

Coursera Capstone: IBM Data Science Professional Certificate.

Project Report

By Preet Mishra

1. Introduction

Business Problem:

People often find themselves in a situation where they are looking for a place to move to and they don't have any reliable source to look for the best neighbourhood. People in that situation tend to go to online forums, ask someone for suggestions and use other means to conclude their decisions. For solving this problem, I will use the Foursquare location data to get the data of each neighbourhood and cluster them together.

Objective:

To find a neighbourhood for person X who wants to move to Toronto based on his/her preferences of venues.

Target audience:

A potential client could be someone who is looking for a neighbourhood to move to in Toronto.

Why to care:

With this project I will try to ease the whole process by providing reliable data and a mean that they can use to select a neighbourhood for the people of Toronto on the basis of their preferences of venues.

Stakeholders:

1. Government of Canada (to help an immigrant or a person who is looking for a neighbourhood to settle in Toronto).
2. End user.

2. Data Section

Following sources of data are used in the Project:

- **Data title:**

List of postal codes of Canada: M

Type of data:

HTML.


Description of the dataset:

This is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. Only the first three characters are listed, corresponding to the Forward Sortation Area.

A python script was developed to harness the data and use it in the project.

Source:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M



WIKIPEDIA
The Free Encyclopedia

[Main page](#)
[Contents](#)
[Featured content](#)
[Current events](#)
[Random article](#)
[Donate to Wikipedia](#)
[Wikipedia store](#)

Interaction

[Help](#)
[About Wikipedia](#)
[Community portal](#)
[Recent changes](#)
[Contact page](#)

Tools

[What links here](#)
[Related changes](#)

Article [Talk](#)

List of postal codes of Canada: M

From Wikipedia, the free encyclopedia

This is a list of [postal codes in Canada](#) where the first letter is M. Postal codes beginning with M are located within corresponding to the Forward Sortation Area.

[Canada Post](#) provides a free postal code look-up tool on its website,^[1] via its [applications](#) for such [smartphones](#) a: also sell validation tools, which allow customers to properly match addresses and postal codes. Hard-copy directo

Toronto - FSAs [\[edit \]](#)

Note: There are no rural FSAs in Toronto, hence no postal codes start with M0.

Postcode ↕	Borough ↕	Neighbourhood ↕
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village

- **Data title:**

Toronto Neighborhood coordinates

Type of data:

CSV.

Description of the dataset:

It contains the coordinates of the neighbourhoods in Toronto which is used to make the API calls to Foursquare.

Source:

http://cocl.us/Geospatial_data

Now that we have our dataframe intact. Let's add their longitude and latitude from a csv file.

```
In [12]: geo_data = pd.read_csv('http://cocl.us/Geospatial_data')
         geo_data.head()
```

```
Out[12]:
```

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

- **Data title:**

Foursquare's Location data

Type of data:

JSON.

Description of the dataset:

Foursquare provides the details about particular location including the places one can explore, venues, tips and photos.

We have used Foursquare's API to get the venues and their category available in a particular neighbourhood.

Source:

<http://foursquare.com/>

Getting the venues

```
In [35]: toronto_venues = getNearbyVenues(names=df['Neighbourhood'],
                                         latitudes=df['Latitude'],
                                         longitudes=df['Longitude'])
```

```
In [38]: print(toronto_venues.shape)
         toronto_venues.head()
```

(2235, 7)

```
Out[38]:
```

	Neighbourhood	Neighbourhood Latitude	Neighbourhood 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	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
2	Guildwood, Morningside, West Hill	43.763573	-79.188711	Swiss Chalet Rotisserie & Grill	43.767697	-79.189914	Pizza Place
3	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	Marina Spa	43.766000	-79.191000	Spa

