Analyzing the neighborhood of Toronto, Canada for setting up a new business

By Md Zulfikar Ali

Introduction

Toronto, the capital city of Canadian province of Ontario, is one of the populous cities in Canada. It has a diverse population with people migrating from different parts of the world. It is a cultural and historical hub and a tourist attraction. Toronto is also considered to be the financial capital of Canada. Because of the multicultural background the place is filled with restaurants having cuisines from all over the world. Also, being a tourist attraction, the city requires restaurants serving different varieties of food. Keeping this in mind, in this project, we explore the neighborhood of Toronto for a finding a suitable neighborhood for opening new restaurant. The stakeholders for this purpose could be people interested in investing or opening a new business.

Data acquisition

For this project, we need the neighborhood data, geographical data for the neighborhood, and the venue information. Neighborhood data is downloaded from the Wikipedia page "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M". We use the subroutine read_html() in pandas to scrape data from the Wikipedia page and stored into a dataframe for further investigations. The geographical coordinates in principle can be obtained by using the GeoPy library in python but for this particular project, we used the geographical coordinated provided by Coursera whose link is given provided here (http://cocl.us/Geospatial_data). The file provides latitude and longitude information for each neighborhood. However, Geopy is used wherever required. Finally, the venue information is obtained by using FourSquare API. Foursquare provides, the information the venue information for a given location (latitude, longitude) that is used to build a database about what is popular in a neighborhood etc. We use, the obtained data as explained before, to find neighborhoods that are popular for food lovers and in the process suggest which neighborhood is ideal for opening new restaurants or Cafe etc.

Data Cleaning and analysis

Many of the cells of data downloaded from wikipedia had no assigned values. Before compiling the neighborhood data, the cells that are not assigned a borough were removed. Furthermore the cells that had boroughs but not assigned neighborhood were replaced with boroughs as neighborhood. The final cleaned file looks like Fig. 1 below.

	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	Regent Park, Harbourfront

	PostalCode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

Figure 1. Top panel, neighborhood data obtained from wikipedia for Toronto, Canada. Bottom panel, data after cleaning.

Next, we obtained the location data using the link (http://cocl.us/Geospatial_data) and Geopy library (a sample of the location data along with the neighborhood on a map is shown in Fig. 2, 3). This location data was used to gathered information from Foursquare API and stored in dataframe as shown below. There are 240 unique categories of venues. A sample of the top 10 venues for some neighborhoods are shown below in Fig. 4.

	PostalCode	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

	PostalCode	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

Figure 2. Top panel, first five lines of the geo-location data containing longitude and latitude for the postal codes in Toronto city. Bottom panel, combined with boroughs, neighborhood and geolocations.

