

Capstone Project - The Battle of the Neighbourhood

Introduction: Business Problem

In this project we will try to Explore and cluster the neighbourhoods in Toronto. We are analysing any neighbourhood or city using location data. However, we decided to explore neighbourhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York or Toronto. To be able to do that, we have to cluster data which is a form of unsupervised machine learning. Done using popular clustering algorithm, called k-means. You will use the explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters.

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Data

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.

[Canada Post](#) provides a free postal code look-up tool on its website via its [applications](#) for such [smartphones](#) as the [iPhone](#) and [BlackBerry](#) and sells hard-copy directories and [CD-ROMs](#). Many vendors also sell validation tools, which allow customers to properly match addresses and postal codes. Hard-copy directories can also be consulted in all post offices, and some libraries.

Following data sources will be needed to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Geocoders**
- number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**
- geopy and folium also used

Foursquare

Now that we have our location candidates, let's use Foursquare API to get info on restaurants in each neighborhood.

We're interested in venues in 'food' category, but only those that are proper restaurants - coffee shops, pizza places, bakeries etc. are not direct competitors so we don't care about those. So we will

include in our list only venues that have 'restaurant' in category name, and we'll make sure to detect and include all the subcategories of specific 'Italian restaurant' category, as we need info on Italian restaurants in the neighborhood.

Methodology

1. Make dataframe from data on website.

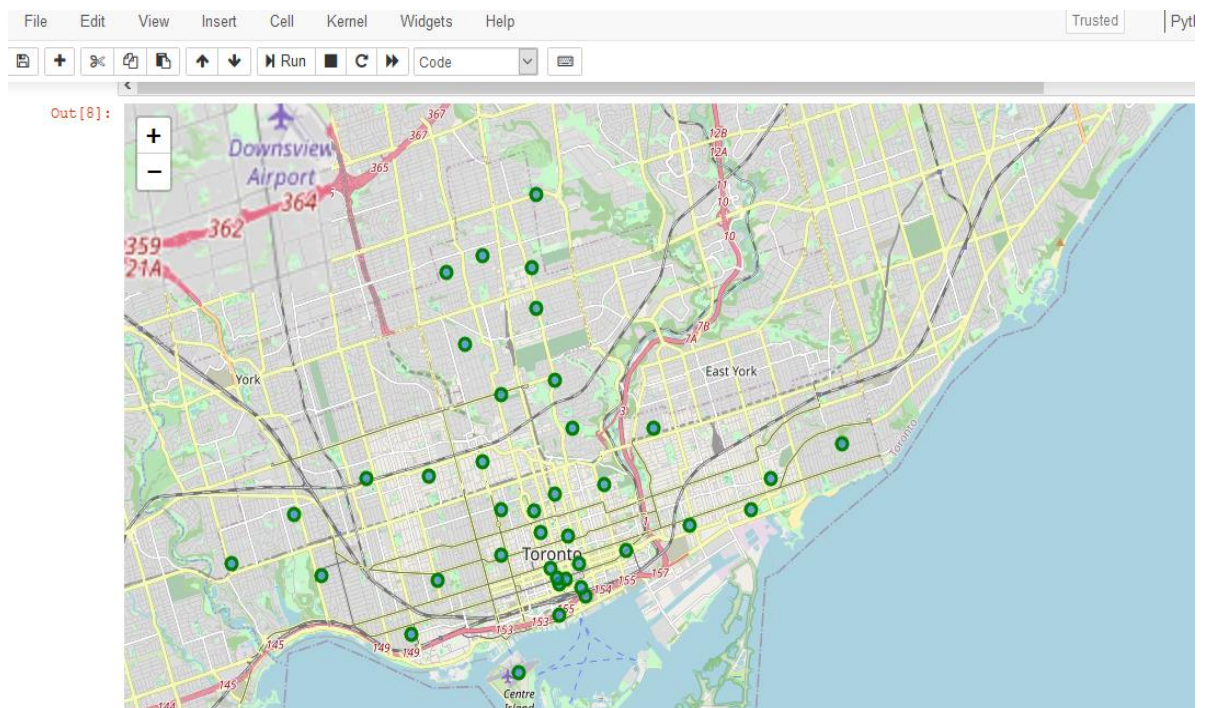
```
['Postal code', 'Borough', 'Neighborhood']
Postal code      Borough \
0      M3A      North York
1      M4A      North York
2      M5A      Downtown Toronto
3      M6A      North York
4      M7A      Downtown Toronto
..      ...      ...
98     M8X      Etobicoke
99     M4Y      Downtown Toronto
100    M7Y      East Toronto
101    M8Y      Etobicoke
102    M8Z      Etobicoke
```

2. Add geospatial information i.e. latitude and longitude
 - Now that we have built a dataframe of the postal code of each neighborhood along with the borough name and neighborhood name, in order to utilize the Foursquare location data, we need to get the latitude and the longitude coordinates of each neighborhood.
 - We will use a link to a csv file that has the geographical coordinates of each postal code: http://cocl.us/Geospatial_data to get the latitude and the longitude coordinates of each neighborhood.

