Does Global Covid-19 Pandemic and Toronto Lockdown Policy Decrease the Crime Rate?*

A study on crime rate change from covid-19 with the City of Toronto's Open Data

Chengle Yang

April 1st, 2022

Abstract

This report presents a secondary analysis of the change of the City of Toronto's crime rate change from 2016 to 2021. The results indicate that the Covid-19 Pandemic and corresponding government restrictions policy could affect the occurrence number and types of Toronto's crime rate. In detail, covid will cause a decrease in Assault but an increase in AutoTheft. The analysis consists of logistic regression performed with the statistical programming language R. The results contribute to our understanding of how to lower the crime rate, and help the Police department optimize the decision of crime elimination. Also, it can provide some useful insights for the government and policymakers regarding future strategic planning.

Keywords: Covid-19, Toronto crime rate, city of Toronto neighborhood, crime type, open data toronto

Contents

*Code and data are available in this GitHub repository: chle1999/Covid_Influence_Crime_Toronto.

Introduction

Crime rate is an important standard factor used to measure the safety of the city. Its types and occurrence number rapidly change between 2019 and 2020 as the Covid-19 pandemic profoundly impacts Canada's economy, health care system, and society. Government also enacted policies used to contain the spread of the virus, which also change people's lifestyles including socialization, learning, and working. Since March 2020, the majority of Canada's population were spending more time at home. These changes bring an influence on the crime patterns across the country. Police-reported crime in Canada, as measured by the Crime Severity Index (CSI), decreased 8% from 79.8 in 2019 to 73.4 in 2020.(Moreau 2021) And the total major crime indicators for Toronto decreased from 2019 to 2020, down about 14% from a rate of 1,332.1 occurrences per 100,000 people in 2019 to 1,146.5 per 100,000 people in 2020.(Rakowska 2022) But, do all types of crime really decrease under the covid-19 pandemic and government lockdown policy? To answer this question, the following pages present a secondary analysis of survey data collected by the City of Toronto, about neighborhood crime rate change during 2019-2021.

Although the overall decreasing trend seems pretty well, we want to focus on the detailed change of distinct crime types. According to Greg Moreau(Moreau 2021), in 2020 Canada, there were decreases in the rates of police-reported breaking and entering (-16%), theft of $5,000 or under (-20%), robbery (-18%), shoplifting of $5,000 or under (-36%), administration of justice violations (-17%) and sexual assault (level 1) (-9%). Thus, we are more curious about the change in the city we live in - Toronto. Some of the reports mention that there exists an unexpected increase in some of the crime types in Toronto. For example, Toronto police reported 151%, 222% jump in stunt driving charges and speeding tickets respectively, compared to 2019.(D'ANDREA 2021) And the attacks on Asian people have risen significantly across North America since the onset of the pandemic amid false allegations the virus was deliberately unleashed by China.(Perkel 2021) In order to solve the corresponding issue, the Toronto city council is asking the police services board to add more resources and officers to the Toronto Police Hate Crime Unit in response to an "unprecedented" rise in incidents of hatred in the city.(Dunn 2021)

Thus, statistical support may be required for analyzing the crime rate change and used for the targeted crime eradication.

The paper first presents an overview of the original dataset(Toronto Public Safety Data Portal 2022) and explains the essential variables inside and some key terminology and definitions. And it also describes the advantage and disadvantages of the collection methods of the dataset. Next, the Methodology section will talk about the strengths and weaknesses of the model used in analyzing the data. It includes details on how the data were re-coded to suit the model. Then, the Result section will summarize and explains the results based on the visual figures and table which are modeled in the previous section. Finally, the Discussion section will conclude the limitations and ethics of the research.

To analyze the data in this report, the R statistical programming language(R Core Team (2022)) will be applied. And the following packages are also required for extracting, cleaning, analyzing, and drawing data: here(Müller 2020), janitor(Firke 2021), kableExtra(Zhu 2020), knitr(Xie 2022), opendatatoronto(Gelfand 2020), readxl(Wickham and Bryan 2019), tidyverse(Wickham et al. 2019),lme4(Bates 2022), rvest(Wickham

2021), polite(Perepolkin 2019), lmtest(Hothorn 2022), lmerTest(Kuznetsova 2020) and ggplot2(Wickham 2016).

Data

Data Source and Characteristics

Toronto Public Safety Data Portal is the website store of all the reported crime cases in Toronto. It includes various datasets including Major Crime Indicators(MCI), Homicide, Traffic, and so on. As we focus on the main crime type change in the city of Toronto, we use the MCI(Toronto Public Safety Data Portal 2022) as our data resource. We downloaded data as a CSV file from the website(Toronto Public Safety Data Portal 2022), imported it into R studio, and extracted the data. There was no other dataset can that described all the details of each criminal case like the MCI.

Here are the explanations of some crime types and keywords in the dataset, which we will use in the following analysis:
- Neighbourhood: Name of City of Toronto neighborhood.
- Assault: The act of inflicting physical harm or unwanted physical contact upon a person.
- AutoTheft: The criminal act of stealing or attempting to steal a motor vehicle.
- BreakAndEnter: Someone enters a residence, business or other property without permission and commits or intends to commit a crime.
- Robbery: The theft or extortion of property, and the use of a weapon, violence or threats of violence.
- TheftOver: Take or convert anything that belongs to another person with the intent to deprive that person temporarily or permanently of that good.
- Homicide: A person commits homicide when, directly or indirectly, by any means, he causes the death of a human being.

