IBM Data Science Specialization

Applied Data Science Capstone Project

The Battle of Neighbourhoods: New York vs Toronto

Author: **Musah Abdul-Mumuni**

September, 2019

TABLE OF CONTENTS

1. Executive Summary

The Battle of Neighbourhoods: New York vs Toronto: One of the biggest challenges in life is making a decision to setup a business in an unknown region. i.e diving into unchartered waters. The risk is not only that the business can collapse and the capital lost, but also ones reputation to take important decisions would be tarnished by investors and family members alike. But when properly done the reward is immense. A family that operates a Doner shop in Europe, precisely Berlin, the capital of Germany is hoping to achieve a similar success in a North American financial capital. The decision on which city to choose to launch the business in order to avoid failure, survive and grow is not an easy task.
For such an investment to succeed, it requires very good understanding of the business location. The decision of choosing which city in North America to launch the business can be time consuming, tedious and costly because it requires good knowledge of the popular venues, foot traffic and certain establishments with which one can gauge the level of patronage of a similar business.

An important method that will immensely assist the decision making process is the application of Data Science methods which includes the use of Python Jupyter Notebook and Foursquare API for collecting available data online for analysis.

A presentation of the process of collecting, preprocessing and analyzing Neighbourhood data of North America´s two richest financial capitals, New York and Toronto in order to make a decision on where to start an exotic fast food business is outlined in this report.

2. Contents

a. BUSINESS UNDERSTANDING

A family that operates a Doner (a very popular dish in Germany consisting of spiced lamb cooked on a spit and served in slices, typically with pitta bread) shop in Berlin the capital of Germany in Europe is considering introducing its business to the North American market. The family is facing the choice of either establishing its business in New York or Toronto. The reason for choosing one of the two cities is that, Toronto being the Commercial capital of Canada and New York the commercial capital of the United States of America will offer the much needed foot traffic and also the exotic (Italian, Chinese, Indian, Jamaican Restaurants) eating habits that is associated with cosmopolitan neighbourhoods, a necessary ingredient for the survival and growth of such a Business venture.

On the basis of the fact that The US and Canada fall in the top 3 countries with the highest consumption of fast foods in the world, with the U.S coming first and Canada placing 3rd behind Japan, makes North America a reasonable location for establishing a Doner business. It is also important to consider that, more than 1 out of every 3 American adult eats fast food in a given day and a similar thing exist in Canada.

The presence of other similar exotic businesses and Beach, touristic, sports, leisure and recreational facilities are positive indication of a future for the business.

Data Science methodology is one of the modern world´s effective and efficient method of making decisions such as making an investment decision far away on another continent. To choose between New York and Toronto on the basis of where is more suitable to establish a Doner Business can be done using Python Jupyter Notebook and Foursquare API for collecting location data and required libraries are imported for preprocessing, clustering and analyzing the data.

Clustering, a machine learning technique is used to segment the neighborhoods with similar objects on the basis of each neighborhood data. These objects will be given priority on the basis of exotic restaurants and foot traffic in their respective neighborhoods.

From the work done, any exotic fast food entrepreneur outside North America wishing to invest in one of the top Financial Capitals in North America can find the model useful in deciding either to choose New York or Toronto.

b. ANALYTIC APPROACH

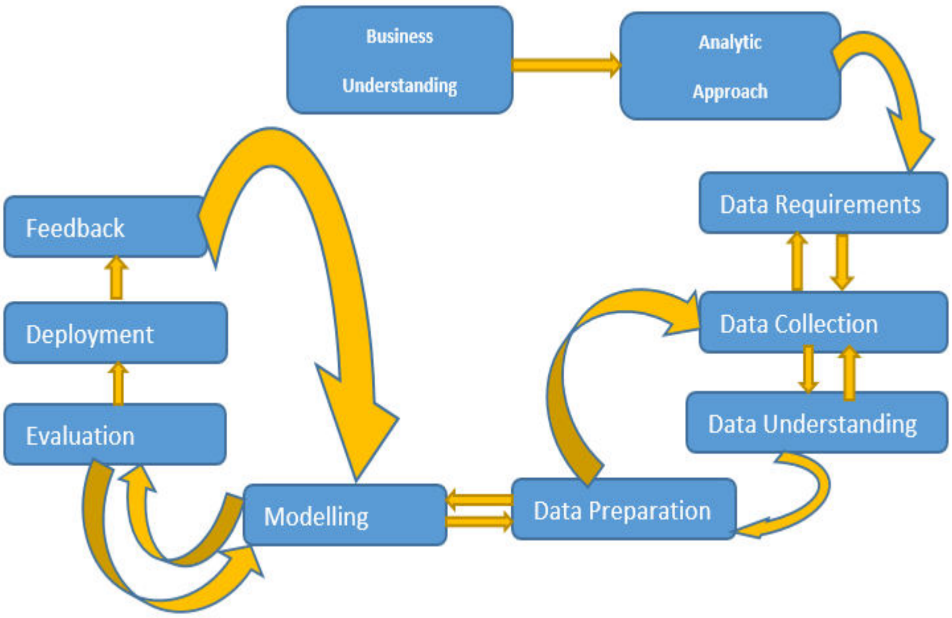

Fig. 1

The chart above summarizes the Data Science methodology to be applied in solving the problem of choosing between New York and Toronto for Doner Business establishment. It starts with Business Understanding, where a definition of the goal is establishing a Doner Business which will be successful in a North American financial capital is defined. It is followed by Analytical Approach where the guidelines or patterns for achieving the goal is defined. In this case, with the use of Python Jupyter notebook and Foursquare API, Neighbourhood data are collected online of the two cities and analyzed.

Data Requirement, collection, understanding and preparation stages are done iteratively. Questions are continuously asked in the process and more or similar and better data are required, collected and prepared for analysis to create a model from which solutions concerning establishing other related businesses in a similar setting can be found by manipulating the model. The model is then evaluated by checking other Neighbourhood venue within the model. Model evaluation goes hand-in-hand with model building as such, the modeling and evaluation stages are done iteratively. When satisfied of its performance it is deployed by establishing the business in the chosen city. Continuous feedback is fed into the model to achieve optimum functioning of the model.

c. DATA ANALYSIS AND VISUALISATION

i. Data Requirement

The data to be used are the Neighbourhood data of New York and that of Toronto. The data contains the postcodes, Boroughs and the corresponding neighbourhoods from which the venues where establishments like restaurants, café etc can be found for use in the analysis.

ii. Data Collection & Understanding

By making use of Foursquare API, the opportunity to explore the data of two cities, in terms of their neighborhoods can be made. The data sets are imported from Wikipedia and IBM open data. The data also include the information about the places around each neighborhood like restaurants, hotels, coffee shops, parks, theaters, art galleries, museums and many more. A selection of one Borough from each city is done to analyze their neighborhoods. Queens from New York is chosen for the U.S. and Scarborough in Toronto is chosen for Canada.

