Moving into a Toronto neighborhood

Guilherme Gonçalves Pessoa

I. Introduction

When looking for a new house to live in, choosing a new neighborhood to move into can be a quite difficult experience. There are many factors that should be considered: is it a quiet neighborhood? Are there nearby places to buy food, clothes, to go out? And those factors all differ from one person to another.

In my Capstone Project, I'm going to evaluate which neighborhoods are the best options for different people and families when searching for a new place to live.

II. Data

There were three steps to acquire the data necessary for the project:

1) Scraping a Wikipedia page [1] with a list of postal codes of Toronto with their respective Boroughs and Neighborhoods. The postal codes with no borough or neighborhood assigned where removed from the data.
2) Merging the scraped data of the postal codes with the .csv file [2] that contained their geographical coordinates.
3) Using the Foursquare API to get the nearby venues in all the neighborhoods. The search was limited to 100 venues for each neighborhood, since there is a limited number of calls that can be made to the Foursquare API daily.

III. Methodology

In order to find the most suitable neighborhood, I came up with four different cases:

1) **The family** A family consisting of four people: the mother, the father and their two children, one of them a baby.
2) **The business person** A travelling business person staying for some time in Toronto.
3) **The young woman** An art college student looking for a new place to live.
4) **The young man** A recently graduated man also looking for a new neighborhood.

Five venues were defined as the priorities for each case: the family is looking for a place which has one or more: baby store, pet store, shopping mall, convenience store and supermarket; for the business person: hotel, airport, café, pub and coworking space; for the young woman: college arts building, art gallery, arts and crafts store, yoga studio and asian restaurant; and finally, for the young man: college gym, bar, video game store, park and concert hall. Since those twenty venues were the only necessary ones, the rest of the venues were dropped for the data, in order to make it easier the analysis. A column with the sum of the frequencies of all the venues per neighborhood was created to be used as the primary source for the ordering the neighborhood, since a greater sum frequency means there are more venues in that neighborhood.

IV. Results

For each case, it will be presented a dataframe with the five most suitable neighborhoods based on the total frequency. Since for none of the cases there was a neighborhood that featured all the venues desired, some results will be presented based on the frequency of one of the venues that could be a priority for the case.

For the family, the five best neighborhoods would be as follows:

Neighborhood	Baby Store	Pet Store	Shopping Mall	Convenience Store	Supermarket	Frequency
Weston	0.0	0.000	0.0	0.500000	0.0	0.500000
York Mills West	0.0	0.000	0.0	0.333333	0.0	0.333333
Runnymede, The Junction North	0.0	0.000	0.0	0.250000	0.0	0.250000
Eringate, Bloordale Gardens, Old Burnhamthorpe...	0.0	0.125	0.0	0.125000	0.0	0.250000
East Toronto, Broadview North (Old East York)	0.0	0.000	0.0	0.250000	0.0	0.250000

Fig. 1. Top 5 neighborhoods for the family

As we can see, if we consider just the total frequency, the Weston neighborhood would be the best option, but it only has the convenience store. If the the family has a priority among the five venues, for example the baby store, we could reorder the neighborhoods as such:

Neighborhood	Baby Store	Pet Store	Shopping Mall	Convenience Store	Supermarket	Frequency
Christie	0.0625	0.000000	0.000000	0.0	0.000000	0.062500
Bathurst Manor, Wilson Heights, Downsview North	0.0000	0.000000	0.047619	0.0	0.047619	0.095238
India Bazaar, The Beaches West	0.0000	0.050000	0.000000	0.0	0.000000	0.050000
Willowdale, Willowdale East	0.0000	0.029412	0.029412	0.0	0.000000	0.058824
Weston	0.0000	0.000000	0.000000	0.5	0.000000	0.500000

Fig. 2. Top 5 neighborhoods for the family based on baby stores

Now, if the baby store is something the family can't go without, the Christie neighborhood would be their only option.

