Suggesting best venues for stay in Toronto for Tourism

Navaneeth Krishna Vadapalli

August 15, 2019

1. Introduction

1.1. Background

Tourism is a social, cultural and economic phenomenon which entails the movement of people to countries or places outside their usual environment for personal or business/professional purposes. When looking at tourism it's important to consider the term hospitality. Some define hospitality as "the business of helping people to feel welcome and relaxed and to enjoy themselves" Simply put, the hospitality industry is the combination of the accommodation and food and beverage groupings, collectively making up the largest segment of the industry.

Tourism generates about **Half a Billion Dollar revenue** worldwide. World Tourism Organization (UNWTO) is the most widely recognized and the leading international organization in travel and tourism today. It is a specialized agency of the United Nations. It serves as a global forum for tourism policy and a practical source of tourism know-how. With its headquarters in Madrid, Spain the World Tourism Organization plays a central and decisive role in promoting development of responsible, sustainable and universally accessible tourism.

1.2. Problem Statement

Looking back into time there is a significant growth the rate of development in Tourism and so has the rate of exploitation by the local neighborhoods when a tourist comes to an unknown place. Many a times Tourists end up staying in places where the venues are very fat apart causing a lot of inconvenience. So, wouldn't it be better if we had means and the ability to suggest best neighborhoods based on the Tourist basic amenities like café, diner, gym, hotels, restaurants etc. The scope of venues in this project is restricted to Toronto, Canada.

1.3. Interest or Target Audience

With the explosion of technology in the modern era, the ease of connectivity to internet and data is made a lot easier. But still customer satisfaction is a factor which can never be determined. If this solution is integrated with an API and website to display the search results based on user input. It can tap the untapped market and could very well be a means of generating large sums of revenue. Majority of the target audience are the end users of mobile, laptop, iPad, tablets etc.

2. Data acquisition and cleaning

2.1. Data Sources

The source of the data is Wikipedia and the links are attached below:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

The link contains data about the neighborhoods in Toronto.

https://foursquare.com/

To get the venues around a location.

http://cocl.us/Geospatial_data

To get the co-ordinates of various neighborhoods of Toronto.

2.2. Data Cleaning

The final data was obtained by combining data from multiple sources into one. The data from Wikipedia has to be scrapped cleaned and obtain the columns PostalCode, Borough, and Neighborhood. Later these had to be integrated with the co-ordinates (Latitudes and Longitudes). The processing involved working with the cells that have an assigned borough. Ignore cells with a borough that is Not assigned. More than one neighborhood can exist in one postal code area. For example, in the table on the Wikipedia page, you will notice that M5A is listed twice and has two neighborhoods: Harbourfront and Regent Park. These two rows will be combined into one row with the neighborhoods separated with a comma. a cell has a borough but a **Not assigned** neighborhood, then the neighborhood will be the same as the borough.

3. Methodology

3.1. Exploratory data Analysis

3.1.1. Part 1

The preliminary data to create the Folium Map was prepared and the screenshot of the dataframe is below.

The data has 103 records and 3 columns namely PostalCode, Borough, and Neighborhood.