The reason for choosing Queens is that it is the Borough with the highest Neighbourhoods and it has almost the largest population in New York. Scarborough is only second to North York in terms of highest neighbourhoods and population in Toronto and unlike North York it is located near a beach.

IBM Watson Jupyter Notebook was used for the exercise. First of all the raw data had to be imported. For the Toronto data it was scraped from Wikipedia. Later coordinates data for the city of Toronto are collected and added in order to support finding the venues. The original data of venues collected for the analysis was not uniform for both cities since Queens has a larger population than Scarborough and data returned for Scarborough within a radius of 500m was not satisfactory. Therefore the radius for collection of data was increased for both cities from 500m to 1000m.

Neighborhoods Segmentation

Neighborhoods Segmentation is done by importing the Boroughs and neighborhood list of Toronto from Wikipedia and converting it to data frame using pandas package in python. Then, another data set comprised of location data of neighborhood and boroughs was imported. It was in .csv format and then converted to data frame. After Cleaning the data set, two tables were merged to get the final Toronto neighborhood data set. The url of the Toronto data is as follows:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

A list of the libraries imported and downloaded for the project are as follows

LIBRARY	USAGE
pandas	For data analysis. For data processing, CSV files
numpy	To handle data in a vectorized manner. For scientific computing, arrays, algebra, matrices
beautiful soup	It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping
json	library to handle JSON files. is a very common data format used for asynchronous browser–server communication, including as a replacement for XML in some AJAX-style systems.
request	to handle requests. It is to make HTTP requests simpler and more human-friendly.
matplotlib	It provides an object-oriented API for embedding plots into applications
scikit-learn	Simple and efficient tools for data mining and data analysis
cluster	allows to create several groups (clusters) of objects from a list.
folium	Helps create several types of Leaflet maps
geopy	makes it easy to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources

Table. 1

a. *Toronto Data*

Looking at the content and Structure of the data: The first 15 rows of the Toronto data making up of Postcode, Borough and Neighbourhood data as presented in a pandas dataframe. Putting together neighborhoods which have the same postcode in one row under the particular postcode

	Postcode	Borough	Neighbourhood
0	M1B	Scarborough	Rouge,Malvern
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union
2	M1E	Scarborough	Guildwood,Morningside,West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	East Birchmount Park,Ionview,Kennedy Park
7	M1L	Scarborough	Clairlea,Golden Mile,Oakridge
8	M1M	Scarborough	Cliffcrest,Cliffside,Scarborough Village West
9	M1N	Scarborough	Birch Cliff,Cliffside West
10	M1P	Scarborough	Dorset Park,Scarborough Town Centre,Wexford He...
11	M1R	Scarborough	Maryvale,Wexford
12	M1S	Scarborough	Agincourt
13	M1T	Scarborough	Clarks Corners,Sullivan,Tam O'Shanter
14	M1V	Scarborough	Agincourt North,L'Amoreaux East,Milliken,Steel...
15	M1W	Scarborough	L'Amoreaux West

Fig. 2 initial dataframe of Toronto

Neighbourhood with missing names "Not assigned" are given the corresponding Borough names.

After importing csv file into a pandas dataframe containing the corresponding coordinates of the city of Toronto merging the coordinates together with the city names and postcode in Toronto

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park,Ionview,Kennedy Park	43.727929	-79.262029
7	M1L	Scarborough	Clairlea,Golden Mile,Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffcrest,Cliffside,Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff,Cliffside West	43.692657	-79.264848
10	M1P	Scarborough	Dorset Park,Scarborough Town Centre,Wexford He...	43.757410	-79.273304
11	M1R	Scarborough	Maryvale,Wexford	43.750072	-79.295849
12	M1S	Scarborough	Agincourt	43.794200	-79.262029
13	M1T	Scarborough	Clarks Corners,Sullivan,Tam O'Shanter	43.781638	-79.304302
14	M1V	Scarborough	Agincourt North,L'Amoreaux East,Milliken,Steel...	43.815252	-79.284577
15	M1W	Scarborough	L'Amoreaux West	43.799525	-79.318389

Fig. 3 dataframe of Toronto data without missing values

Summary of Boroughs and their Neighbourhoods in Toronto

```
Borough
Central Toronto      9
Downtown Toronto    18
East Toronto         5
East York            5
Etobicoke           12
Mississauga          1
North York          24
Queen's Park         1
Scarborough         17
West Toronto         6
York                 5
```

Fig 4 summary of Boroughs and their number of neighbourhoods in Toronto

Collecting and analyzing data for North York.

North York was first considered to be the Neighbourhood in Toronto for the establishment of the business. It has the highest number of Neighbourhoods, 24 with a population of 656 thousand compared to Scarborough having 17 Neighbourhoods and a population of 632 thousand.

Its data was collected and analysed in an attempt to pick it as the Main contender for Queens in New York but after looking closely at its location on the map as shown below, Scarborough was chosen instead.

