A Recommender Systems for;

1) Idea of opening new Restaurant

2) Best area suggestion to contractor for opening new business.

Applied Data Science Capstone Project

PALLAVI VISHWANATH KULKARNI

Use Case : A Recommender System for Best area to open new Restaurant and New Business

Introduction and Business problem :

This project is about **finding the best neighborhoods in the city of Toronto to open a new restaurant of a specific type** (for example : Chinese, or Italian restaurant).

This project would interest anyone which wants to open a new restaurant in the city of Toronto, and seeks the best neighborhoods where the habitants would likely eat in this kind of restaurant, and where the competition is limited (e.g. there is a reasonable number of existing restaurants of the same type in the neighborhood).

A) **<u>Cluster Toronto neighborhoods using demographic data:</u>**

- When we look for the best place to open a new restaurant in a city like Toronto, **we have to gauge people's taste in each neighborhood of the city**. We will then know in what neighborhood of the city people will likely come and spend money in our new restaurant.
- A good way to gauge people's taste in a specific area is to **look into the demographic data of this area**. For example, areas with a majority of Chinese people would be good for Chinese restaurants, and areas with a majority of Italian people would be good for opening an Italian restaurant, etc.
- With this kind of demographic data associated with different neighborhoods of Toronto, **we can cluster neighborhoods by demographic data**. Thus, we will be able to distinguish the areas where a lot of Chinese people live, the areas where a lot of Italian people live, and so on, based on the clustering.

Introduction and Business problem :

B) **<u>Find the best neighborhoods within a cluster to open the restaurant :</u>**

- Once the neighborhoods have been categorized into clusters, and we've got a list of neighborhoods where people living there would likely want to eat in the restaurant we want to open, **we need to find out in which neighborhoods there is less competition**. It means that we have to find out **what neighborhoods contain the lowest number of existing restaurants of the same type** as the one we want to open.
- In order to count the number of existing restaurants of the same type in a neighborhood, we perform a **FoursquareAPI explore query**. Like that, we obtain the list of venues of each neighborhood, and we can count the number of restaurants of each type.

Data :

A) Demographic data from the City of Toronto's open data

- The list of neighborhoods, and the demographic data associated to each neighborhood, has been made available by the city of Toronto here :
- https://www.toronto.ca/ext/open_data/catalog/data_set_files/2016_neighbourhood_profiles.csv
- The Toronto demographic dataset contains multiple features such as :
- Citizenship
- **Ethnic origin**
- Income
- Languages / Mother tongue
- Marital status
- **Neighborhood information**
- Work activity
- Etc.
- For this project, we use the **Ethnic origin** and the **Neighborhood information** for each neighborhood, in order to cluster the neighborhoods of Toronto.

Data:

Examples of data from the dataset :

- *Neighborhood information data :*

We can see :

1. We have the **name of each neighborhood** in each column name (starting at position 6)
2. The **neighborhood number** (also called CDN number) in the first row (starting at position 6)

Neighbourhood

Category	Topic	Data Source	Characteristic	City of Toronto	Agincourt North	Agincourt South-Malvern West	Alderwood	Annex	Banbury-Don Mills	Bathurst Manor	Bay Street Corridor	Ba Vi
Neighbourhood Information	Neighbourhood Information	City of Toronto	Neighbourhood Number	n/a	129	128	20	95	42	34	76	52

Neighbourhood number (or CDN)

Data:

Examples of data from the dataset :

- *Ethnic origin data :*

We can see :

1. We have the **name of each neighborhood** in each column name (starting at position 6)
2. We have the **name of each ethnic origin** in the Characteristic column
3. The **number of people living in each neighborhood**, associated to each ethnic origin name.

Neighbourhood

Category	Topic	Data Source	Characteristic	City of Toronto	Agincourt North	Agincourt South-Malvern West	Alderwood	Annex	Banbury-Don Mills	Bat Ma
Ethnic origin	Ethnic origin population	Census Profile 98-316-X2016001	North American Aboriginal origins	"35,630"	40	105	305	475	230	90
Ethnic origin	Ethnic origin population	Census Profile 98-316-X2016001	First Nations (North American Indian)	"27,610"	25	90	200	345	175	75
Ethnic origin	Ethnic origin population	Census Profile 98-316-X2016001	Inuit	515	0	0	15	20	10	0
Ethnic origin	Ethnic origin population	Census Profile 98-316-X2016001	Métis	"8,465"	10	25	100	115	60	0

Ethnic origin

Number of people by neighbourhood and ethnic origin

Data :

B) List of venues by neighborhood using the FoursquareAPI :

