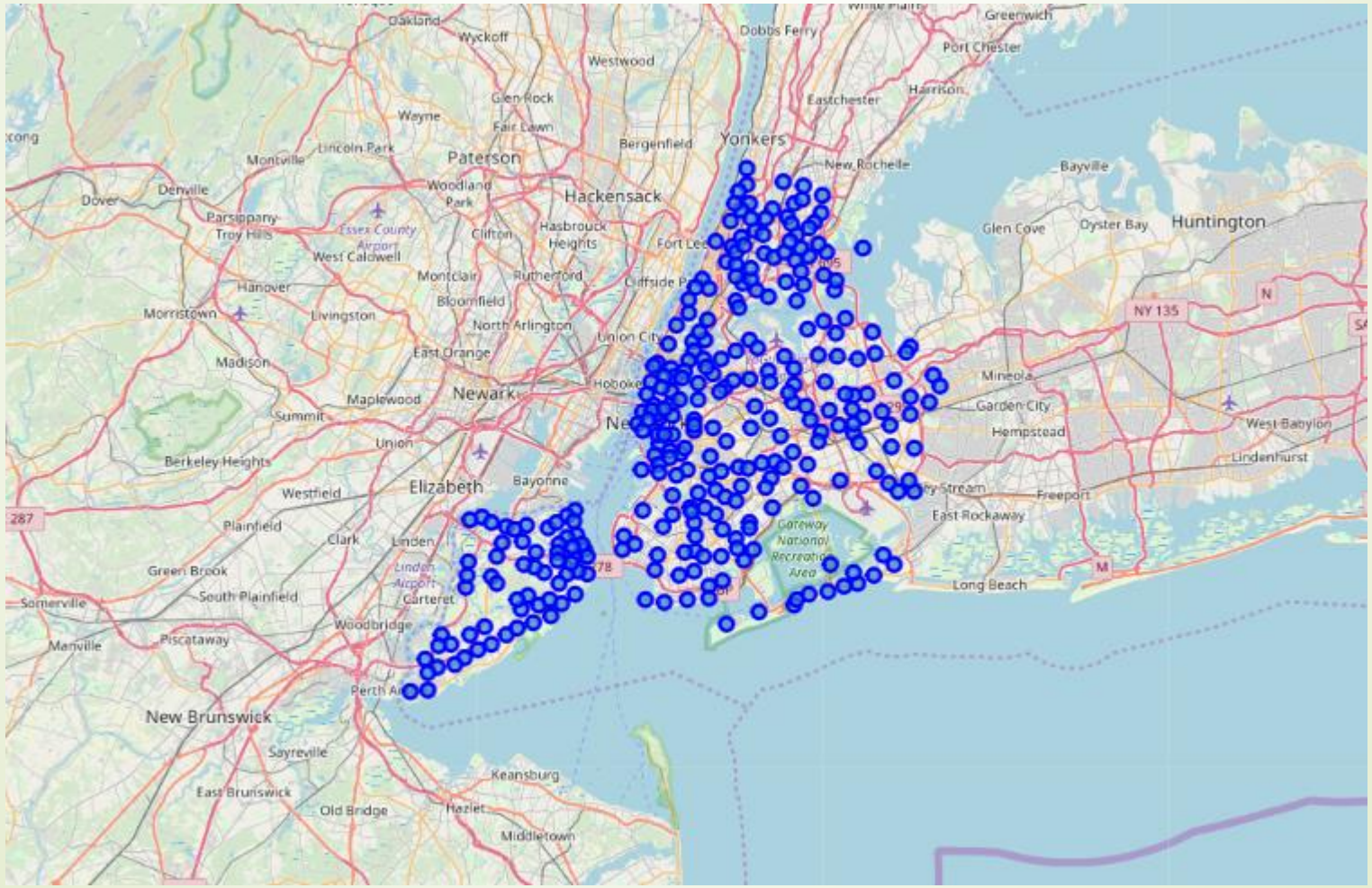
Using foursquare API

- After getting the data of every Borough and Neighborhood on New York with Their Latitude and Longitude, we'll use Foursquare API to get venues in every Neighborhod. The result of calling the Api is shown here :

