Top 10 safest neighborhoods in Toronto (2014-2019)*

Based on police reports for six major crimes

Babak Mokri

31 January 2021

Abstract

Crime rate statistics in Toronto neighbourhoods can be interesting for current residence of it as well as for those who are planning to move to Toronto. In this paper I have compared six types of major crimes including Assault, Auto Theft, Break & Enter, Robbery, Theft Over, and Homicide, happened in 140 neighbourhoods in Toronto during 2014 to 2019. Firstly, I tried to list top ten and last five safest Toronto neighborhoods, however, it became so controversial as I looked into that from 3 various perspectives. Then, I made an analogy of this contradictions to the biased results from AI algorithm, and finally, how safe neighborhoods can be victimized to the existing systematic violence.

1 Introduction

Choosing a safe neighborhood in Toronto has always been a top concern for a newcomer parent like me, especially when dealing with two curious and challenging early teenagers who prefer not to stay at home with parents but explore the whole neighborhood with their own pals. Tough, I personally encourage them find themselves on their own feet, I should always take some provisions on the background like choosing a safer environment for them to live and flourish. To choose a safe neighborhood, I have compared the crime rates for different areas in Toronto during 2014-2019.

Although the conclusions are prone to initiate some biases, I am just providing information to grab the readers' attention to the fact that what numbers are covering for the different crime rates in each Toronto neighborhood. And also, how these numbers can change the results if you impose your own subjective opinions. In addition, different population densities may result in different conclusions as well. Other demographic factors may also raise this subjectivity, prejudice and partiality.

In this report I will also elaborate more on selecting different types of variables which initiates biases among people, leading to the same issues when they are chosen for the machine learning practices.

For coding, analysis and data visualization, I used R (R Core Team 2020) programing language, and some of its libraries including tidyverse (Wickham et al. 2019), opendatatoronto (Gelfand 2020), ggplot2(Wickham 2016), and bookdown(Xie 2016). This report was then compiled using R markdown(Xie, Dervieux, and Riederer 2020). I acknowledge that I am not concerned with the reasons behind these numbers nor am I pointing to any specific ward to name it dangerous.

I start to download the required dataset, make some quantitative analysis from different viewpoints, and finally make some analogies between human rationale and what may happen in AI algorithms as well as unanticipated issues with categorizations in a dataset.

*Code and data are available at: https://github.com/datababak/2178-A1

2 Data

I used Toronto Police Reports data, taken from the Open Data Toronto Portal (Toronto 2019), to explore how crime rates vary between different neighborhoods in Toronto. Among many existing datasets, I chose the one which contains data for six Major Crime Indicators (MCI) including Assault, Auto Theft, Break & Enter, Robbery, Theft Over, and Homicide occurred between the years 2014 to 2019. This dataset is updated annually, and the last update was at Sept 18, 2020 by its publisher, Toronto Police Services. This dataset is available in SHP, CSV, GEOJSON, GPKG formats and it can be found in this LINK.

In this dataset, there are 62 attributes in separate columns, providing data for 140 Toronto neighborhoods in distinct rows. The list of attributes for this dataset is appended to this report.

2.1 Part I

For the purpose of choosing a safe neighborhood for my residence, I considered all reported crimes for each neighborhood, for which I summed up the 6 MCI averages, occurred during 2014-2019.

(Table1 **??**) shows the first top 10 and the last 5 safest neighborhoods in Toronto.

Neighborhood	Average_Crimes
Lambton Baby Point	**58.8**
Woodbine-Lumsden	**63.4**
Maple Leaf	**68.4**
Yonge-St.Clair	**68.6**
Markland Wood	**68.8**
Guildwood	**68.8**
Casa Loma	**79.9**
Old East York	**80.1**
Forest Hill South	**82.4**
Kingsway South	**82.6**
Moss Park	**800.2**
West Humber-Clairville	**951.8**
Church-Yonge Corridor	**1040.6**
Bay Street Corridor	**1137.6**
Waterfront Communities-The Island	**1292.2**

Then I realized that, crimes like Auto-theft, Break & Enter, and Theft Over are less likely to happen to a teenager compared to Homicide, Robbery and Assault for which my kids would be more vulnerable. So, I decided to give some weights to the 3 latter crimes to make the results more accurate for my case. Based on my own reasoning, I gave 5 weights to Homicide (as this was scarier to the parents as well), 2 to Robbery and Assault, and 3 to Assault (to which kids are very sensitive and usually victimized). (Table **??**) shows the new results after my modifications.