For the business person staying in Toronto for a while:

Neighborhood	Hotel	Airport	Café	Pub	Coworking Space	Frequency
Bayview Village	0.0	0.0	0.25	0.00	0.0	0.25
Glencairn	0.0	0.0	0.00	0.25	0.0	0.25
Birch Cliff, Cliffside West	0.0	0.0	0.25	0.00	0.0	0.25
The Beaches	0.0	0.0	0.00	0.25	0.0	0.25
The Annex, North Midtown, Yorkville	0.0	0.0	0.15	0.05	0.0	0.20

Fig. 3. Top 5 neighborhoods for the business person

The first four neighborhoods each have only one of the venues wanted, while the fifth one has two, both cafés and pubs. Though both of them have a lower frequency, The Annex could be a good choice.

Neighborhood	College Gym	Bar	Video Game Store	Park	Concert Hall	Frequency
York Mills West	0.0	0.333333	0.0	0.333333	0.0	0.666667
Caledonia-Fairbanks	0.0	0.000000	0.0	0.500000	0.0	0.500000
East Toronto, Broadview North (Old East York)	0.0	0.000000	0.0	0.500000	0.0	0.500000
Parkwoods	0.0	0.000000	0.0	0.500000	0.0	0.500000
Rouge Hill, Port Union, Highland Creek	0.0	0.500000	0.0	0.000000	0.0	0.500000

Fig. 4. Top 5 neighborhoods for recently graduated man

For the recently graduated young man we have:

York Mills West, the first neighborhood, seems like the best option since it has both bars and parks, while the other neighborhoods only have one of each venue.

As for the college woman:

Neighborhood	College Arts Building	Art Gallery	Arts & Crafts Store	Yoga Studio	Asian Restaurant	Frequency
Little Portugal, Trinity	0.000000	0.021739	0.000000	0.021739	0.065217	0.217391
Don Mills	0.000000	0.041667	0.000000	0.000000	0.041667	0.166667
Queen's Park, Ontario Provincial Government	0.000000	0.000000	0.030303	0.030303	0.000000	0.121212
Business reply mail Processing Centre, South C...	0.000000	0.000000	0.000000	0.058824	0.000000	0.117647
University of Toronto, Harbord	0.028571	0.000000	0.000000	0.028571	0.000000	0.114286

Fig. 5. Top 5 neighborhoods for the college woman

In her case, the Little Portugal/Trinity neighborhood is a very good option, since she can find art galleries, yoga studios and Asian restaurants. Though if she would like to live closest to the college, the Harbord neighborhood would be a better choice.

V. Discussion

Based on the results of the previous section, there are a few observations that can be noted:

- For the family, if we consider just the total frequency, the Weston neighborhood would be the best option, but it only has the convenience store. If the the family has a priority among the five venues, the results would be different.
- For the business person, also considering just the total frequency, the first four neighborhoods each have only one of the venues wanted, while the fifth one has two, both cafés and pubs. Though both of them have a lower frequency, The Annex could be a good choice.
- For the young man, York Mills West, the first neighborhood based on total frequency, seems like the best option since it has both bars and parks, while the other neighborhoods only have one of each venue.
- For the young woman, considering the total frequency, the Little Portugal/Trinity neighborhood is a very good option, since she can find art galleries, yoga studios and Asian restaurants. Though if she would like to live closest to the college, the Harbord neighborhood would be a better choice.
- As we can see, for each case there are many different ways the available data can be analysed in order to find the most suitable neighborhood. There's no right or wrong answer, it's all based on preferences, and the analysis can be adapted to work with all different cases.

VI. Conclusion

Overall, the project worked as intended. It was possible to use many of the content learnt throughout the data science course. Working with the Toronto neighborhood data and the Foursquare API made it possible to get all the information needed and analyse it. There are still a few points that could be improved, such as raising the limit of venues for each neighborhood, as it would allow an even better analysis. Unfortunately the daily calls limitation to the Foursquare API make it a little harder to do so. Still, with the 100 venues limit used, it was already possible to get some good insight into the neighborhoods and chose the best options for each case, whilst also being possible to expand the analysis for more cases.

References

[1] https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M acessed in August, 10th 2020.
[2] http://cocl.us/Geospatial_data acessed in August, 10th 2020.