- In order to obtain the list of venues, and especially the list of restaurants with the same type as the one we want to open, we are going to request FoursquareAPI with an Explore query. The documentation for the Explore query can be found here :
- https://developer.foursquare.com/docs/api/venues/explore
- We query FoursquareAPI supplying the neighborhood's information (coordinates calculated with the **Geocoder** package), the radius of scan, and the limit of number of venues we want to retrieve.
- Here is an example of a place in a specific neighborhood retrieved from a FoursquareAPI call :
- Link Of page :

We can see :

1. The place comes from the *Lawrence Park South* : **headerLocation tag**
2. The name of the place is *Wooden Woodworking Canada Inc.* : **Group -> Items -> Venue -> Name tag**
3. The category of the place is *Service à domicile* (English translation : Home service) : **Group -> Items -> Venue -> Categories -> Name tag**

Methodology :

As we previously saw, we use the following datasets :

- A list of general information about the neighborhoods (Neighborhood name, Number, and coordinates calculated using the Geocoder package)
- A list of demographic data about the neighborhoods, with the number of people of each ethnic origin living in these neighborhoods.

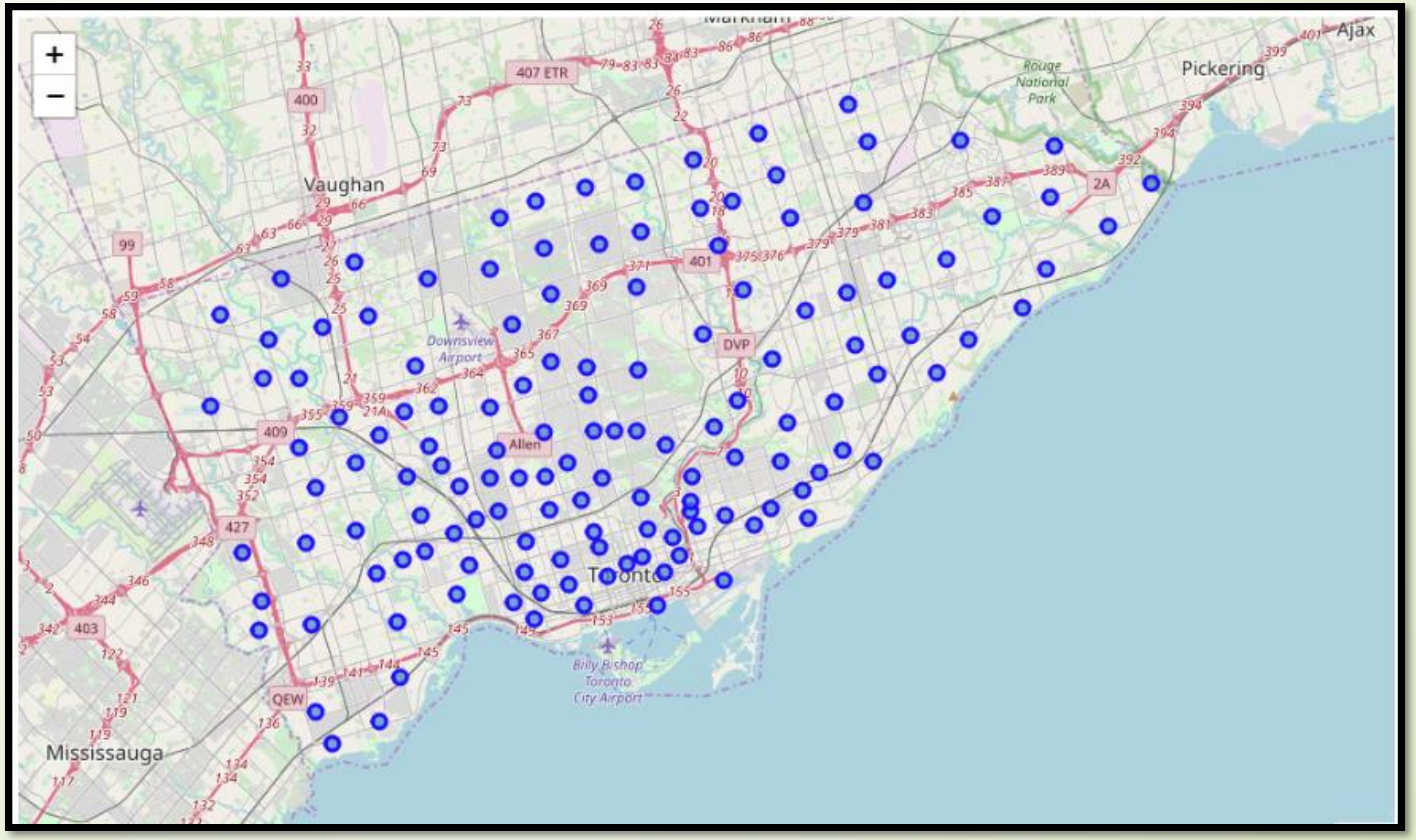
A good way to start our analysis is to draw each neighborhood over the map of the city of Toronto, in order to check if the dataset with the list of neighborhoods is complete and covers the whole city. For that, we need each neighborhood' coordinates.