North York latitude 43.7708175 & longitude -79.4132998

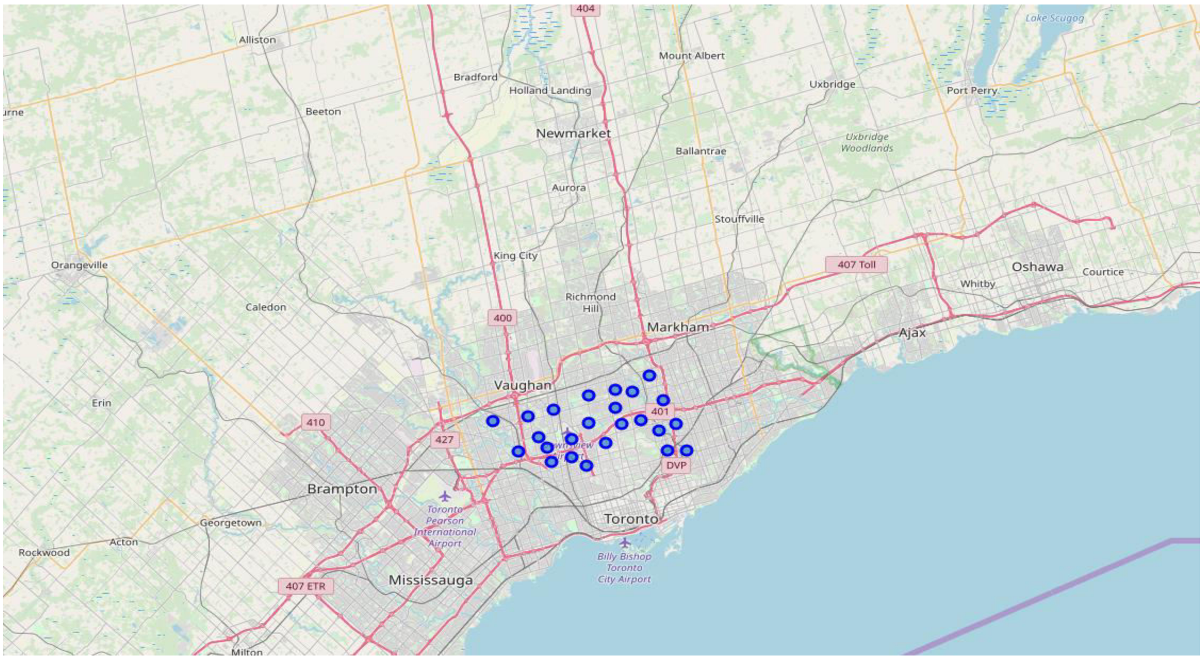

Fig. 5 map of North York

Map of North York above clearly shows it is away from the coast which was the reason Scarborough was chosen instead of North York.

Most tourists and foot traffic are heavily concentrated around beaches thereby increasing the buying of fast and exotic food especially in Summer.

Down Town Toronto was not chosen despite its higher Neighbourhoods of 18 because its population is just 199 thousand which is less than that of Scarborough

Scarborough

Scarborough which has almost the biggest Boroughs (17 Neighbourhoods) in Toronto with almost the highest population in Toronto of 632 thousand was chosen. Among the reasons why Scarborough was chosen ahead of North York in Toronto is that Scarborough just like Queens is closer to the coast than North York, despite North York having slightly higher population and more Neighbourhoods

The first five results of the Neighbourhoods in Scarborough with Rouge, Malvern being the first Neighbourhood data is given in the dataframe below

	Borough	Neighborhood	Latitude	Longitude
0	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497
2	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711
3	Scarborough	Woburn	43.770992	-79.216917
4	Scarborough	Cedarbrae	43.773136	-79.239476