The following figure1 present the monthly number of seven crime types in the city of Toronto from 2014 to 2020. It is very useful for us to obtain the changing trend of different crimes before the covid. From the figure, we can observe that only “Assault” and “AutoTheft” have an obvious change in 2020 when compared with the previous 6 years. The rest of the crime types, such as BreakAndEnter and Robbery, have slightly increased or decreased in 2020, which can be considered a normal fluctuation. Therefore, our main research topic will focus on the effect of COVID-19 Pandanmic and its corresponding limitation policy on the Assault and AutoTheft. And we will also consider building the appropriate generalized linear regression model for finding the relationship between the crime rate with covid and lockdown.

We would like to separate the modeling section into two parts as we want to separately find the most appropriate model for these two topics: Assault and AutoTheft.

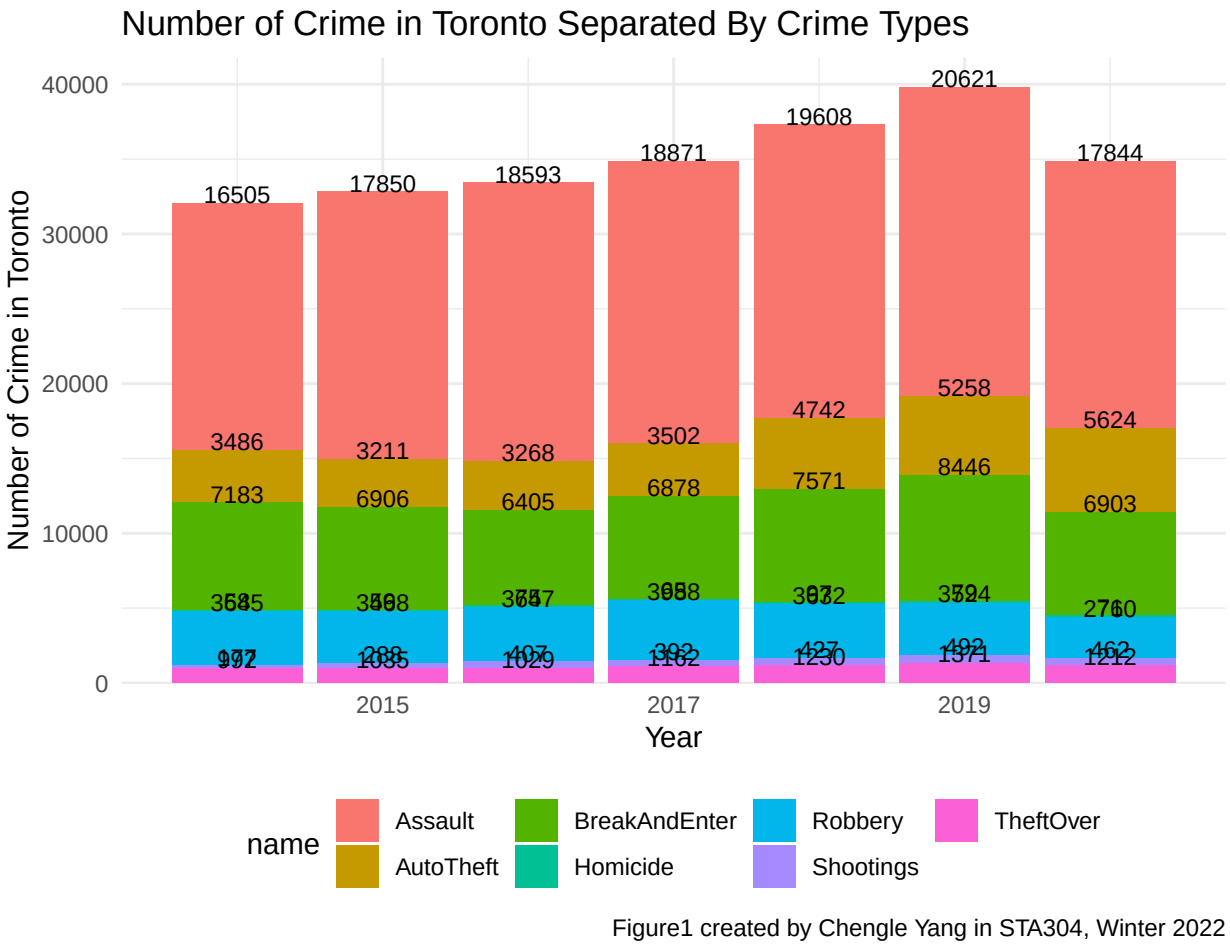

Figure 1: Annual Crime Number in Toronto From 2014 to 2020 Separated by Crime Types

Table 1: Indicator and ratio of deviance to residual degrees-of-freedom for two fitted models

models	ratio
Assault ~ Covid	23.10377
Assault ~ Covid + If_Lockdown	21.77206

Survey Methodology

Analysis for Assault

Firstly, we clean the data from the raw data by extracting and counting the monthly Assault number from the dataset(Toronto Public Safety Data Portal 2022). We completed this process in the script. then we add the monthly covid case number as an additional column in the table, which is extracted from the Toronto Open Date website(City of Toronto 2022) and counted by the script. we also add another column to the table which describes the lockdown situation of the city of Toronto. the "Lockdown" column is a boolean indicator, which can be used as a YES-NO factor in the linear model.

