2020

To relocate or not?

Coursera Capstone | Places in Toronto

INDEX

Introduction

Business problem and intended audience

Data

Data sources, nature and profile

Methodology

Data analysis methodology, visualization and machine learning methods

Results

Discuss the result

Discussion

Observations and recommendations

Conclusion

Concluding remarks

1. Introduction

Should you relocate from west to east of Toronto?

Suppose you live in west side of Toronto, Canada. Recently, you have got a better job offer on the other side of the city, that is, the east. You have a good life in the west within a lively community and all the amenities a metro city can offer.

Before you are ready to accept the job offer, you want to evaluate if you will have same level of quality of life in the east. If not, does it worth compromise your quality of life for the increased paycheck?

Things to consider:

- Identify a few neighbourhoods which may be as good as the present location, that is, west of Toronto
- Check the venues around those neighbourhoods and decide if those are as good as the west

2. Data

Data overview

- Neighbourhoods of Toronto
  Sourced from Wikipedia
- Geocoded data
  Sourced from Google Maps API
- Venues around neighbourhoods
  Sourced from Foursquare API

```
import pandas as pd
raw_data = pd.read_html("https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M")[0]
raw_data.head()
```

	Postal Code	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

```
raw_data.describe(include = 'all')
```

	Postal Code	Borough	Neighbourhood
count	180	180	180
unique	180	11	100
top	M7Z	Not assigned	Not assigned
freq	1	77	77

It can further be noted that there are 180 postal codes, 10 boroughs and 100 neighbourhoods in Toronto.

3. Methodology

1. Data Cleansing

Removed records without assigned Neighbourhoods

2. Identify East

Visualise the boroughs on map and shortlist (G Maps | Folium)

3. East Neighbourhoods

Visualise neighbourhoods within those boroughs (G Maps | Folium)

4. Venues in the East

Venues, their types and characteristics (G Maps | Folium)

5. Clustering Venues

Clustering the venues on their common themes (KMeans cluster analysis)

6. Decisioning

3.1. Data Cleansing

Removed the records without any assigned borough

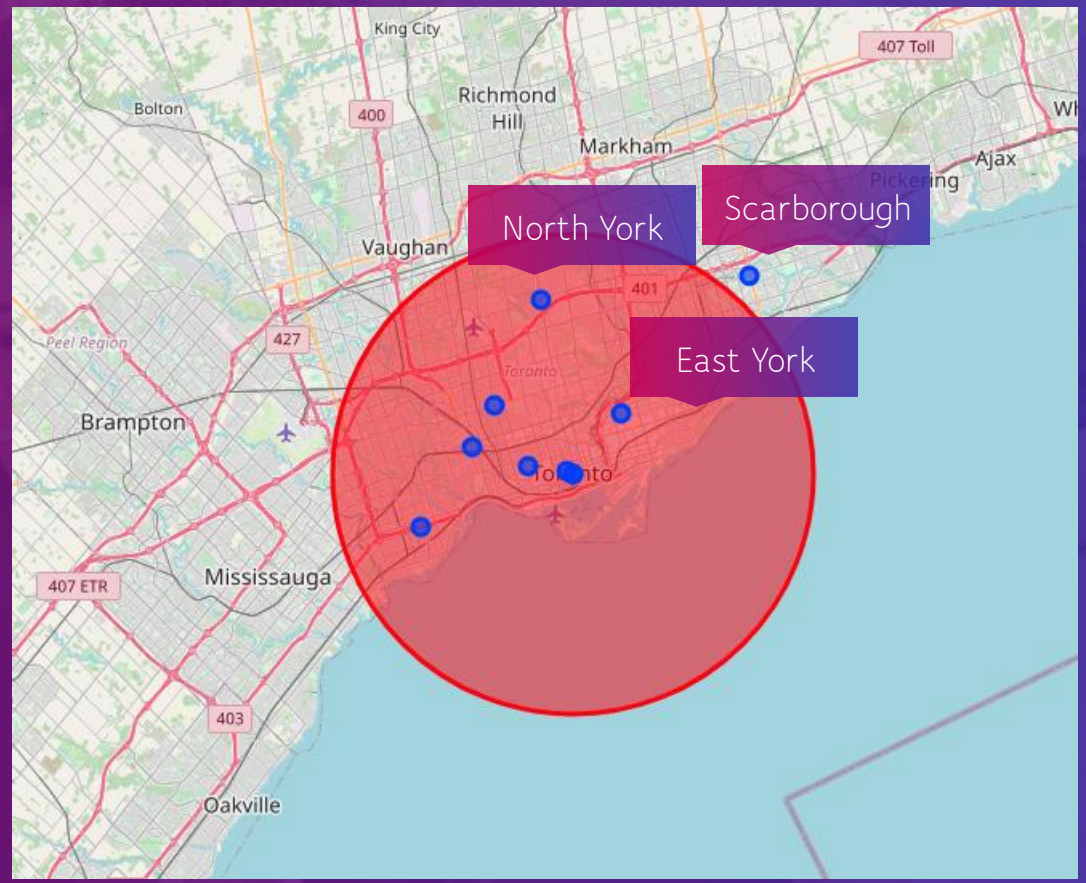
	Postal Code	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

```
df = raw_data.loc[raw_data['Borough'] != "Not assigned"].reset_index(drop=True)
df.head()
```

