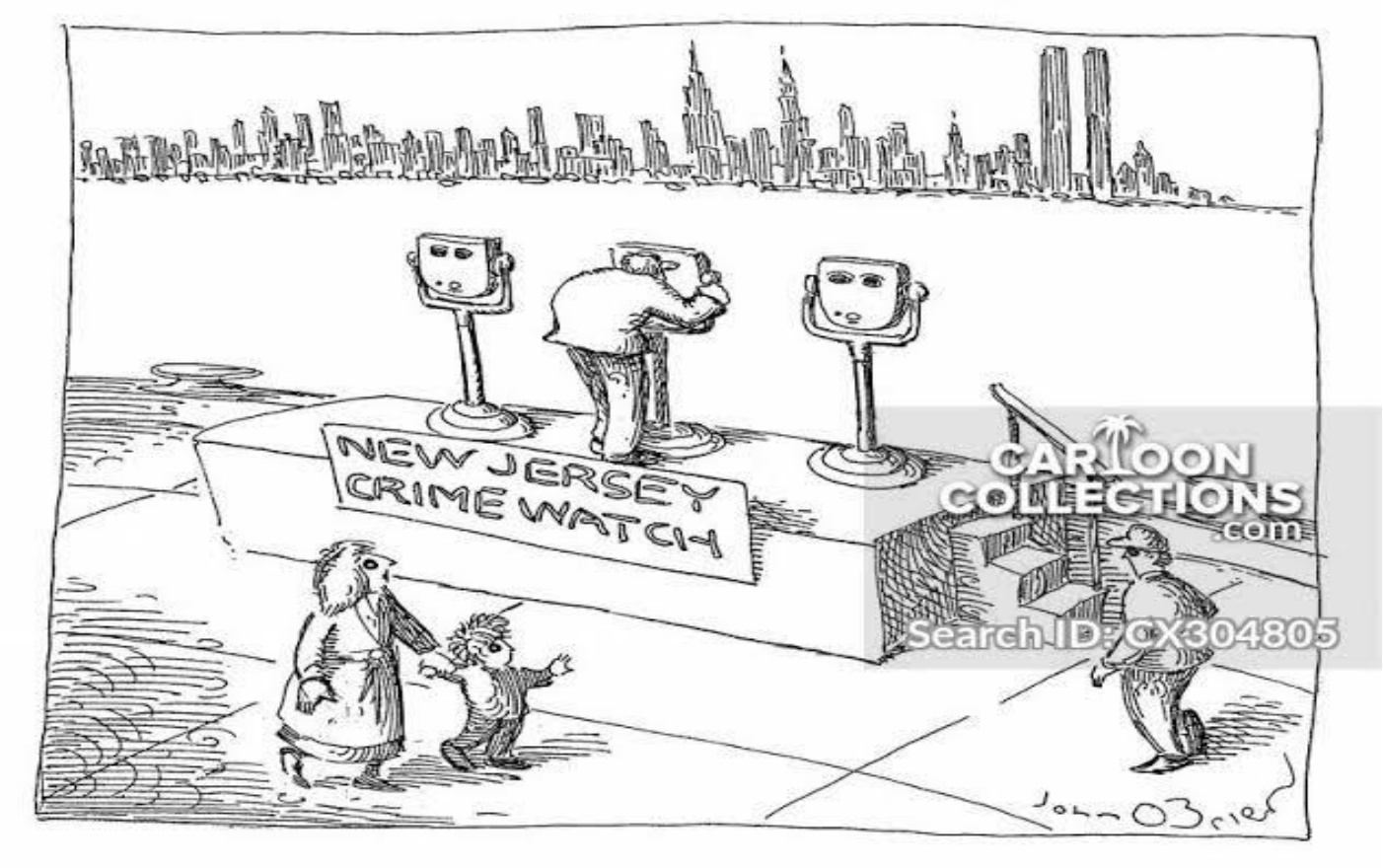


Capstone Project: The Battle of Neighborhoods (Week2)

A. **Introduction to the business problem**

b. **Data used to solve problem**

c. **Methodology section**

d. **Result Section**

e. **Discussion Section**

e. **Conclusion Section**

A. Introduction to the business problem

In this project, I chose to explore New York City and the city of Toronto and segmented and clustered their neighborhoods. Both cities are very diverse and are the financial capitals of their respective countries. One interesting idea would be to compare the neighborhoods of the two cities and determine how similar or dissimilar they are. New York City more like Toronto or Paris or some other multicultural city because The city is divided into 62 neighborhoods in total.

B. Introduction to the business problem

B.1 Analyzing Crime Data

Therefore, stakeholders require some advice on where to settle and launch their new businesses. Experience from local residents is always useful, but it is interesting to also rely on data to take a final decision.

