

The Battle of Neighborhoods

LAL BABU SAH

DEC 07, 2019

1.Introduction

1.1 Background

In the neighborhoods of Toronto, there are a great number of different type of venue categories, but if someone wants to know which type of venue category is more common in neighborhoods of Toronto and which will be more suitable(in business) venue categories for the stakeholders. Here, we will use clustering models to group similar type of neighborhoods having more common venue categories are similar, this will help stakeholders to make a better decision for choosing the best venue category.

1.2 Business Problem

In this project, we will compare the most common venues category in the neighborhoods of Toronto. Since there are lots of venues category in neighborhoods of Toronto we will try to detect which type of venue category are more common in Neighborhoods.

Specifically, this report will be targeted to stakeholders interested in finding the optimal venue category and which will be more suitable(in business) venue category in the neighborhoods of Toronto.

1.3 Interest

We will use data science to generate a few most promising venues category based on these criteria. The advantages of venue category in each neighborhood will then be clearly expressed so that the best possible final venue can be chosen by stakeholders.

2.Data acquisition and Cleaning

2.1 Data Sources

Based on the definition of our problem, factors that will influence our decision are:

- 1.number of different types of venue categories in the neighborhood of Toronto.
- 2.venue category that is most common in the neighborhoods of Toronto.

Following data sources will be needed for analysis of Neighborhoods of Toronto:

1. For all the information we need to explore and cluster the neighborhoods in Toronto a Wikipedia page exists, here is the link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
2. For scraping the table, we can simply use pandas to read the table into a pandas dataframe.
3. For the geographical coordinates of each postal code, read a CSV file (link:https://cocl.us/Geospatial_data) using pandas dataframe.
- 4.explore the neighborhoods and segment them using Foursquare API.

2.2 Data Cleaning

Data downloaded and scraped from multiple sources were combined into one table.

When we scrap the table using pandas dataframe, there were a lot of Not assigned 'borrow' and a lot of Not assigned 'neighborhoods' in the table, we drop cells with a 'borough' that is Not assigned.

There is more than one neighborhood that can exist in one postcode area. These rows will be combined into one row with the neighborhoods separated with a comma using 'groupby'.

3.Methodology

In this project, we will direct our efforts on detecting the most common venue category of each neighborhood of Toronto.

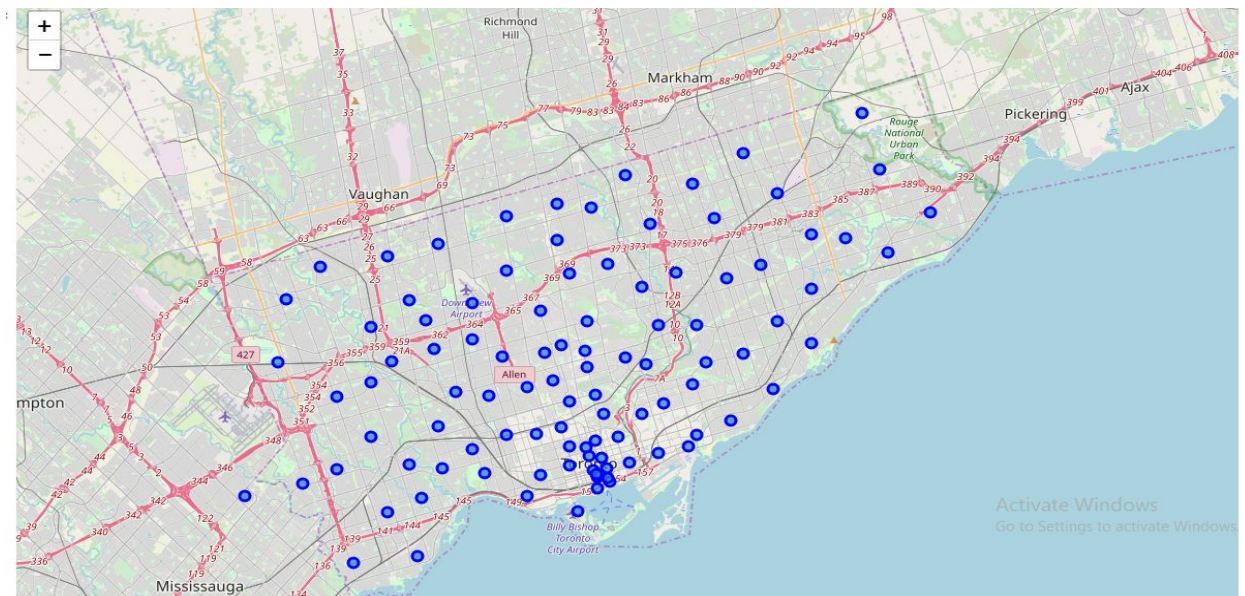
In first step, we have collected the required data, using given wikipedia page link and geographical coordinates of each code using csv file. And explore the neighborhoods and segment them using Foursquare API.

Second step in our analysis will be exploration of venues across different areas of Toronto. we will use Folium library to identify a few promising areas close to center with most common venues.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders.

