



Abstract : To come up with appropriate location to set up a resto-bar in the city of Toronto

Table of Contents:

1. Business problem
2. Data description
 - 2.1Data Requirements
 - 2.2Data Collection
 - 2.3Data Understanding
3. Methodology
 - 3.1Methodology used
 - 3.2Data Analysis
 - 3.3Clustering Analysis
4. Results
5. Conclusion
6. References

1. Business Problem

An established resto-bar owner wants to expand his business by setting up another branch in Toronto. He approaches a data analytics firm where as a data scientist we will help him gain some insights on location preferences and scope out the competition. He would also like to know the existing places in and around Toronto where he can access foot traffic, parking and accessibility.

2. Data Description

2.1 Data requirements

For the above business problem we will collect the data of all the bars and restaurants in Toronto. The data would contain longitude and latitude information of each location and this can be obtained from foursquare API. This collected data will be used for clustering analysis to find the density of bars and help the client in finding the right place to set up his new branch of resto-bar.

2.2 Data collection

- I. Postal codes of Canada from Wikipedia
(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- II. Longitude and Latitude data for Canadian postal codes.
- III. Geolocations of venues in Toronto.

2.3 Data Understanding

Data	Description	Source
Postal Code	Postal Codes of Canada	https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
Borough	A town or district which is an administrative unit	https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
Neighbourhood	Neighbourhoods near to the area of interest	https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
Latitude	Latitude Co-ordinates of neighbourhoods	Foursquare API
Longitude	Longitude Co-ordinates of neighbourhoods	Foursquare API

Venue	Names of various Venues	Foursquare API
Venue Latitude	Latitude Co-ordinates of Venues	Foursquare API
Venue Longitude	Longitude Co-ordinates of Venues	Foursquare API
Top 10 Most common Venues	Top 10 most common venues categories in 10 columns	Foursquare API

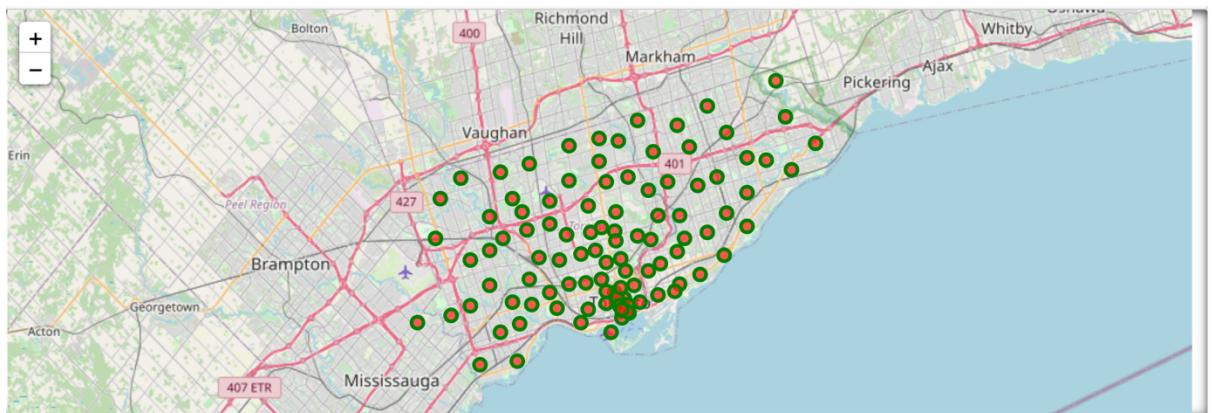
3. Methodology

3.1 Methodology used

- I. EDA(Exploratory Data Analysis)
- II. Clustering Analysis

3.2 Data analysis

- I. Get the postal codes, boroughs, neighbourhoods of Canada from the Wikipedia 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
- II. From this wiki link we are transferring this data to dataframe for better understanding using BeautifulSoup library available in python.
- III. We also collected the latitude and longitude locations of Postal codes of Canada, which is csv file named as 'Geospatial_Coordinates.csv' and read these coordinates into an another dataframe.
- IV. We joined the two dataframes on postal code into a single dataframe.
- V. The data set has various boroughs like 'North York', 'Downtown Toronto', "Queen's Park", 'Etobicoke', 'Scarborough', 'East York', 'York', 'East Toronto', 'West Toronto', 'East York/East Toronto', 'Central Toronto', 'Mississauga', 'Downtown Toronto Stn A', 'Etobicoke Northwest', 'East Toronto Business'.
- VI. Since, we are only interested in the Toronto region, we filtered out the areas which has Toronto in its borough name and named it as geo_data.
- VII. After analysis we figured out that geo_data dataframe has 7 boroughs and 39 neighbourhoods.
- VIII. We, visualized the same using geopy.geocoders library on a map as shown below.