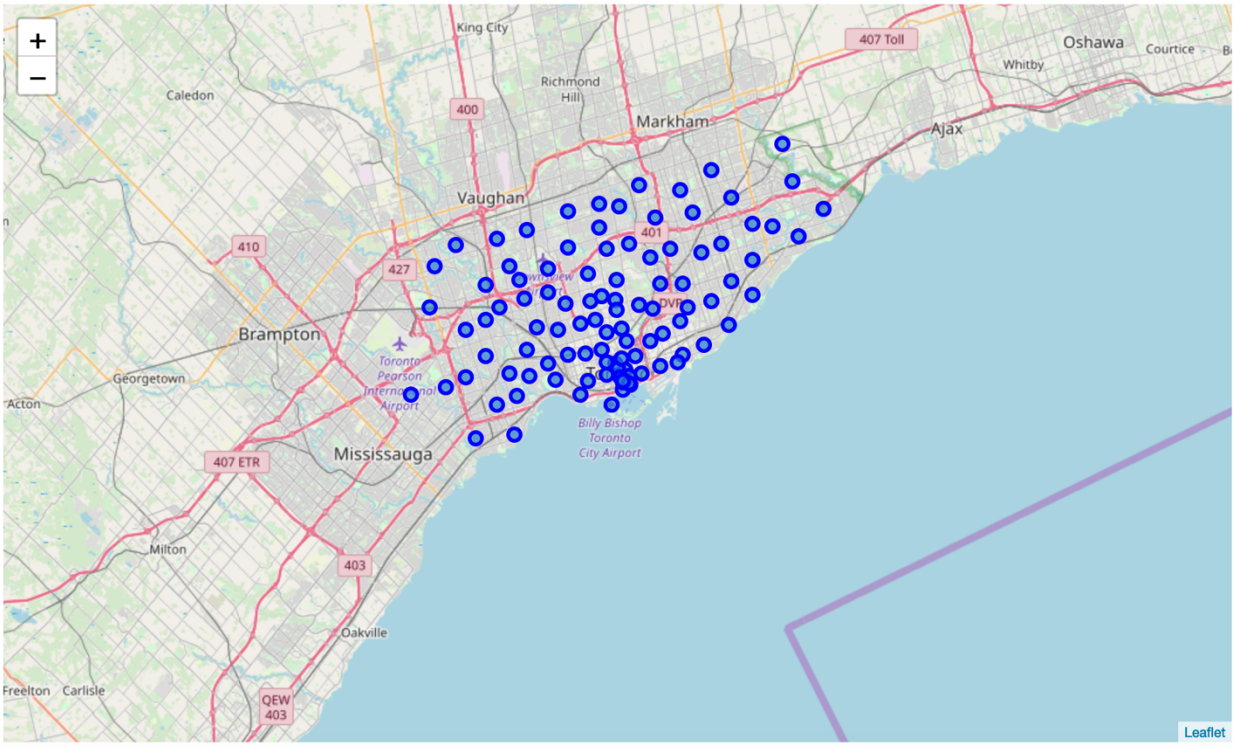
	Postalcode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront, Regent Park
3	M6A	North York	Lawrence Heights, Lawrence Manor
4	M7A	Queen's Park	Queen's Park
5	M9A	Etobicoke	Islington Avenue
6	M1B	Scarborough	Rouge, Malvern
7	M3B	North York	Don Mills North
8	M4B	East York	Woodbine Gardens, Parkview Hill
9	M5B	Downtown Toronto	Ryerson, Garden District
10	M6B	North York	Glencairn
11	M9B	Etobicoke	Cloverdale, Islington, Martin Grove, Princess ...

3.1.2. Part 2

After obtaining data from the Wikipedia, the data had to be merged with the co-ordinates; Latitude and Longitude so that this could be plotted on a Folium map.

The merged dataframe has a size of (103,5) and the dataframe is attached below along with the Folium map visualization:

	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	Harbourfront, Regent Park	43.654260	-79.360636
3	M6A	North York	Lawrence Heights, Lawrence Manor	43.718518	-79.464763
4	M7A	Queen's Park	Queen's Park	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242
6	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
7	M3B	North York	Don Mills North	43.745906	-79.352188
8	M4B	East York	Woodbine Gardens, Parkview Hill	43.706397	-79.309937
9	M5B	Downtown Toronto	Ryerson, Garden District	43.657162	-79.378937
10	M6B	North York	Glencairn	43.709577	-79.445073
11	M9B	Etobicoke	Cloverdale, Islington, Martin Grove, Princess ...	43.650943	-79.554724
12	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
13	M3C	North York	Flemingdon Park, Don Mills South	43.725900	-79.340923
14	M4C	East York	Woodbine Heights	43.695344	-79.318389
15	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
16	M6C	York	Humewood-Cedarvale	43.693781	-79.428191



3.1.3. Part 3

Using Foursquare API to get nearby venues:

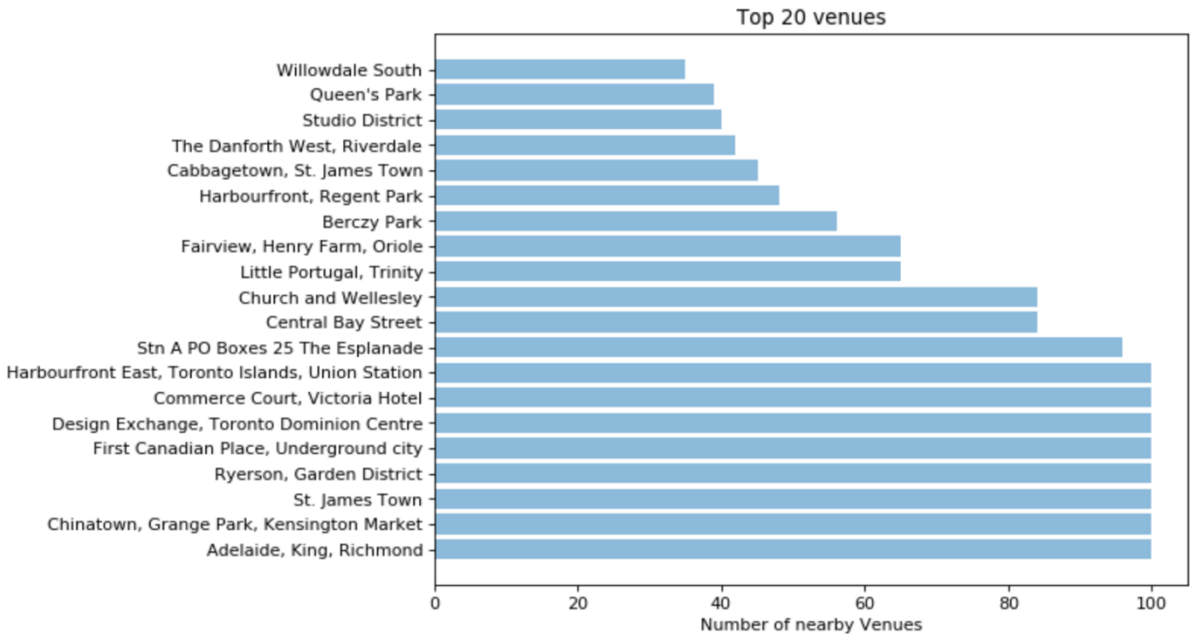
	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	KFC	43.754387	-79.333021	Fast Food Restaurant
2	Parkwoods	43.753259	-79.329656	TTC stop #8380	43.752672	-79.326351	Bus Stop
3	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
4	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
5	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
6	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
7	Victoria Village	43.725882	-79.315572	Eglinton Ave E & Sloane Ave/Bermondsey Rd	43.726086	-79.313620	Intersection
8	Victoria Village	43.725882	-79.315572	Cash Money	43.725486	-79.312665	Financial or Legal Service
9	Harbourfront, Regent Park	43.654260	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery

```
Adelaide, King, Richmond
                              Venue  frequency
0           Venue Category_Coffee Shop       0.08
1                  Venue Category_Café       0.05
2                   Venue Category_Bar       0.04
3            Venue Category_Steakhouse       0.04
4       Venue Category_Thai Restaurant       0.04
5        Venue Category_Breakfast Spot       0.03
6  Venue Category_American Restaurant       0.03
7                 Venue Category_Hotel       0.03
8                   Venue Category_Gym       0.03
9            Venue Category_Restaurant       0.03


Agincourt
                                       Venue  frequency
0                   Venue Category_Breakfast Spot       0.25
1                          Venue Category_Lounge       0.25
2                   Venue Category_Clothing Store       0.25
3                     Venue Category_Skating Rink       0.25
4               Venue Category_Miscellaneous Shop       0.00
5                    Venue Category_Movie Theater       0.00
6                           Venue Category_Motel       0.00
7             Venue Category_Monument / Landmark       0.00
8  Venue Category_Molecular Gastronomy Restaurant       0.00
9       Venue Category_Modern European Restaurant       0.00


Agincourt North, L'Amoreaux East, Milliken, Steeles East
                                       Venue  frequency
0                       Venue Category_Playground        0.5
1                             Venue Category_Park        0.5
2               Venue Category_Mexican Restaurant        0.0
3                           Venue Category_Motel        0.0
4             Venue Category_Monument / Landmark        0.0
5  Venue Category_Molecular Gastronomy Restaurant        0.0
6       Venue Category_Modern European Restaurant        0.0
```