This project will be targeted to new investors trying to find an optimal location for a new Pub

B.2 Identifying neighborhood

Harbourfront Queen's Park Ryerson, Garden District St. James Town The Beaches Berczy Park Central Bay Street Christie Adelaide, King, Richmond Dovercourt Village, Dufferin Harbourfront East, Toronto Islands, Union Station Little Portugal, Trinity The Danforth West, Riverdale Design Exchange, Toronto Dominion Centre Brockton, Exhibition Place, Parkdale Village The Beaches West, India Bazaar Commerce Court, Victoria Hotel Studio District Lawrence Park Roselawn Davisville North Forest Hill North, Forest Hill West High Park, The Junction South North Toronto West The Annex, North Midtown, Yorkville Parkdale, Roncesvalles Davisville Harbord, University of Toronto Runnymede, Swansea Moore Park, Summerhill East Chinatown, Grange Park, Kensington Market Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara Rosedale Stn A PO Boxes 25 The Esplanade Cabbagetown, St. James Town First Canadian Place, Underground city Church and Wellesley Business Reply Mail Processing Centre 969 Eastern

B.3 Common Venues

For Example 1st Most Common VenueCoffee Shop Café 2nd Most Common Venue Steakhouse 3rd Most Common Venue Gastropub 4th Most Common Venue Hotel 5th Most Common Venue Restaurant and Theater , Cosmetics Shop or Gym

C. Methodology section

Website scraping with BeautifulSoup

Create the DataFrame

The dataframe will consist of three columns: PostalCode, Borough, and Neighborhood Only process the cells that have an assigned borough. Ignore cells with a borough that is Not assigned. More than one neighborhood can exist in one postal code area. For example, in the table on the Wikipedia page, you will notice that M5A is listed twice and has two neighborhoods: Harbourfront and Regent Park. These two rows will be combined into one row with the neighborhoods separated with a comma as shown in row 11 in the above table.

Clean the DataFrame and show the first 5 rows of the resulting DataFrame

If a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough. So for the 9th cell in the table on the Wikipedia page, the value of the Borough and the Neighborhood columns will be Queen's Park. Clean your Notebook and add Markdown cells to explain your work and any assumptions you are making. In the last cell of your notebook, use the .shape method to print the number of rows of your dataframe.

Use the Geocoder package or the csv file to create the following dataframe:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M5A	Downtown Toronto	Harbourfront	43.654260	-79.360636
1	M7A	Downtown Toronto	Queen's Park	43.662301	-79.389494
2	M5B	Downtown Toronto	Ryerson, Garden District	43.657162	-79.378937
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
4	M4E	East Toronto	The Beaches	43.676357	-79.293031

Explore and cluster the neighborhoods in Toronto.

We can decide to work with only boroughs that contain the word Toronto and then replicate the same analysis we did to the New York City data. It is up to you.

to add enough Markdown cells to explain what you decided to do and to report any observations you make. to generate maps to visualize your neighborhoods and how they cluster together. Once you are happy with your analysis, submit a link to the new Notebook on your Github repository.