As we saw, the neighborhoods' coordinates are not available in the **Neighborhood information data** dataset. So we are going to retrieve them using the **Geocoder** package. We then store each neighborhood's coordinates into a data frame, like this :

CDN	City_Area	Latitude	Longitude
129	Agincourt North	43.80930	-79.26707
128	Agincourt South-Malvern West	43.78735	-79.26941
20	Alderwood	43.60496	-79.54116
95	Annex	43.66936	-79.40280
42	Banbury-Don Mills	43.74041	-79.34852

Methodology :

The map of the city is displayed using the **Folium package**. On this map, we draw a blue circle for each neighborhood, using the neighborhoods' coordinates. It is a good way to visualize the position of each neighborhood in our dataset. It also confirms that the different neighborhoods are well distributed within the city, and that our dataset covers the whole city (no missing neighborhood).



Methodology:

- Once the map with the neighborhoods is displayed, we need to find out what are the top most common ethnic origins for each neighborhoods, in order to prepare our data for clustering by demographic data.
  We do this by counting the number of occurrences of each ethnic origin for each neighborhood, and sorting the ethnic origins by number of occurrence descending.
- Here are two examples of neighborhoods, with their top 5 most common ethnic origins of habitants, sorted by count descending :

Agincourt North

Origin	Count
Chinese	16950.0
Sri Lankan	2230.0
East Indian	2090.0
Filipino	1465.0
Canadian	1295.0

Alderwood

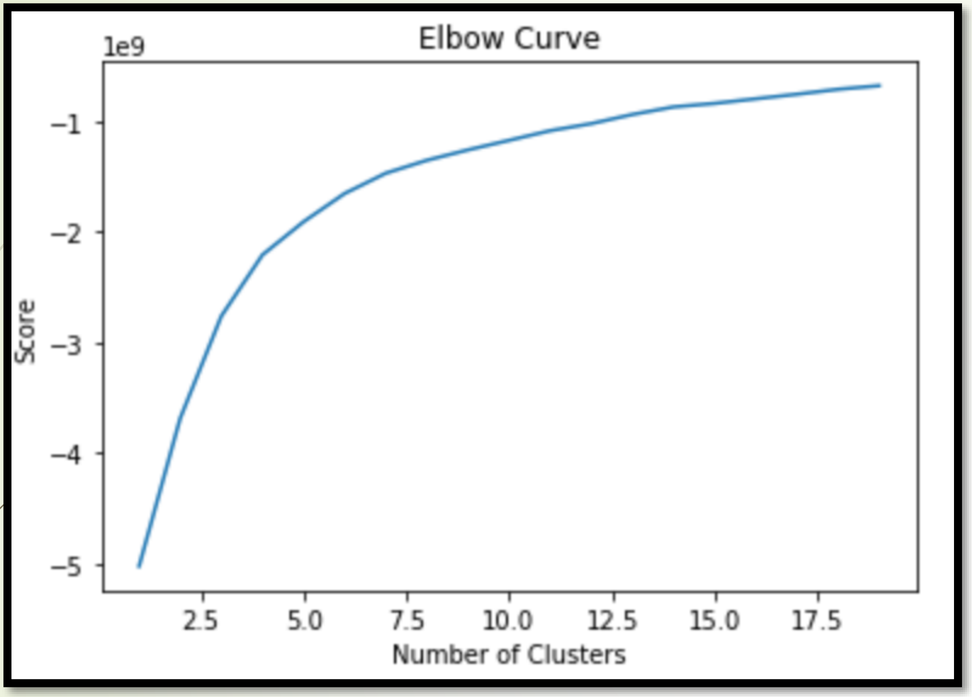
Origin	Count
English	2320.0
Canadian	2245.0
Irish	1900.0
Scottish	1720.0
Italian	1275.0

- This will allow us to perform a clustering based on demographic data.

