

Capstone Project

Hotel Recommendation for Toronto Tourists

Zhanweidian - December 19, 2019

Introduction

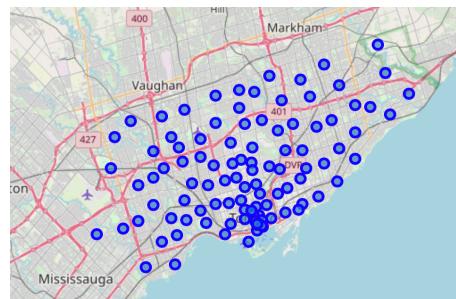
Image there is a tourist who is new to Toronto and he would like some advice on hotels. He may want to visit some cultural attractions in Toronto or taste French food etc.

Anyway, the idea is that, by using the location data of the trending venues he/she may interested in and the hotel rating data from Foursquare.com, we can figure out which hotels are better for him by clustering or some other models and generate a list of hotels for recommendation.

Data

The data we would use to recommend hotels to tourists:

1. The location data of boroughs of Toronto from wikipedia;
2. The trending venues and their categories of each borough from Foursquare.com;
3. The location and rating data of the hotels nearby the cluster centre from Foursquare.com.

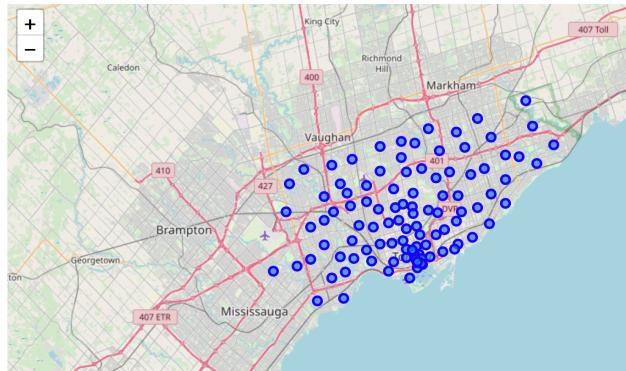


Methodology

Firstly, we get the list of boroughs of Toronto from wikipedia, and get the corresponding location data based on the postal-codes by Geopy library.

PostalCode	Borough	Neighbourhood	Postal Code	Latitude	Longitude
0 M3A	North York	Parkwoods	0 M1B	43.806686	-79.194353
1 M4A	North York	Victoria Village	1 M1C	43.784535	-79.160497
2 M5A	Downtown Toronto	Regent Park	2 M1E	43.763573	-79.188711
3 M6A	North York	Lawrence Heights	3 M1G	43.770992	-79.216917
4 M6A	North York	Lawrence Manor	4 M1H	43.773136	-79.239476
5 M7A	Queen's Park (Toronto)	Queen's Park (Toronto)
6 M9A	Downtown Toronto	Queen's Park (Toronto)	98 M9N	43.706876	-79.518188
7 M1B	Scarborough Toronto	Rouge Toronto	99 M9P	43.696319	-79.532242
8 M1B	Scarborough Toronto	Malvern Toronto	100 M9R	43.688905	-79.554724
9 M3B	North York	Don Mills North			

We can draw the locations of all the boroughs to take a look at these places on the map, the figure below shows how they spread in Toronto City.

