Choosing the best location to open a bar in Toronto

Crapha

November, 2019

1. Introduction

I have always dreamed of running a bar in Toronto. But I am not sure in which neighborhood I should open my bar.

I would therefore like to determine where the most appreciated bars are located in Toronto (more precisely, in which neighborhoods), and then group those neighborhoods using some clustering algorithm to see if this can help me decide where to open my bar.

2. Data

The following Wikipedia page is used to obtain the list of neighborhoods in Toronto :

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

The location data provided in the previous labs (longitudes and latitudes) are also used in order to retrieve the longitude and latitude values of each neighborhood.

Finally, we use the Foursquare database in order to retrieve information about the various venues that can be found in the different neighborhoods of Toronto, with a focus on bars.

3. Methodology

We first construct a dataframe listing all neighborhoods in Toronto, along with their latitude and longitude values.

We then use the Foursquare explore function to retrieve the most liked bars in each neighborhood.

Initially, we wanted to then run some clustering algorithm on the data available from Foursquare to get some insight on the best neighborhoods to open a bar. However the standard « explore » request returns only very few useful information about bars (no rating, no price range, etc.). It seems that the only way to obtain this kind of information would be to use « Get Details of a venue » kind of requests. However, these are premium requests, which

are limited to 500 daily. We were therefore not able to use them / obtain the information we needed.

As a way around, we used the « section » parameter in our « explore » requests, where we chose « topPicks ». This is a cheap way of keeping only bars that are qualified as « topPicks » (instead of having a proper rating and other useful information).

4. Results

Scraping the Wikipedia website for neighborhoods allowed us to construct a data frame with the following structure :

	Postcode	Borough	Neighbourhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Using the latitudes and longitudes by postal codes provided in a CSV file (see previous labs), we managed to add longitude and latitude values to each neighborhood in our data frame, using the postal code as a common key :

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

We then made some requests to Foursquare in order to locate the most popular bars around each of these neighborhoods. The map below was created using the Folium Python library. Red circles show bars, whereas blue circles show neighborhoods.
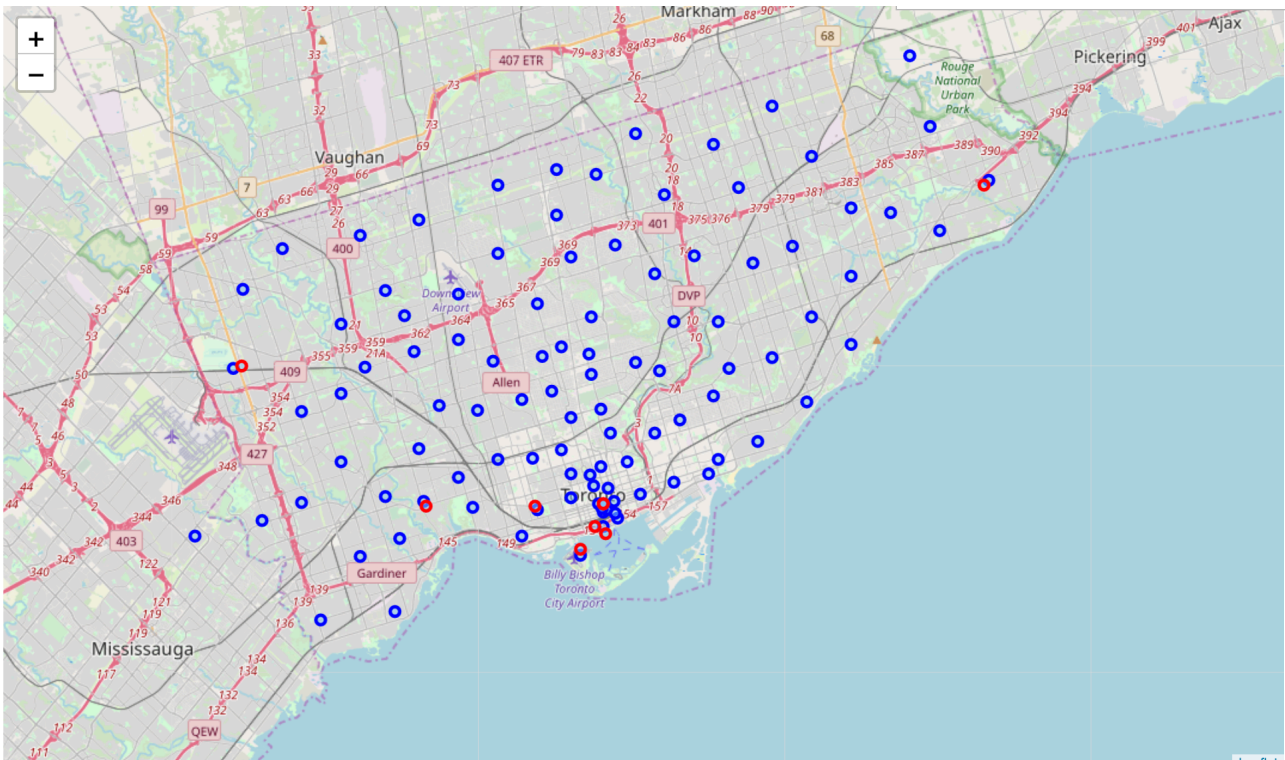
We can see that most popular bars are grouped within a few neighborhoods in the city center, except for a few outliers situated almost at the borders of Toronto. This suggests that the best place to open a bar would be the central neighborhoods :

- Harbourfront
- Cityplace
- Financial District
- Entertainment District
- Fashon District

5. Discussion

Determining the location of popular bars is only one useful information among others when deciding where to open a new bar. Unfortunately, we did not have other information at hand to perform deeper analysis (clustering according to various other factors, as planned initially).

We saw that the most popular bars are located mostly in the city center. This makes sense, as it is the densest part of the city, where people gather - not only to work, but also to have drinks afterwards…

Opening a bar in the city center would be a good option since it is « the place to be », however it is also where there is the most competition and

probably where the rents are the most expensive. Also, opening a bar in this region means following the trend.

This statistical analysis should therefore be balanced by some human exploration of Toronto in person, in order to possibly identify other neighborhoods where a bar would be welcome, even if no such bar exists at the moment yet.

6. Conclusion

This final project was interesting as it was a good opportunity to apply data analysis techniques to a real world business case. We were not able to go as deep as we wanted in the analysis due to lacking data, but are sure interesting insights could have been obtained using some unsupervised algorithm to sort the Toronto bars out.