Neighborhood	Average_Crimes_w
Markland Wood	**118.0**
Forest Hill South	**125.1**
Lambton Baby Point	**126.1**
Yonge-St.Clair	**136.3**
Maple Leaf	**138.6**
Woodbine-Lumsden	**141.8**
Lawrence Park North	**145.7**
Kingsway South	**146.0**
Bridle Path-Sunnybrook-York Mills	**152.2**
Lawrence Park South	**155.7**
West Humber-Clairville	**1653.2**
Moss Park	**1885.1**
Church-Yonge Corridor	**2469.9**
Bay Street Corridor	**2806.9**
Waterfront Communities-The Island	**3082.0**

Having said that I am organizing these neighborhoods for my own study, and probably would recommend it to other parents, you may notice how a subjective weighting system could label some other safe areas as a *not safe* area. This tag may then become a viral term for that neighborhood and hence, initiate an unfavorable prejudice among people.

Further to the safety, there are other factors such as affordability, school ranks, and the proximity to my relatives which impact my decision making. So, calling the safety factor as my first concern, I would like to have the first 20 safest neighborhoods in a graph as shown in (Figure 1), and apply the other factors to each subsequent candidate neighborhoods consequently.

The above graph provides a visual information representation for the different safety levels at a glance, for which in the first two tables, I had to calculate them manually.

In this story, I would like to draw your attention to how subjectivity it may lead to emerging future biases. This is not much far from what may happen in Machine Learning algorithms. For example, in ML algorithms where experts may hire some external heuristic algorithms to reduce the "worthless" CPU energy consumption, improve productivity, and help the machine algorithms to converge faster, they may accidentally contribute to generating unforeseen biased outcomes, like how our outcomes tagged some neighborhoods as more dangerous ones. As Karen Hao claims "The introduction of bias isn't always obvious during a model's construction because you may not realize the downstream impacts of your data and choices until much later. Once you do, it's hard to retroactively identify where that bias came from and then figure out how to get rid of it" (Hao 2019). Likewise, I agree the experts' intention, in our example, was positively tuning the results for their own objectives, however, their producing unwilling biases at the background may remain hidden for years (or forever) as "unknown-unknowns" (Hao 2019) leading to controversial results and antagonistic disputes.

2.2 Part II

Like other users, I just relied on the existing data, which is generated by a valid source (police), and it is announcing the crime rates for each neighborhood, regardless of the density of population and/or the size of each individual neighborhood. Let me put it in this way; Does a low-rate crime in a small area or very less crowded neighborhood mean that it is safe?

Fortunately, this dataset has provided one column indicating the rate of 6 above-mentioned MCIs for 2019 per 100,000 population. (Table **??**) shows the top 10 and last 5 safest Toronto neighborhoods considering the population in each district.

