A Comparative Study between Toronto & Mumbai

OMAR LINARES

Date : 10 May 2021

1. Introduction/Background

Toronto and Mumbai are the financial capitals of Canada & India respectively. Both the cities are very different in many aspects like demography, climate, culture etc. At the same time both the cities are prominent tourist attractions due to their diverse, multicultural and wide range of experiences. In this study we have grouped the neighbourhoods of Toronto & Mumbai to compare the similarities and differences.

2. Business Problem

Our main aim is to group the neighbourhoods of Toronto & Mumbai to help the stakeholders take informed decision while planning to travel or relocate in either of the cities. There are two types of stakeholders in this case –

a) Tourists/Travel Agents – who can look around in the venues, categories and locations and plan accordingly as per their choice of experience.
b) Migrants – people who want to migrate or relocate in these cities can find all details in their choice of neighbourhood, like departmental store, Bank, Park, medical shop etc.

3. Data

We have obtained Neighbourhood data along with postal codes, Latitude and Longitude for both the cities from websites. Also we have used FourSquare API to get the venues corresponding to each neighbourhood. Below are the sources of our data collection –

a) Toronto Data – Wikipedia page https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=1011037969 has the details of all the Borough, Neighbourhoods along with postal codes of Toronto. We have obtained latitude and longitude of those places from https://cocl.us/Geospatial_data .

b) Mumbai Data – We have obtained all the Neighbourhoods of Mumbai along with postal codes, latitude & longitude from the website https://geographic.org/streetview/india/maharashtra/konkan/mumbai.html . However, I have saved the data in excel format after cleaning & sorting in the Github location

c) FourSquare API Data – To obtain the venues of a neighbourhood, we have used the FourSquare API, which is a location data provided with all kinds of information on venues and events within an area. This includes venue name, category, reviews, images etc. We have obtained list of venues around 500 meters of each neighbourhood. After obtaining the information, we have arranged the data by venue name & category for our exploratory data analysis and clustering of the neighbourhoods. We have also used the same data to find out top 5 categories of venue for each neighbourhood.

4. Methodology

We have used python programming to solve our problem. We will briefly describe the method as follows :

a) Packages used – Requests: reads http data, pandas: to manipulate data and use data frames for analysis, numpy: to handle arrays of data, matplotlib: to generate maps, folium: to generate geospatial maps of Toronto & Mumbai, Sklearn: to use K-means for clustering machine learning.

b) Steps involved – we have done this analysis in 9 steps. First, downloading the data for each city neighbourhood from the website. Second, obtaining the latitude and longitude of these locations. Third, getting a map of the city using folium. Fourth, use Foursquare API to obtain venues near to each neighbourhood. Fifth, sort the data by venue names and categories. Sixth, create dummy variables for all the venue categories and take the means for each neighbourhood. Seventh, find out top 5 venue categories for each neighbourhood. Eighth, cluster the neighbourhoods using K-means machine learning and the venue categories. Ninth, create a map using folium showing all the neighbourhoods grouped in different colours.

c) Exploratory Data Analysis – In this process, we have explored the top 5 venue category for each neighbourhood. This data is view like below

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Agincourt	Latin American Restaurant	Lounge	Skating Rink	Breakfast Spot	Clothing Store
1	Alderwood, Long Branch	Pizza Place	Coffee Shop	Pub	Sandwich Place	Pool
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Bank	Shopping Mall	Ice Cream Shop	Sushi Restaurant
3	Bayview Village	Café	Chinese Restaurant	Japanese Restaurant	Bank	Department Store
4	Bedford Park, Lawrence Manor East	Sandwich Place	Italian Restaurant	Coffee Shop	Sushi Restaurant	Juice Bar
5	Berczy Park	Coffee Shop	Cocktail Bar	Beer Bar	Farmers Market	Seafood Restaurant
6	Birch Cliff, Cliffside West	Café	College Stadium	Skating Rink	General Entertainment	Dance Studio
7	Brockton, Parkdale Village, Exhibition Place	Café	Coffee Shop	Breakfast Spot	Nightclub	Bar
8	Business reply mail Processing Centre, South C...	Light Rail Station	Skate Park	Garden Center	Fast Food Restaurant	Farmers Market