3.1 Analysis

The dataframe has 11 boroughs and 103 neighborhoods, the geographical coordinates of Toronto are 43.653963, -79.387207. Create a map of Toronto with neighborhoods superimposed on top using Folium library and the map of Toronto will be:



Now, using Foursquare API explore the neighborhoods and get the nearby venues, we can see Here that for the same Neighborhood 'Guildwood, Morningside, West Hill' in row 2, 3, 4 have different venues, so we group these rows by Neighborhood using 'groupby'.

```
print(toronto_venues.shape)
toronto_venues.head()
```

(2235, 7)

	Neighborhood	Neighborhood Latitude	Neighborhood 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

Now create the new dataframe and display the top 10 venues for each neighborhood and in the following table, neighborhoods Albion Gardens,Beaumont Heights,Humbergate in row 3 have same 1st most common venues, 2nd most common venue up to 10th most common venues, so there are different neighborhoods have same 1st most common venues, 2nd most common venue up to 10th most common venues.

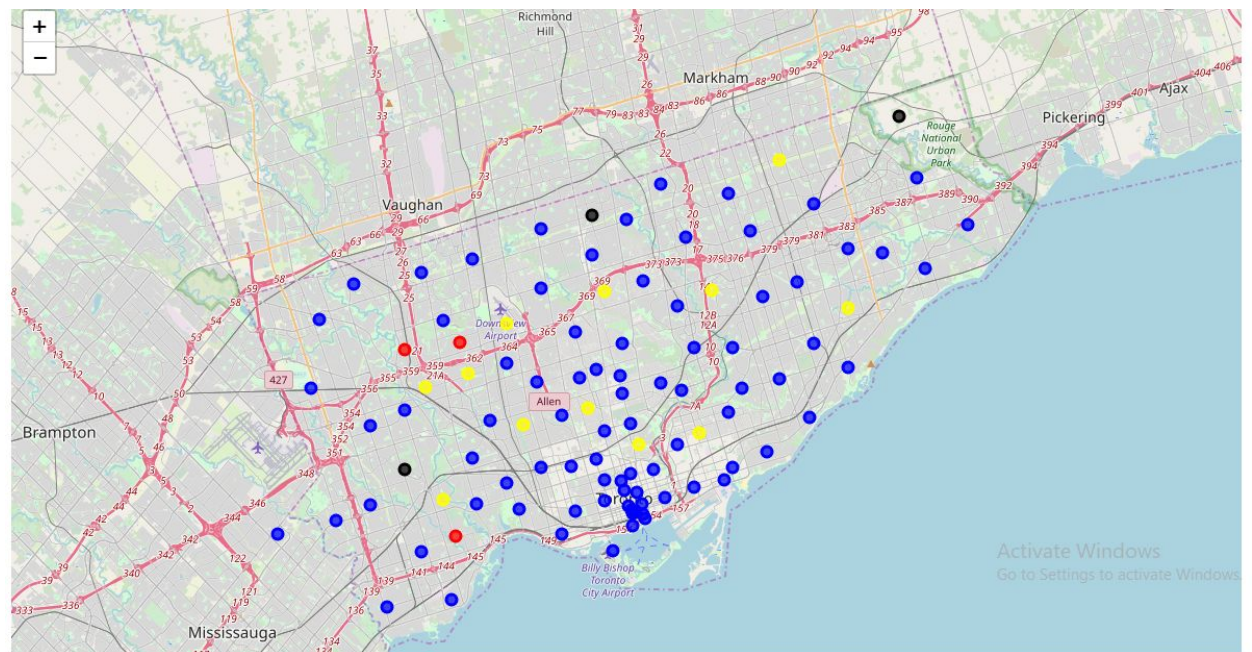
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide,King,Richmond	Coffee Shop	Café	Thai Restaurant	Steakhouse	Salad Place	Restaurant	Burger Joint	Bar	Bakery	Sushi Restaurant
1	Agincourt	Lounge	Latin American Restaurant	Skating Rink	Breakfast Spot	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant
2	Agincourt North,L'Amoreaux East,Miliken,Steel...	Park	Playground	Women's Store	Doner Restaurant	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run
3	Albion Gardens,Beaumont Heights,Humbergate,Jam...	Pizza Place	Fried Chicken Joint	Grocery Store	Pharmacy	Beer Store	Fast Food Restaurant	Sandwich Place	Discount Store	Department Store	Dessert Shop
4	Alderswood,Long Branch	Pizza Place	Skating Rink	Sandwich Place	Coffee Shop	Gym	Pub	Pharmacy	Pool	Deli / Bodega	Department Store

3.2 Clustering Models

Here we have to group similar type venue categories so, we have used K-means clustering models to cluster similar types of venues. We have Taken k=3 (where k is number of cluster) for cluster analysis and the table are:

Postcode	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
M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353	1.0	Fast Food Restaurant	Women's Store	Donut Shop	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Drugstore
M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	1.0	Bar	Women's Store	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Drugstore	Department Store
M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	1.0	Rental Car Location	Mexican Restaurant	Spa	Breakfast Spot	Medical Center	Intersection	Electronics Store	Pizza Place	Concert Hall	Construction & Landscaping
M1G	Scarborough	Woburn	43.770992	-79.216917	1.0	Coffee Shop	Korean Restaurant	Women's Store	Donut Shop	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant
M1H	Scarborough	Cedarbrae	43.773136	-79.239476	1.0	Fried Chicken Joint	Gas Station	Hakka Restaurant	Bakery	Bank	Athletics & Sports	Caribbean Restaurant	Thai Restaurant	Dessert Shop	Dim Sum Restaurant

Finally, let's visualize the resulting clusters. Same color shows the same type of cluster, cluster 1 of red color, cluster 2 of blue color and cluster 3 of yellow color. Here, after visualizing the map We analyze that cluster 1 is smaller than cluster 2 and cluster 3, cluster 2 is larger than cluster 1 and cluster 3, and cluster 3 is larger than cluster 1 and smaller than cluster 2.



In cluster 1, we have analyzed that the 1st most common venue category is Baseball Field.

Cluster 1

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 0, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	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
32	North York	0.0	Baseball Field	Food Truck	Women's Store	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Department Store
91	Etobicoke	0.0	Deli / Bodega	Baseball Field	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Department Store
97	North York	0.0	Baseball Field	Women's Store	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Department Store

In cluster 2, we have analyzed that the 1st most common venue category is Fast Food Restaurant, Coffee Shop, Pizza Place.

Cluster 2

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	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	Scarborough	1.0	Fast Food Restaurant	Women's Store	Donut Shop	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Drugstore
1	Scarborough	1.0	Bar	Women's Store	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Drugstore	Department Store
2	Scarborough	1.0	Rental Car Location	Mexican Restaurant	Spa	Breakfast Spot	Medical Center	Intersection	Electronics Store	Pizza Place	Concert Hall	Construction & Landscaping
3	Scarborough	1.0	Coffee Shop	Korean Restaurant	Women's Store	Donut Shop	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant
4	Scarborough	1.0	Fried Chicken Joint	Gas Station	Hakka Restaurant	Bakery	Bank	Athletics & Sports	Caribbean Restaurant	Thai Restaurant	Dessert Shop	Dim Sum Restaurant
6	Scarborough	1.0	Hobby Shop	Coffee Shop	Discount Store	Convenience Store	Department Store	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Deli / Bodega
7	Scarborough	1.0	Bakery	Bus Line	Park	Metro Station	Soccer Field	Bus Station	Fast Food Restaurant	Intersection	Diner	Discount Store
8	Scarborough	1.0	Motel	American Restaurant	Women's Store	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant

In cluster 3, we have analyzed that the 1st most common venue category is Park.

Cluster 3

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	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
5	Scarborough	2.0	Playground	Women's Store	Doner Restaurant	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Donut Shop
14	Scarborough	2.0	Park	Playground	Women's Store	Doner Restaurant	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run
23	North York	2.0	Park	Convenience Store	Bank	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Women's Store
25	North York	2.0	Park	Food & Drink Shop	Women's Store	Doner Restaurant	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Donut Shop
30	North York	2.0	Park	Airport	Women's Store	Donut Shop	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant

4.Results and Discussion

Our analysis shows that although there is a great number of different types of venue category in neighborhoods of Toronto, In this project, we have direct our efforts on detecting the most common venue category of each neighborhood of Toronto.

For Neighborhood Borrow 'Scarborough' most common venue category is 'Fast Food Restaurant'.

For Neighborhood Borrow 'Etobicoke' most common venue category is 'Pizza Place Store'.

For Neighborhood Borrow 'East Toronto' most common venue category is 'Coffee Shop'.

For Neighborhood Borrow 'Central Toronto' most common venue category is Gym'.

For Neighborhood Borrow 'Downtown Toronto' most common venue category is 'Coffee Shop'.

For Neighborhood Borrow 'North York' most common venue category are 'Baseball Field', 'Grocery Store', 'Park'.

We have seen that the 1st most common venue category in cluster 1 is Baseball Field, in cluster 2 are Fast Food Restaurant, Coffee Shop, Pizza Place, and in cluster 3 is Park.

We have collected the required data, using a given Wikipedia page link and geographical coordinates of each code using csv file. And explored the neighborhoods and segment them using Foursquare API. And we have explored the venues across different areas of Toronto.

After creating the dataframe to display the top 10 venues for each neighborhood, we have analyzed that there are different neighborhoods in the same row have the same 1st most common venues, 2nd most common venue up to 10th most common venues.

5.Conclusion

The purpose of this project was to identify the most common venue category in neighborhoods of Toronto, in order to aid stakeholders in narrowing down the search for the optimal venue category in the neighborhoods of Toronto.

Final decision on optimal neighborhoods venue category will be made by stakeholders based on specific characteristics of neighborhoods, taking into consideration additional factors like the attractiveness of each neighborhood, levels of noise/proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood, etc.