As the Covid happened after January 2020, so we will only consider the monthly data, including monthly AutoTheft Number, Covid Case Number, and If_Lockdown, after January 2020 when we need to build the linear regression model. As our after-process dataset contain the boolean factor - "If_Lockdown", we start by trying the generalized linear model with Poisson Regression.

From the Table1, we can get the ratio of deviance to residual degrees-of-freedom of the fitted model after fitting the data with Poisson Regression Analysis. However, both model's value is much larger than 1, which means the data, especially the response "IfLockdown", is not satisfy the Poisson Distribution and we should not used the generalized linear model based on Poisson Regression.

Thus, we should consider the simple regression linear model $AssaultNumber \sim CovidNumber$ as model3 and multiple regression linear model $AssaultNumber \sim CovidNumber * IfLockdown$ as model4.

Then, we used likelihood ratio test(LRT) to compare these models and this pair of model are nested models.

Table 2: Likelihood Ratio test between model3 and model4

#Df	LogLik	Df	Chisq	Pr(>Chisq)
3	-158.5936	NA	NA	NA
5	-155.8979	2	5.3914	0.0675

For model 4, we add a new indicator IfLockdown as a boolean factor, then we compared model 3 and model 4. From the table 2, the p-value is which is 0.0675 larger than $\alpha = 0.05$, it means that we do not have evidence against the null hypothesis that the larger model explains data better, so we prefer the larger model which is model 4. Thus it is necessary to add IfLockdown as a new indicator.

The linear equation for the model looks like this:

$$Assault = \beta_0 + \beta_1 * CovidNumber + \beta_2 * Lockdown + \beta_3 * [CovidNumber * Lockdown] + \epsilon$$

where:
- `Assault` is the response variable of interest
- β_0 is the intercept term for the average monthly Assault number in the city of Toronto
- $\beta_1 * CovidNumber$ is the variable that accounts for the relationship between monthly Assault number and monthly covid case number. To interpret this, β_1 represents the change in monthly Assault number for every one unit increase in monthly covid case number.
- $\beta_2 * Lockdown$) represents the indicator/dummy term that accounts for the relationship between monthly Assault number and Toronto lockdown situation.
- $\beta_3 * [CovidNumber * Lockdown]$ is the interaction term between monthly covid case number and the lockdown situation.
- ϵ is the error term.

Then, let's consider the relationship between covid and monthly AutoTheft number.

Table 3: Indicator and ratio of deviance to residual degrees-of-freedom for two fitted models

models	ratio
AutoTheft ~ Covid	14.29809
AutoTheft ~ Covid + IfLockdown	10.26365

Analysis for AutoTheft

Firstly, we clean the data from the raw data by extracting and counting the monthly AutoTheft number from the dataset(Toronto Public Safety Data Portal 2022). We completed this process in the script. then we add the monthly covid case number as an additional column in the table, which is extracted from the Toronto Open Date website(City of Toronto 2022) and counted by the script. we also add another column to the table which describes the lockdown situation of the city of Toronto. the "Lockdown" column is a boolean indicator, which can be used as a YES-NO factor in the linear model.

As the Covid happened after January 2020, so we will only consider the monthly data, including monthly AutoTheft Number, Covid Case Number, and If_Lockdown, after January 2020 when we need to build the linear regression model. As our after-process dataset contain the boolean factor - "If_Lockdown", we start by trying the generalized linear model with Poisson Regression.

From the Table3, we can get the ratio of deviance to residual degrees-of-freedom of the fitted model after fitting the data with Poisson Regression Analysis. However, both model's value is much larger than 1, which means the data, especially the response "IfLockdown", is not satisfy the Poisson Distribution and we should not used the generalized linear model based on Poisson Regression.

Thus, we should consider the simple regression linear model $AutoTheftNumber \sim CovidNumber$ as model3 and multiple regression linear model $AutoTheftNumber \sim CovidNumber * IfLockdown$ as model4

We used likelihood ratio test to compare these models and each pair of model are nested models

Table 4: Likelihood Ratio test between model3 and model4

#Df	LogLik	Df	Chisq	Pr(>Chisq)
3	-140.0134	NA	NA	NA
5	-135.5251	2	8.9766	0.0112

For model 4, we add a new indicator IfLockdown as a boolean factor, then we compared model 3 and model 4. From the table 4, the p-value is 0.0112 which is smaller than $\alpha = 0.05$, it means that we have very strong evidence against the null hypothesis that the larger model explains data better, so we prefer the larger model which is model 4. Thus it is necessary to add IfLockdown as a new indicator.