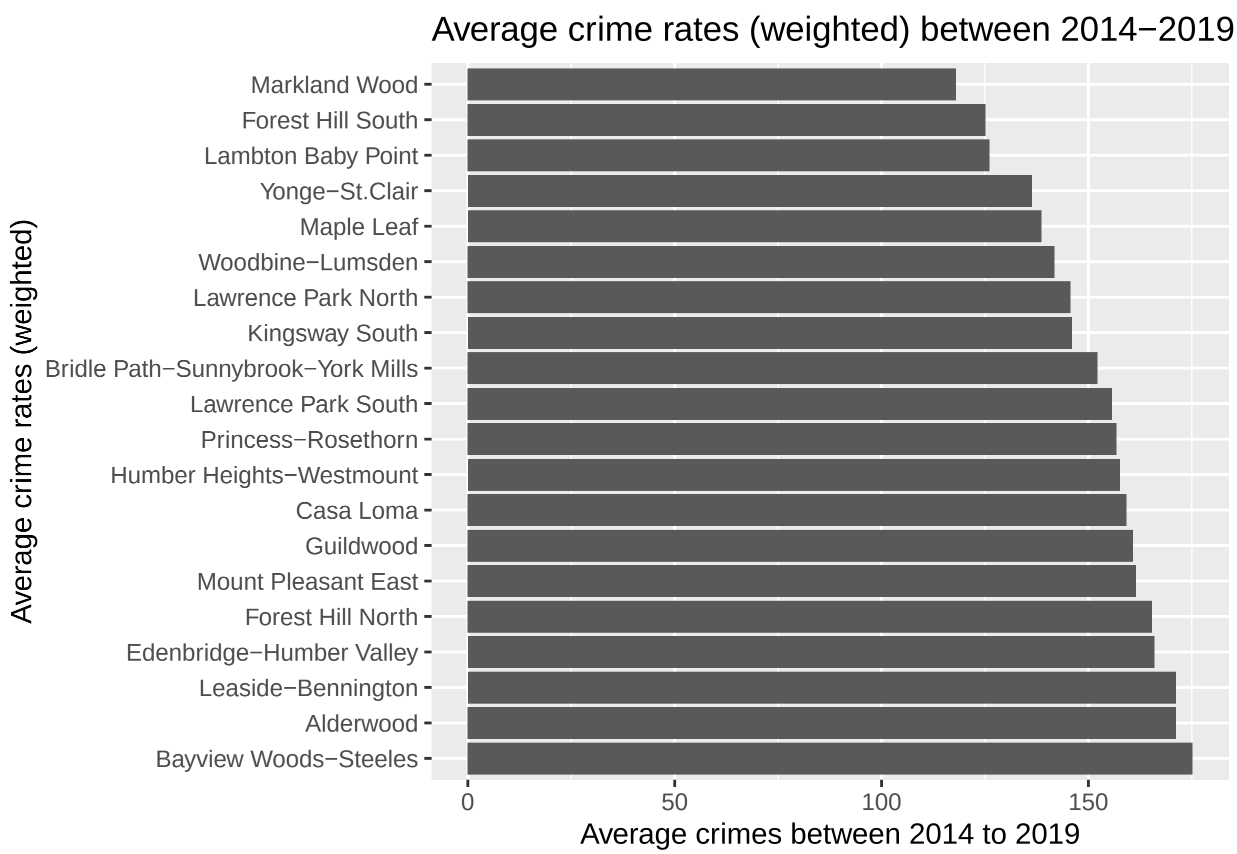

Figure 1: The top 20 safest Toronto neighborhoods

Neighborhood	total_Rate2019
Guildwood	**504.2**
Lawrence Park North	**595.5**
Pleasant View	**625.9**
Steeles	**629.6**
Markland Wood	**634.9**
Yonge-St.Clair	**646.5**
Eringate-Centennial-West Deane	**661.8**
Leaside-Bennington	**689.4**
Bayview Woods-Steeles	**691.7**
Woodbine-Lumsden	**699.2**
University	**3194.4**
Kensington-Chinatown	**4307.6**
Church-Yonge Corridor	**4537.3**
Moss Park	**5281.3**
Bay Street Corridor	**5314.5**

let's look into the results. (Table **??**) compares Table 1 and Table 3 side by side; And you may notice, a complete change in the results.

Neighborhood	Average_Crimes	Neighborhood	total_Rate2019
Lambton Baby Point	58.8	Guildwood	504.2
Woodbine-Lumsden	63.4	Lawrence Park North	595.5
Maple Leaf	68.4	Pleasant View	625.9
Yonge-St.Clair	68.6	Steeles	629.6
Markland Wood	68.8	Markland Wood	634.9
Guildwood	68.8	Yonge-St.Clair	646.5
Casa Loma	79.9	Eringate-Centennial-West Deane	661.8
Old East York	80.1	Leaside-Bennington	689.4
Forest Hill South	82.4	Bayview Woods-Steeles	691.7
Kingsway South	82.6	Woodbine-Lumsden	699.2
Moss Park	800.2	University	3194.4
West Humber-Clairville	951.8	Kensington-Chinatown	4307.6
Church-Yonge Corridor	1040.6	Church-Yonge Corridor	4537.3
Bay Street Corridor	1137.6	Moss Park	5281.3
Waterfront Communities-The Island	1292.2	Bay Street Corridor	5314.5

Considering that safe/dangerous labels directly impact the real estate prices in each neighborhoods, in my opinion, for areas which are systematically labeled as the most dangerous ones in our first table, I claim they are the victim to systematic violence which XXX explains in his article .... in This article he covers the systematic violence for the absence of some existing attributes like which we have not divided neighborhoods equally to realize where you can find most criminals among different Toronto neighborhoods.

2.3 Paer III

Toronto is receiving number of immigrants each year. So, we should expect demographic changes in neighborhoods. There is also on variable in this dataset which shows the trend of crimes change between 2018 to 2019. Figure (Figure1 **??**) shows this trend for the 15 neighborhoods introduced in Table 1

If you see a sharp increase/decrease in ... it is because of lack os data.

Appendix

Apendix 1 List of attributes of the main dataset (https://open.toronto.ca/dataset/neighbourhood-crime-rates/).

