Robbery Cases in Toronto in December 2021*

Relationship Between Occurrence Time And Location Types of Robbery

Shengyi Dai

27 April 2022

Abstract

This paper is about the robbery in December 2021, we are looking at the robbery in Toronto. In order to prevent crime happen in our world, analyzing the data and the region where crime occurs more frequently is a really important thing since the robbery may not only cause the victims to lose property, they also may get seriously injured. Even though we cannot prevent the future, the analysis can tell you where and when has a higher chance of a robbery that may happen. From the analysis, we used graphs and a logistic regression model, we found out that the out area and commercial areas are the places that are most likely to get robbed and the time of the robbery is likely to happen in the afternoon to midnight on weekdays.

Keywords: crime, robbery, toronto police, toronto residents

Contents

*Code and data are available at: https://github.com/celinadai/sta304-finall.git

1 Introduction

2021 had been a really tough year for people because of COVID-19. A lot of people are losing jobs or making less money. Crimes may occur more frequently since people may feel harder these days, robbery is one of the major crimes in the world. Hence in this paper, we will be looking at the robbery data in Toronto from 2021 to analyze the frequency of the robbery happening. Hence we can provide advice to the police on where and when do they need to focus to prevent the crime of robbery to protect our community.

This paper analyzed the relationship between the robbery in the occurrence time and the location type. The city we are mainly focusing on is Toronto. The whole analysis had been done using R studio (R Core Team 2020), which is a really helpful tool for analyzing and predicting data. It is important to analyze this for us to prevent crime and losing money from it and save people from danger. The way to find the data can be found and the structure of this project is in "README". From the analysis, we found that the open areas and business areas are the ones that have a higher chance that being robbed which is understandable by our common sense. Similarly, the weekday afternoon to midnight is also a higher chance time period that people got robbed.

The data is from the Toronto police services (*Robbery* 2022) which have a lot of open data about crimes, but this time we are just going to focus on the robbery. This paper is separated into a few different parts. Firstly is the data part which introduced the information about data, and the characteristics and relationships between the data. The next part is the method and results in part which introduced the method this paper will be using and the results I got from the method. The last part is the discussion which talks about the results I got and connecting to the world.

2 Data

2.1 Data Collection

This data is the robbery data from Toronto that was collected from each report party, and the collection process was started in 2014 and the recent update date is April 4th, 2022. The data is found on the website of the Toronto Police Service under open data. This data includes the occurrence dates, times, regions, offence types, etc. The accuracy of the dataset may not be that well since this data was preliminary while the reporting parties were sent to the Toronto police service, hence some of the data may not be fully validated, plus, this data also did not include the robbery that reportedly did not occur, nor was it attempted. This dataset is collected and published by the Toronto Police Service, so the dataset is unique and cannot find any similar dataset.

2.2 Data Characteristic

Before the cleaning process, there are 27820 observations and 30 variables. Since the observations are too large, and the variables are not only the ones that we are going to use for our analysis, hence we cleaned the dataset to 212 observations and six variables. In the data cleaning process, the package tidyverse (Wickham et al. 2019) had been used. Firstly, because we want to focus on the recent data, we choose the robbery time as December 2021. Then, to help our analysis, mutate the days of the weeks into weekdays and weekends, then created a new variable that changed the days of the week from one to seven which can help us in the next step. Also, the hours of each day had been changed into morning, afternoon, and night. The occurrence day was mutated into the beginning, middle, and end of the month to help us to centralize the data. Next, in order to make the dataset look nicer, we changed each name of the location type to a cleaner name. Finally, we selected the variables that are only relative to our analysis.

The variables' names and the explanation of each name can be found in the [Table: 1]. Also, in the [Table: 2], it shows the mean of the occurrence time of the day for robbery is 14.86321, the standard deviation is 6.238177, and the median is 16. From the result we got, we can see in the afternoon and night robbery is more likely to happen. Hence the police should be more focused while they are on the street in the afternoon and night, especially at night since the median is 16 which means there are half of the robbery cases happen at the time after 16:00.

2.3 Data Relationship

Graphics are always a great way to visualize the relationships between data. In order to use graphics in R, the package ggplot2 (Wickham 2016) had been used while writing the codes. There are several graphs we made to see the relationships between data. Firstly is the [Graph: 1], it is obvious, that robbery is more likely to occur in commercial places and outside, and the education places have a lower chance to have robbery happening. The commercial places include places like banks, bars, restaurants, convenience stores, gas stations, etc. The outside includes places like open areas, bus stops, parking lots, etc. From this graph, we know that we need to protect our belongings more carefully in these kinds of areas.