Machine learning algorithm:

- For the clustering, we use a **K-Means algorithm**. I chose to use a K-Means algorithm, as it is one on the most used algorithm for unsupervised learning and clustering. It is typically used for scenarios like understanding the population demographics, market segmentation, social media trends, anomaly detection, etc... where the clusters are unknown to begin with. It is exactly our scenario, as we want to understand how the neighborhoods of Toronto are segmented, and the clusters to begin with are unknown in this situation.
  Also, K-Means is one of the simplest clustering algorithm to implement and to run, and is less time consuming than other, more complex algorithms.
- As we can see, the number of occurrences in the demographic data is a numerical value. This means that we can directly use this data for the clustering, we don't need to use any One Hot Encoding.
- In order to determine the best number of clusters for our dataset, which is the optimal K for our K-means algorithm, we are going to use the **Elbow method** as described here : https://blog.cambridgespark.com/how-to-determine-the-optimal-number-of-clusters-for-k-means-clustering-14f27070048f
- The Elbow method is a method to find the most appropriate number of clusters in a dataset, by **running several K-means algorithm and comparing the sum of squared distances of samples to the nearest cluster center**. The more the sum of squared distance is, the further the data points are globally from their cluster center. But we don't have to set K too high, as if K is set to the number of data points, then each sample will form its own cluster meaning sum of squared distances equals zero, which is not a good clustering.

Machine learning algorithm:



We are going to use **K = 5**, as the elbow is highly visible for this value.
Once the clustering is done, we obtain a dataset like this (example with two neighborhoods from two different clusters) :

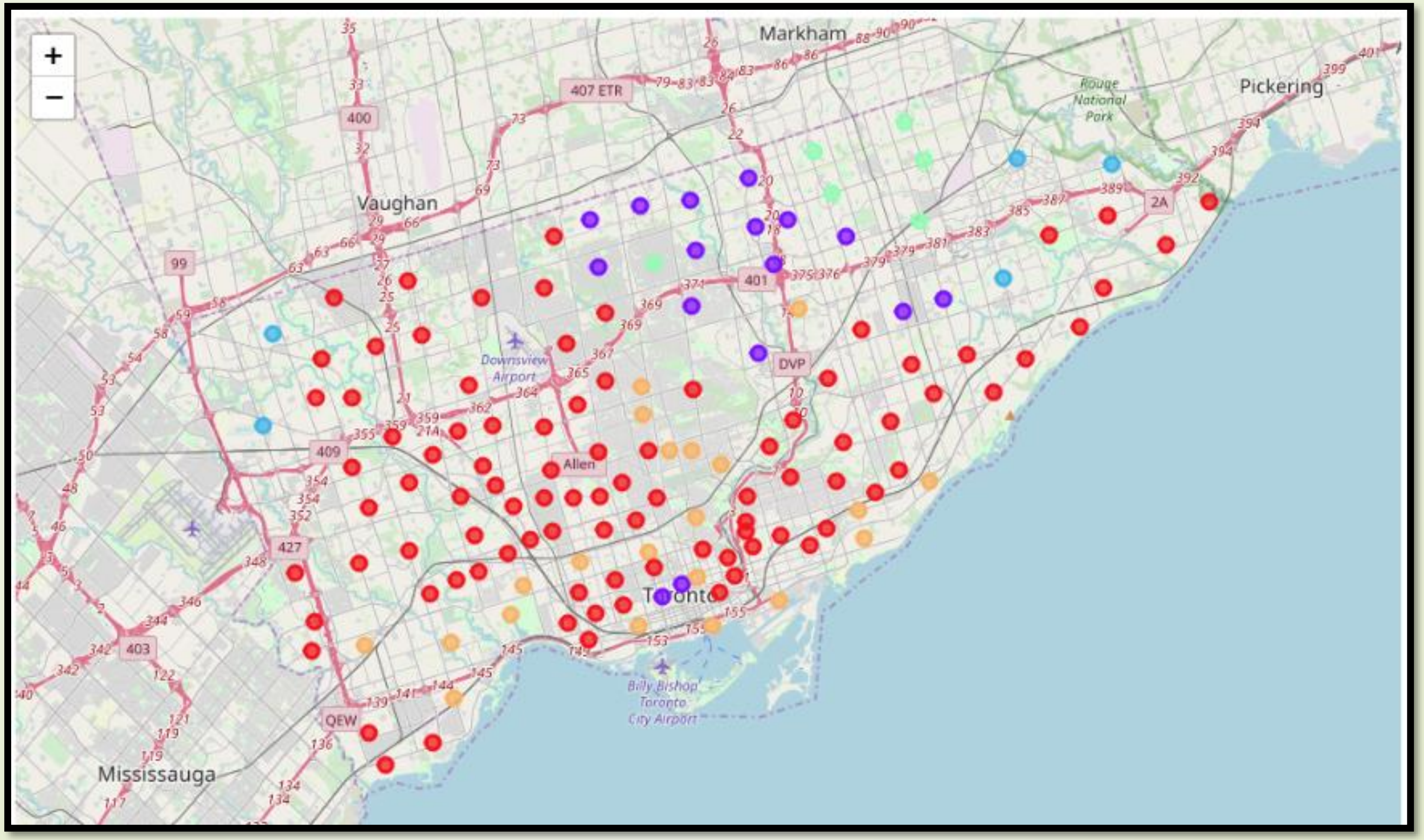
We have the CDN number, the name of the neighborhood, the coordinates, the cluster label obtained by the K-means algorithm, and the top most common ethnic origins by neighborhood.

