Crime in the city:
Spatial planning for crime reduction

Ruben Sharpe

March 16, 2021

Contents

1 Introduction

It is intuitively clear that public venues, such as bars and restaurants, but also public service locations such as transport hubs or banks impact the functioning of local communities. However, it is not straightforward to predict whether this impact will be positive or negative.Bars and restaurants can enliven a neighbourhood and may contribute to greater safety because of increased presence of people in the streets as potential witnesses deter crime. On the other hand, the presence of bars and restaurants is also correlated with alcohol abuse related crime. Similarly, there is a duality in the social impact for the presence of shops and banks. These may have a positive impact in the daytime, but may also attract criminal activity at night because of reduced traffic compared to residential areas.

The complexity and the interrelatedness of these issues carries over to the complexity of city planning. The difficulty for city planners, moreover, is that they rarely get to design a neighbourhood from scratch. Typically, they have to deal with neighbourhoods with a mostly fixed set of public facilities, and which may be suboptimal from a design point of view. The question for a city planner, therefore, is often : "What would be, *for a certain neighbourhood* and given all its other attributes, the impact of building (or removing) a shopping mall? Or a park? Or adding, or removing a bus stop?"

There is, of course, a lot of scientific literature on this topic but that is out of the scope of this project. This report is aimed at city planners and is intended to show the feasibility of machine learning as a tool for spatial planning.
Machine learning allowes an automated approach to compare neighbourhoods both in terms of their public venues and in terms of social markers, such as crime rates. This provides city planners, apart from their professional expertise, with examples of good and bad practices that may help them in their decision making.
For reasons of convenience, the analysis in this report will look at the city of Toronto (Ontario, Canada). Crime rates will be used as the proxy for the health status of communities, although other markers such as employment rate or physical health could also be used.

2 Data

Location data, information about public venues and crime statistics will be used to classify neighbourhoods (k-means classifier) and map them with respect to the prevalance of crime (choropleth). From this, generalised conclusions will be drawn with respect to the makeup of 'high-crime' and 'low-crime' neighbourhoods. An example of a map of the geographical location of similar neighbourhoods, which can be mapped onto crime information, is given in Figure 1. An example of the characterisation of one of such clusters is given in Figure 2

2.1 Location data

Location data for the city of Toronto can be acquired by scraping postal codes, boroughs and neighbourhoods from Wikipedia.[1] This data needs to be enriched with the GPS coordinates of the respective neighbourhoods. These coordinates can be obtained from the internet in the form of a '.csv' file from the Coursera website for the Data Science Capstone.[2] Examples of both freshly scraped and enriched data are given in Figures 3 and 4.

2.2 Public venues

Foursquare is used to obtain information about the public venues, specifically their category, in each neighbourhood . For this purpose, the neighbourhood is

[1] https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
[2] https://cocl.us/Geospatial_data

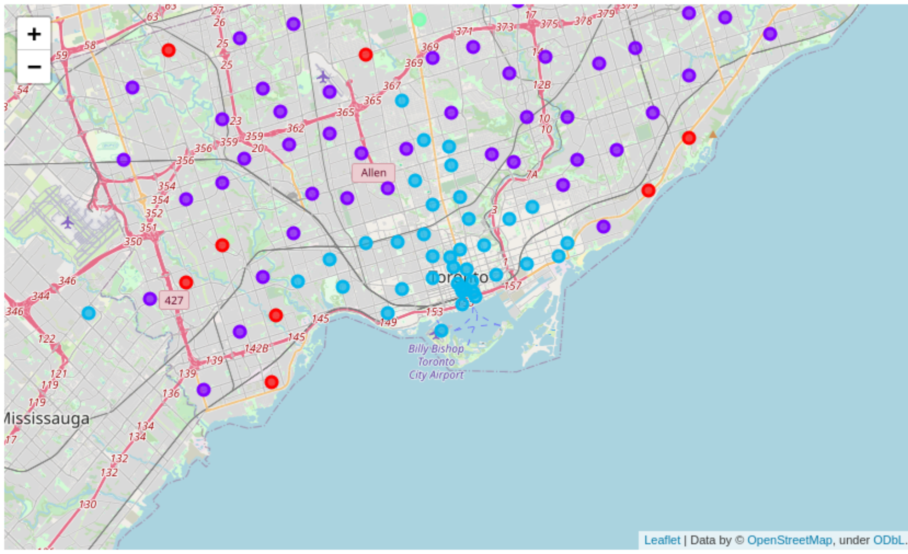

Figure 1: Example the clustering of similar neighbourhoods in Toronto. Similar colours denote similar neighbourhoods.

arbitrarily defined as the area within a 1.500 m radius around a postal code's geographical centre. An example of what this looks like is shown in Figure 5. The top 10 most prevalent venues will be used to characterise the neighbourhood.