Using
Foursquare's
API

3. Methodology

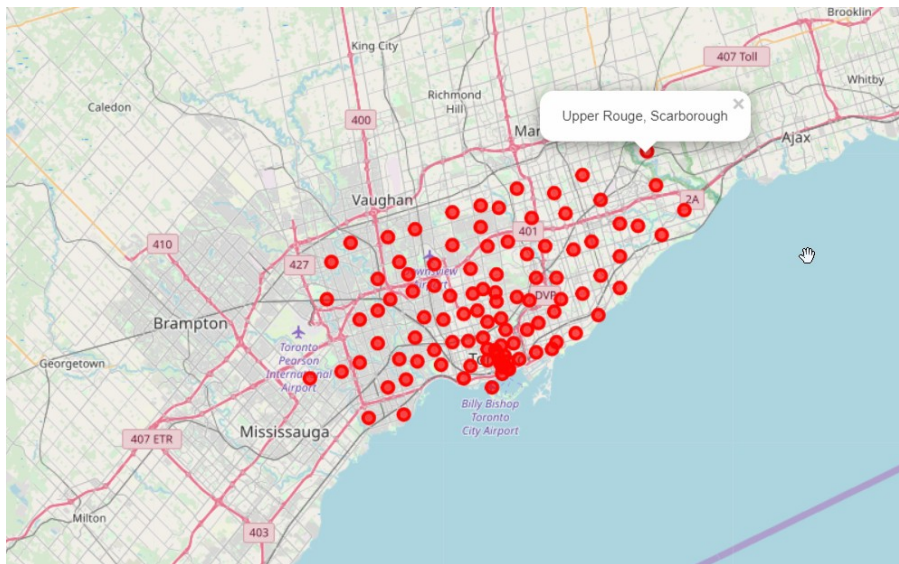
- The postal codes were firstly scraped from the Wikipedia's page and converted to a dataframe which was used in the project.

	PostalCode	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront

- The Borough and Neighbourhood which were 'Not assigned' weren't used and hence deleted.

	PostalCode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront
3	M5A	Downtown Toronto	Regent Park
4	M6A	North York	Lawrence Heights

- This was later merged with 'Toronto Neighborhood coordinates' and was visualized on a map.



- Using the Foursquare's API the list of venues and their categories were fetched.
- Using one hot encoding the categorical values were changed.

	Neighbourhood	Accessories Store	Adult Boutique	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium
0	Rouge, Malvern	0	0	0	0	0	0	0	0	0	0	0	0
1	Highland Creek, Rouge Hill, Port Union	0	0	0	0	0	0	0	0	0	0	0	0
2	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	0	0
3	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	0	0
4	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	0	0

- These neighbourhoods were then grouped by taking the mean of the frequency of each category.

	Neighbourhood	Accessories Store	Adult Boutique	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Garage	Auto Workshop	BBQ Joint
0	Adelaide, King, Richmond	0.01	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.03	0.0	0.0	0.01	0.01	0.0	0.03	0.0	0.0	0.0	0.0
1	Agincourt	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.00	0.00	0.0	0.00	0.0	0.0	0.0	0.0
2	Agincourt North, L'Amoreaux East, Milliken, St...	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.00	0.00	0.0	0.00	0.0	0.0	0.0	0.0
3	Albion Gardens, Beaumont Heights, Humbergate, ...	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.00	0.00	0.0	0.00	0.0	0.0	0.0	0.0
4	Alderwood, Long Branch	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.00	0.00	0.0	0.00	0.0	0.0	0.0	0.0

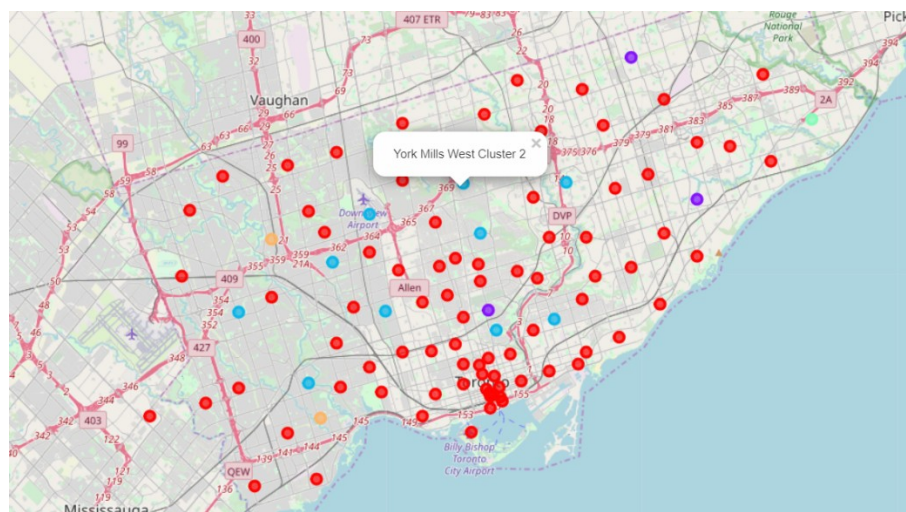
- Using this a dataframe containing top 10 most common venues of each neighbourhood was created.

	Neighbourhood	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	Adelaide, King, Richmond	Coffee Shop	Café	Steakhouse	Bar	Thai Restaurant	Burger Joint	Gym	Asian Restaurant	Hotel	Bakery
1	Agincourt	Lounge	Sandwich Place	Breakfast Spot	Skating Rink	Dumpling Restaurant	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Electronics Store
2	Agincourt North, L'Amoreaux East, Milliken, St...	Playground	Park	Yoga Studio	Drugstore	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop
3	Albion Gardens, Beaumont Heights, Humbergate, ...	Grocery Store	Pizza Place	Fried Chicken Joint	Coffee Shop	Sandwich Place	Fast Food Restaurant	Beer Store	Liquor Store	Pharmacy	Dance Studio
4	Alderwood, Long Branch	Pizza Place	Skating Rink	Dance Studio	Pharmacy	Coffee Shop	Pool	Pub	Sandwich Place	Gym	Airport Terminal

- For clustering the neighbourhoods based on the basis of common venues I used k-means clustering Machine Learning, which is unsupervised, in order to cluster the neighbourhoods together.
- I decided to go with k = 5. Thus the total neighbourhoods were clustered into 5 using k-means clustering.
- The cluster labels were assigned to the neighbourhoods.

	PostalCode	Borough	Neighbourhood	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	43.806686	-79.194353	0	Fast Food Restaurant	Dumpling Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Eastern European Restaurant	Hardware Store
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	3	Bar	Yoga Studio	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Eastern European Restaurant	Field
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	0	Pizza Place	Breakfast Spot	Rental Car Location	Intersection	Mexican Restaurant	Medical Center	Electronics Store	Spa	Eastern European Restaurant	Dumpling Restaurant
3	M1G	Scarborough	Woburn	43.770992	-79.216917	0	Coffee Shop	Korean Restaurant	Yoga Studio	Eastern European Restaurant	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	0	Hakka Restaurant	Athletics & Sports	Caribbean Restaurant	Thai Restaurant	Bakery	Bank	Fried Chicken Joint	Yoga Studio	Donut Shop	Dog Run
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476	1	Playground	Yoga Studio	Dumpling Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Eastern European Restaurant
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029	0	Discount Store	Chinese Restaurant	Bus Station	Department Store	Coffee Shop	Convenience Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577	0	Bus Line	Bakery	Fast Food Restaurant	Intersection	Bus Station	Soccer Field	Park	Ethiopian Restaurant	Empanada Restaurant	Electronics Store

- Using folium, the five clusters were visualized on the map.



- After assigning the labels to each neighbourhood a representative of each cluster was chosen by finding the closest neighbourhood to the centroid of each of the clusters.

Cluster 1:

	Neighbourhood	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	Rouge, Malvern	0	Fast Food Restaurant	Dumpling Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Eastern European Restaurant	Hardware Store
2	Guildwood, Morningside, West Hill	0	Pizza Place	Breakfast Spot	Rental Car Location	Intersection	Mexican Restaurant	Medical Center	Electronics Store	Spa	Eastern European Restaurant	Dumpling Restaurant
3	Woburn	0	Coffee Shop	Korean Restaurant	Yoga Studio	Eastern European Restaurant	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant
4	Cedarbrae	0	Hakka Restaurant	Athletics & Sports	Caribbean Restaurant	Thai Restaurant	Bakery	Bank	Fried Chicken Joint	Yoga Studio	Donut Shop	Dog Run
6	East Birchmount Park, Ionview, Kennedy Park	0	Discount Store	Chinese Restaurant	Bus Station	Department Store	Coffee Shop	Convenience Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore
7	Clairlea, Golden Mile, Oakridge	0	Bus Line	Bakery	Fast Food Restaurant	Intersection	Bus Station	Soccer Field	Park	Ethiopian Restaurant	Empanada Restaurant	Electronics Store

Representative of Cluster 1

```
get_venues_by_freq(20, closest[0])
```

RYERSON, GARDEN DISTRICT:

	venue	freq
0	Coffee Shop	0.08
1	Clothing Store	0.06
2	Cosmetics Shop	0.04
3	Café	0.04
4	Middle Eastern Restaurant	0.03
5	Thai Restaurant	0.02
6	Pizza Place	0.02
7	Bar	0.02
8	Tea Room	0.02
9	Bubble Tea Shop	0.02
10	Theater	0.02
11	Japanese Restaurant	0.02
12	Lingerie Store	0.02
13	Diner	0.02
14	Italian Restaurant	0.02
15	Ramen Restaurant	0.02
16	Restaurant	0.02
17	Fast Food Restaurant	0.01
18	Chinese Restaurant	0.01
19	Dessert Shop	0.01

Cluster 2:

	Neighbourhood	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
5	Scarborough Village	1	Playground	Yoga Studio	Dumpling Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Eastern European Restaurant
14	Agincourt North, L'Amoreaux East, Milliken, St...	1	Playground	Park	Yoga Studio	Drugstore	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop
45	Moore Park, Summerhill East	1	Playground	Gym	Yoga Studio	Dumpling Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore

Representative of Cluster 2

```
get_venues_by_freq(20, closest[1])
```

AGINCOURT NORTH, L'AMOREAUX EAST, MILLIKEN, STEELES EAST:

	venue	freq
0	Playground	0.5
1	Park	0.5
2	Mexican Restaurant	0.0
3	Motel	0.0
4	Monument / Landmark	0.0
5	Molecular Gastronomy Restaurant	0.0
6	Modern European Restaurant	0.0
7	Miscellaneous Shop	0.0
8	Middle Eastern Restaurant	0.0
9	Men's Store	0.0
10	Museum	0.0
11	Mediterranean Restaurant	0.0
12	Medical Center	0.0
13	Massage Studio	0.0
14	Martial Arts Dojo	0.0
15	Market	0.0
16	Movie Theater	0.0
17	Music Store	0.0
18	Health Food Store	0.0
19	Music Venue	0.0

Cluster 3:

	Neighbourhood	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
20	York Mills West	2	Park	Bank	Yoga Studio	Dumpling Restaurant	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Eastern European Restaurant
22	Parkwoods	2	Fast Food Restaurant	Park	Food & Drink Shop	Dumpling Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore
27	CFB Toronto, Downsview East	2	Bus Stop	Park	Airport	Yoga Studio	Eastern European Restaurant	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant
37	East Toronto	2	Park	Coffee Shop	Convenience Store	Yoga Studio	Dumpling Restaurant	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore
41	Lawrence Park	2	Gym / Fitness Center	Park	Bus Line	Swim School	Yoga Studio	Drugstore	Discount Store	Dog Run	Doner Restaurant	Donut Shop
47	Rosedale	2	Park	Trail	Playground	Yoga Studio	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop
71	Caledonia-Fairbanks	2	Park	Fast Food Restaurant	Market	Pharmacy	Women's Store	Grocery Store	Department Store	Gym	Empanada Restaurant	Electronics Store
76	Downsview, North Park, Upwood Park	2	Basketball Court	Park	Construction & Landscaping	Bakery	Electronics Store	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Eastern European Restaurant
87	The Kingsway, Montgomery Road, Old Mill North	2	Park	River	Smoke Shop	Pool	Yoga Studio	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run
95	Kingsview Village, Martin Grove Gardens, Richv...	2	Pizza Place	Park	Bus Line	Drugstore	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant

Representative of Cluster 3

```
get_venues_by_freq(20, closest[2])
```

```
CALEDONIA-FAIRBANKS:
      venue  freq
0      Park  0.33
1  Women's Store  0.17
2      Pharmacy  0.17
3      Market  0.17
4  Fast Food Restaurant  0.17
5  Accessories Store  0.00
6  Mexican Restaurant  0.00
7  Monument / Landmark  0.00
8  Molecular Gastronomy Restaurant  0.00
9  Modern European Restaurant  0.00
10 Miscellaneous Shop  0.00
11 Middle Eastern Restaurant  0.00
12 Medical Center  0.00
13 Men's Store  0.00
14 Mediterranean Restaurant  0.00
15 Movie Theater  0.00
16 Massage Studio  0.00
17 Martial Arts Dojo  0.00
18 Motel  0.00
19 Music Store  0.00
```

Cluster 4:

	Neighbourhood	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	Highland Creek, Rouge Hill, Port Union	3	Bar	Yoga Studio	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Eastern European Restaurant	Field

Representative of Cluster 4

```
get_venues_by_freq(20, closest[3])
```

```
HIGHLAND CREEK, ROUGE HILL, PORT UNION:
      venue  freq
0      Bar  1.0
1  Accessories Store  0.0
2  Middle Eastern Restaurant  0.0
3      Motel  0.0
4  Monument / Landmark  0.0
5  Molecular Gastronomy Restaurant  0.0
6  Modern European Restaurant  0.0
7  Miscellaneous Shop  0.0
8  Mexican Restaurant  0.0
9      Museum  0.0
10 Men's Store  0.0
11 Mediterranean Restaurant  0.0
12 Medical Center  0.0
13 Massage Studio  0.0
14 Martial Arts Dojo  0.0
15 Market  0.0
16 Movie Theater  0.0
17 Music Store  0.0
18 Mac & Cheese Joint  0.0
19 Organic Grocery  0.0
```

Cluster 5:

	Neighbourhood	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
88	Humber Bay, King's Mill Park, Kingsway Park So...	4	Baseball Field	Pool	Yoga Studio	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Eastern European Restaurant
93	Emery, Humberlea	4	Baseball Field	Paper / Office Supplies Store	Yoga Studio	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Eastern European Restaurant

Representative of Cluster 5

```
get_venues_by_freq(20, closest[4])
```

EMERY, HUMBERLEA:

	venue	freq
0	Paper / Office Supplies Store	0.5
1	Baseball Field	0.5
2	Accessories Store	0.0
3	Middle Eastern Restaurant	0.0
4	Motel	0.0
5	Monument / Landmark	0.0
6	Molecular Gastronomy Restaurant	0.0
7	Modern European Restaurant	0.0
8	Miscellaneous Shop	0.0
9	Mexican Restaurant	0.0
10	Museum	0.0
11	Men's Store	0.0
12	Mediterranean Restaurant	0.0
13	Medical Center	0.0
14	Massage Studio	0.0
15	Martial Arts Dojo	0.0
16	Market	0.0
17	Movie Theater	0.0
18	Music Store	0.0
19	Mac & Cheese Joint	0.0

4. Result

- The neighbourhoods were successfully divided in five clusters based on the venues they had.
- For each cluster I also came up with a representative by finding the neighbourhood which was closest to the centroid of that particular cluster.
 - *Representatives are the neighbourhoods whose venues are most likely to resemble with each of the neighbourhoods in that particular cluster.*

5. Discussion

The clusters formed using k-means had the similar common venue types in their neighbourhood.

Recommendation:

Using this results a person X who wants to move to Toronto can choose a neighbourhood from different clusters. For choosing a particular cluster, X can analyze provided cluster's *representative* based on his/her preferences of venues.

- Each neighbourhood of the cluster is *most likely* to have similar characteristics as the cluster's representative.

6. Conclusion

Finding a perfect place to live can sometimes be overwhelming with lack of proper resources to look up to. With this project I have tried to gather the reliable information about the neighbourhoods of Toronto and the most common venues they have and clustered them together using Machine Learning. I have also provided a representative of each of the clusters.

So looking for new place in Toronto to move to? Jot down your venue preferences, analyze which of the available representative suits your preferences, choose your desired neighbourhood from that representative's cluster and you are done.