

# **IBM Data Science Professional Certificate - Capstone Project**

Mohammed Adi





# Introduction

In this presentation, I will try to use the data science techniques that I have learned throughout the IBM Data Science Professional Certificate course, to solve a hypothetical business problem.

The business problem is a question that can be faced by F&B business in New York City (NYC), which is deciding where to open a new restaurant.



## The Data

- The data will be comprised of a JSON file which contains a list of NYC's boroughs and neighborhoods, retrieved from IBM's server.
- The geo-coordinates of the city and neighborhood will be acquired using the Nominatim API from Python's Geopy client.
- Venues data will be acquired using Foursquare's API.



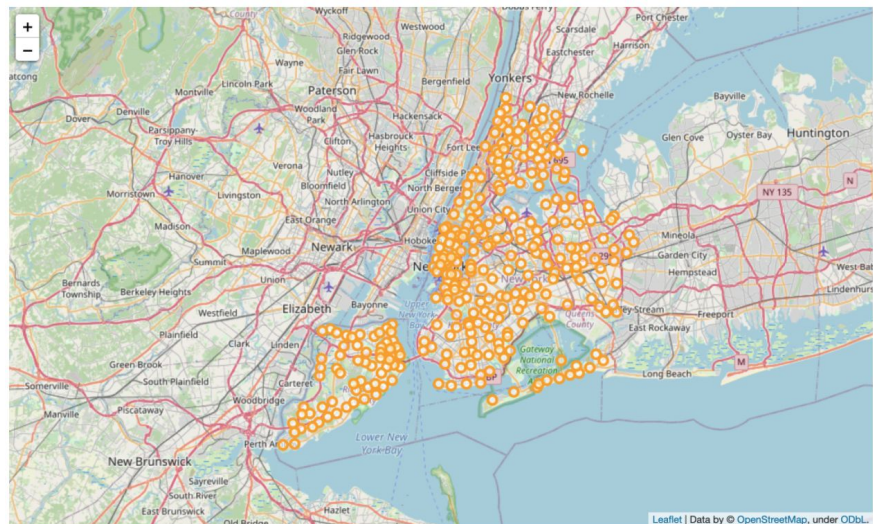
# Methodology

- We will begin by importing the necessary libraries.
- Then, data will be acquired from a json file and will be cleaned up and rearranged using Python's Pandas library.
- Next, we will use Foursquare's API to acquire the top venues.
- An unsupervised machine learning clustering algorithm, using k-means technique will be used on the grouped and cleaned data.
- Lastly, clusters will be created to divide the neighborhoods and we will perform our final analysis



# Data Gathering

The initial shape of the data frame is (5, 306), which is 5 boroughs and 306 neighborhoods in NYC and we can create a folium map to visualize.





## Exploring the top venues using Foursquare API and merging data with original dataframe

	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	Ripe Kitchen & Bar	40.898152	-73.838875	Caribbean Restaurant
2	Wakefield	40.894705	-73.847201	Ali's Roti Shop	40.894036	-73.856935	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Jackie's West Indian Bakery	40.889283	-73.843310	Caribbean Restaurant
4	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy



**One Hot Encoding is performed to analyze neighborhoods, and data will be grouped and mean of occurrence frequency of each category will be calculated**

	Neighborhood	Zoo Exhibit	ATM	Accessories Store	Adult Boutique	Adult Education Center	Afghan Restaurant	African Restaurant	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Amphitheater	Animal Shelter
0	Allerton	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.016667	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	0.117647	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	0.000000	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	0.043478	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	0.000000	0.0	0.0



## Now we find top 5 venues in each neighborhood

----Erasmus----

	venue	freq
0	Caribbean Restaurant	0.11
1	Deli / Bodega	0.07
2	Discount Store	0.07
3	Mobile Phone Shop	0.05
4	Pizza Place	0.05

----Far Rockaway----

	venue	freq
0	Pizza Place	0.09
1	Chinese Restaurant	0.09
2	Deli / Bodega	0.06
3	Sandwich Place	0.06
4	Caribbean Restaurant	0.03