Using Foursquare API, we got details abouts nearby venues like names, latitudes, longitudes, Venue Categories etc as shown below.

```

Parkwoods
Victoria Village
Regent Park, Harbourfront
Lawrence Manor, Lawrence Heights
Ontario Provincial Government
Islington Avenue
Malvern, Rouge
Don Mills North
Parkview Hill, Woodbine Gardens
Garden District, Ryerson
Glencairn
West Deane Park, Princess Gardens, Martin Grove, Islington, Cloverdale
Rouge Hill, Port Union, Highland Creek
Don Mills South
Woodbine Heights
St. James Town
Humewood-Cedarvale
Eringate, Bloordale Gardens, Old Burnhamthorpe, Markland Wood
Guildwood, Morningside, West Hill
The Beaches
Berczy Park
Caledonia-Fairbanks
Woburn
Leaside
Central Bay Street
Christie
Cedarbrae
Hillcrest Village
Bathurst Manor, Wilson Heights, Downsview North
Thorncliffe Park
Richmond, Adelaide, King
Dufferin, Dovercourt Village
Scarborough Village
Fairview, Henry Farm, Oriole
Northwood Park, York University
The Danforth East
Harbourfront East, Union Station, Toronto Islands
Little Portugal, Trinity
Kennedy Park, Ionview, East Birchmount Park
Bayview Village
Downsview East

```

Using onehot encoding technique, we gave weightage for various venues in each neighbourhood as shown below.

```
In [20]: # one hot encoding
tor_onehot = pd.get_dummies(tor_near_values[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
tor_onehot['Neighborhood'] = tor_near_values['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [tor_onehot.columns[-1]] + list(tor_onehot.columns[:-1])
tor_onehot = tor_onehot[fixed_columns]

tor_onehot.head()
```

Out[20]:

	Yoga Studio	Accessories Store	Adult Boutique	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	...	Turkish Restaurant	Vegetarian / Vegan Restaurant	Video Game Store	Video Store	Vietnamese Restaurant	Ware
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

5 rows × 237 columns

Depending on these values, top 5 and top 10 most common venue categories is determined.

```
In [22]: top_ven = 5

for hood in toronto_grp['Neighborhood']:
    print("----"+hood+"----")
    temp = toronto_grp[toronto_grp['Neighborhood'] == hood].T.reset_index()
    temp.columns = ['venue', 'freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(top_ven))
    print('\n')
```

----Agincourt----

	venue	freq
0	Breakfast Spot	0.2
1	Lounge	0.2
2	Clothing Store	0.2
3	Latin American Restaurant	0.2
4	Skating Rink	0.2

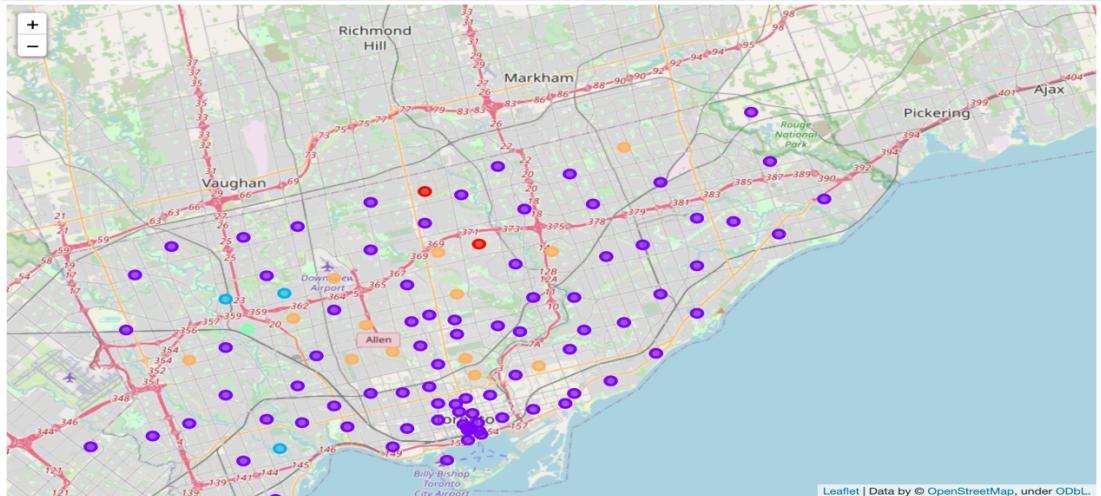
----Alderwood, Long Branch----

	venue	freq
0	Pizza Place	0.29
1	Sandwich Place	0.14
2	Gym	0.14
3	Pub	0.14
4	Coffee Shop	0.14

	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	Agincourt	Lounge	Breakfast Spot	Clothing Store	Latin American Restaurant	Skating Rink	Department Store	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop
1	Alderwood, Long Branch	Pizza Place	Gym	Pub	Skating Rink	Sandwich Place	Coffee Shop	Dog Run	Donut Shop	Distribution Center	Dance Studio
2	Bathurst Manor, Wilson Heights, Downsview North	Bank	Coffee Shop	Pizza Place	Deli / Bodega	Sushi Restaurant	Supermarket	Bridal Shop	Diner	Shopping Mall	Sandwich Place
3	Bayview Village	Café	Chinese Restaurant	Bank	Japanese Restaurant	Department Store	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop
4	Bedford Park, Lawrence Manor East	Sandwich Place	Coffee Shop	Italian Restaurant	Pharmacy	Pub	Juice Bar	Butcher	Fast Food Restaurant	Café	Indian Restaurant

3.3 Clustering Analysis

- I. We have used k-Means Clustering Machine learning Algorithm for finding the clusters in Toronto for different venues.
- II. We have divided the region and venues into 5 clusters (k=5) with top 10 venues.
- III. We have added a new column named ‘Cluster Labels’ to identify and retrieve data related to specific clusters as depicted below.