Cluster

	PostalCode	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	9th Most Common Venue	10th Most Common Venue
0	M1B	Scarborough	Rouge, Malvern	43.811650	-79.195561	2	Zoo Exhibit	Fast Food Restaurant	Coffee Shop	Spa	Hobby Shop	Food	Falafel Restaurant	Eastern European Restaurant	Electronics Store	Empanada Restaurant
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.785605	-79.158701	1	Italian Restaurant	Playground	Breakfast Spot	Burger Joint	Farmers Market	Electronics Store	Empanada Restaurant	Ethiopian Restaurant	Event Space	Exhibit
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.765690	-79.175299	4	Pharmacy	Fast Food Restaurant	Pizza Place	Coffee Shop	Grocery Store	Sports Bar	Discount Store	Bank	Bar	Supermarket
3	M1G	Scarborough	Woburn	43.768216	-79.217610	4	Pizza Place	Fast Food Restaurant	Coffee Shop	Indian Restaurant	Bank	Sandwich Place	Supermarket	Juice Bar	Park	Thrift / Vintage Store
4	M1H	Scarborough	Cedarbrae	43.769608	-79.239440	1	Indian Restaurant	Bakery	Coffee Shop	Fried Chicken Joint	Hakka Restaurant	Supplement Shop	Sandwich Place	Caribbean Restaurant	Fish & Chips Shop	Athletics & Sports

Machine learning algorithm:

- We can then visualize the clusters on a Folium map. We display each neighborhood as a circle on the map, each circle will be colored according to the cluster they have been categorized into.



Discussion :

A) Clusters :

We obtain the following results :

- **Cluster 0 : European & Canadian (Red color)**

The **Cluster 0** regroups areas highly habited by **European and Canadian people**. We can see English, Italian, Portuguese, French people .
Most of them are positioned in almost all the south of Toronto, and in the downtown.

For example :

	PostalCode	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	9th Most Common Venue	10th Most Common Venue
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.696690	-79.260069	0	Park	American Restaurant	Skating Rink	General Entertainment	Café	Gym	Gym Pool	College Stadium	Cuban Restaurant	Dumpling Restaurant
19	M2K	North York	Bayview Village	43.781015	-79.380542	0	Park	Café	Bank	Chinese Restaurant	Trail	Japanese Restaurant	Zoo Exhibit	Eastern European Restaurant	Electronics Store	Empanada Restaurant
20	M2L	North York	Silver Hills, York Mills	43.757177	-79.379940	0	Park	Gym	Intersection	Gym / Fitness Center	Bus Stop	Furniture / Home Store	Candy Store	Farmers Market	Electronics Store	Donut Shop
23	M2P	North York	York Mills West	43.747810	-79.400062	0	Coffee Shop	Park	Gym	Tennis Court	French Restaurant	Restaurant	Seafood Restaurant	Intersection	Thai Restaurant	Gym / Fitness Center
31	M3L	North York	Downsview West	43.740945	-79.505004	0	Park	Spa	Coffee Shop	Vietnamese Restaurant	Pizza Place	Grocery Store	Bank	Exhibit	Dumpling Restaurant	Eastern European Restaurant
32	M3M	North York	Downsview Central	43.733610	-79.496750	0	Restaurant	Food Truck	Playground	Park	Baseball Field	Falafel Restaurant	Eastern European Restaurant	Electronics Store	Empanada Restaurant	Ethiopian Restaurant
44	M4N	Central Toronto	Lawrence Park	43.728135	-79.387090	0	Park	Coffee Shop	Café	Gym / Fitness Center	Trail	Bookstore	College Gym	College Quad	Flea Market	Falafel Restaurant

This cluster would interest anyone which wants to open a **European oriented restaurant**, for example an Italian or a French restaurant, as it contains the neighborhoods with the strongest European tendency within their habitants.

Discussion :

A) Clusters :

Cluster 1 : Asian (Purple color)

The **Cluster 1** regroups areas highly habited **Chinese people, and people from others countries in Asia**.
We can see that most of them are positioned at the north of Toronto.