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
2	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
3	Wakefield	40.894705	-73.847201	Cooler Runnings Jamaican Restaurant Inc	40.898276	-73.850381	Caribbean Restaurant
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop
5	Wakefield	40.894705	-73.847201	SUBWAY	40.890656	-73.849192	Sandwich Place
6	Wakefield	40.894705	-73.847201	Pitman Deli	40.894149	-73.845748	Food
7	Wakefield	40.894705	-73.847201	Baychester Avenue Food Truck	40.892293	-73.843230	Food Truck
8	Wakefield	40.894705	-73.847201	Koss Quick Wash	40.891147	-73.850230	Laundromat
9	Co-op City	40.874294	-73.829939	Capri II Pizza	40.876374	-73.829940	Pizza Place

Exploratory Data Analysis

- Let's first create a map where we can render the neighborhoods in New York :



Exploratory Data Analysis

- We'll group rows by neighborhood and by taking the mean of the frequency of occurrence of each category where we'll find the name 'Bar'

	Neighborhood	Bar	Beer Bar	Cocktail Bar	Dive Bar	Gay Bar	Hookah Bar	Hotel Bar	Juice Bar	Karaoke Bar	Sake Bar	Salon / Barbershop	Sports Bar	Whisky Bar	Wine Bar
0	Allerton	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000	0.0	0.0
1	Annadale	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.125	0.0	0.0
2	Arden Heights	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000	0.0	0.0
3	Arlington	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000	0.0	0.0
4	Arrochar	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000	0.0	0.0

Exploratory Data Analysis

- Let's create the new dataframe and display the top 10 venues that contain Bars.

	Neighborhood	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 Commo Venue
0	Allerton	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop	Sake Bar	Karaoke Bar	Juice Bar	Hotel Bar	Hookah Bar	Gay Bar
1	Annadale	Sports Bar	Wine Bar	Whisky Bar	Salon / Barbershop	Sake Bar	Karaoke Bar	Juice Bar	Hotel Bar	Hookah Bar	Gay Bar
2	Arden Heights	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop	Sake Bar	Karaoke Bar	Juice Bar	Hotel Bar	Hookah Bar	Gay Bar
3	Arlington	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop	Sake Bar	Karaoke Bar	Juice Bar	Hotel Bar	Hookah Bar	Gay Bar
4	Arrochar	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop	Sake Bar	Karaoke Bar	Juice Bar	Hotel Bar	Hookah Bar	Gay Bar
5	Arverne	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop	Sake Bar	Karaoke Bar	Juice Bar	Hotel Bar	Hookah Bar	Gay Bar
6	Astoria	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop	Sake Bar	Karaoke Bar	Juice Bar	Hotel Bar	Hookah Bar	Gay Bar
7	Astoria Heights	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop	Sake Bar	Karaoke Bar	Juice Bar	Hotel Bar	Hookah Bar	Gay Bar
8	Auburndale	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop	Sake Bar	Karaoke Bar	Juice Bar	Hotel Bar	Hookah Bar	Gay Bar
9	Bath Beach	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop	Sake Bar	Karaoke Bar	Juice Bar	Hotel Bar	Hookah Bar	Gay Bar

Modeling

K-means Algorithm

- Let's use k-means to cluster the neighborhood into 5 clusters. It means that the value of k for this exemple is 5.
- k-means is one of the simplest and popular unsupervised machine learning algorithms. In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