The linear equation for the model looks like this:

$$AutoTheft = \beta_0 + \beta_1 * CovidNumber + \beta_2 * Lockdown + \beta_3 * [CovidNumber * Lockdown] + \epsilon$$

where:
- `AutoTheft` is the response variable of interest
- β_0 is the intercept term for the average monthly AutoTheft number in the city of Toronto
- $\beta_1 * CovidNumber$ is the variable that accounts for the relationship between monthly AutoTheft number and monthly covid case number. To interpret this, β_1 represents the change in monthly AutoTheft number for every one unit increase in monthly covid case number.
- $\beta_2 * Lockdown$) represents the indicator/dummy term that accounts for the relationship between monthly AutoTheft number and Toronto lockdown situation.
- $\beta_3 * [CovidNumber * Lockdown]$ is the interaction term between monthly covid case number and the lockdown situation.
- ϵ is the error term.

Result

Overview

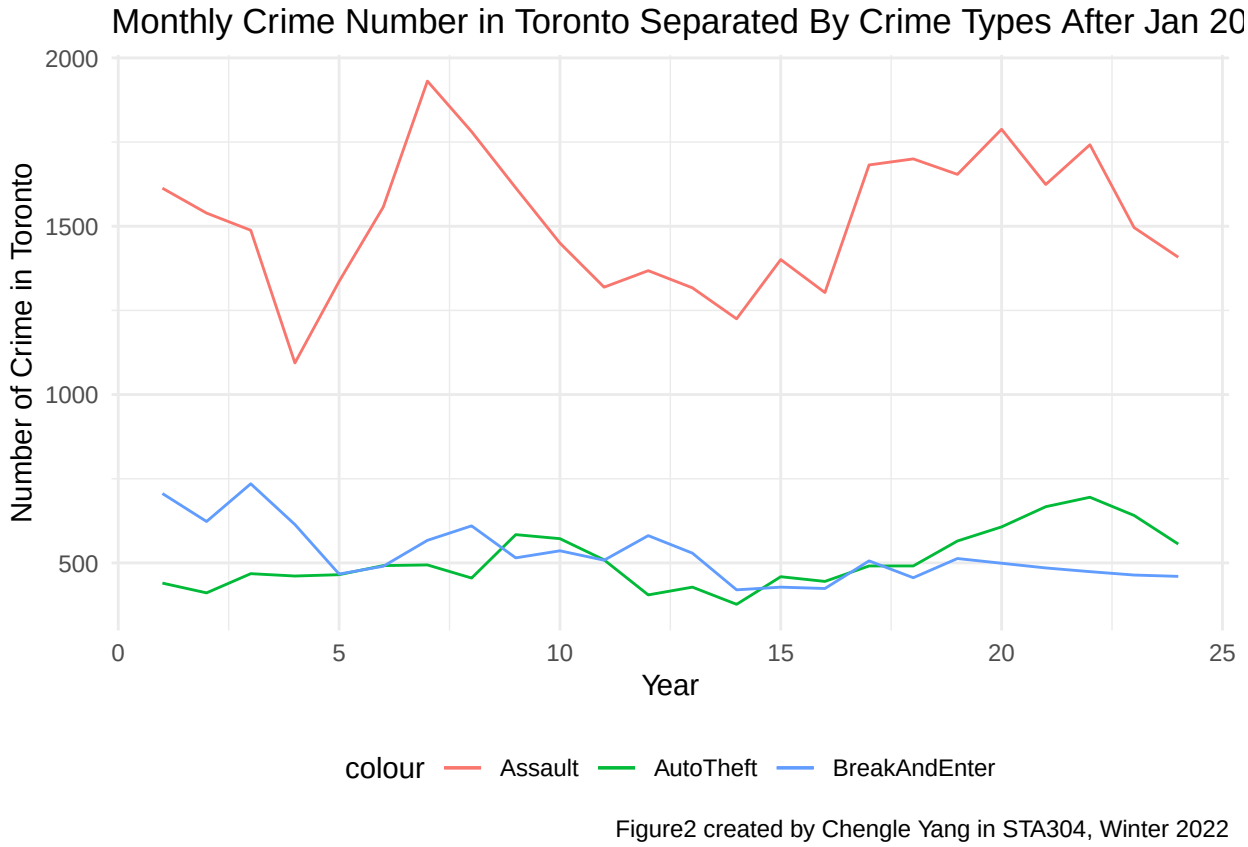

Figure 2: Annual Crime Number in Toronto From 2014 to 2020 Separated by Crime Types

Racism and corresponding assualt increased during covid lockdown

According to Vanessa Balintec(Balintec 2022), the hate incidents reported by South Asian and Southeast Asian people increased by 318 per cent and 121 per cent respectively in 2021 when compared with previous years. From the figure2, we can obviously see that the assault rapidly increase during the lock down period, which is from Mar 2020 to Jul 2021. After around 7 months decreasing, the monthly assault number increase again, which happened in the end second lock down period, and keep high in the following months. Also, based on the reported from University of Victoria(Lou 2022), many racist attacks is caused by the COVID-19 pandemic and those discrimination worsened Chinese Canadian's well-being and sense of belonging.

AutoTheft keep increasing during the whole COVID-19 pandemic period until now

Based on the report of Becky Robertson(Robertson 2022), stay-at-home orders and business closures made for a prime time for not just dangerous driving, but also auto theft. From the figure2, we can clearly observe that Autotheft has a slightly increasing trend from January 2020 to December 2021. Although there exists 3 months decrease starting from September 2020, it quickly springs back and stays at a high level in the following months.