For example :

	PostalCode	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	9th Most Common Venue	10th Most Common Venue
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.785605	-79.158701	1	Italian Restaurant	Playground	Breakfast Spot	Burger Joint	Farmers Market	Electronics Store	Empanada Restaurant	Ethiopian Restaurant	Event Space	Exhibit
4	M1H	Scarborough	Cedarbrae	43.769608	-79.239440	1	Indian Restaurant	Bakery	Coffee Shop	Fried Chicken Joint	Hakka Restaurant	Supplement Shop	Sandwich Place	Caribbean Restaurant	Fish & Chips Shop	Athletics & Sports
18	M2J	North York	Fairview, Henry Farm, Oriole	43.780810	-79.347782	1	Clothing Store	Coffee Shop	Japanese Restaurant	Sandwich Place	Bakery	Electronics Store	Caribbean Restaurant	Juice Bar	Supermarket	Bank
21	M2M	North York	Newtonbrook, Willowdale	43.791300	-79.413546	1	Korean Restaurant	Coffee Shop	Café	Middle Eastern Restaurant	Pizza Place	Sandwich Place	Shopping Mall	Grocery Store	Park	Dessert Shop
22	M2N	North York	Willowdale South	43.768165	-79.407420	1	Japanese Restaurant	Coffee Shop	Ramen Restaurant	Fast Food Restaurant	Korean Restaurant	Pizza Place	Sandwich Place	Café	Pharmacy	Sushi Restaurant
26	M3B	North York	Don Mills North	43.749055	-79.362212	1	Coffee Shop	Japanese Restaurant	Restaurant	Electronics Store	Bank	Dim Sum Restaurant	Fast Food Restaurant	Diner	Bagel Shop	Discount Store
27	M3C	North York	Flemingdon Park, Don Mills South	43.721375	-79.343415	1	Japanese Restaurant	Grocery Store	Gym	Coffee Shop	Asian Restaurant	Chinese Restaurant	Sandwich Place	Beer Store	Middle Eastern Restaurant	Fast Food Restaurant

This cluster would interest anyone which wants to open an **Asian restaurant**, for example a Chinese restaurant, as it contains the neighborhoods with the strongest Asian tendency within their habitants.

Discussion :

A) Clusters :

- **Cluster 2 : Indian (Dark green color)**

The **Cluster 2** concentrates areas highly habited by Indian people.
We can see that these areas are located at the edges of Toronto.

For example :

	PostalCode	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	9th Most Common Venue	10th Most Common Venue
0	M1B	Scarborough	Rouge, Malvern	43.81165	-79.195561	2	Zoo Exhibit	Fast Food Restaurant	Coffee Shop	Spa	Hobby Shop	Food	Falafel Restaurant	Eastern European Restaurant	Electronics Store	Empanada Restaurant

This cluster would interest anyone which wants to open an **Indian restaurant**.

Discussion :

A) Clusters :

- **Cluster 3 : Chinese (Light green color)**

The **Cluster 3** also regroups areas highly habited by Asian people, the most common ethnic origin is Chinese.
We can see that most of them are positioned at the north east of Toronto, next to the cluster 1. This cluster is highly similar to the cluster 1.

For example :

	PostalCode	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	9th Most Common Venue	10th Most Common Venue
16	M1X	Scarborough	Upper Rouge	43.834215	-79.216701	3	Bakery	Zoo Exhibit	Farm	Eastern European Restaurant	Electronics Store	Empanada Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant

This cluster would interest anyone which wants to open a **Chinese restaurant**, or an Asian restaurant in general.

Discussion :

A) Clusters :

- **Cluster 4 : Irish, Scottish & English (Yellow color)**

The **Cluster 4** regroups areas highly habited by **English, Irish, Scottish and Canadian people**.
We can also see that there are a lot of people from other European countries as well, such as French, German, Polish, ...
Most of these neighborhoods are positioned at the south and in the downtown of Toronto.

For example :

	PostalCode	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	9th Most Common Venue	10th Most Common Venue
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.765690	-79.175299	4	Pharmacy	Fast Food Restaurant	Pizza Place	Coffee Shop	Grocery Store	Sports Bar	Discount Store	Bank	Bar	Supermarket
3	M1G	Scarborough	Woburn	43.768216	-79.217610	4	Pizza Place	Fast Food Restaurant	Coffee Shop	Indian Restaurant	Bank	Sandwich Place	Supermarket	Juice Bar	Park	Thrift / Vintage Store
5	M1J	Scarborough	Scarborough Village	43.743085	-79.232172	4	Fast Food Restaurant	Sandwich Place	Big Box Store	Restaurant	Train Station	Indian Restaurant	Chinese Restaurant	Coffee Shop	Pizza Place	Convenience Store
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.726260	-79.263670	4	Discount Store	Coffee Shop	Chinese Restaurant	Grocery Store	Metro Station	Department Store	Sandwich Place	Asian Restaurant	Bus Line	Light Rail Station
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.713213	-79.284910	4	Convenience Store	Bakery	Coffee Shop	Fast Food Restaurant	Bus Station	Bus Line	Intersection	Soccer Field	Grocery Store	Bank
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.723575	-79.234976	4	Fast Food Restaurant	Discount Store	Bistro	Park	Sporting Goods Shop	Furniture / Home Store	Burger Joint	Liquor Store	Bank	Flower Shop
10	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...	43.759975	-79.268974	4	Pizza Place	Fast Food Restaurant	Chinese Restaurant	Coffee Shop	Light Rail Station	Automotive Shop	Indian Restaurant	Park	Electronics Store	Wings Joint

The cluster would interest anyone which wants to open an **English or Irish pub**, or any UK/Ireland related type of restaurant.