Table for Toronto

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	A I Staff Colony	Athletics & Sports	Department Store	Snack Place	Lounge	Pool
1	Aareymilk Colony	Dance Studio	Mountain	Event Space	Zoo	Duty-free Shop
2	Agripada	Athletics & Sports	Platform	Coffee Shop	Bank	Bakery
3	Airport	Hotel	Coffee Shop	Bar	Café	Indian Restaurant
4	Ambewadi	Juice Bar	Fast Food Restaurant	Indian Restaurant	Breakfast Spot	Chinese Restaurant
5	Andheri	Yoga Studio	Shopping Mall	Vegetarian / Vegan Restaurant	Sandwich Place	Indian Restaurant
6	Andheri East	Hotel	Hotel Bar	Vegetarian / Vegan Restaurant	Pizza Place	Bistro
7	Andheri Railway Station	Indian Restaurant	Fast Food Restaurant	Paper / Office Supplies Store	Salon / Barbershop	Burger Joint
8	Antop Hill	Music Venue	Zoo	Donut Shop	Flea Market	Fish Market
9	Asvini	Soccer Field	Café	Zoo	Convenience Store	Flea Market
10	Audit Bhavan	Pizza Place	Brazilian Restaurant	Food Court	Gym	Fast Food Restaurant
11	Azad Nagar	Indian Restaurant	Snack Place	Vegetarian / Vegan Restaurant	South Indian Restaurant	Café
12	B P T Colony	Diner	Music Venue	Zoo	Duty-free Shop	Flea Market
13	B.N. Bhavan	Indian Restaurant	Gym	Zoo	Donut Shop	Flea Market
14	B.P.Lane	Indian Restaurant	Indian Sweet Shop	American Restaurant	Rest Area	BBQ Joint

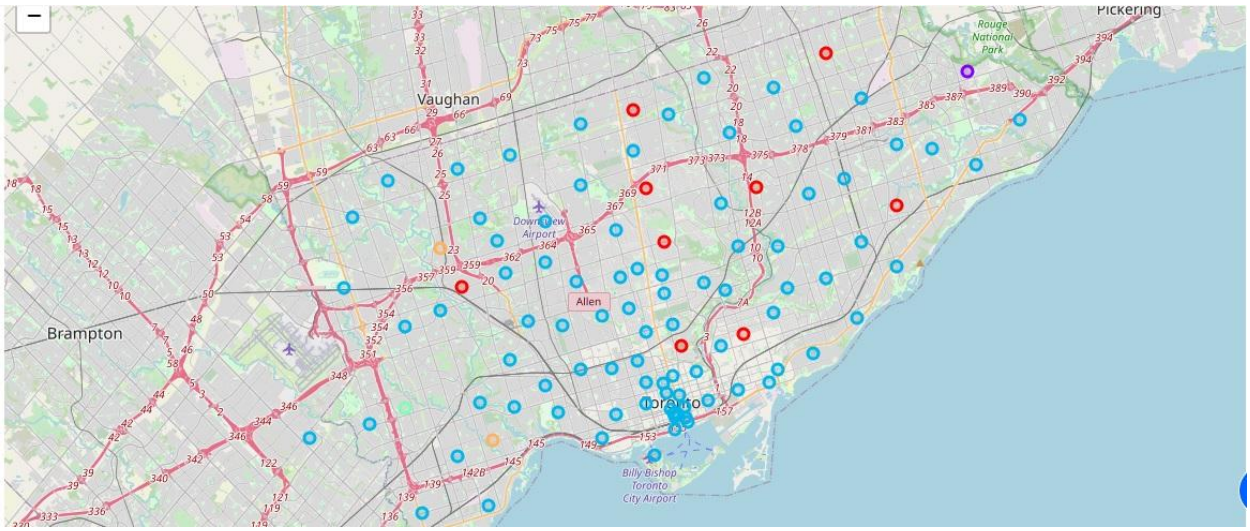
Table for Mumbai

d) Machine Learning Technique used – We have used K-means clustering from Sklearn package. This is a unsupervised machine learning technique. We have fixed the number of clusters to 5 for both the cities.