Cluster 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	
45	North York	0	Park	Women's Store	Dance Studio	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center	Discount Store
52	North York	0	Park	Women's Store	Dance Studio	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center	Discount Store

Cluster 2:

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	
1	North York	1	Intersection	Portuguese Restaurant	Coffee Shop	French Restaurant	Hockey Arena	Diner	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant
2	Downtown Toronto	1	Coffee Shop	Bakery	Park	Breakfast Spot	Gym / Fitness Center	Historic Site	Restaurant	Pub	Performing Arts Venue	Mexican Restaurant
3	North York	1	Clothing Store	Furniture / Home Store	Accessories Store	Event Space	Boutique	Vietnamese Restaurant	Coffee Shop	Distribution Center	Dim Sum Restaurant	Diner
4	Queen's Park	1	Coffee Shop	Sushi Restaurant	Gym	Hobby Shop	Bar	Beer Bar	Spa	Smoothie Shop	Burrito Place	Sandwich Place
5	Etobicoke	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...
97	Downtown Toronto	1	Café	Coffee Shop	Restaurant	Seafood Restaurant	Hotel	Gym / Fitness Center	Art Gallery	Speakeasy	Japanese Restaurant	Steakhouse
98	Etobicoke	1	Pool	River	Women's Store	Dance Studio	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center
99	Downtown Toronto	1	Bookstore	Hobby Shop	Dance Studio	Sushi Restaurant	Ice Cream Shop	Indian Restaurant	Martial Arts School	Restaurant	Italian Restaurant	Beer Bar
100	East Toronto Business	1	Gym / Fitness Center	Spa	Auto Workshop	Brewery	Burrito Place	Comic Shop	Farmers Market	Fast Food Restaurant	Garden	Garden Center
102	Etobicoke	1	Grocery Store	Tanning Salon	Convenience Store	Sandwich Place	Discount Store	Burrito Place	Burger Joint	Supplement Shop	Bakery	Fast Food Restaurant

84 rows x 12 columns

Cluster 3:

	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
53	North York	2	Baseball Field	Food Truck	Women's Store	Falafel Restaurant	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run
57	North York	2	Baseball Field	Women's Store	Farmers Market	Event Space	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run
101	Etobicoke	2	Baseball Field	Women's Store	Farmers Market	Event Space	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run

Cluster 4:

	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
64	York	3	Convenience Store	Women's Store	Deli / Bodega	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center

Cluster 5:

	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	North York	4	Bus Stop	Fast Food Restaurant	Park	Food & Drink Shop	Women's Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center
10	North York	4	Playground	Park	Bakery	Japanese Restaurant	Deli / Bodega	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run
16	York	4	Field	Park	Trail	Hockey Arena	Women's Store	Discount Store	Department Store	Dessert Shop	Dim Sum Restaurant	Diner
21	York	4	Park	Women's Store	Pool	Dance Studio	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center
35	East York/East Toronto	4	Coffee Shop	Park	Convenience Store	Women's Store	Deli / Bodega	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run
40	North York	4	Airport	Park	Women's Store	Deli / Bodega	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run
49	North York	4	Basketball Court	Bakery	Park	Trail	Construction & Landscaping	Dog Run	Diner	Discount Store	Distribution Center	Women's Store
61	Central Toronto	4	Park	Swim School	Dim Sum Restaurant	Bus Line	Women's Store	Department Store	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop
66	North York	4	Park	Convenience Store	Women's Store	Deli / Bodega	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center
77	Etobicoke	4	Park	Mobile Phone Shop	Sandwich Place	Women's Store	Dance Studio	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center
83	Central Toronto	4	Lawyer	Park	Tennis Court	Women's Store	Deli / Bodega	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run
85	Scarborough	4	Playground	Intersection	Park	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center	Discount Store

4. Results

By observing the above clusters, cluster 2 with cluster label 1 has highest foot fall and also have Bars and restaurants as common places. Cluster 5 is also a prominent cluster which has next height common venues. Cluster 1,3 and 4 has not much of common venues. Generally for the resto-bar, we need venues which has hight foot fall and most diverse places.

5. Conclusion

From the above clustering analysis we can conclude that cluster 2 has highest foot fall and has few bars and restaurants. We can suggest cluster 2 for the client to have better competitive advantage. Considering the cost if client requests for an alternative location we can suggest the cluster 5 which is also having diversified common venue's.

6. References

- List of postal codes of Canada:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Foursquare : <https://foursquare.com/developers/apps>
- Notebook :
https://github.com/Haripawan/Coursera_Capstone/blob/709a4fb8184b2de895d5cbf5e9168f7e57116764/Capstone_find_right_place%20for_restobar.ipynb