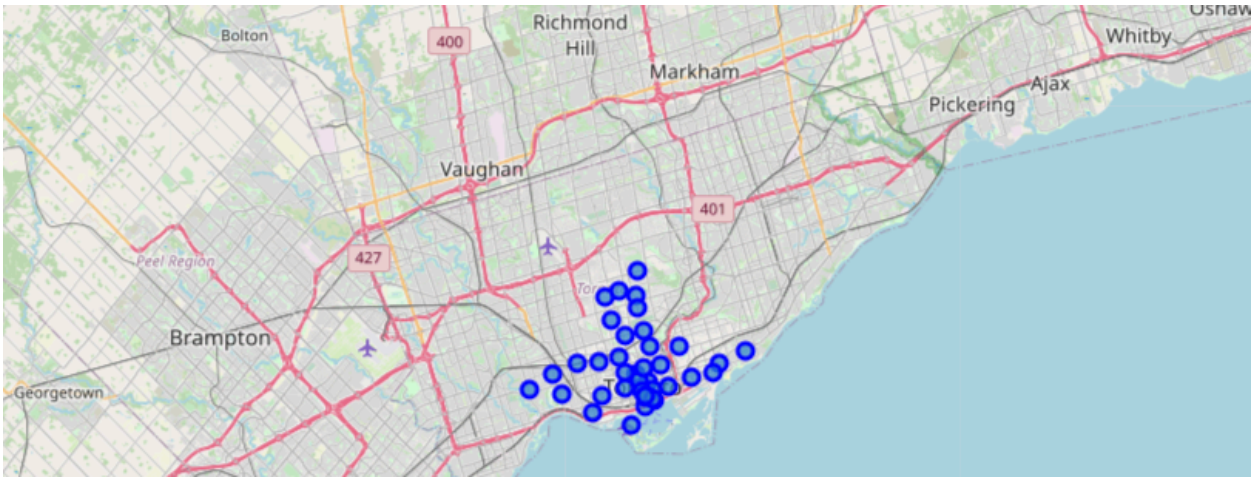

Figure 3 Map showing all the neighborhoods in Toronto city, Canada

	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	Berczy Park	Coffee Shop	Farmers Market	Bakery	Beer Bar	Cocktail Bar	Cheese Shop	Seafood Restaurant	Restaurant	Grocery Store	Pub
1	Brockton, Parkdale Village, Exhibition Place	Café	Coffee Shop	Breakfast Spot	Nightclub	Climbing Gym	Bar	Bookstore	Burrito Place	Restaurant	Playground
2	Business reply mail Processing Centre, South C...	Park	Recording Studio	Restaurant	Light Rail Station	Auto Workshop	Fast Food Restaurant	Farmers Market	Burrito Place	Pizza Place	Butcher
3	CN Tower, King and Spadina, Railway Lands, Har...	Airport Lounge	Airport Service	Plane	Harbor / Marina	Boutique	Boat or Ferry	Rental Car Location	Bar	Historic Site	Coffee Shop
4	Central Bay Street	Coffee Shop	Café	Sandwich Place	Italian Restaurant	Department Store	Japanese Restaurant	Thai Restaurant	Burger Joint	Bubble Tea Shop	Salad Pl

Figure 4 Top 10 venues for some of the neighborhoods

Methodology: Clustering the venues using k-means algorithm

In order to do machine learning, the venue category was transformed into numerical values. This was done using the One-hot encoding method. Basically, what it does is for a given neighborhood each category gets a value of 1 if the category exists in that neighborhood or 0 otherwise. A matrix containing 0s and 1s is then formed which is further used for unsupervised learning, k-means algorithm. The neighborhood is clustered into 5 clusters using k-means algorithm on the category data matrix.

Results and Observations

Using k-means algorithm the neighborhoods are divided into 5 clusters as shown in Fig. 5. The map in this figure is color coded, where the neighborhoods having same color fall in the same cluster. We further show the top ten venues of each neighborhood in all 5 clusters in Fig. 6.