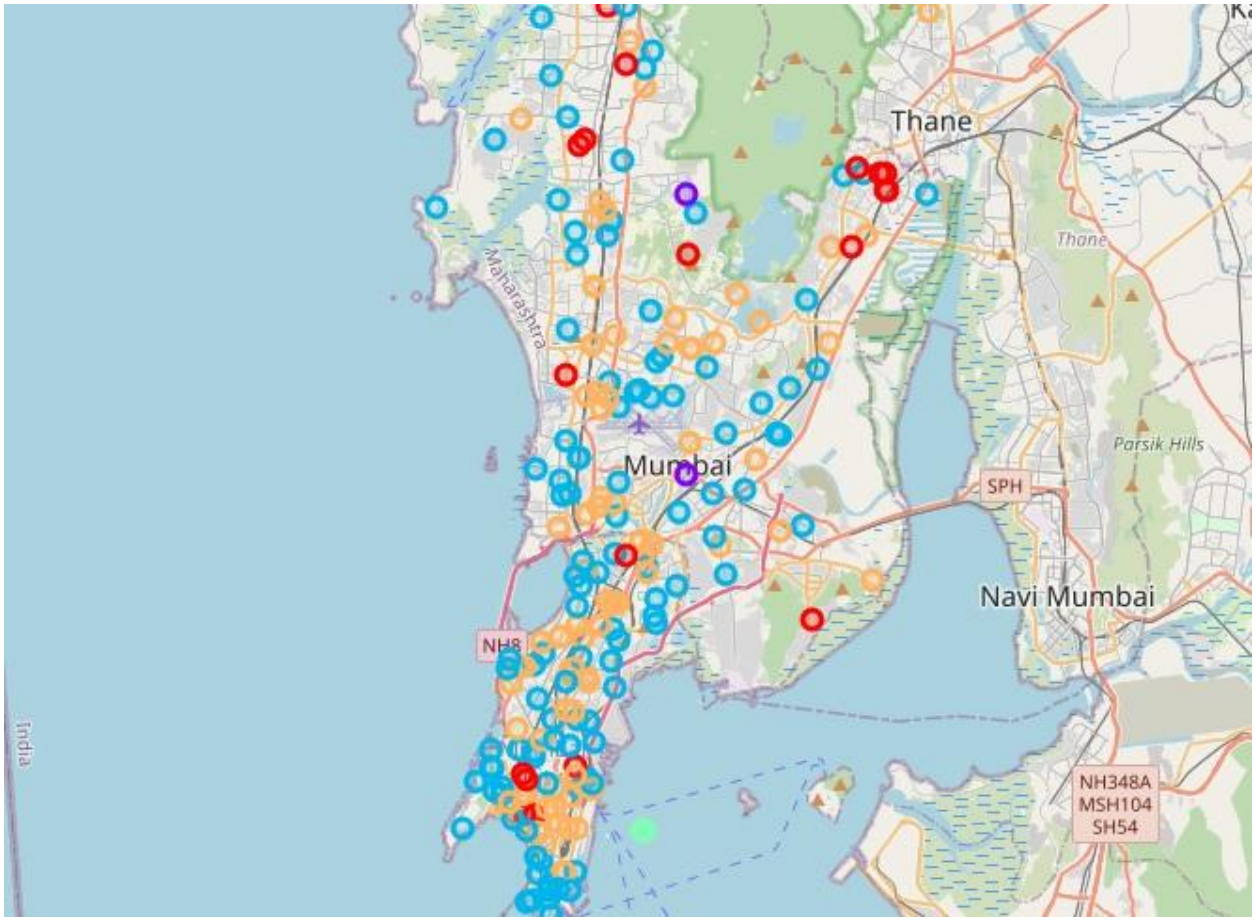
5. Results

	TORONTO	MUMBAI
Cluster 1	*Food and Entertainment cluster: 89 neighbourhoods. Mainly coffee shops, various restaurants, cafe, bar, pub, bakery, park, gym etc.*	*Fast food cluster: 20 neighbourhoods. Mostly Fast food restaurants, snacks, ice cream shops*
Cluster 2	*Relaxation cluster: 1 neighbourhood. Bar, falafel restaurant and escape roooms at top 3*	*Explore cluster: 2 neighbourhoods. Mainly bar, zoo and food courts.*
Cluster 3	*Life Style Cluster: 7 neighbourhoods. Mainly parks, convenience stores, yoga studios at the top*	*Entertainment cluster: 119 neighbourhoods. Mainly dance studios, sport clubs, music cafe, lounges, multiplexes, theatres, markets, gym and spa.*
Cluster 4	*Fitness Cluster: 2 neighbourhoods. Mainly basketball fields, yoga studios, falafel restaurants etc.*	*Life Style cluster: 5 cluster. Mostly ferry, departmental stores, electronics shops, food and flower markets.*
Cluster 5	*Activity cluster: 1 neighbourhood. Dance studio, restaurant and electronics shop at the top.*	*Ethnic cluster: 8 neighbourhoods. Mainly Indian restaurants, hotels and cafes.*

Toronto clusters are seen in the map as follows:



Mumbai clusters are seen in the map as follows:



6. Discussion

We can see from the analysis that both the cities have food, entertainment and life style in common. On the other hand Toronto has more of fitness and relaxation venues, whereas Mumbai has more of ethnic and explore venues. As explained before both the cities are commonly known for its multicultural and diverse range of experiences, however they have certain differences and that mainly because of demographic and cultural backgrounds.

7. Conclusion

The purpose of the study was to group the neighbourhoods of Toronto and Mumbai and find out similarities and differences which could provide a lot of information to tourists and migrants. We have explored both the cities using postal codes of the neighbourhoods and listed out various types of venues and activities, ranging from food, entertainment, fitness, departmental store, park, zoo, ferry and many more. We have clustered the neighbourhoods into 5 categories based on the venues. Now it's upto the stake holders to use this data and analysis to find out which cluster or neighbourhood suits their interest and decide accordingly.