Modeling

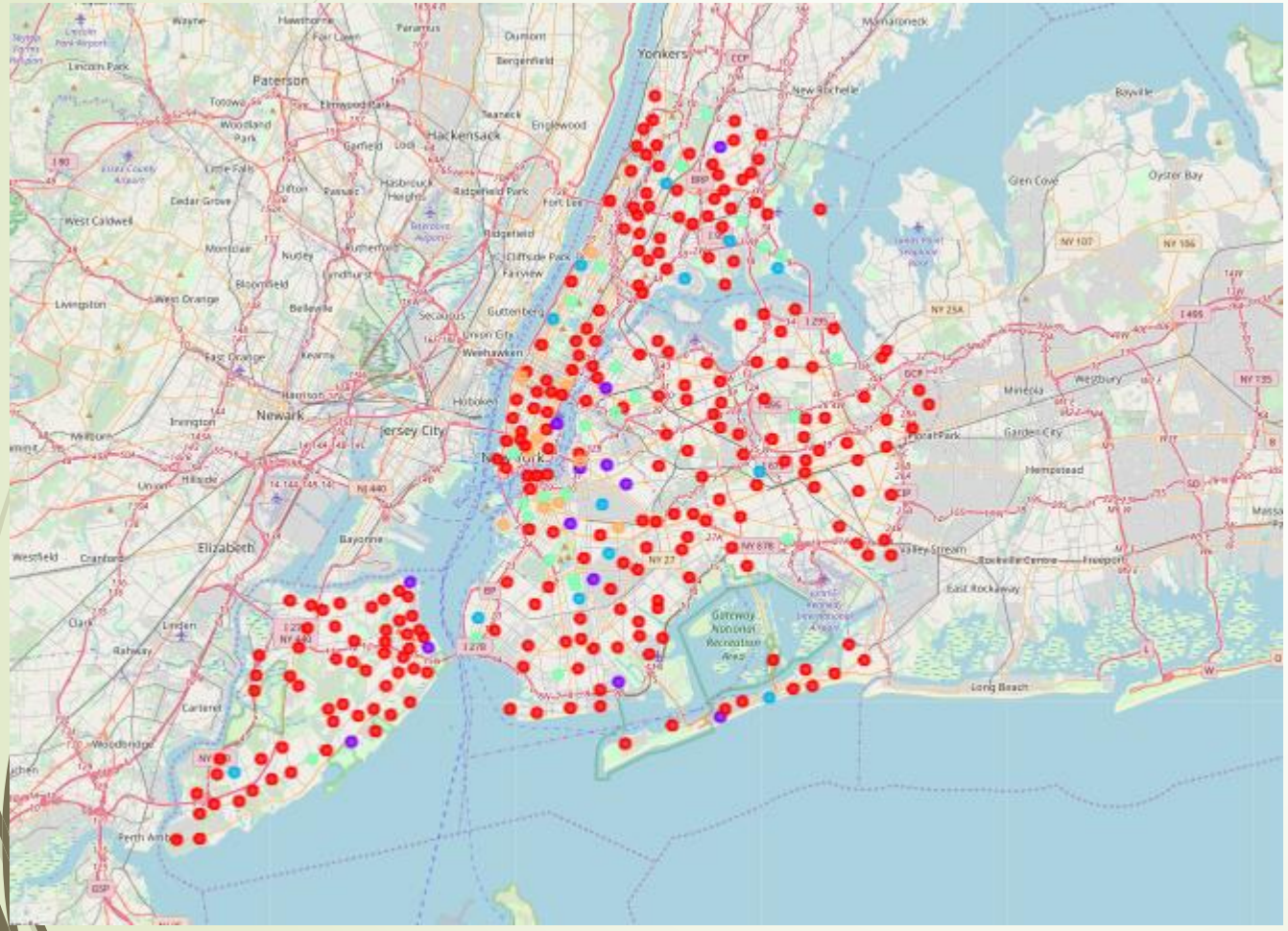
K-means Algorithm

- Here is the result of clustering with 5 labels after fiting our k-means algorithm :

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bronx	Wakefield	40.894705	-73.847201	0.0	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop	Sake Bar
1	Bronx	Co-op City	40.874294	-73.829939	0.0	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop	Sake Bar
2	Bronx	Eastchester	40.887556	-73.827806	0.0	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop	Sake Bar
3	Bronx	Fieldston	40.895437	-73.905643	0.0	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop	Sake Bar
4	Bronx	Riverdale	40.890834	-73.912585	0.0	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop	Sake Bar
5	Bronx	Kingsbridge	40.881687	-73.902818	0.0	Beer Bar	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop
6	Manhattan	Marble Hill	40.876551	-73.910660	0.0	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop	Sake Bar
7	Bronx	Woodlawn	40.898273	-73.867315	3.0	Bar	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop
8	Bronx	Norwood	40.877224	-73.879391	0.0	Wine Bar	Whisky Bar	Sports Bar	Salon / Barbershop	Sake Bar

Result

- Let's show our clusters on a map using Folium to get a visual version of the clusters



As you can see on the map, we have 5 clusters of every type of bars. This clusters can be labeled depending on the density of bars in the cluster and help the person who want to open a bar to find the best neighborhood for that