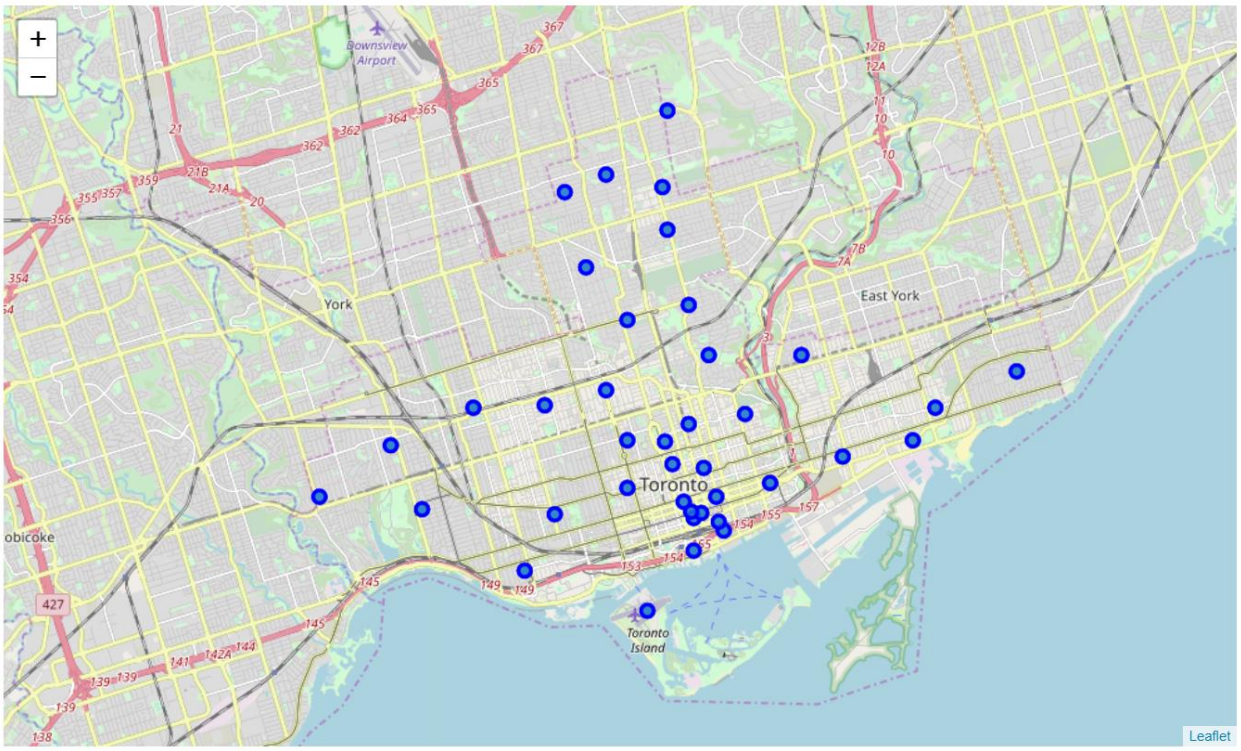
D. Results

Explore the Dataset

Replicate the same analysis we did to the New York City data

`The geograpical coordinate of Toronto are 43.653963, -79.387207.`

Create a map of Toronto with neighborhoods superimposed on top.



Define Foursquare Credentials and Version

Let's explore the first Neighborhood in our dataframe

```
https://api.foursquare.com/v2/venues/explore?&client_id=xxxxZ34GZ1DOLZOCN10I52Y2331VNHXLKKNCCGIQBXWYxxxx&client_secret=xxxxJEML3LVGAE1UZZSYQAZEVTEKBPBGR2YD5Q1PSTRIMPxxxx=20180605&ll=43.6542599,-79.3606359&radius=800&limit=100'
```

From the Foursquare lab we know that all the information is in the items key.

Before we proceed, let's borrow the get_category_type function from the Foursquare lab.

	name	categories	lat	lng
0	Roselle Desserts	Bakery	43.653447	-79.362017
1	Tandem Coffee	Coffee Shop	43.653559	-79.361809
2	Cooper Koo Family YMCA	Gym / Fitness Center	43.653191	-79.357947
3	Impact Kitchen	Restaurant	43.656369	-79.356980
4	Body Blitz Spa East	Spa	43.654735	-79.359874

```
82 venues were returned by Foursquare.
```

Cluster Neighbourhood and Examine Clusters

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	**Adelaide, King, Richmond**	**Coffee Shop**	**Café**	**Steakhouse**	**Gastropub**	**Hotel**	**Restaurant**	**Sushi Restaurant**	**Theater**	**Cosmetics Shop**	**Gym**
1	**Berczy Park**	**Coffee Shop**	**Café**	**Beer Bar**	**Hotel**	**Japanese Restaurant**	**Restaurant**	**Italian Restaurant**	**Cocktail Bar**	**Park**	**Cheese Shop**
2	**Brockton, Exhibition Place, Parkdale Village**	**Coffee Shop**	**Café**	**Furniture / Home Store**	**Arts & Crafts Store**	**Bakery**	**Bar**	**Restaurant**	**Gift Shop**	**Italian Restaurant**	**Park**
3	**Business Reply Mail Processing Centre 969 Eastern**	**Fast Food Restaurant**	**Light Rail Station**	**Grocery Store**	**Brewery**	**Burrito Place**	**Bar**	**Bakery**	**Pub**	**Pizza Place**	**Park**
4	**CN Tower, Bathurst Quay, Island airport, Harbo...**	**Harbor / Marina**	**Boat or Ferry**	**Airport Service**	**Airport Terminal**	**Sculpture Garden**	**Airport Lounge**	**Rental Car Location**	**Tunnel**	**Coffee Shop**	**Bar**

E. Discussion

The aim of this project is to help people who want to relocate to the safest borough in London, expats can chose the neighborhoods to which they want to relocate based on the most common venues in it. For example if a person is looking for a neighborhood with good connectivity and public transportation we can see that Clusters 3 and 4 have Train stations and Bus stops as the most common venues. If a person is looking for a neighborhood with stores and restaurants in a close proximity then the neighborhoods in the first cluster is suitable. For a family I feel that the neighborhoods in Cluster 4 are more suitable dues to the common venues in that cluster, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Bus Stops, Restaurants, Electronics Stores and Soccer fields which is ideal for a family.

F. Conclusion

This project helps a person get a better understanding of the neighborhoods with respect to the most common venues in that neighborhood. It is always helpful to make use of technology to stay one step ahead i.e. finding out more about places before moving into a neighborhood. We have just taken safety as a primary concern to shortlist the borough of London. The future of this project includes taking other factors such as cost of living in the areas into consideration to shortlist the borough based on safety and a predefined budget.