	Postal Code	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

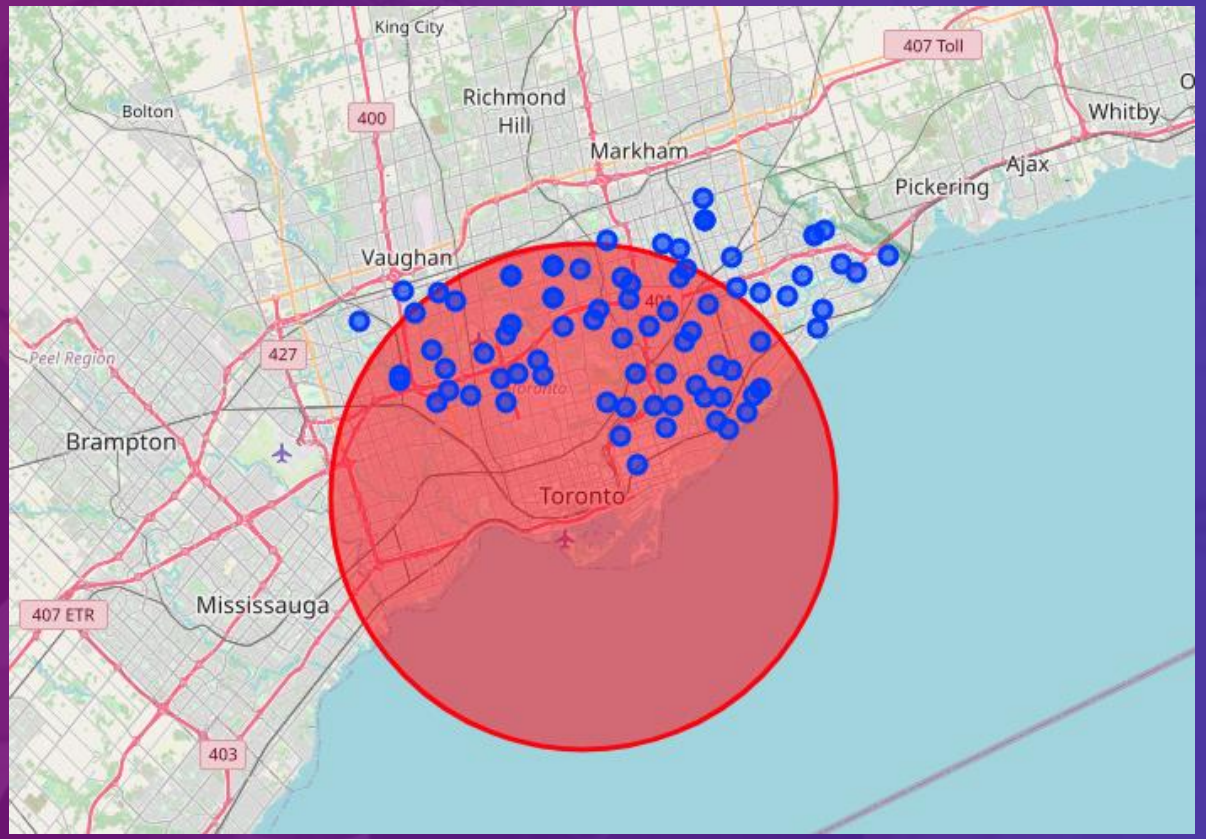
3.2. Identify East

- Visualised the boroughs on map and shortlisted the target boroughs:
  - North York
  - East York
  - Scarborough
- G Maps API was used for geocoding
- Folium was used for map



3.3. East Neighbourhoods

- Visualised neighbourhoods within those boroughs
- 83 neighbourhoods are present
- G Maps API was used for geocoding
- Folium was used for map



3.4. Venues in the East

- 786 attractive venues were accessible

```
print(toronto_venues.shape)
toronto_venues.head()
```

