The Battle of Neighborhood

Data Science Capstone Project

By: Aditya Prakash

Date: 18th Feb 2021

CONTEXT:

- Introduction
- Data
- Methodology
- Results and discussion
- Conclusion

INTRODUCTION

Would you like to find a perfect location for business living or traveling, using just a few mouse clicks?

MAKIBO LTD CAN HELP YOU!

We have a special application, which will help you to find a perfect area for your goals due some moments, even accelerator needs more time than you with us!

DATA

- Wikipedia helps to get Postal Code, Borough and Neighborhood in Toronto
- Geospatial data for Toronto contains the geographical coordinates of each postal code
- Foursquare API allows to obtain the data on what venues are located at each neighborhood
- Random user data, with a random number (from 1 to 10) of preferences to check, how our system works

METHODOLOGY:

- Pandas library to scrap the table from HTML page

	Postcode	Borough	Neighbourhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M5A	Downtown Toronto	Regent Park
6	M6A	North York	Lawrence Heights

METHODOLOGY:

- Geospatial data for Toronto contains the geographical coordinates of each postal code

	PostalCode	Borough	Neighborhood	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
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029

METHODOLOGY:

Foursquare API allows to obtain the data on what venues are located at each neighborhood

- Apply OneHot encode to get a new table with the neighborhood as the index and percentage of each category available in that neighborhood

METHODOLOGY:

Random User – Generate a random User to use for a test of the system

select a random number from 1 to 10 to represent the amount of categories selected by the user

create a table with the categories as the columns and one row, where the values are 1 if the user has that category in his list and 0 for vice versa

METHODOLOGY:

Recommendation system

Multiply the user profile with the table that has the neighborhood and the weight of each category

Result is a matrix with the score of each neighborhood

The higher score, the better the neighborhood fit the user interest

METHODOLOGY:

Recommendation system

Multiply the user profile with the table that has the neighborhood and the weight of each category

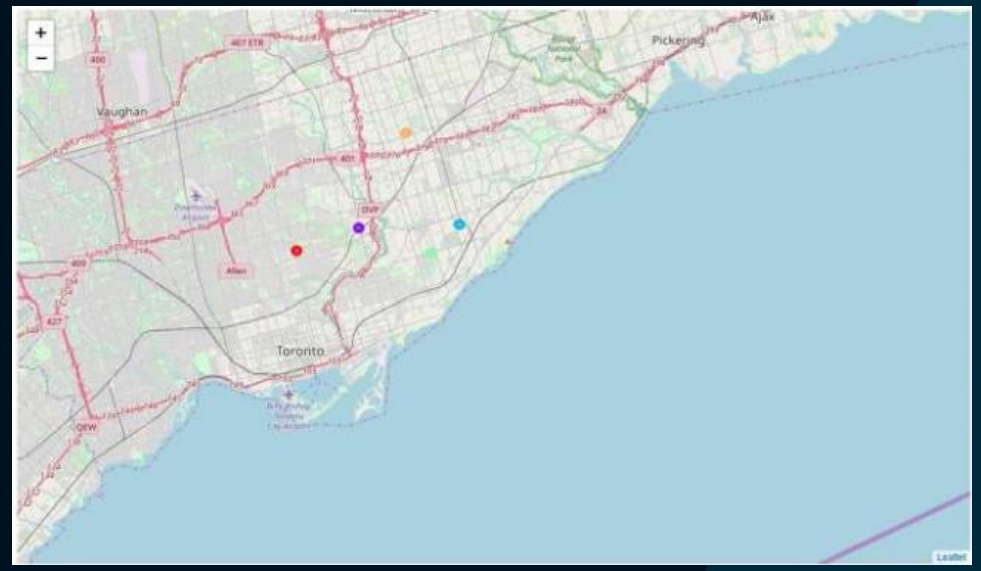
Result is a matrix with the score of each neighborhood

The higher score, the better the neighborhood fit the user interest

RESULTS AND DISCUSSION

- In this prototype we generated some random categories for our user

```
['Italian Restaurant',
 'Asian Restaurant',
 'Bus Station',
 'Tanning Salon',
 'Sake Bar']
```



	PostalCode	Borough	Neighborhood	Latitude	Longitude	Score
0	M3C	North York	Flemingdon Park, Don Mills South	43.725900	-79.340923	0.142857
1	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029	0.142857
2	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577	0.111111
3	M1T	Scarborough	Clarks Corners, Sullivan, Tam O'Shanter	43.781638	-79.304302	0.100000
4	M4P	Central Toronto	Davisville North	43.712751	-79.390197	0.090909

RESULTS AND DISCUSSION

- 2 best neighborhoods for our user are “North York Flemingdon Park, Don Mills South” and “East Birchmount Park, Ionview, Kennedy Park”
- Difference of the score amount the 5 neighborhoods is not big. A probable reason is that categories, which our user chose are more or less common, they don’t include anything extraordinary as “Airport Food Court”.

CONCLUSION:

- This is a sample content-based recommendation system that still need to be improved.
- The data and algorithm need more date and accuracy, especially for some small towns with a few venues.
- There are more parameters, which could be use for the search, as distance to the work place, bus station and etc.

THANK YOU!

Aditya Prakash

Phone
9284703012

Email
adityaprakash.2009@gmail.com