As we discussed before, we are curious why Assault and AutoTheft have such a decrease/increase respectively during the COVID-19 pandemic when compared with previous years. With the multiple linear regression, we can find the quantitative relationship between crime and covid and its corresponding result - Lockdown. Firstly, let's talk about the assault.

Table 5: Estimate and P-values for responses in linear fitted model 4

	Estimate	Standard.Error	Test.Statistic	P.value
(Intercept)	1695.9949121	68.8168320	24.645059	0.0000000
Covid	-0.0226129	0.0122419	-1.847165	0.0795718
LockdownYES	-221.8127075	99.8795822	-2.220801	0.0380769
Covid:LockdownYES	0.0196528	0.0126443	1.554279	0.1358002

Analysis for Assault

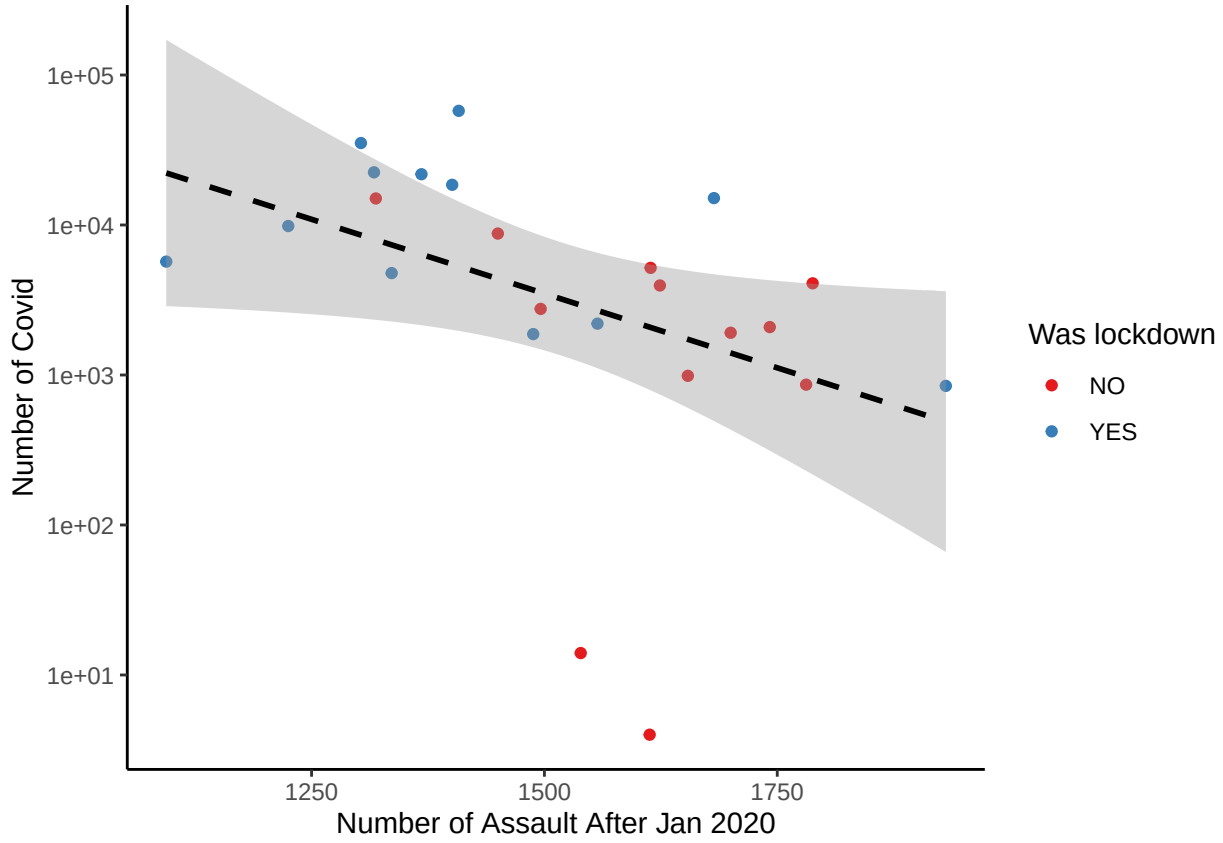

Figure 3: Multiple linear regression with Monthly Assault Numer after Jan 2020 and related Covid Number, with a binary variable for whether Toronto was lockdown

From the figure3, we can see that Number of Covid and monthly Assaults number have a reverse relationship. And there exist several points, especially the blue points in the graph, not inside the confidential interval of the prediction line, which means the lockdown policy does not reduce the assault number in our research period.

Then, get the estimate and p-values for responses in linear fitted model 4(table5):

From the table 5, we can find that interactive factor with $p = 0.1358002 > 0.05$, which mean it fail to reject $H_0 : \beta_3 = 0$. Thus, there is no interaction relationship between CovidNumber and IfLockdown. And we should use addition for multiple linear regression model rather than the model with interaction relationship.