Out[4]:

	Borough	Neighborhood	Latitude	Longitude
Postal code				
M3A	North York	Parkwoods	43.753259	-79.329656
M4A	North York	Victoria Village	43.725882	-79.315572
M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
M9A	Etobicoke	Islington Avenue	43.667856	-79.532242
M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
M3B	North York	Don Mills	43.745906	-79.352188
M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937
M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
M6B	North York	Glencairn	43.709577	-79.445073

3. Explore and cluster the neighborhoods in Toronto and generate map to visualize the neighborhoods and how they are clustered together
 - Filter Canada data to only use boroughs in Toronto
 - Show an initial map of the neighborhoods in Toronto
 - Utilizing the Foursquare API to explore the neighborhoods and segment them.
4. Get Toronto data and generate map for neighbourhood



5. We are going to start utilizing the Foursquare API to explore the neighborhoods and segment them.

Out[11]:

	name	categories	lat	lng
0	Roselle Desserts	Bakery	43.653447	-79.362017
1	Tandem Coffee	Coffee Shop	43.653559	-79.361809
2	Morning Glory Cafe	Breakfast Spot	43.653947	-79.361149
3	Cooper Koo Family YMCA	Distribution Center	43.653249	-79.358008
4	Body Blitz Spa East	Spa	43.654735	-79.359874

6. Generalize Venue Data Collection

Out[15]:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Berczy Park	58	58	58	58	58	58
Brockton, Parkdale Village, Exhibition Place	23	23	23	23	23	23
Business reply mail Processing CentrE	15	15	15	15	15	15
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	14	14	14	14	14	14
Central Bay Street	62	62	62	62	62	62
Christie	18	18	18	18	18	18
Church and Wellesley	74	74	74	74	74	74
Commerce Court, Victoria Hotel	100	100	100	100	100	100
Davisville	39	39	39	39	39	39
Davisville North	7	7	7	7	7	7
Dufferin, Dovercourt Village	15	15	15	15	15	15
First Canadian Place, Underground city	100	100	100	100	100	100
Forest Hill North & West	4	4	4	4	4	4
Garden District, Ryerson	100	100	100	100	100	100
Harbourfront East, Union Station, Toronto Islands	100	100	100	100	100	100
High Park, The Junction South	23	23	23	23	23	23

7. Analyse and cluster neighbourhood.

```

----Berczy Park----
      venue  freq
0   Coffee Shop 0.07
1   Cocktail Bar 0.05
2  Farmers Market 0.03
3    Restaurant 0.03
4     Beer Bar 0.03

----Brockton, Parkdale Village, Exhibition Place----
      venue  freq
0      Café 0.13
1 Breakfast Spot 0.09
2   Coffee Shop 0.09
3    Nightclub 0.09
4      Bakery 0.04

```

Results and Discussion

Clustering Approach:

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

Out[22]:

	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	
Postal code														
M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	0	Coffee Shop	Park	Pub	Bakery	Breakfast Spot	Café	Theater	Dessert Shop	F
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	0	Coffee Shop	Sushi Restaurant	Diner	Yoga Studio	Creperie	Beer Bar	Sandwich Place	Burger Joint	
M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	0	Clothing Store	Coffee Shop	Café	Restaurant	Japanese Restaurant	Bubble Tea Shop	Cosmetics Shop	Middle Eastern Restaurant	
M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	0	Coffee Shop	Café	Gastropub	Italian Restaurant	American Restaurant	Cocktail Bar	Cosmetics Shop	Farmers Market	F
M4E	East Toronto	The Beaches	43.676357	-79.293031	3	Trail	Health Food Store	Pub	Women's Store	Deli / Bodega	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	F

Work Flow:

Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

The major purpose of this project, is to suggest a better neighborhood in a new city for the person who are shifting there. Social presence in society in terms of like-minded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

Conclusion

This Capstone Project concludes to create an analysis of features for a people migrating to Toronto to search a best neighborhood as a comparative analysis between neighborhoods.

We have gathered neighborhood according to postal code and we have also gathered most common venues in that area.