Discussion :

B) Analyze each neighborhood's competition:

Let's say we want to open an Irish pub. We are going to use the **cluster 4** in order to find the best neighborhood for this will.
In order to analyze the competition for each neighborhood, we are going to retrieve the list of existing venues of the type **pub**, in the neighborhoods categorized as **cluster 4**. For this task, we use **FoursquareAPI**.

We build a data frame as such (top 5 rows which represent 5 venues with the Pub category) :

CDN	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
95	43.66936	-79.40280	The Madison Avenue Pub	43.667947	-79.403486	Pub
95	43.66936	-79.40280	Duke of York	43.669186	-79.397527	Pub
75	43.66024	-79.37868	Churchmouse & Firkin	43.664632	-79.380406	Pub
62	43.68415	-79.29911	Grover Pub and Grub	43.679181	-79.297215	Pub
62	43.68415	-79.29911	Mullins Irish Pub	43.680348	-79.289370	Pub

Discussion :

B) Analyze each neighborhood's competition:

We can then count the number of existing Pubs for each neighbourhood, and sort these neighborhoods by count ascending :

	CDN	City_Area	Latitude	Longitude	Cluster Labels	1st Most Common Origin	2nd Most Common Origin	3rd Most Common Origin	4th Most Common Origin	5th Most Common Origin	6th Most Common Origin	7th Most Common Origin	8th Most Common Origin	9th Most Common Origin	10th Most Common Origin	C
113	16	Stonegate-Queensway	43.63718	-79.50058	4	English	Canadian	Irish	Scottish	Polish	Italian	Ukrainian	German	French	Portuguese	
12	122	Birchcliffe-Cliffside	43.69472	-79.26460	4	English	Irish	Canadian	Scottish	French	German	Chinese	Italian	Filipino	British Isles origins; n.i.e.	
95	45	Parkwoods-Donalda	43.75613	-79.32880	4	Canadian	English	Chinese	Scottish	Irish	Filipino	East Indian	German	Jamaican	French	
110	70	South Riverdale	43.65221	-79.33820	4	Chinese	English	Irish	Scottish	Canadian	French	German	Italian	British Isles origins; n.i.e.	Polish	
77	17	Mimico (includes Humber Bay Shores)	43.61729	-79.49885	4	English	Canadian	Irish	Scottish	Italian	Polish	Ukrainian	German	French	Chinese	
69	103	Lawrence Park South	43.71853	-79.40574	4	English	Scottish	Canadian	Irish	German	Polish	Russian	French	Italian	Chinese	
58	14	Islington-City Centre West	43.63608	-79.54296	4	English	Canadian	Italian	Irish	Scottish	Polish	Ukrainian	East Indian	German	Chinese	
104	98	Rosedale-Moore Park	43.68191	-79.37937	4	English	Scottish	Irish	Canadian	German	French	Chinese	Polish	Italian	British Isles origins; n.i.e.	
70	56	Leaside-Bennington	43.70071	-79.36758	4	English	Scottish	Irish	Canadian	German	French	Italian	Chinese	British Isles origins; n.i.e.	Polish	
35	62	East End-Danforth	43.68415	-79.29911	4	English	Irish	Scottish	Canadian	French	German	Chinese	Italian	Polish	British Isles origins; n.i.e.	
83	104	Mount Pleasant West	43.70541	-79.39256	4	English	Scottish	Canadian	Irish	German	French	Polish	Russian	East Indian	Chinese	
50	87	High Park-Swansea	43.64704	-79.47114	4	English	Irish	Scottish	Canadian	German	Polish	French	Italian	Ukrainian	British Isles origins; n.i.e.	
68	105	Lawrence Park North	43.72830	-79.40642	4	English	Scottish	Canadian	Irish	German	Chinese	French	Polish	Italian	Russian	
116	63	The Beaches	43.67413	-79.29644	4	English	Irish	Scottish	Canadian	French	German	Italian	Polish	British Isles origins; n.i.e.	Dutch	
82	99	Mount Pleasant East	43.70563	-79.38193	4	English	Irish	Scottish	Canadian	German	French	Polish	Chinese	Italian	Russian	