Finally, by data from table 6, the equaion for the fitted model of Assault(terms explain before):

$$Assault = 1625.9382 - 0.0042 * CovidNumber - 131.6490 * Lockdown + \epsilon$$

Table 6: Estimate for fitted model 5

	x
(Intercept)	1625.938204
Covid	-0.004191
LockdownYES	-131.649002

Analysis for AutoTheft

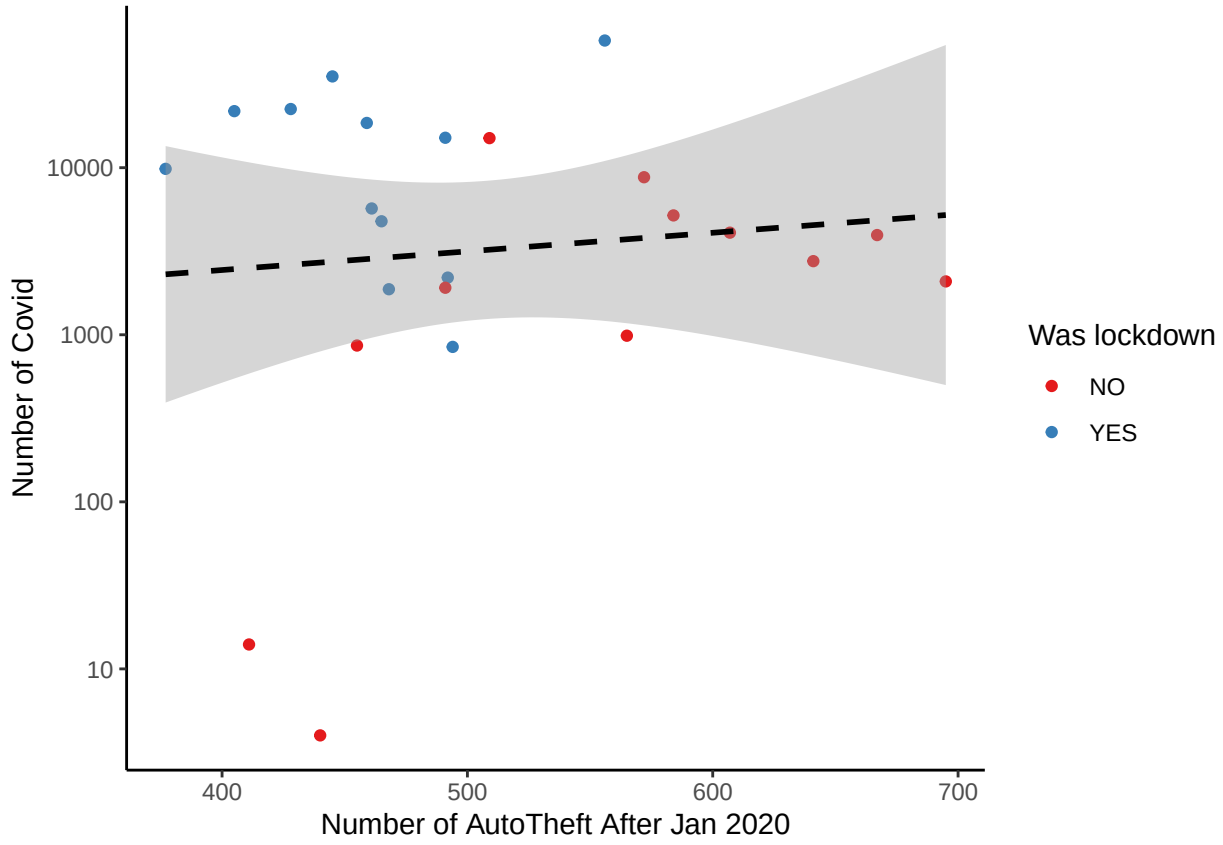

Figure 4: Multiple linear regression with Monthly AutoTheft Numer after Jan 2020 and related Covid Number, with a binary variable for whether Toronto was lockdown

From the figure4, we can see that Number of Covid and monthly AutoTheft number have a positive relationship. And there exist several points, especially the blue points in the graph, not inside the confidential interval of the prediction line, which means the lockdown policy does reduce the assault number in our research period.

Then, get the estimate and p-values for responses in linear fitted model 4(table7):

From the table 7, we can find that interactive interactive factor with $p = 0.652 > 0.05$, which means we fail to reject $H_0 : \beta_3 = 0$. Thus, there does not exist interaction relationship between CovidNumber and Lockdown. And we should use addition for multiple linear regression model rather than the model with interaction relationship.

Finally, by data from table 8, the equaion for the fitted model of AutoTheft(terms explain before):

$$AutoTheft = 549.3700 - 0.0010 * CovidNumber - 103.5692 * Lockdown + \epsilon$$

Table 7: Estimate and P-values for responses in linear fitted model 4

	Estimate	Standard.Error	Test.Statistic	P.value
(Intercept)	540.5396299	29.4466974	18.3565452	0.0000000
Covid	0.0032984	0.0052383	0.6296756	0.5360346
LockdownYES	-92.2044239	42.7384368	-2.1574122	0.0433159
Covid:LockdownYES	-0.0024772	0.0054105	-0.4578410	0.6520018

Table 8: Estimate for fitted model 5

	x
(Intercept)	549.3699704
Covid	0.0009765
LockdownYES	-103.5691631

Discussion

Findings

Overall, through the analysis conducted in this report, there are three general findings:

1. The auto theft number quickly increased during the global pandemic period caused by the local lockdown policy. But the serious situation of covid reduces the monthly assault number in Toronto.