2.3 Crime data

The city of Toronto provides a lot of open data related to the city.[3] This data ranges from polls conducted by the government, to inventory lists of street furniture, to 'bicycle count and locations' and is ever increasing. From this repository, a list of crime rates can be obtained.[4] This list contains, for each neighbourhood the number of incidences of certain types of crime (Table 1). Population information from the 2016 census, included in the file, can be used to normalise crime rates relative to population size. Conveniently, the file also contains geographical information so that this information can be easily mapped. The crime categories that are accessible from the data files are given in Table 1. A graphical example of how this data can be visualised in relation to geographical location is given in Figure 6.

[3]https://www.toronto.ca/city-government/data-research-maps/open-data/
[4]https://open.toronto.ca/dataset/neighbourhood-crime-rates/

	Neighbourhood	Cluster Labels	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
12	Agincourt	3	Chinese Restaurant	Shopping Mall	Breakfast Spot	Park	Coffee Shop	Cantonese Restaurant	Caribbean Restaurant	Bakery	Hong Kong Restaurant	Gym / Fitness Center
14	Milliken, Agincourt North, Steeles East, L'Amo...	3	Chinese Restaurant	Dessert Shop	Korean Restaurant	Coffee Shop	Intersection	Vietnamese Restaurant	Pizza Place	Pharmacy	Noodle House	Bakery
21	Willowdale, Newtonbrook	3	Korean Restaurant	Coffee Shop	Café	Bubble Tea Shop	Restaurant	Shopping Mall	Bank	Gas Station	Japanese Restaurant	Middle Eastern Restaurant
22	Willowdale, Willowdale East	3	Korean Restaurant	Bubble Tea Shop	Coffee Shop	Grocery Store	Ramen Restaurant	Japanese Restaurant	Pizza Place	Sushi Restaurant	Café	Sandwich Place

Figure 2: Example of a cluster of similar neighbourhoods. Note that the most prevalent venues are Asian restaurants.

	Postal Code	Borough	Neighbourhood
count	103	103	103
unique	103	11	99
top	M2K	North York	Downsview
freq	1	24	4

Figure 3: Example of cleaned data after scraping from Wikipedia.

3 Methodology

4 Results

5 Discussion

6 Conclusion

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	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

Figure 4: Example of data enriched with coordinates.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern, Rouge	43.806686	-79.194353	Images Salon & Spa	43.802283	-79.198565	Spa
1	Malvern, Rouge	43.806686	-79.194353	Harvey's	43.800020	-79.198307	Restaurant
2	Malvern, Rouge	43.806686	-79.194353	Canadiana exhibit	43.817962	-79.193374	Zoo Exhibit
3	Malvern, Rouge	43.806686	-79.194353	RBC Royal Bank	43.798782	-79.197090	Bank
4	Malvern, Rouge	43.806686	-79.194353	Staples Morningside	43.800285	-79.196607	Paper / Office Supplies Store

Figure 5: Example of which venue information can be obtained.

Assault
Auto Theft
Breaking and Entering
Homicide
Robbery
Theft

Table 1: Crime data categories. Data is available for each year from 2014–2019.

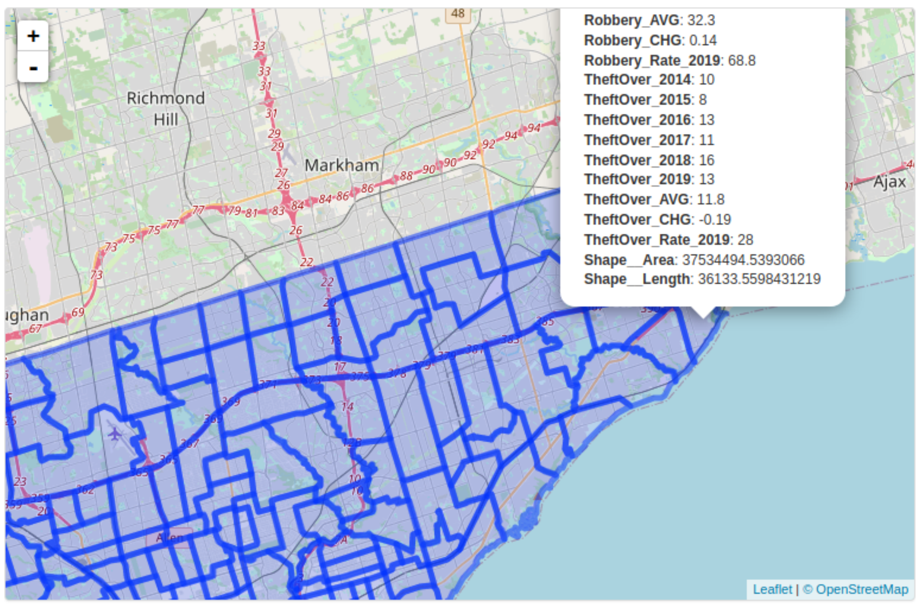

Figure 6: Example of crime information can be obtained in conjunction with a geographical location. Image is generated from the Toronto city government website.