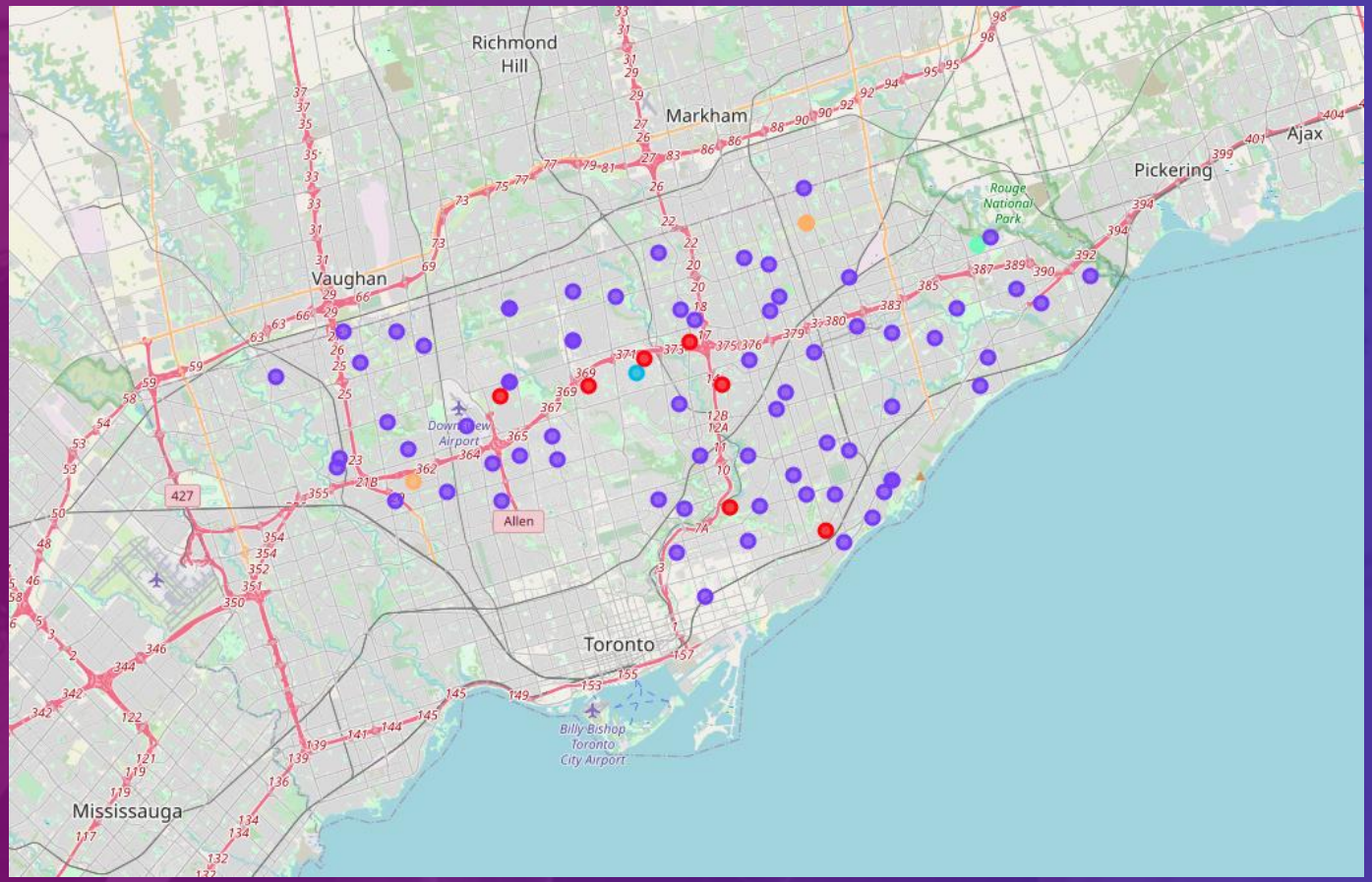
```
(786, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
2	Rouge Hill	43.794719	-79.134478	Rouge Bowl	43.797993	-79.137013	Bowling Alley
3	Rouge Hill	43.794719	-79.134478	Bikram Yoga East	43.798124	-79.137291	Yoga Studio
4	Port Union	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar

- Foursquare API was used for venues

3.5. Cluster Analysis

- The venues were clustered in 5 clusters, of which, 3 were big
- Of these 3, one cluster was concentrated in **Scarbourough**, one was in **North York** and one was divided in these two
- KMeans was used



3.6. Decisioning

- With 500+ attractive venues, East is a good place to live
- **Scarborough** and **North York** are most attractive boroughs

4. Results

You should relocate to Scarborough or North York in the East

Basis the 6 broad steps of the methodology, the results can be summarised as:

- North York, East York and Scarborough are the boroughs in the East Toronto
- There are 83 neighbourhoods within these boroughs
- There are 500+ attractive venues
- North York or Scarborough is recommended to live given the nearest venues and their themes

Discussion and Conclusion

5. Discussion

- The map visualisation was an effective method to identify the target locations and their attributes
- Various map APIs, like, Google and Foursquare, are powerful tools to obtain important information about those locations
- Datascience methods, like, cluster analysis, was effective to show the common themes of various venues

6. Conclusion

- Datascience analysis using Python, Folium maps, Google maps API, Foursquare API and Scikit Learn was an effective proccess to solve a real-life dilemma of whether to relocate to a specific part of a city
- This method can be used in a similar situation in other cities of the world
- Thank you Coursera, IBM and instructors from IBM for creating this exercise!

2020