Fig. 6 dataframe of Scarborough data with coordinates

Map of Scarborough

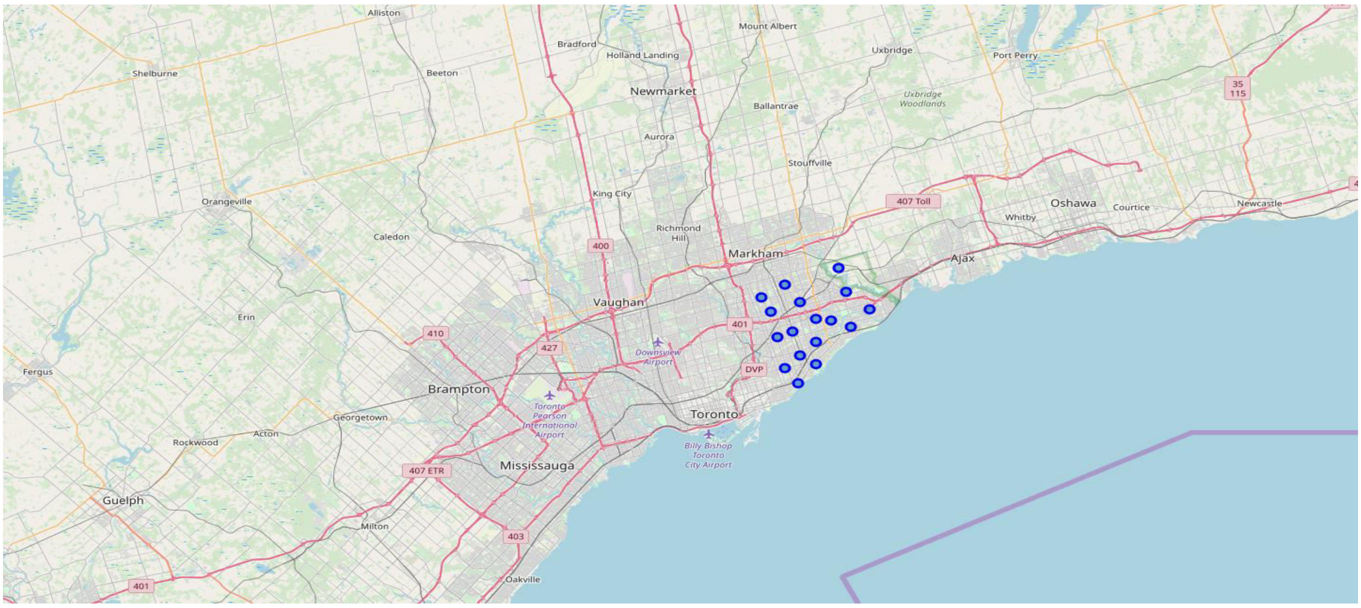

Fig. 7 map of scarborough

Cluster map of Scarborough

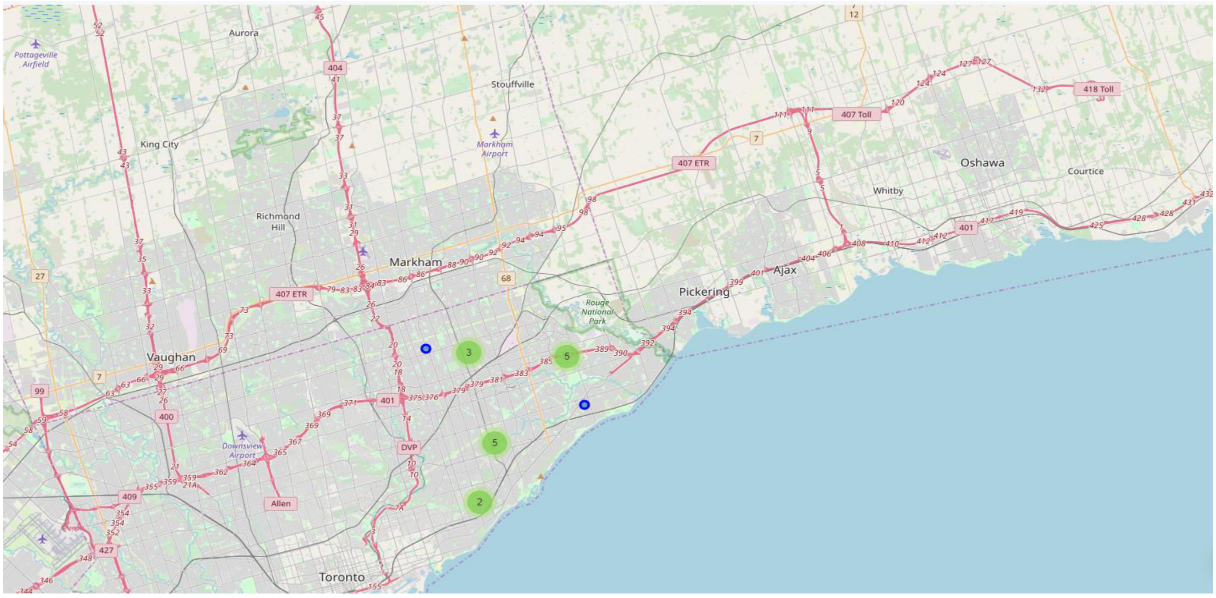

Fig. 8 cluster map of Scarborough

1st Data Collection

Getting the top 100 venues within a radius of 500 meters of Rouge, Malvern

After cleaning the json and structuring it into a pandas dataframe.

	name	categories	lat	lng
0	Wendy's	Fast Food Restaurant	43.807448	-79.199056

Fig. 9 Only one result was returned for the 500m radius

2nd Data Collection:

Due to the very low number of venues returned in the first data collection, a second data collection is made where the radius is increased from 500 to 1000

Getting the top 100 venues that are in Rouge, Malvern within a radius of 1000 meters.

After cleaning the json and structuring it into a pandas dataframe

And checking how many venues were returned by Foursquare

	name	categories	lat	lng
0	Images Salon & Spa	Spa	43.802283	-79.198565
1	Staples Morningside	Paper / Office Supplies Store	43.800285	-79.196607
2	Caribbean Wave	Caribbean Restaurant	43.798558	-79.195777
3	Wendy's	Fast Food Restaurant	43.802008	-79.198080
4	Wendy's	Fast Food Restaurant	43.807448	-79.199056

Fig. 10 First five results of 19 venues returned for radius 1000

b. New York Data

importing data for preprocessing and understanding the New York City data and putting it into a dataframe. The following was used:

!wget -q -O 'newyork_data.json' https://cocl.us/new_york_dataset

	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
10	Bronx	Baychester	40.866858	-73.835798
11	Bronx	Pelham Parkway	40.857413	-73.854756
12	Bronx	City Island	40.847247	-73.786488
13	Bronx	Bedford Park	40.870185	-73.885512
14	Bronx	University Heights	40.855727	-73.910416

Fig. 11 imported data are placed in a dataframe with its coordinates

Summary of Boroughs and their neighbourhoods in New York

```
Borough
Bronx            52
Brooklyn         70
Manhattan        40
Queens           81
Staten Island    63
```

Fig. 12

Queens

Queens Borough, New York has the most Neighbourhoods, 81 and a population of about 2.4 million (wikipedia) was chosen as the City in New York to contend for the Doner business establishment.

Getting the geographical coordinates of Queens, New York and part of its Neighbourhood data

Queens latitude 40.6504178 & longitude -73.7971341

	Borough	Neighborhood	Latitude	Longitude
0	Queens	Astoria	40.768509	-73.915654
1	Queens	Woodside	40.746349	-73.901842
2	Queens	Jackson Heights	40.751981	-73.882821
3	Queens	Elmhurst	40.744049	-73.881656
4	Queens	Howard Beach	40.654225	-73.838138
5	Queens	Corona	40.742382	-73.856825
6	Queens	Forest Hills	40.725264	-73.844475
7	Queens	Kew Gardens	40.705179	-73.829819
8	Queens	Richmond Hill	40.697947	-73.831833
9	Queens	Flushing	40.764454	-73.831773
10	Queens	Long Island City	40.750217	-73.939202
11	Queens	Sunnyside	40.740176	-73.926916
12	Queens	East Elmhurst	40.764073	-73.867041
13	Queens	Maspeth	40.725427	-73.896217
14	Queens	Ridgewood	40.708323	-73.901435