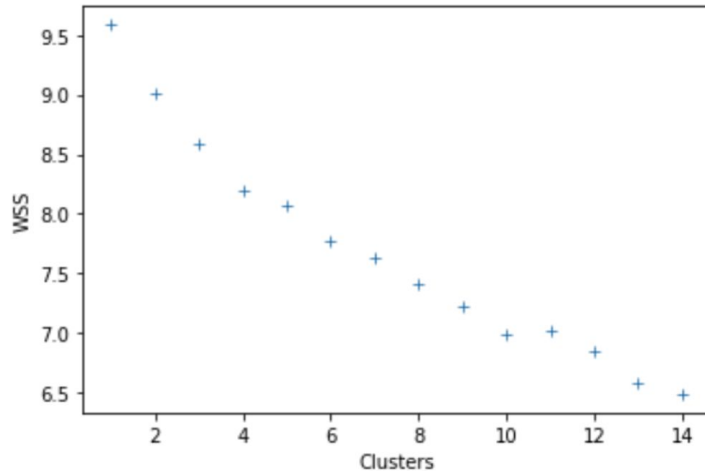
Now we get top 10 venues for all neighborhoods and create a new dataframe with merged data

	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 Common Venue
0	Allerton	Pizza Place	Donut Shop	Sandwich Place	Deli / Bodega	Caribbean Restaurant	Fast Food Restaurant	Mexican Restaurant	Pharmacy	Fried Chicken Joint	Food
1	Annadale	American Restaurant	Trail	Home Service	Restaurant	Pizza Place	Diner	Train Station	Park	Pub	Liquor Store
2	Arden Heights	Park	Italian Restaurant	Dog Run	Sushi Restaurant	Optical Shop	Food	Bus Stop	Spa	Gift Shop	Mexican Restaurant
3	Arlington	Bus Stop	Deli / Bodega	Hardware Store	Discount Store	American Restaurant	Check Cashing Service	Snack Place	Fast Food Restaurant	Coffee Shop	Donut Shop
4	Arrochar	Bus Stop	Italian Restaurant	Deli / Bodega	Baseball Field	Beach	Bagel Shop	Mediterranean Restaurant	Middle Eastern Restaurant	Nail Salon	Outdoors & Recreation

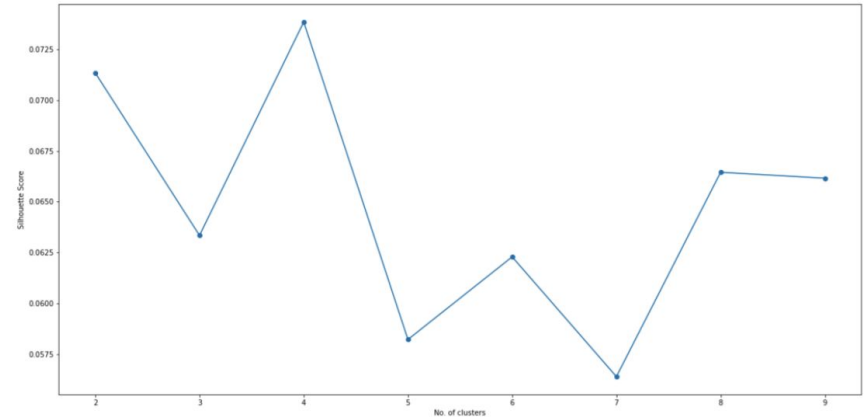
# Machine Learning

We will begin by trying to find the optimal number of clusters to be used.