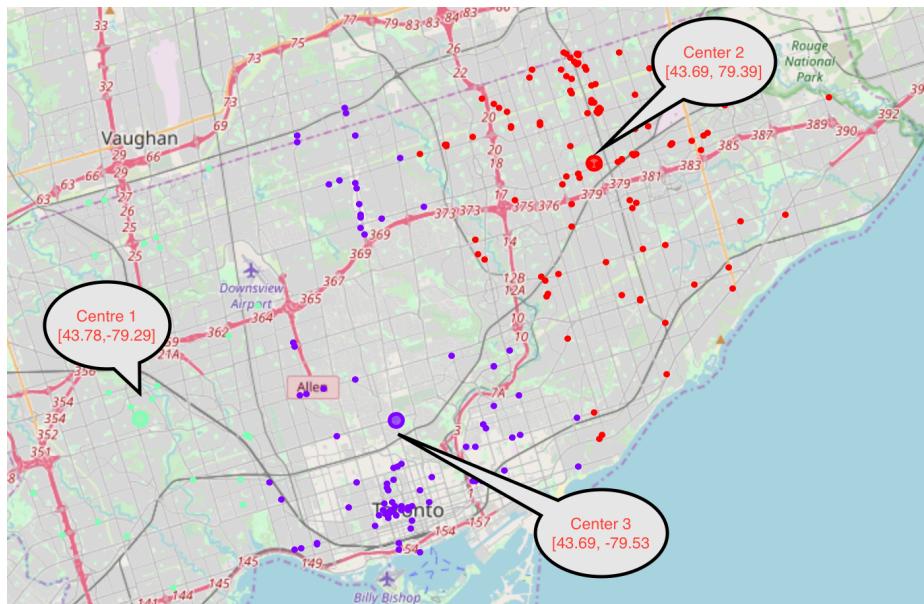


id	ir	address	categories	cc	city	country	crossStreet	distance	formattedAddress
17	4c674ced7abde21e94536668	2873 lawrence avenue east	Chinese Restaurant	CA	Toronto	Canada	NaN	1647	[2873 lawrence avenue east, Toronto ON, Canada]
27	52e2afb0498e24a2f325a2d2	4400 Brimley Road	Chinese Restaurant	CA	Toronto	Canada	NaN	988	[4400 Brimley Road, Toronto ON, Canada]
3	4b7054e3f964a5204d132de3	4379 Kingston Road	None	CA	Toronto	Canada	NaN	2211	[4379 Kingston Road, Toronto ON, Canada]
104	58c38b0dc05d12d5e823ee3	4750 Yonge St. (Unit 2103), Toronto ON M2N 5M...	Chinese Restaurant	CA	Toronto	Canada	Unit 2103	1281	[4750 Yonge St. (Unit 2103), Toronto ON M2N 5M...]
1	5345f39a498e016dd2d7ebf2	NaN	Chinese Restaurant	CA	Toronto	Canada	NaN	2252	[Toronto ON, Canada]
26	4b92b14ff964a5204a1234e3	NaN	Church	CA	NaN	Canada	NaN	702	[Canada]
160	4c3a5d9e1e06d13a1d437b3e	NaN	Chinese Restaurant	CA	NaN	Canada	NaN	2159	[Canada]

Secondly, we use these location data of boroughs to get the trending venues of each boroughs using Foursquare.com. Based on the preference of the tourist, filter the data according to the categories of the venues.

Now that we have the location data of each borough, we can find all the nearby venues using the Places API of foursquare.com. Supposed our tourist want to taste some Chinese (forget about why he/she do this in Toronto), we now query “Chinese” and a radius of 2000 meters on each boroughs and get all the Chinese restaurants nearby.

Thirdly, use clustering algorithms to cluster the venues and find out the clustering centre points. We want to know how these restaurants spread in Toronto, and by using clustering algorithm, we can find out the center point of each clusters.



These center points of the clusters help us determine the location of hotels we recommend the tourist to live. We use the foursquare.com's Places API again to search for hotels nearby these center points and sort the hotels in descending order to find out the better hotels with relatively higher rating.

	name	address	distance	rating
0	Four Seasons Hotel Toronto	60 Yorkville Avenue	1827	8.7
11	dbar	60 Yorkville Ave	1855	8.3
9	The Hazelton Hotel	118 Yorkville Avenue	1913	8.2
5	The Spa at Four Seasons Hotel Toronto	60 Yorkville Ave	1828	8.1
12	ONE Restaurant/Lounge	116 Yorkville Ave	1908	7.5

Finally, based on the centre points, search for hotels with relatively higher ratings nearby these location points from Foursquare.com and generate the hotel list for the tourist.

Result

We successfully generated a list of hotels that may satisfy the specific tourist we imagine to visit Toronto based on his preferences. The hotels are located by the location data of the venues he may be interested in and calculated by Clustering algorithm to make sure that they were near those sights.

Discussion

1. We only use the tourist's food preference to search for specific venues. Yet in the real world, many aspects will change one's decision on where he live. For example, if someone wants to visit some cultural sights, he/she may would like to live near a museum or a famous church etc. The public transportation situation of the city also influences the decision of tourists.

2. We sort the list of the hotels by their ratings, which in some cases may not be a good way because there are other things affecting the tourists' decisions. Price is of great importance and people may have their own favorite hotel brand.

Conclusion

There are many other things that we can do based on the location data. This report shows a method to make recommendations for tourists on their trip plan using data analysis techniques and machine learning algorithms.

The result can be a great reference to a tourist new to Toronto if he don't know where to live and confused on what hotel is better for him. But the methodology we deploy here is quite simple then that in the real world, and it still need to be improved. Anyway, we show the reader briefly on how to collect data and use in a specified scenario to do meaningful work.