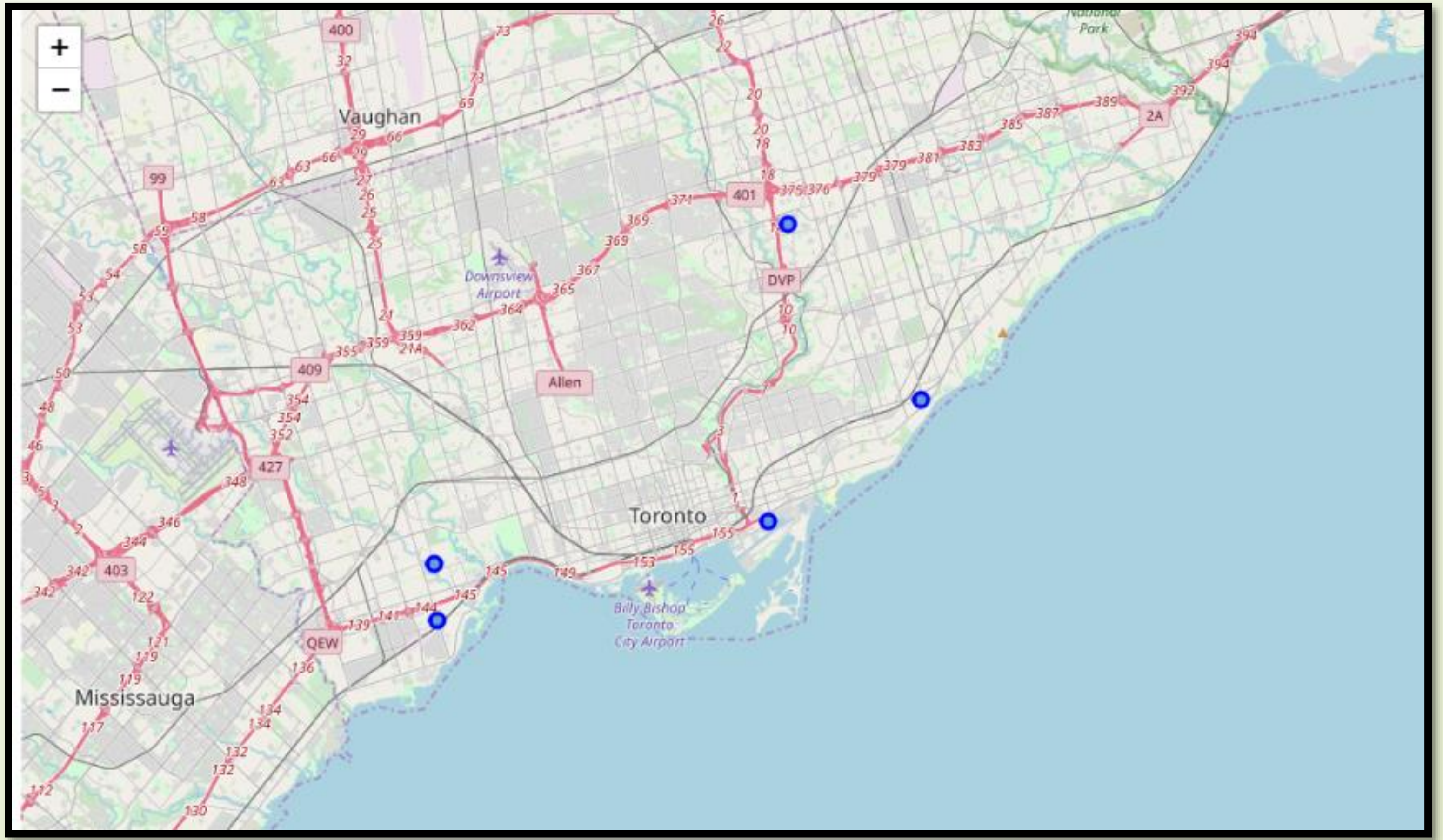
Discussion :

B) **Analyze each neighborhood's competition:**

We assume that the **top 5 neighborhoods** from this list represent the best places to open a new pub, as :

- The demographic data shows that people will likely come in this kind of venue in these neighborhoods
- The competition is limited in these neighborhoods

We can draw these 5 neighborhoods on the map of Toronto :



Conclusion :

- In this project, we managed to cluster the city of Toronto using **demographic data by neighborhoods**. This helps us identify which neighborhoods are the most adequate for opening a new restaurant of a specific type.
- Then, we managed to **identify the neighborhoods with the less competition** within these adequate neighborhoods, in order to optimize the performance of this new business.
- Best **New Business** will be **“Irish pub”**
- And **Best Locations** are Highlighted in Map.

Tableau Dashboard: