Capstone Project

PROJECT REPORT

Author: Heta Patel

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 in Toronto.

Why 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 based on 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.

Author: Heta Patel

Data Section:
The following sources of the data are used in the Project.

* Data Title: List of postal codes in 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

Author: Heta Patel

* Data Title: Toronto Neighbourhood coordinates

* Type of Data: CSV

Description of the dataset:
The dataset contains the coordinates of the neighbourhood in Toronto which is used to make the API calls to Foursquare.

*SOURCE:
http://cocl.us/Geospatial_data

In [11]:
```
# load the coordinates from the csv file on Coursera
coordinates = pd.read_csv("http://cocl.us/Geospatial_data")
coordinates.head()
```

Out[11]:

	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 a 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

* METHODOLOGY

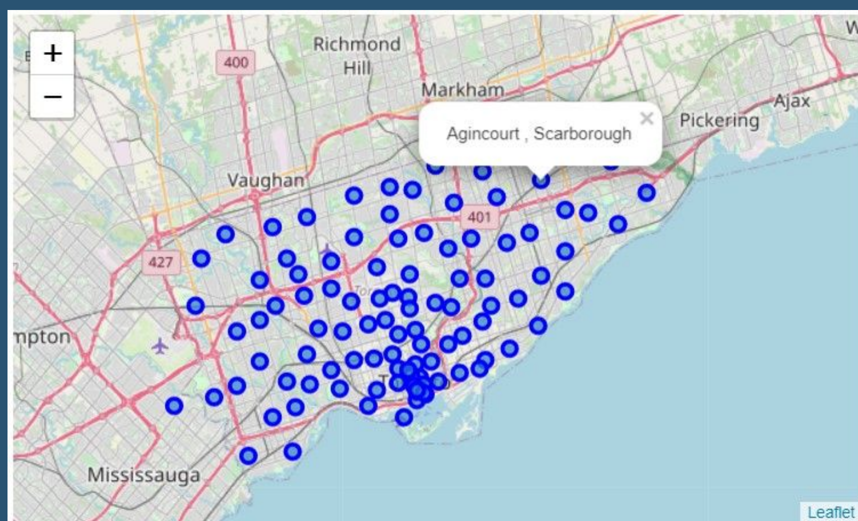
* The postal codes were firstly scraped from the Wikipedia's page and converted to a dataframe which was used in the project.

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

	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

	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 Neighbourhood 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

* METHODOLOGY

* Their 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, Beaumond 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	Agincourt	Clothing Store	Lounge	Skating Rink	Breakfast Spot	Latin American Restaurant	Eastern European Restaurant	Distribution Center	Dog Run	Doner Restaurant	Donut Shop
1	Alderwood / Long Branch	Pizza Place	Pharmacy	Athletics & Sports	Dance Studio	Coffee Shop	Pub	Sandwich Place	Skating Rink	Gym	Antique Shop
2	Bathurst Manor / Wilson Heights / Downsview No...	Coffee Shop	Bank	Fried Chicken Joint	Bridal Shop	Sandwich Place	Diner	Restaurant	Middle Eastern Restaurant	Supermarket	Sushi Restaurant
3	Bayview Village	Café	Bank	Chinese Restaurant	Japanese Restaurant	Yoga Studio	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Drugstore
4	Bedford Park / Lawrence Manor East	Coffee Shop	Restaurant	Sandwich Place	Italian Restaurant	Sushi Restaurant	Greek Restaurant	Thai Restaurant	Cosmetics Shop	Pharmacy	Pizza Place

* 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 Mos Commo Venu
0	M1B	Scarborough	Malvern / Rouge	43.806686	-79.194353	1	Fast Food Restaurant	Print Shop	Diner	Discount Store	Distribution Center	Dog Ru
1	M1C	Scarborough	Rouge Hill / Port Union / Highland Creek	43.784535	-79.160497	1	Construction & Landscaping	History Museum	Bar	Yoga Studio	Eastern European Restaurant	Distributio Cente
2	M1E	Scarborough	Guildwood / Morningside / West Hill	43.763573	-79.188711	1	Rental Car Location	Breakfast Spot	Medical Center	Intersection	Electronics Store	Mexica Restaurar
3	M1G	Scarborough	Woburn	43.770992	-79.216917	4	Coffee Shop	Korean Restaurant	Convenience Store	Yoga Studio	Distribution Center	Dog Ru
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	1	Hakka Restaurant	Athletics & Sports	Fried Chicken Joint	Bank	Bakery	Gas Statio

* Using Folium, the five clusters were visualised on the map.
* After assigning the labels to each neighbourhood a representative of each cluster was chosen by finding the closet neighbourhood to the centroid of each of the clusters.

* RESULT

* The neighbourhoods were successfully divided into five clusters based on the venues they had.
* For each cluster I also came up with a representative by finding the neighbourhood which was closet 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.

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 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

A Project By HETA PATEL