Clustering Neighbourhoods in Toronto

IBM Applied Data Science Capstone Week 5

26 December 2019

Business requirement

There are more than 1,500 coffee shops in Toronto. The business problem is to identify which neighbourhoods in Toronto have a high number of parks and green spaces, but low number of coffee shops – leaving a potential gap in the market.

The output would benefit coffee chains that are all about the location – those renounced for being situated in beautiful places, such as Benugo's in the UK. The research would drive value by promoting the use of data-led decision making – to identify green spaces without an abundance of competitors (i.e. other coffee shops). Without data science, it would be difficult and time consuming for an individual to gather and analyse various disparate sources of information.

Data

The project will be performed using the following datasets:

(a) Open data from Wikipedia on Toronto neighbourhoods; this includes a list of each postcode, and their corresponding borough and neighbourhood.
(b) Open data containing the geo-spatial data (longitude and latitude) of each postcode area.
(c) Foursquare API data, containing data on venues and their geo-spatial location.

Methodology

We performed the following steps:

- Use BeautifulSoup (python library) to web-scrape the Toronto neighbourhoods data from the following Wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Perform data pre-processing, by formatting the data to support our analysis. For cells where the "neighbourhood" is blank, we replaced these with the name of the borough.
- After cleansing the data, we merged the Wikipedia neighbourhoods data with geo-spatial data, to identify the longitude and latitude of each neighbourhood. These were added as extra columns to our dataframe.
- We then retrieved the corresponding venues in each neighbourhood using FourSquare API, based on the geo-spatial data. We applied professional judgement and applied limits of 500 venues, and within 500 metre radius of the neighbourhood.
- We then grouped the results by neighbourhood and venue type. The dataframe is 101 rows (including header) by 280 columns. An illustrative row is shown below:

```
In [59]: toronto_grouped = toronto_onehot.groupby('Neighborhood').mean().reset_index()
         toronto_grouped
```

Out[59]:

	Neighborhood	Yoga Studio	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	...	Tra
0	Agincourt	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.0

- Next, we summarised the output by identifying the top 5 venues in each neighbourhood and their prevalance. Collectively, this information would enable us to analyse the results to identify neighbourhoods which may have potential for opening a new coffee shop.

Results

The results enabled us to describe quantitatively:

- Neighbourhoods in Toronto
- Venue types by neighbourhood
- Venue frequency in each neighbourhood

Based on this, we were able to identify neighbourhoods that have a park, and their composition. The results are highlighted below.

----Agincourt North, L'Amoreaux East, Milliken, Steeles East----

	venue	freq
0	Park	0.5
1	Playground	0.5

----CFB Toronto, Downsview East----

	venue	freq
0	Airport	0.5
1	Park	0.5

----Rosedale----

	venue	freq
0	Park	0.4
1	Playground	0.2
2	Building	0.2
3	Trail	0.2

----Davisville North----

	venue	freq
0	Dance Studio	0.11
1	Park	0.11
2	Dog Run	0.11
3	Breakfast Spot	0.11
4	Gym	0.11

----East Toronto----

	venue	freq
0	Park	0.4
1	Coffee Shop	0.2
2	Pizza Place	0.2
3	Convenience Store	0.2

----York Mills West----

	venue	freq
0	Convenience Store	0.33
1	Park	0.33
2	Bank	0.33

----Weston----

	venue	freq
0	Park	0.5
1	Convenience Store	0.5

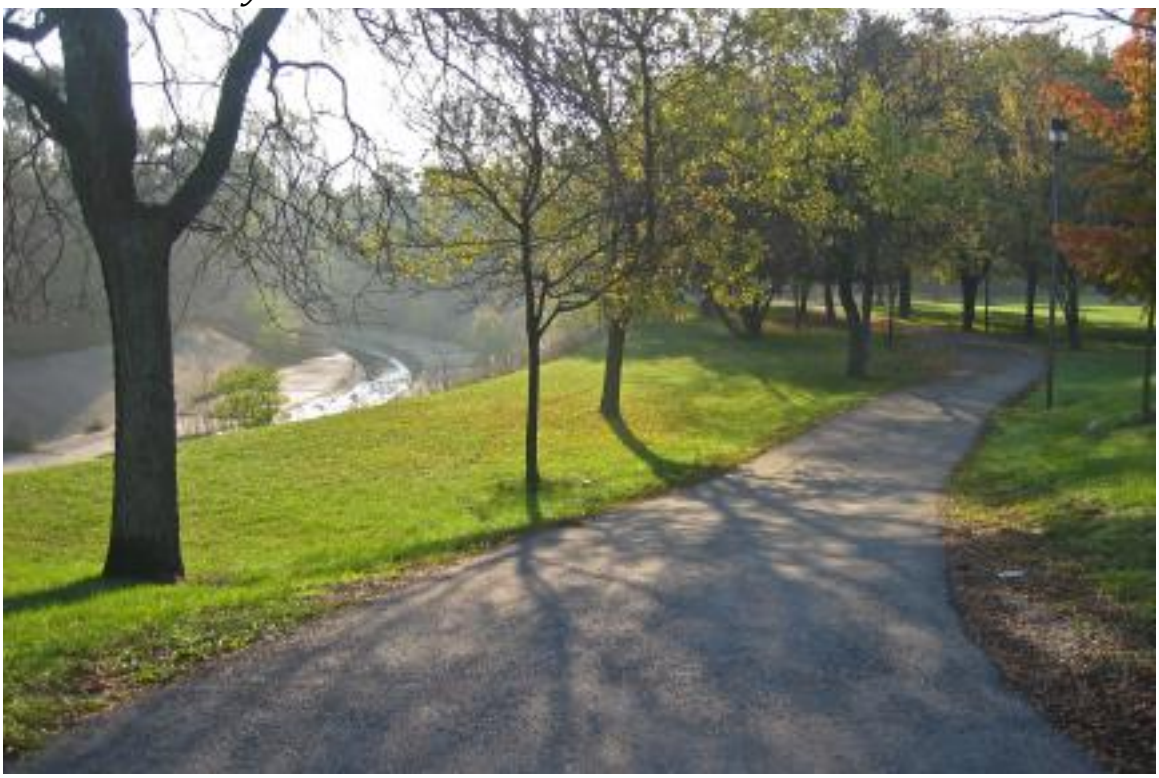
----The Kingsway, Old Mill North, Montgomery Road----

	venue	freq
0	River	0.5
1	Park	0.5

Discussion

Weston and York Mills West – These neighbourhoods comprise both parks and convenience stores, which indicates that there is sufficient footfall to warrant such facilities. Upmarket coffee shops provide a different service offering that may compliment, rather than directly compete, with convenience stores. As such, there may be value in opening coffee shops in these neighbourhoods.

York Mills Valley Park



We used Google Maps and Google Images to inspect the parks located in these neighbourhoods, to check if the type of park would attract upmarket coffee shop go-ers – for example, the prescence of trees, benches or water features.

Cruickshank Park in Weston

Kingsway – This neighbourhood comprise both parks and rivers, which is a suitable location for upmarket coffee shops such as Benugos in the UK, which has the strapline "Coffee and food in the most beautiful places". As such, Kingsway may be an appropriate location for such upmarket coffee shops.

We used Google Maps and Google Images to inspect the parks located in this neighbourhoods, and note that it is beautifully presented. See overleaf for a photo of Magwood Park in Kingsway.

Others – Weston, York Mills West and Kingsway therefore appear to be the prime locations, and present the most attractive options. We had considered the other neighbourhoods but noted the following drawbacks:

- Downsview East – located with airport which impacts ambience and envrionment
- Agincourt North and Rosedale – located with playground, so likely to attract families and a different clientelle

- Davisville North – located with breakfast spot, dance studio and gym. This does does not indicate a clear character or type of clientelle for the neighbourhood
- East Toronto – located with coffee shop, resulting in direct competition

Magwood Park in Kingsway

Recommendation

It is recommended that further analysis is performed on Weston, and Kingsway York Mills West, to understand the following:

- Footfall
- Retail space cost per sqm
- Retail space availability

Further, it is recommended that focus groups are held to understand (a) the demand for a coffee shop, and (b) preferences for the coffee shop in that neighbourhood.

Analysis performed

https://github.com/jc125/sugoi/blob/master/Capstone%20Week%203.ipynb