Fig. 13 Queens Borough was chosen to represent New York

creating map of Queens using latitude and longitude values

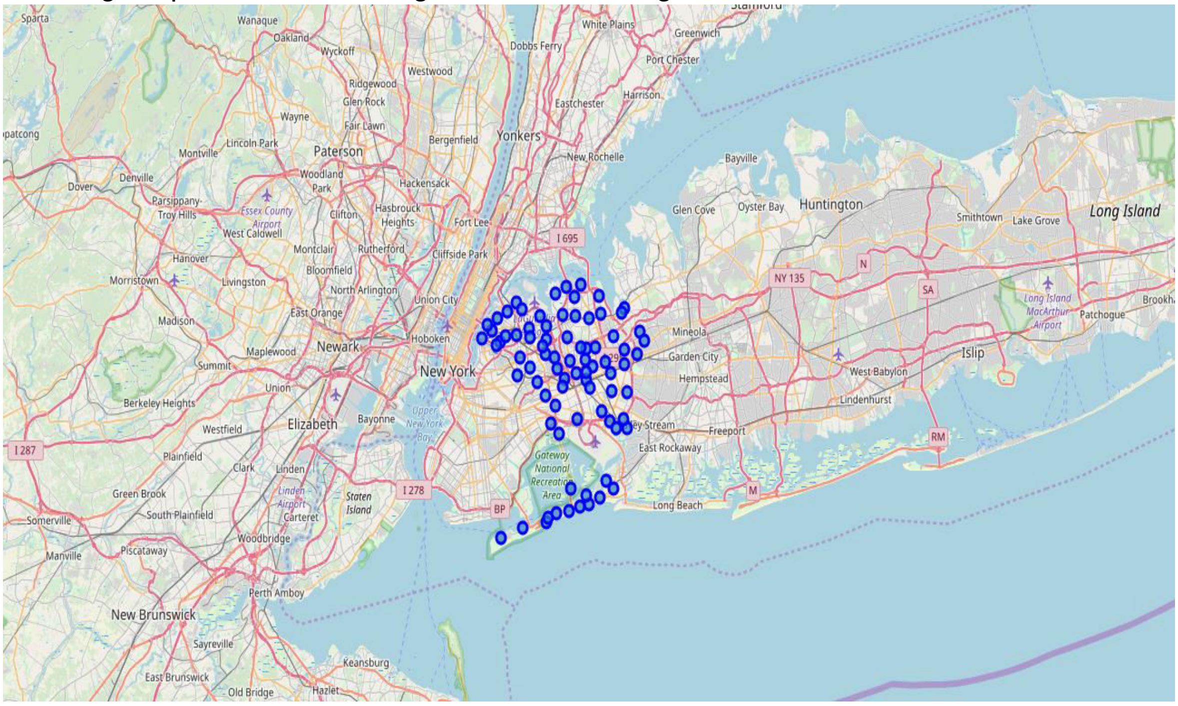

Fig. 14 map of queens with its Neighbourhoods

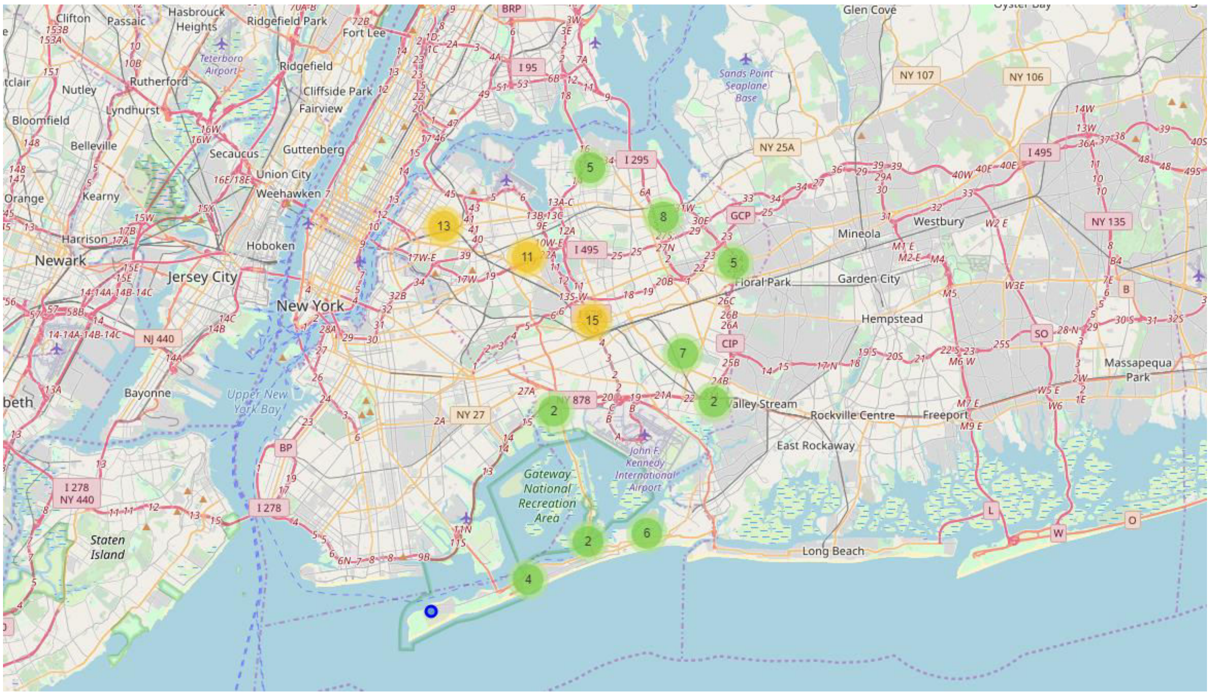

Fig. 15 visualising the cluster map of Queens

Exploring the first Neighborhood in our Queens dataframe by getting the neighborhood's name which is “Astoria”

Getting the top 100 venues that are in Astoria within a radius of 1000 meters.

Exploring the first Neighborhood in our dataframe by getting the neighborhood's name, Astoria. Getting the top 100 venues that are in Astoria within a radius of 1000 meters.

The dataframe for the first 5 results are as follows

	name	categories	lat	lng
0	Favela Grill	Brazilian Restaurant	40.767348	-73.917897
1	Titan Foods Inc.	Gourmet Shop	40.769198	-73.919253
2	CrossFit Queens	Gym	40.769404	-73.918977
3	Sitan Muay Thai	Martial Arts Dojo	40.766108	-73.913224
4	Al-sham Sweets and Pastries	Middle Eastern Restaurant	40.768077	-73.911561

Fig. 16 top 100 venues within a radius of 1000m from Astoria

checking how many venues were returned by Foursquare.100 venues were returned by Foursquare.

iii. Data Preparation

a. Toronto Data

II Exploring Neighborhoods in Scarborough by creating a function to repeat the same process to all the neighborhoods in Scarborough

checking the size of the resulting dataframe (90, 7) i.e 90 rows and 7 columns

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rouge,Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge,Malvern	43.806686	-79.194353	Interprovincial Group	43.805630	-79.200378	Print Shop
2	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
3	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	Affordable Toronto Movers	43.787919	-79.162977	Moving Target
4	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	Scarborough Historical Society	43.788755	-79.162438	History Museum

Fig. 17 exploring neighbourhoods in Scarborough checking how many venues were returned for each neighborhood

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Agincourt	4	4	4	4	4	4
Agincourt North,L'Amoreaux East,Milliken,Steeles East	3	3	3	3	3	3
Birch Cliff,Cliffside West	4	4	4	4	4	4
Cedarbrae	8	8	8	8	8	8
Clairlea,Golden Mile,Oakridge	9	9	9	9	9	9
Clarks Corners,Sullivan,Tam O'Shanter	13	13	13	13	13	13
Cliffcrest,Cliffside,Scarborough Village West	2	2	2	2	2	2
Dorset Park,Scarborough Town Centre,Wexford Heights	7	7	7	7	7	7
East Birchmount Park,Ionview,Kennedy Park	5	5	5	5	5	5
Guildwood,Morningside,West Hill	7	7	7	7	7	7
Highland Creek,Rouge Hill,Port Union	3	3	3	3	3	3
L'Amoreaux West	12	12	12	12	12	12
Maryvale,Wexford	6	6	6	6	6	6
Rouge,Malvern	2	2	2	2	2	2
Scarborough Village	2	2	2	2	2	2
Woburn	3	3	3	3	3	3

Fig. 18 Scarborough Neighbourhoods and corresponding venues

b. New York Data

Exploring Neighborhoods in Queens by creating a function to repeat the same process to all the neighborhoods in Queens, running a function on each neighborhood and create a new dataframe called Queens_venues

checking the size of the resulting dataframe (2154, 7) i.e 2154 rows and 7 columns