1. Elbow Method:



2. Silhouette Scores



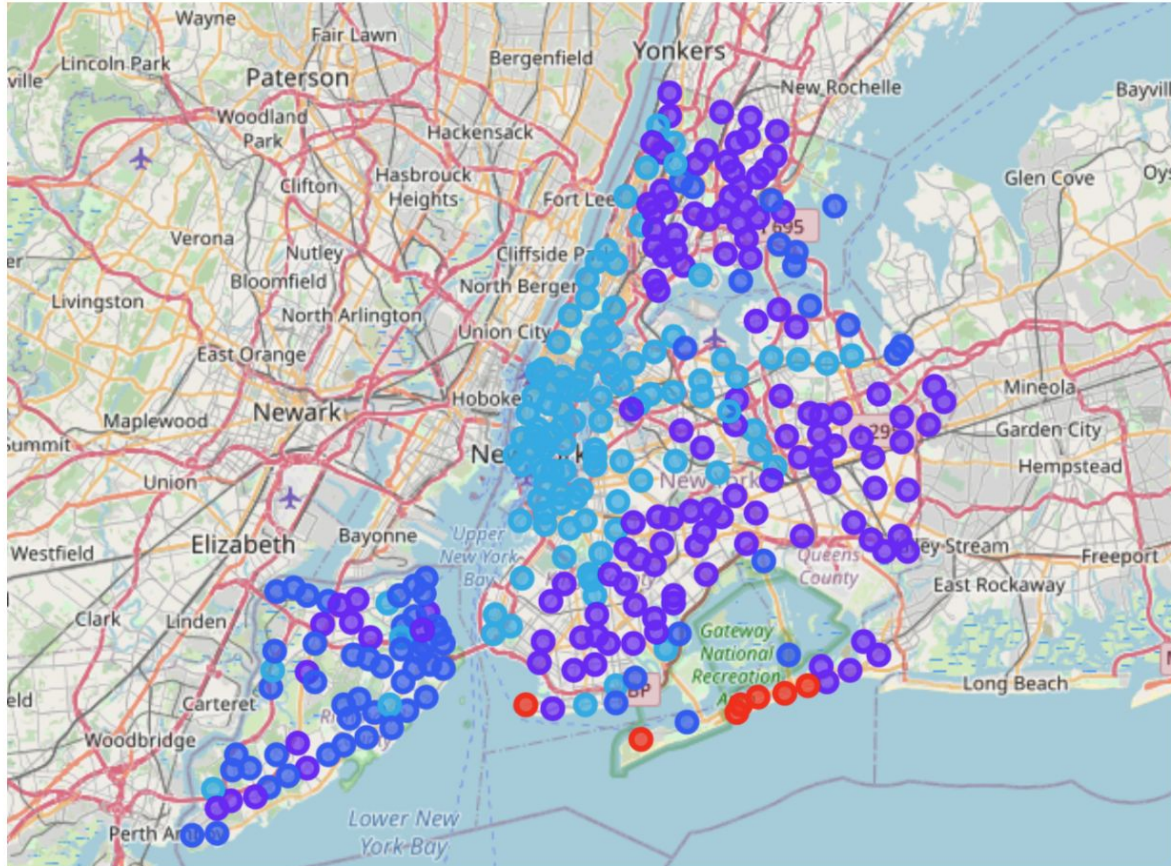
We can see that the elbow method was not able to provide a clear answer, but the silhouette was able to provide an accurate number of clusters to be used which is 4.



# Label Cluster and create a dataframe

	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	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9 C
0	Bronx	Wakefield	40.894705	-73.847201	1	Pharmacy	Caribbean Restaurant	Supermarket	Donut Shop	Fast Food Restaurant	Fried Chicken Joint	Bank	Pizza Place	
1	Bronx	Co-op City	40.874294	-73.829939	1	Pizza Place	Department Store	Mobile Phone Shop	Shopping Mall	Shoe Store	Pharmacy	Clothing Store	Bakery	
2	Bronx	Eastchester	40.887556	-73.827806	1	Caribbean Restaurant	Shopping Mall	Diner	Fast Food Restaurant	Gas Station	Supplement Shop	Grocery Store	Pharmacy	
3	Bronx	Fieldston	40.895437	-73.905643	1	Deli / Bodega	Bus Station	Pizza Place	Bar	Plaza	Sandwich Place	Park	Mexican Restaurant	
4	Bronx	Riverdale	40.890834	-73.912585	3	Bank	Pizza Place	Bar	Park	Diner	Deli / Bodega	Japanese Restaurant	Coffee Shop	

# Map the clusters





# Examine Clusters

## Cluster 1:

	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 Common Venue
85	Sea Gate	Beach	Supermarket	American Restaurant	Pharmacy	Park	Donut Shop	Chinese Restaurant	Bus Stop	Video Store	Spa
172	Breezy Point	Surf Spot	American Restaurant	Beach	Monument / Landmark	Park	Trail	Deli / Bodega	Exhibit	Factory	Falafel Restaurant
178	Rockaway Beach	Beach	Bar	Pharmacy	Boat or Ferry	Latin American Restaurant	Ice Cream Shop	Pizza Place	Eastern European Restaurant	Arepa Restaurant	Supermarket
179	Neponsit	Beach	Spa	Deli / Bodega	Pub	Mexican Restaurant	Italian Restaurant	Restaurant	Harbor / Marina	Bakery	Boutique
190	Belle Harbor	Beach	Pub	Donut Shop	Bakery	Mexican Restaurant	Bagel Shop	Chinese Restaurant	Harbor / Marina	Pharmacy	Smoke Shop
191	Rockaway Park	Beach	Pizza Place	Donut Shop	Liquor Store	Bagel Shop	Bar	Pharmacy	Deli / Bodega	Pub	Supermarket
302	Hammels	Beach	Surf Spot	Donut Shop	Fast Food Restaurant	Coffee Shop	Pizza Place	Taco Place	Bar	Supermarket	Wine Bar