2. There is a negative relationship between Covid-19 Padanmic and Toronto's monthly assault number - the increase in covid cases in the city of Toronto will reduce the local assault number. But, the lockdown policy will also cause a rapid increase in local monthly assault numbers.

3. There is a negative relationship between Covid-19 Padanmic and Toronto's monthly auto theft number - the increase in covid cases in the city of Toronto will reduce the local assault number. However, the lockdown policy will also cause a rapid increase in local monthly auto theft numbers.

Ethics & Implications

Government should foucs on the auto theft when apply lockdown policy

From Finding 1 and 3, we find that the lockdown policy has a huge influence on auto theft crime. In other words, the lockdown policy can be considered as the direct reason which causes the increasing auto theft case. The government and the police department should consider adding more patrolling police and traffic camera which can be used for tracking losing cars.

Racist Attacks, as part of Assault, need to be taken seriously

By external reports mentioned previously and the analysis for assault, I believe the lockdown policy cause negative feeling for people and indirectly cause discrimination. From Finding 2 and the annual crime rate graph, we can obvious that covid have a negative influence on the assault number. When compared with the overall decreasing trend for assault cases, the racist attack should be taken seriously as it is obviously increasing.

Limitations & Future Work

For the limitation of the model, we should include more data and factors as the current model's coefficient of determination is not high. We only contain 24 months, which can be considered as 24 data points, for evaluating the model. Narrowing the time interval may be a better idea. To be more specific, we can use weekly data as the data point.

For future research, we should improve the model and find an accurate linear relationship between the crime rate and the factors by collecting additional data points. More factors can be used for building the model, such as neighborhood characteristics, age demographics, and economic conditions.

Appendix

A. Datasheet

Extract of the questions from Gebru et al. (2021)

A.1 Motivation

Q: For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

A: In the same manner as previous surveys, it was designed to provide policymakers and managers of police Department with comprehensive information regarding Assault and AutoTheft.

Q: Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

A: By Toronto Police Department and OpenToronto database.

Q: Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.

A: The Government of Toronto funded all costs of the data collection and summary.

Q: Any other comment?

A: N/A

A.2 Composition

Q: What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.

A: The observations are constituted by recorded crimes in the city of Toronto. Crime are classified by different types including Assault, AutoTheft, BreakAndEnter, and other types.

Q: How many instances are there in total (of each type, if appropriate)?

A: The dataset record total 281,692 crimes from Jan 2014 to Dec 2021 in city of Toronto.

Q: Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).

A: The dataset contain all possible instances

Q: What data does each instance consist of? "Raw" data (e.g., unprocessed text or images) or features? In either case, please provide a description.

A: The data all the detail of the each crime, including location, reported date, crime type, and so on.

Q: Is there a label or target associated with each instance? If so, please provide a description.

A: Yes, the second column of the dataset clearly labelled and classified each instance by event_unique_id.

Q: Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.

A: No.

Q: Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.

A: There exists no explicit relationship between individual instances.

Q: Are there recommended data splits (e.g., training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.

A: There are no recommended data splits.

Q: Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.

A: There are no errors, sources of noise, or redundancies in the dataset presented.

Q: Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they existed at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.

A: The dataset is self-contained without the link to external sources.

Q: Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.

A: Yes, the data in the dataset of the individual level are all confidential and strictly applied to the privacy restrictions.

Q: Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.

A: No.

Q: Does the dataset identify any subpopulations (e.g., by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset. Is it possible to identify individuals (i.e., one or more natural per- sons), either directly or indirectly (i.e., in combination with other data) from the dataset? If so, please describe how.

A: No.

Q: Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.

A: No.

Q: Any other comments?

A: N/A.

A.3 Collection Process

Q: How was the data associated with each instance acquired? Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.

A: The data was collected by the police officer in city of Toronto, it can be considered as directly observable.

Q: What mechanisms or procedures were used to collect the data (e.g., hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?

A: The Toronto Police Department does not mention what mechanism or precedures are used during the data collection.

Q: If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?

A: N/A.

Q: Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?

A: Data were collected by the police officer in city of Toronto.

Q: Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.

A: The data collection began on January 2014 and was completed on December 2021.

Q: Were any ethical review processes conducted (e.g., by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.

A: N/A

Q: Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (e.g., websites)?

A: The data was accessed from the Toronto Public Safety Data Portal.

Q: Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.

A: It was unclear whether individuals are notified about the data collection.

Q: Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.

A: No information was provided by the dataset.

Q: If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).

A: N/A

Q: Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.

A: N/A

Q: Any other comments?

A: N/A

A.4 Preprocessing/cleaning/labelling

Q: Was any preprocessing/cleaning/labelling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.

A: Yes, we use only extract three crime types from the dataset, which are Assault, AutoTheft, and BreakAndEnter.

Q: Was the "raw" data saved in addition to the preprocessed/cleaned/labelled data (e.g., to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.

A:https://github.com/chle1999/Covid_Influence_Crime_Toronto/tree/main/inputs/data

Q: Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.

A: https://www.rstudio.com

Q: Any other comments?