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Astoria	40.768509	-73.915654	Favela Grill	40.767348	-73.917897	Brazilian Restaurant
1	Astoria	40.768509	-73.915654	CrossFit Queens	40.769404	-73.918977	Gym
2	Astoria	40.768509	-73.915654	Titan Foods Inc.	40.769198	-73.919253	Gourmet Shop
3	Astoria	40.768509	-73.915654	Orange Blossom	40.769856	-73.917012	Gourmet Shop
4	Astoria	40.768509	-73.915654	Off The Hook	40.767200	-73.918104	Seafood Restaurant

Fig. 19 first five results of venues in Queens checking how many venues were returned for each Neighbourhood

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Arverne	18	18	18	18	18	18
Astoria	100	100	100	100	100	100
Astoria Heights	13	13	13	13	13	13
Auburndale	17	17	17	17	17	17
Bay Terrace	40	40	40	40	40	40
Bayside	74	74	74	74	74	74
Bayswater	4	4	4	4	4	4
Beechhurst	13	13	13	13	13	13
Bellaire	11	11	11	11	11	11
Belle Harbor	19	19	19	19	19	19
Bellerose	19	19	19	19	19	19
Blissville	21	21	21	21	21	21
Breezy Point	5	5	5	5	5	5
Briarwood	11	11	11	11	11	11
Broad Channel	5	5	5	5	5	5
Brookville	1	1	1	1	1	1
Cambria Heights	13	13	13	13	13	13
College Point	45	45	45	45	45	45
Corona	21	21	21	21	21	21
Douglaston	18	18	18	18	18	18
East Elmhurst	11	11	11	11	11	11

Fig. 20

Finding out how many unique categories can be curated from all the returned venues. There are 270 unique categories.

iv. Modelling and Evaluation

a. Toronto Data

Analyzing Each Neighborhood and checking the size of the new dataframe `(90, 55)` i.e 90 rows and 55 columns. Then grouping rows by neighborhood and by taking the mean of the frequency o f occurrence of each category. checking the new size `(16, 55)`i.e 16 rows and 55 columns

creating the new dataframe and display the top 10 venues for each neighborhood.

	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Clothing Store	Skating Rink	Breakfast Spot	Lounge	Coffee Shop	Grocery Store	General Entertainment	Gaming Cafe	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant
1	Agincourt North,L'Amoreaux East,Milliken,Steel...	Playground	Park	Coffee Shop	Vietnamese Restaurant	Clothing Store	Grocery Store	General Entertainment	Gaming Cafe	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant
2	Birch Cliff,Cliffside West	General Entertainment	Skating Rink	Café	College Stadium	Vietnamese Restaurant	Coffee Shop	Grocery Store	Gaming Cafe	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant
3	Cedarbrae	Thai Restaurant	Athletics & Sports	Bakery	Bank	Fried Chicken Joint	Lounge	Caribbean Restaurant	Hakka Restaurant	Hakka Restaurant	Department Store	Gym Pool
4	Clairlea,Golden Mile,Oakridge	Bakery	Fast Food Restaurant	Soccer Field	Metro Station	Intersection	Park	Bus Line	Bus Station	Bus Station	Electronics Store	Department Store

Fig. 21 first five results of top 10 venues running k-means to cluster the neighborhood into 5 clusters and merging data for each neighborhood

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Scarborough	Rouge,Malvern	43.806686	-79.194353	0.0	Fast Food Restaurant	Print Shop	Vietnamese Restaurant	Clothing Store
1	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	1.0	History Museum	Bar	Moving Target	Coffee Shop
2	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711	1.0	Intersection	Electronics Store	Rental Car Location	Breakfast Spot
3	Scarborough	Woburn	43.770992	-79.216917	3.0	Coffee Shop	Korean Restaurant	Vietnamese Restaurant	Grocery Store
4	Scarborough	Cedarbrae	43.773136	-79.239476	1.0	Thai Restaurant	Athletics & Sports	Bakery	Bank

Fig. 22 initial clustering

Examining the Clusters:

each cluster is examined to determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, a name for each cluster can be assigned.

first cluster (fast food, Vietnamese restaurant, fried chicken joint)

	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	Rouge,Malvern	Fast Food Restaurant	Print Shop	Vietnamese Restaurant	Clothing Store	Grocery Store	General Entertainment	Gaming Cafe	Fried Chicken Joint	Electronics Store	Discount Store

Fig. 23 1st cluster Fast food Restaurant is the 1st most common venue

(Thai, Indian, Chinese, Caribbean restaurants, pizza place)

	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
1	Highland Creek,Rouge Hill,Port Union	History Museum	Bar	Moving Target	Coffee Shop	Grocery Store	General Entertainment		Fried Chicken Joint	Fast Food Restaurant	Electronics Store
2	Guildwood,Morningside,West Hill	Intersection	Electronics Store	Rental Car Location	Breakfast Spot	Pizza Place	Medical Center	Mexican Restaurant	Vietnamese Restaurant	Coffee Shop	Gaming Cafe
4	Cedarbrae	Thai Restaurant	Athletics & Sports	Bakery	Bank	Fried Chicken Joint	Lounge	Caribbean Restaurant	Hakka Restaurant	Department Store	Gym Pool
6	East Birchmount Park,Ionview,Kennedy Park	Coffee Shop	Discount Store	Department Store	Convenience Store	Chinese Restaurant	Vietnamese Restaurant	Gym Pool	Grocery Store	General Entertainment	Gaming Cafe
7	Clairlea,Golden Mile,Oakridge	Bakery	Fast Food Restaurant	Soccer Field	Metro Station	Intersection	Park	Bus Line	Bus Station	Electronics Store	Department Store
9	Birch Cliff,Cliffside West	General Entertainment	Skating Rink	Café	College Stadium	Vietnamese Restaurant	Coffee Shop		Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant
10	Dorset Park,Scarborough Town Centre,Wexford He...	Indian Restaurant	Vietnamese Restaurant	Gaming Cafe	Latin American Restaurant	Pet Store	Chinese Restaurant		General Entertainment	Fried Chicken Joint	Fast Food Restaurant
11	Maryvale,Wexford	Smoke Shop	Auto Garage	Shopping Mall	Sandwich Place	Middle Eastern Restaurant	Breakfast Spot	Restaurant	Discount Store	College Stadium	Convenience Store
12	Agincourt	Clothing Store	Skating Rink	Breakfast Spot	Lounge	Coffee Shop	Grocery Store	General	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant
13	Clarks Corners,Sullivan,Tam O'Shanter	Pizza Place	Breakfast Spot	Noodle House	Fried Chicken Joint	Thai Restaurant	Fast Food Restaurant	Shopping Mall	Bank	Rental Car Location	Chinese Restaurant
14	Agincourt North,L'Amoreaux East,Milliken,Steel...	Playground	Park	Coffee Shop	Vietnamese Restaurant	Clothing Store	Grocery Store	General	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant
15	L'Amoreaux West	Coffee Shop	Chinese Restaurant	Fast Food Restaurant	Grocery Store	Gym Pool	Pharmacy		Breakfast Spot	Sandwich Place	Electronics Store

