



The Battle of Neighborhoods

ABSTRACT

An analysis of the city of Toronto, CA with a comparison to the city of Paris, FR.

Ljiljana Nikolajev

IBM Capstone Project

1 Introduction

1.1 Background

For this project I am a real estate agent recruited by a family of four, about to relocate from Paris, France to Toronto, Canada. The client is looking to relocate in a neighborhood that is more secure than their current city. Ideally, the client wishes to live in a family friendly average neighborhood, where schools (elementary and high schools) are at walking distance from their homes. As they are big foodies and love discovering new types of food, they would like to have great diversity in restaurant options close to their home. Both parents will be working at the Michael Garron hospital and wished for their commute to be no longer than 20 minutes by car.

1.2 Problematic

Moving from one country to another is not an easy task, particularly when one does not have a point of reference. This project will help to determine the best borough in Toronto for this family to relocate and ensure that their transition into this new life is as smooth as possible

2. Data Acquisition and Cleansing

2.1 Data Source

- To assess the level of criminality in Toronto, we will scrap and analyze data from the [open data portal](#) from the city of Toronto.
- To compare these data, criminality in Paris was extracted from [this article](#), sourced from the Ministre de L'interieur
- To help define the boroughs and neighborhood of Toronto, the [open data portal](#) for the city of Toronto was accessed again.

2.2 Data Cleaning

Toronto crime dataset first had to be cleaned before it could be used. First, this data set contained all major crime incidents in the city of Toronto from 2014-2019. Since we are only interested in comparing the data from 2019, all previous years had to be dropped.

	X	Y	Index_	event_unique_id	occurrence	reported	premise	ucr_code	ucr_ext	offence	...	occurrence	occurrence	occurrence	MCI	Division	Hood_ID	Neighbourhood	Long	Lat	Objectid
0	-8.816401e+06	5.434587e+06	701	GO-20141756319	2014/03/24 00:00:00+00	2014/03/24 00:00:00+00	Commercial	1430	100	Assault	...	83.0	Monday	1	Assault	D42	132	Malvern (132)	-79.199081	43.800281	1
1	-8.837252e+06	5.413357e+06	901	GO-20143006885	2014/09/27 00:00:00+00	2014/09/29 00:00:00+00	Other	2120	200	B&E	...	270.0	Saturday	16	Break and Enter	D52	76	Bay Street Corridor (76)	-79.386383	43.662472	2
2	-8.862433e+06	5.422276e+06	702	GO-20141756802	2014/03/24 00:00:00+00	2014/03/24 00:00:00+00	Commercial	2120	200	B&E	...	83.0	Monday	6	Break and Enter	D23	1	West Humber-Clairville (1)	-79.612595	43.720406	3
3	-8.833104e+06	5.431887e+06	703	GO-20141760570	2014/03/24 00:00:00+00	2014/03/24 00:00:00+00	Apartment	2120	200	B&E	...	83.0	Monday	15	Break and Enter	D33	47	Don Valley Village (47)	-79.349121	43.782772	4
4	-8.845311e+06	5.413667e+06	902	GO-20142004859	2014/05/03 00:00:00+00	2014/05/03 00:00:00+00	Commercial	1610	210	Robbery - Business	...	123.0	Saturday	2	Robbery	D11	90	Junction Area (90)	-79.458778	43.664490	5

5 rows x 29 columns

First, all non-relevant columns were dropped. Second, because there were null values in the dataset, all row that contained a non-value for the “occurrenceyear” column were dropped. Once this was done, it was possible to isolate all incidents from year 2019.

	MCI	occurrenceyear	Lat	Long	Neighbourhood
0	Assault	2019.0	43.810932	-79.227135	Malvern (132)
1	Assault	2019.0	43.663906	-79.384155	Church-Yonge Corridor (75)
2	Assault	2019.0	43.655777	-79.380676	Church-Yonge Corridor (75)
3	Assault	2019.0	43.723015	-79.415932	Bedford Park-Nortown (39)
4	Break and Enter	2019.0	43.648773	-79.528748	Islington-City Centre West (14)

Unfortunately, there are no open source for crime data for the Paris metro area. Data were extracted from the aforementioned article, for which the data were provided by the French government. To standardize the categories, all assaults were grouped together. The same was for robberies and thefts. Data are from Jan - Sept inclusively. Since the data were in an image and not in a table, the values were added to a panda dictionnary that was further transformed in a dataframe.

2.3 Techniques

Techniques that will be used through this project are as such:

Data Analysis will help us narrow down the safest neighborhood

- we will create a new dataframe that will contain all crimes by types.
 - Pie charts will then be used to compare Paris and Toronto criminality (example shown)
- We will use the panda library again to isolate the 10 more dangerous and the 10 safest neighborhoods of Toronto. Graph bars will then be use to illustrate the difference.
- Using folium and geopy, all crimes will be transposed onto a map of Toronto to provide an additional visual.
- All neighborhoods within 10-15 km distance from the hospital will be selected for further analysis

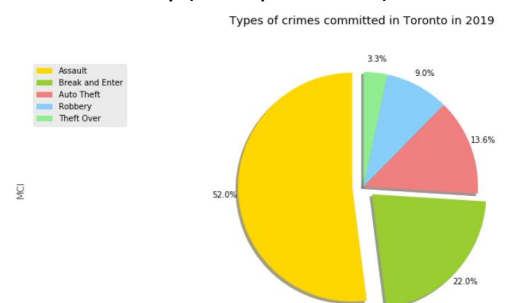
Foursquare API will help us find the venues in each of these neighborhoods and help us narrow down the list even more

- Foursquare API will be used to first, determine if each neighborhood contained both a elementary and high schools that are close to each other.
- Foursquare API will be used to retrieve all restaurants venues in the selected neighborhoods.

Evolution des crimes et délits à Paris Le Parisien
 Comparaison janvier à septembre 2018 et 2019

Catégories	Janv. à sept. 2019	Evolution 2018/2019
• Vols - dont cambriolages	144 641 13 743	+ 12,4 % + 7,9 %
• Délits économiques et financiers (escroqueries)	29 909	+ 4,6 %
• Violences	17 784	+ 7,1 %
• Destructions et dégradations	12 757	+ 15,4 %
• Stups	9 157	- 10,6 %
• Outrage et violence sur personnes dépositaires de l'autorité	4 307	+ 20,4 %
• Violences sexuelles	2 690	+ 4,9 %
• Port ou détention d'armes prohibées	2 662	+ 15 %
• Crimes et délits contre mineurs - dont viols	1 518 837	+ 7,4 % + 11,5 %
• Autres	8 786	+ 11,2 %
TOTAL	234 211	+ 10,5 %

SOURCE : MINISTÈRE DE L'INTÉRIEUR L'INFORMAGRAPHIE



- 7) For the selected neighborhoods, the one with the overall best school scores and with the greatest variety of restaurants will be selected.

3. Methodology

This is the methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

3.1 Data Analysis

To be filled in week 5

3.2 Foursquare API

To be filled in week 5

4. Results

This is the results section – this section will be filled during Week 5

5. Discussion

This is the discussion section here you discuss any observations you noted and any recommendations you can make based on the results. – this section will be filled during Week 5

6. Conclusion

This is the conclusion section – this section will be filled during Week 5