# Examine Clusters

Cluster 2: (first 7 rows shown)

	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 Common Venue
0	Wakefield	Pharmacy	Caribbean Restaurant	Supermarket	Donut Shop	Fast Food Restaurant	Fried Chicken Joint	Bank	Pizza Place	Bakery	Bus Station
1	Co-op City	Pizza Place	Department Store	Mobile Phone Shop	Shopping Mall	Shoe Store	Pharmacy	Clothing Store	Bakery	Liquor Store	Mattress Store
2	Eastchester	Caribbean Restaurant	Shopping Mall	Diner	Fast Food Restaurant	Gas Station	Supplement Shop	Grocery Store	Pharmacy	Donut Shop	Discount Store
3	Fieldston	Deli / Bodega	Bus Station	Pizza Place	Bar	Plaza	Sandwich Place	Park	Mexican Restaurant	Bank	Chinese Restaurant
6	Marble Hill	Park	Pizza Place	Mexican Restaurant	Café	Donut Shop	Spanish Restaurant	Sandwich Place	Grocery Store	Scenic Lookout	Supermarket
7	Woodlawn	Pub	Bar	Deli / Bodega	Pizza Place	Bakery	Discount Store	Donut Shop	Bank	Baseball Field	Fast Food Restaurant



# Examine Clusters

Cluster 3: (first 7 rows shown)

	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 Common Venue
12	City Island	Harbor / Marina	Seafood Restaurant	Boat or Ferry	Italian Restaurant	Bar	Thrift / Vintage Store	American Restaurant	Music Venue	French Restaurant	Bank
16	Fordham	Italian Restaurant	Pizza Place	Spanish Restaurant	Deli / Bodega	Shoe Store	Mobile Phone Shop	Diner	Coffee Shop	Supplement Shop	Dessert Shop
27	Clason Point	Pool	Boat or Ferry	Event Space	Discount Store	Gym / Fitness Center	Park	Wings Joint	Bus Stop	Farm	Falafel Restaurant
28	Throgs Neck	Italian Restaurant	Deli / Bodega	Donut Shop	Pizza Place	Harbor / Marina	American Restaurant	Asian Restaurant	Coffee Shop	Trail	Bookstore
34	Belmont	Italian Restaurant	Pizza Place	Deli / Bodega	Bakery	Shoe Store	Zoo	Dessert Shop	Latin American Restaurant	Café	Garden
37	Pelham Bay	Italian Restaurant	Deli / Bodega	Fast Food Restaurant	Sandwich Place	Pizza Place	Bank	Bakery	Convenience Store	Donut Shop	Rental Car Location

# Examine Clusters

Cluster 4: (first 7 rows shown)

	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 Common Venue
4	Riverdale	Bank	Pizza Place	Bar	Park	Diner	Deli / Bodega	Japanese Restaurant	Coffee Shop	Health & Beauty Service	Bagel Shop
5	Kingsbridge	Pizza Place	Mexican Restaurant	Sandwich Place	Bar	Diner	Donut Shop	Coffee Shop	Pharmacy	Bank	Park
19	High Bridge	Lounge	Baseball Stadium	Park	Plaza	Burger Joint	Bar	BBQ Joint	Beer Bar	Caribbean Restaurant	Liquor Store
24	Hunts Point	Park	Bakery	Market	Paintball Field	Travel & Transport	Gourmet Shop	Farmers Market	Seafood Restaurant	Mexican Restaurant	Home Service
46	Bay Ridge	Pizza Place	Spa	Italian Restaurant	Coffee Shop	Bagel Shop	Chinese Restaurant	Cosmetics Shop	Greek Restaurant	American Restaurant	Café
48	Sunset Park	Bakery	Mexican Restaurant	Pizza Place	Chinese Restaurant	Bank	Latin American Restaurant	Ice Cream Shop	Asian Restaurant	Sandwich Place	Dumpling Restaurant





## Conclusion & Recommendation

Cluster 1: suitable for seafood, casual dining concepts, and lounges.

Cluster 2: suitable for family diners and foreign cuisines.

Cluster 3: suitable for American cuisine, casual/family diners, and seafood.

Cluster 4: suitable for coffee shops, cafes, and bars.