Fig. 24 2nd cluster has Indian, Thai and Pizza as the 1st most common venues

(Vietnamese, Fast Food restaurant)

	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
5	Scarborough Village	Grocery Store	Playground	Vietnamese Restaurant	Clothing Store	General Entertainment	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store

Fig. 25 3rd cluster

(Korean, coffee shop, fast food restaurant)

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	lost mon e	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Woburn	Coffee Shop	Korean Restaurant	Vietnamese Restaurant	Grocery Store	ral tainment	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store

Fig. 26 4th cluster

(Fried Chicken joint, Fast Food restaurant)

	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
8	Cliffcrest,Cliffside,Scarborough Village West	American Restaurant	Motel	Coffee Shop	Grocery Store	General Entertainment	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store

Fig. 27 5th cluster

b. New York Data

Clustering Neighborhoods

first results

	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	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	Queens	Astoria	40.768509	-73.915654	0	Bar	Middle Eastern Restaurant	Greek Restaurant	Hookah Bar	Seafood Restaurant	Hookah Bar	Seafood Restaurant	Mediterranean Restaurant	Bakery	Ice Cream Shop	Salon / Barbershop	Chinese Restaurant
1	Queens	Woodside	40.746349	-73.901842	0	Grocery Store	Thai Restaurant	Bakery	Filipino Restaurant	Latin American Restaurant	Filipino Restaurant	Latin American Restaurant	American Restaurant	Donut Shop	Pizza Place	Pub	Bar
2	Queens	Jackson Heights	40.751981	-73.882821	0	Latin American Restaurant	Peruvian Restaurant	Bakery	South American Restaurant	Mobile Phone Shop	South American Restaurant	Mobile Phone Shop	Thai Restaurant	Mexican Restaurant	Diner	Spanish Restaurant	Clothing Store
3	Queens	Elmhurst	40.744049	-73.881656	0	Thai Restaurant	Mexican Restaurant	Chinese Restaurant	South American Restaurant	Bubble Tea Shop	South American Restaurant	Bubble Tea Shop	Vietnamese Restaurant	Bar	Sushi Restaurant	Malay Restaurant	Snack Place
4	Queens	Howard Beach	40.654225	-73.838138	0	Italian Restaurant	Ice Cream Shop	Fast Food Restaurant	Sandwich Place	Bagel Shop	Sandwich Place	Bagel Shop	Pharmacy	Deli / Bodega	Jewelry Store	Gym	Breakfast Spot

Fig. 28

a visualization of the resulting clusters

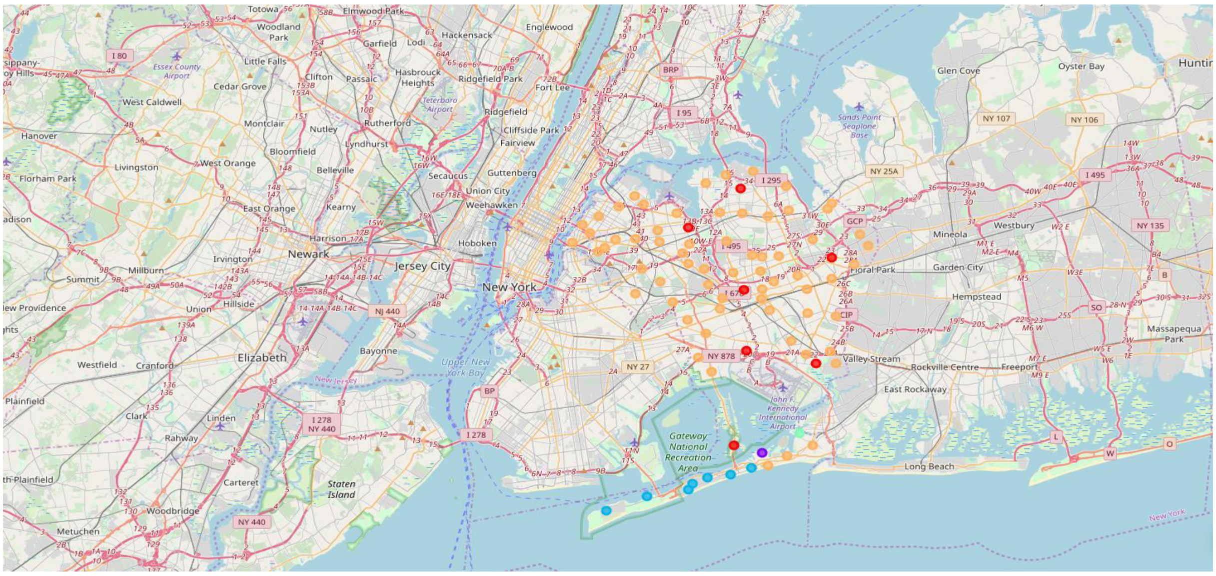

Fig. 29 resulting cluster map of Queens Borough

Examining the Clusters

each cluster is examined to determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, a name for each cluster can be assigned.