A: N/A

A.5 Uses

Q: Has the dataset been used for any tasks already? If so, please provide a description.

A: No, the dataset haven't used for any tasks already.

Q: Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.

A: N/A

Q: What (other) tasks could the dataset be used for?

A: The dataset could potentially be used for comparisons - comparing changes in different crime types in multiple years.

Q: Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labelled that might im- pact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other risks or harms (e.g., legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?

A: The users need to be aware of how the participants were grouped in the dataset. When conducting analysis, be aware of which grouping method was selected and its impact on the analysis.

Q: Are there tasks for which the dataset should not be used? If so, please provide a description.

A: N/A

Q: Any other comments?

A: N/A

A.6 Distribution

Q: Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.

A: N/A

Q: How will the dataset will be distributed (e.g., tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?

A: available on website, it does not have DOI

Q: When will the dataset be distributed?

A: 2022

Q: Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.

A: N/A

Q: Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.

A: N/A

Q: Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.

A: N/A

Q: Any other comments?

A: N/A

A.7 Maintenance

Q: Who will be supporting/hosting/maintaining the dataset?

A: Toronto Public Safety Data Portal

Q: How can the owner/curator/manager of the dataset be contacted(e.g., email address)?

A: https://data.torontopolice.on.ca/datasets/TorontoPS::major-crime-indicators-1/about

Q: Is there an erratum? If so, please provide a link or other access point.

A: N/A

Q: Will the dataset be updated (e.g., to correct labelling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (e.g., mailing list, GitHub)?

A: N/A

Q: If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.

A: N/A

Q: Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers. N/A

Q: If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.

A: N/A

Q: Any other comments?

A: N/A

References

Balintec, Vanessa. 2022. "2 Years into the Pandemic, Anti-Asian Hate Is Still on the Rise in Canada, Report Shows." https://www.cbc.ca/news/canada/toronto/2-years-into-the-pandemic-anti-asian-hate-is-still-on-the-rise-in-canada-report-shows-1.6404034.

Bates, Douglas. 2022. *Lme4: Linear Mixed-Effects Models Using 'Eigen' and S4*. https://cran.r-project.org/web/packages/lme4/index.html.

City of Toronto. 2022. "COVID-19 Cases in Torontos." https://open.toronto.ca/dataset/covid-19-cases-in-toronto/.

D'ANDREA, AARON. 2021. "'It's Scary': Toronto Police Report 222." https://www.toronto.com/news/crime/it-s-scary-toronto-police-report-222-jump-in-stunt-driving-charges-during-covid-19/article_112780ba-0ebc-539e-b6db-e6283c0b4020.html.

Dunn, Trevor. 2021. "Give Toronto Police More Resources to Battle 'Unprecedented' Spike in Hate Crimes, City Council Says." https://www.cbc.ca/news/canada/toronto/give-toronto-police-more-resources-to-battle-unprecedented-spike-in-hate-crimes-city-council-says-1.6338685.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. "Datasheets for Datasets." *Communications of the ACM* 64 (12): 86–92.

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

Hothorn, Torsten. 2022. *Lmtest: Testing Linear Regression Models*. https://cran.r-project.org/web/packages/lmtest/index.html.

Kuznetsova, Alexandra. 2020. *lmerTest: Tests in Linear Mixed Effects Models*. https://cran.r-project.org/web/packages/lmerTest/index.html.

Lou, Nigel Mantou. 2022. "Expert q&a on Anti-Asian Racism During Pandemic." https://www.uvic.ca/news/topics/2022+expert-qa-anti-asian-racism+expert-advisory.

Moreau, Greg. 2021. "Police-Reported Crime Statistics in Canada, 2020." https://www150.statcan.gc.ca/n1/pub/85-002-x/2021001/article/00013-eng.htm.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://cran.r-project.org/web/packages/here/index.html.

Perepolkin, Dmytro. 2019. *Polite: Be Nice on the Web*. https://cran.r-project.org/web/packages/polite/index.html.

Perkel, Colin. 2021. "51." https://www.cbc.ca/news/canada/toronto/51-spike-in-hate-crimes-spurred-partly-by-pandemic-toronto-police-report-finds-1.5998703.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rakowska, Scarlett. 2022. "Analyzing the Change in Toronto Crime Data from COVID-19." https://ecce.esri.ca/uoft-blog/2022/01/06/analyzing-change-in-toronto-crime-from-covid-19/.

Robertson, Becky. 2022. "Car Thefts Way up Toronto." https://www.blogto.com/city/2022/03/car-thefts-way-up-toronto-year-why/.

Toronto Public Safety Data Portal. 2022. "Major Crime Indicators." https://data.torontopolice.on.ca/datasets/TorontoPS::major-crime-indicators-1/explore/.

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

———. 2021. *Rvest: Easily Harvest (Scrape) Web Pages*. https://cran.r-project.org/web/packages/rvest/index.html.

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.

Wickham, Hadley, and Jennifer Bryan. 2019. *Readxl: Read Excel Files*. https://CRAN.R-project.org/package=readxl.

Xie, Yihui. 2022. *Knitr: A General-Purpose Package for Dynamic Report Generation in r. R Package Version 1.38*.

Zhu, Hao. 2020. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.