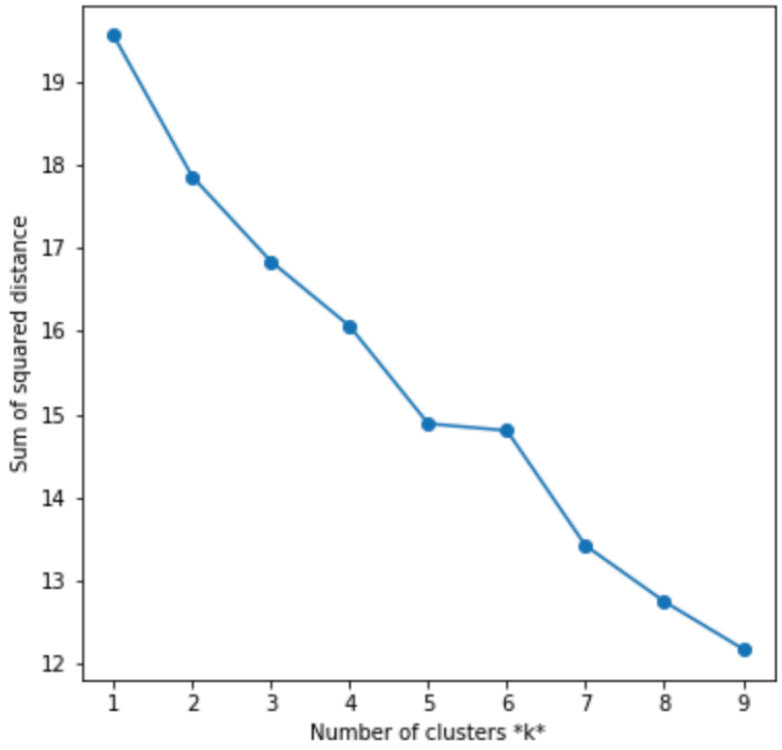
The screenshot above indicates the top 10 venues visited around the neighborhoods and the frequency (% it comprises for).

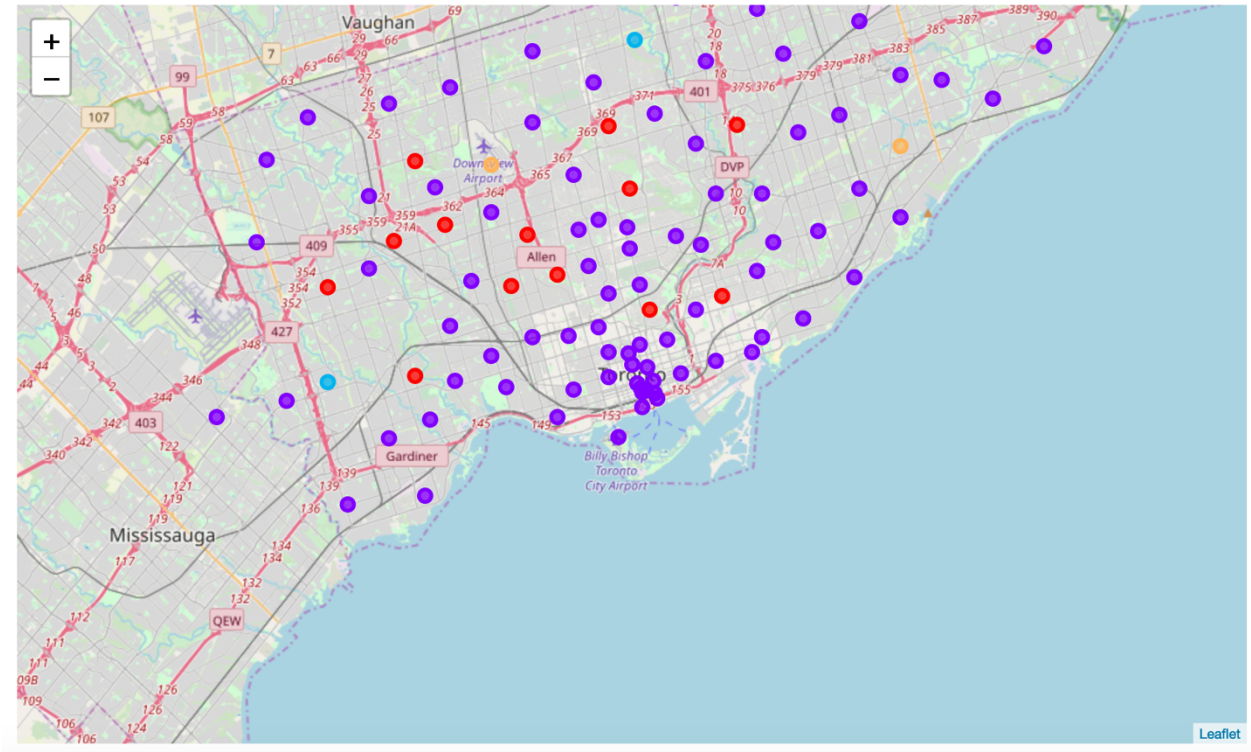


Top 20 neighborhoods which have highest number of venues located within 500 meters of the location.