(Korean, Filipino, Indian, Pizza, Sushi, Vietnamese, Chinese, Italian, Mexican, Greek, Caribbean, Thai restaurants)

	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	Astoria	Bar	Middle Eastern Restaurant	Greek Restaurant	Hookah Bar	Seafood Restaurant	Mediterranean Restaurant	Bakery	Ice Cream Shop	Salon / Barbershop	Chinese Restaurant
1	Woodside	Grocery Store	Thai Restaurant	Bakery	Filipino Restaurant	Latin American Restaurant	American Restaurant	Donut Shop	Pizza Place	Pub	Bar
2	Jackson Heights	Latin American Restaurant	Peruvian Restaurant	Bakery	South American Restaurant	Mobile Phone Shop	Thai Restaurant	Mexican Restaurant	Diner	Spanish Restaurant	Clothing Store
3	Elmhurst	Thai Restaurant	Mexican Restaurant	Chinese Restaurant	South American Restaurant	Bubble Tea Shop	Vietnamese Restaurant	Bar	Sushi Restaurant	Malay Restaurant	Snack Place
4	Howard Beach	Italian Restaurant	Ice Cream Shop	Fast Food Restaurant	Sandwich Place	Bagel Shop	Pharmacy	Deli / Bodega	Jewelry Store	Gym	Breakfast Spot
5	Corona	Bakery	Mexican Restaurant	Ice Cream Shop	Pizza Place	Convenience Store	Deli / Bodega	Park	Restaurant	Donut Shop	Sandwich Place
6	Forest Hills	Gym / Fitness Center	Gym	Yoga Studio	Pizza Place	Convenience Store	Park	Thai Restaurant	Pharmacy	Video Game Store	Optical Shop
7	Kew Gardens	Chinese Restaurant	Donut Shop	Pet Store	Pharmacy	Cosmetics Shop	Bank	Pizza Place	Bar	Park	Indian Restaurant
8	Richmond Hill	Pizza Place	Bank	Latin American Restaurant	Lounge	Caribbean Restaurant	Bakery	Deli / Bodega	Pet Service	Diner	Discount Store
9	Flushing	Hotpot Restaurant	Chinese Restaurant	Korean Restaurant	Bubble Tea Shop	Construction & Landscaping	Asian Restaurant	Gym / Fitness Center	Bakery	Gym	Karaoke Bar
10	Long Island City	Coffee Shop	Hotel	Pizza Place	Bar	Café	Deli / Bodega	Mexican Restaurant	Gym / Fitness Center	Donut Shop	Chinese Restaurant
11	Sunnyside	Pizza Place	Italian Restaurant	South American Restaurant	Grocery Store	Bakery	Hotel	Discount Store	Chinese Restaurant	Coffee Shop	Taco Place
12	East Elmhurst	Donut Shop	Ice Cream Shop	Coffee Shop	Bus Station	Supermarket	Gas Station	Lake	Latin American Restaurant	Rental Car Location	Hotel Bar
13	Maspeth	Pizza Place	Diner	Grocery Store	Mobile Phone Shop	Bank	Ice Cream Shop	Bakery	Sandwich Place	Sushi Restaurant	Lounge
14	Ridgewood	Mexican Restaurant	Bank	Bakery	Pizza Place	Mobile Phone Shop	Greek Restaurant	Grocery Store	Restaurant	Korean Restaurant	Sushi Restaurant
15	Glendale	Arts & Crafts Store	Pizza Place	Brewery	Food & Drink Shop	Food	Flower Shop	Fish Market	Fish & Chip Shop	Filipino Restaurant	Electronics Store

Fig. 30 1st cluster has many exotic restaurants as part of the first most common venue

(park, falafel, European, Eastern European, empanada restaurant)

	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
63	Somerville	Park	Women's Store	Farm	Electronics Store	Empanada Restaurant	Event Space	Eye Doctor	Falafel Restaurant	Farmers Market	Eastern European Restaurant
79	Bayswater	Park	Playground	Women's Store	Farm	Electronics Store	Empanada Restaurant	Event Space	Eye Doctor	Falafel Restauran	Farmers Market

Fig. 31 2nd cluster even though park is the first most common venue there are other exotic restaurants making the top ten cut

(Beach, Fast food, Falafel, Empanada restaurant)

	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
43	Breezy Point	Beach	Monument / Landmark	American Restaurant	Board Shop	Trail	Women's Store	Farmers Market	Event Space	Eye Doctor	Falafel Restaurant
50	Neponsit	Beach	Bus Stop	Beach Bar	Fast Food Restaurant	Eye Doctor	Falafel Restaurant	Farm	Farmers Market	Women's Store	Empanada Restaurant
78	Hammels	Beach	Dog Run	Shoe Store	Southern / Soul Food Restaurant	Deli / Bodega	Building	Diner	Café	Bus Station	Gym / Fitness Center

Fig. 32 3rd cluster even though Beach is the first most common venue in this cluster a couple of exotic food restaurants made the top ten

(Falafel, empanada, Fast Food restaurant)

	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
64	Brookville	Deli / Bodega	Women's Store	Farmers Market	Electronics Store	Empanada Restaurant	Event Space	Eye Doctor	Falafel Restaurant	Farm	Fast Food Restaurant

Fig. 33 4th cluster

(Empanada, Falafel, Filipino, fast food restaurant)

	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
54	Jamaica Estates	Eye Doctor	Intersection	Women's Store	Fast Food Restaurant	Empanada Restaurant	Event Space	Falafel Restaurant	Farm	Farmers Market	Filipino Restaurant

Fig. 34 5th cluster

Deployment and Feedback

When the decision on the city is made and the business is established, say New York is chosen, there should be continuous feedback fed into the model. Questions whose answers should be input into the model are “are the numerous exotic restaurants in the Neighbourhood acting as a factor for high patronage or acting as competition therefore less patronage”

3. Results

After clustering the data of the respective neighborhoods, both cities (Boroughs) have venues which can be explored for exotic restaurants. In the first cluster the first of Queens, the most common venues includes exotic restaurants like Latin American, Thai, Mexican, Chinese and Pizza restaurants. Toronto had its second cluster also made up of a number of exotic restaurants.

The neighborhoods are much similar with regards to them having wide ranging exotic fast food restaurants. The main difference is that New York with its large population and many Neighbourhoods offers more exotic food restaurants than Toronto.

4. Conclusion

The descriptive model created shows the location of exotic restaurants in North America`s two financial capitals New York and Toronto and the top most common venues. From the clusters we can see that exotic restaurants, such as Chinese, Thai, and Greek make up the top most visited places. It is also clear that despite Toronto having its share of exotic restaurants, New York with its large population and many Neighbourhoods offers more exotic restaurants. The decision to choose to establish a Doner business in New York over Toronto can be established. However, Toronto has the advantage of offering less competition but a population that seems very interested in exotic restaurants.

5. Future Directions

A more thorough analysis could be made by comparing income, spending on food especially exotic, fast food, etc. Also compare North York (Toronto) with Queens (New York) or Manhattan in New York with North York. The distance between the venues can also be computed to find a place with the highest number of customers or visits to make the location of business establishment optimum.

References:

1. *Wikipedia*
2. *IBM Cognitive Class Labs*
3. *IBM Data Science Methodology lecture*