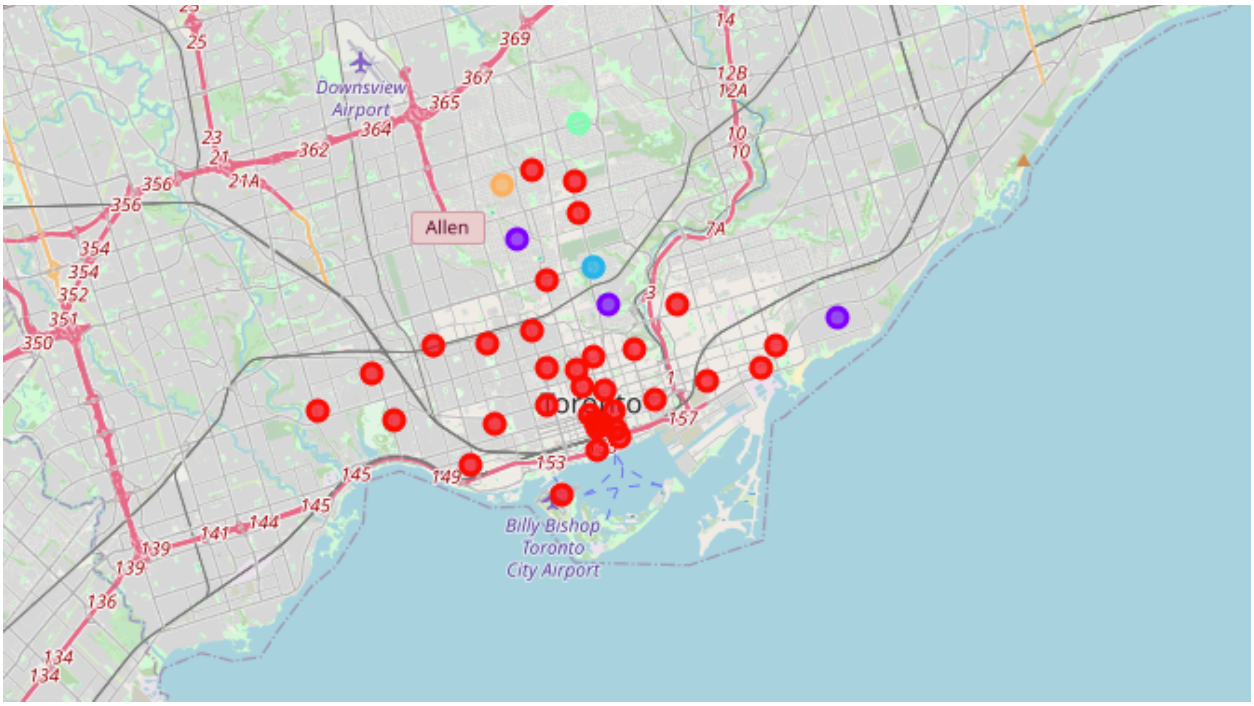

Figure 5 Clusters of neighborhoods in Toronto city. There are five clusters with each the neighborhood in a single cluster is shown in same color.

	Borough	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	Downtown Toronto	0	Coffee Shop	Park	Bakery	Pub	Café	Theater	Breakfast Spot	Shoe Store	Distribution Center	Electronics Store
1	Downtown Toronto	0	Coffee Shop	Yoga Studio	Café	Beer Bar	Smoothie Shop	Italian Restaurant	Sandwich Place	Distribution Center	Restaurant	Diner
2	Downtown Toronto	0	Coffee Shop	Clothing Store	Café	Bubble Tea Shop	Cosmetics Shop	Japanese Restaurant	Furniture / Home Store	Hotel	Pizza Place	Bookstore
3	Downtown Toronto	0	Coffee Shop	Café	Cocktail Bar	Restaurant	Gastropub	Beer Bar	American Restaurant	Japanese Restaurant	Seafood Restaurant	Clothing Store
5	Downtown Toronto	0	Coffee Shop	Farmers Market	Bakery	Beer Bar	Cocktail Bar	Cheese Shop	Seafood Restaurant	Restaurant	Grocery Store	Pub
6	Downtown Toronto	0	Coffee Shop	Café	Sandwich Place	Italian Restaurant	Department Store	Japanese Restaurant	Thai Restaurant	Burger Joint	Bubble Tea Shop	Salad Place
7	Downtown Toronto	0	Grocery Store	Café	Park	Candy Store	Italian Restaurant	Restaurant	Baby Store	Athletics & Sports	Nightclub	Coffee Shop
8	Downtown Toronto	0	Coffee Shop	Café	Restaurant	Gym	Hotel	Bar	Clothing Store	Thai Restaurant	Cosmetics Shop	Concert H

Figure 6 Some of the neighborhoods with top 10 venues in cluster 1.

Most of the neighborhoods were clustered into cluster 1 having 38 neighborhoods (out of 44) suggesting that the neighborhoods are all alike. The other 4 clusters have 3, 1, 1 and 1 neighborhoods, respectively. By analyzing these clusters, we find that cluster 1 (marked in red in Fig. 5) is more suitable for opening restaurants as these clusters are mostly in downtown Toronto; downtowns are in general more favorable for restaurants. The other clusters are mostly situated near parks and trails which are more suitable for fast food type joints.

Conclusion and future direction

In this project, we analyzed Toronto city in Canada to find neighborhoods with potential for a new restaurant setup. Using the neighborhood data of Toronto obtained from Wikipedia, we analyzed the venues for each neighborhood from information gathered through Foursquare API. Our analysis reveals that cluster 1 is more suitable for setting up a new restaurant. A more detailed

analysis can be done using the city crime data and also the living standard of the neighborhood so that we can suggest if the neighborhood is more suited for fine-dining restaurant or fast-food joints etc.

In future, the city crime data, locality information like schools in the area etc. can be used in conjunction with the venue information to build a suggestive model for renting a house or possibly buy a home.