1. *id*: Unique row identifier for Open Data database
2. *OBJECTID*: Autogenerated distinct record identifier
3. *Neighbourhood*: Name of City of Toronto neighbourhood
4. *Hood_ID*: City of Toronto neighbourhood identifier
5. *Population*: 2016 Census Population
6. *Assault_2014*: Count of assaults for 2014
7. *Assault_2015*: Count of assaults for 2015
8. *Assault_2016*: Count of assaults for 2016
9. *Assault_2017*: Count of assaults for 2017
10. *Assault_2018*: Count of assaults for 2018
11. *Assault_2019*: Count of assaults for 2019
12. *Assault_AVG*: Average Assaults from 2014 - 2019
13. *Assault_CHG*: % Change in assaults from 2018-2019
14. *Assault_Rate_2019*: Rate of assaults for 2019 per 100,000 population
15. *AutoTheft_2014*: Count of auto thefts for 2014
16. *AutoTheft_2015*: Count of auto thefts for 2015
17. *AutoTheft_2016*: Count of auto thefts for 2016
18. *AutoTheft_2017*: Count of auto thefts for 2017
19. *AutoTheft_2018*: Count of auto thefts for 2018
20. *AutoTheft_2019*: Count of auto thefts for 2019
21. *AutoTheft_AVG*: Averate auto thefts from 2014-2019
22. *AutoTheft_CHG*: % Change in auto thefts from 2018-2019
23. *AutoTheft_Rate_2019*: Rate of auto thefts for 2019 per 100,000 population
24. *BreakandEnter_2014*: Count of break and enters for 2014
25. *BreakandEnter_2015*: Count of break and enters for 2015
26. *BreakandEnter_2016*: Count of break and enters for 2016
27. *BreakandEnter_2017*: Count of break and enters for 2017
28. *BreakandEnter_2018*: Count of break and enters for 2018
29. *BreakandEnter_2019*: Count of break and enters for 2019
30. *BreakandEnter_AVG*: Averate break and enters from 2014-2019
31. *BreakandEnter_CHG*: % Change in break and enters from 2018-2019
32. *BreakandEnter_Rate_2019*: Rate of break and enters for 2019 per 100,000 population
33. *Homicide_2014*: Count of homicides for 2014
34. *Homicide_2015*: Count of homicides for 2015
35. *Homicide_2016*: Count of homicides for 2016
36. *Homicide_2017*: Count of homicides for 2017
37. *Homicide_2018*: Count of homicides for 2018
38. *Homicide_2019*: Count of homicides for 2019
39. *Homicide_AVG*: Averate homicides from 2014-2019
40. *Homicide_CHG*: % Change in homicides from 2018-2019
41. *Homicide_Rate_2019*: Rate of homicides for 2019 per 100,000 population
42. *Robbery_2014*: Count of robberies for 2014
43. *Robbery_2015*: Count of robberies for 2015
44. *Robbery_2016*: Count of robberies for 2016
45. *Robbery_2017*: Count of robberies for 2017
46. *Robbery_2018*: Count of robberies for 2018
47. *Robbery_2019*: Count of robberies for 2019
48. *Robbery_AVG*: Averate robberies from 2014-2019
49. *Robbery_CHG*: % Change in robberies from 2018-2019

50. *Robbery_Rate_2019*: Rate of robberies for 2019 per 100,000 population
51. *TheftOver_2014*: Count of thefts over for 2014
52. *TheftOver_2015*: Count of thefts over for 2015
53. *TheftOver_2016*: Count of thefts over for 2016
54. *TheftOver_2017*: Count of thefts over for 2017
55. *TheftOver_2018*: Count of thefts over for 2018
56. *TheftOver_2019*: Count of thefts over for 2019
57. *TheftOver_AVG*: Averate thefts over from 2014-2019
58. *TheftOver_CHG*: % Change in thefts over from 2018-2019
59. *TheftOver_Rate_2019*: Rate of thefts over for 2019 per 100,000 population
60. *Shape___Area*: Autogenerated area measurement
61. *Shape___Length*: Autogenerated length measurement
62. *geometry*

References

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Hao, Karen. 2019. *This Is How AI Bias Really Happens—and Why It's so Hard to Fix*. https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Toronto, City of. 2019. "Open Data." *City of Toronto*. https://www.toronto.ca/city-government/data-research-maps/open-data/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2016. *Bookdown: Authoring Books and Technical Documents with R Markdown*. Boca Raton, Florida: Chapman; Hall/CRC. https://github.com/rstudio/bookdown.

Xie, Yihui, Christophe Dervieux, and Emily Riederer. 2020. *R Markdown Cookbook*. Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown-cookbook.