4. Results

Since it is a unsupervised Machine Learning problem, Cluster Analysis is one of the preliminary methods that can be used to group similar entities together. Analyzing the sum of squared distances of the centroids of the cluster. The optimal value of k = 5. The reason being the elbow joint occurring at k = 5. The Cluster 2 is very highly concentrated having 81 Burrows in the group. This implies that the 81 neighborhoods in that area have very high similarity and is the best cluster to live in. The reasons being high availability of the venues around the neighborhood and having a variety of options to explore. The Cluster 1 stands second as there are 13 Burrows in the group.


Sum of squared distance
Number of clusters *k*
19
18
17
16
15
14
13
12
1
2
3
4
5
6
7
8
9


Vaughan
Mississauga
Toronto
Allen
DVP
Gardiner
QEW
401
409
427
403
99
7
107
Billy Bishop Toronto City Airport
Leaflet

5. Discussion

	Neighbourhood	Number of nearby Venues	Cluster Labels	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7	Rank 8	Rank 9	Rank 10
0	Adelaide, King, Richmond	100	1	Coffee Shop	Café	Thai Restaurant	Bar	Steakhouse	Breakfast Spot	Hotel	Restaurant	American Restaurant	Gym
1	Chinatown, Grange Park, Kensington Market	100	1	Café	Vegetarian / Vegan Restaurant	Chinese Restaurant	Mexican Restaurant	Bakery	Vietnamese Restaurant	Dumpling Restaurant	Bar	Coffee Shop	Park
2	St. James Town	100	1	Café	Hotel	Coffee Shop	Restaurant	Italian Restaurant	Gastropub	Cosmetics Shop	Beer Bar	Bakery	Breakfast Spot
3	Ryerson, Garden District	100	1	Coffee Shop	Clothing Store	Cosmetics Shop	Middle Eastern Restaurant	Café	Restaurant	Bookstore	Japanese Restaurant	Diner	Ice Cream Shop
4	First Canadian Place, Underground city	100	1	Coffee Shop	Café	Hotel	Restaurant	Steakhouse	Bar	Seafood Restaurant	American Restaurant	Gym	Gastropub
5	Design Exchange, Toronto Dominion Centre	100	1	Coffee Shop	Café	Hotel	Restaurant	Bakery	Gastropub	Gym	Italian Restaurant	Bar	American Restaurant
6	Commerce Court, Victoria Hotel	100	1	Coffee Shop	Hotel	Café	American Restaurant	Restaurant	Gastropub	Bakery	Deli / Bodega	Steakhouse	Gym
7	Harbourfront East, Toronto Islands, Union Station	100	1	Coffee Shop	Hotel	Aquarium	Café	Italian Restaurant	Sporting Goods Shop	Bakery	Brewery	Scenic Lookout	Pizza Place
8	Stn A PO Boxes 25 The Esplanade	96	1	Coffee Shop	Restaurant	Café	Seafood Restaurant	Italian Restaurant	Cocktail Bar	Beer Bar	Hotel	Bakery	Creperie

Since it is an unsupervised Machine Learning problem, Cluster Analysis is one of the preliminary methods that can be used to group similar entities together. Analyzing the sum of squared distances of the centroids of the cluster. The optimal value of k = 5. The reason being the elbow joint occurring at k = 5. The Cluster 2 is very highly concentrated having 81 Burrows in the group. This implies that the 81 neighborhoods in that area have very high similarity and is the best cluster to live in. The reasons being high availability of the venues around the neighborhood and having a variety of options to explore. The Cluster 1 stands second as there are 13 Burrows in the group.

Based on the use case of **Suggesting best venues for stay in Toronto for Tourism:**

The best neighbourhood could be any one from the top 8 entries. But based on the variety and of-course the case provided, the best Neighborhood would be 'Harbourfront East, Toronto Islands, Union Station'.

6. Conclusion

The data in the scope of the project cannot be decisively state or claim the best neighborhood as the factors such as reviews, number of visits made by people, costs, menu, sanitation and other vital factors are not taken into consideration. For future scenarios if these factors are tapped into and a model is prescribed on it, then the analysis would be more rigid and accurate.