Next is [Graph: 2]. This graph is a graph of the count of cases that occurred in each location type, populated with a variable called "days". In this case, the variable "days" represent cases that occurred on weekdays or weekends. As you can see, the red bars are weekdays, there are far more red bars than the blue bars representing weekends, and it tells us that most robberies occur on weekdays. Places like pharmacies, banks, gas stations, and more have had no robberies over the weekend in December 2021. Streets and other commercial places are much more likely to be robbed on weekdays, but they also have a high chance to be robbed on weekends. Therefore, the protection of such places should be strengthened.

In the [Graph: 3], it is obvious that the cases of robbery are more likely to happen in the afternoon and night. The streets and apartments have a higher chance to get robbed at night. Other commercial places, pharmacies, and streets have a high chance of robbery occurring in the afternoon. The [Graph: 4] is a graph about the count of offences with occurrence days of the month filled in. The days of the month had been separated into beginning, middle, or end the month.

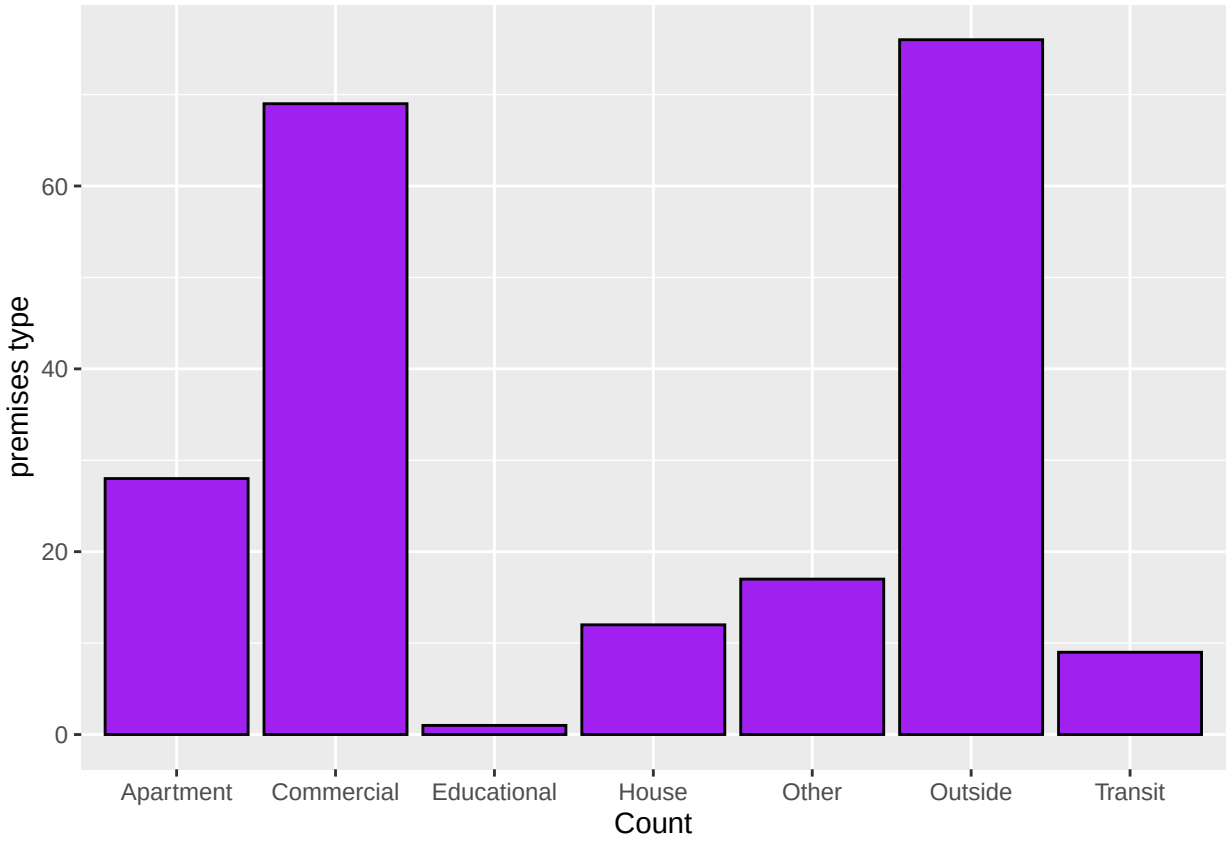

Figure 1: Count of robbery happens in different premises type

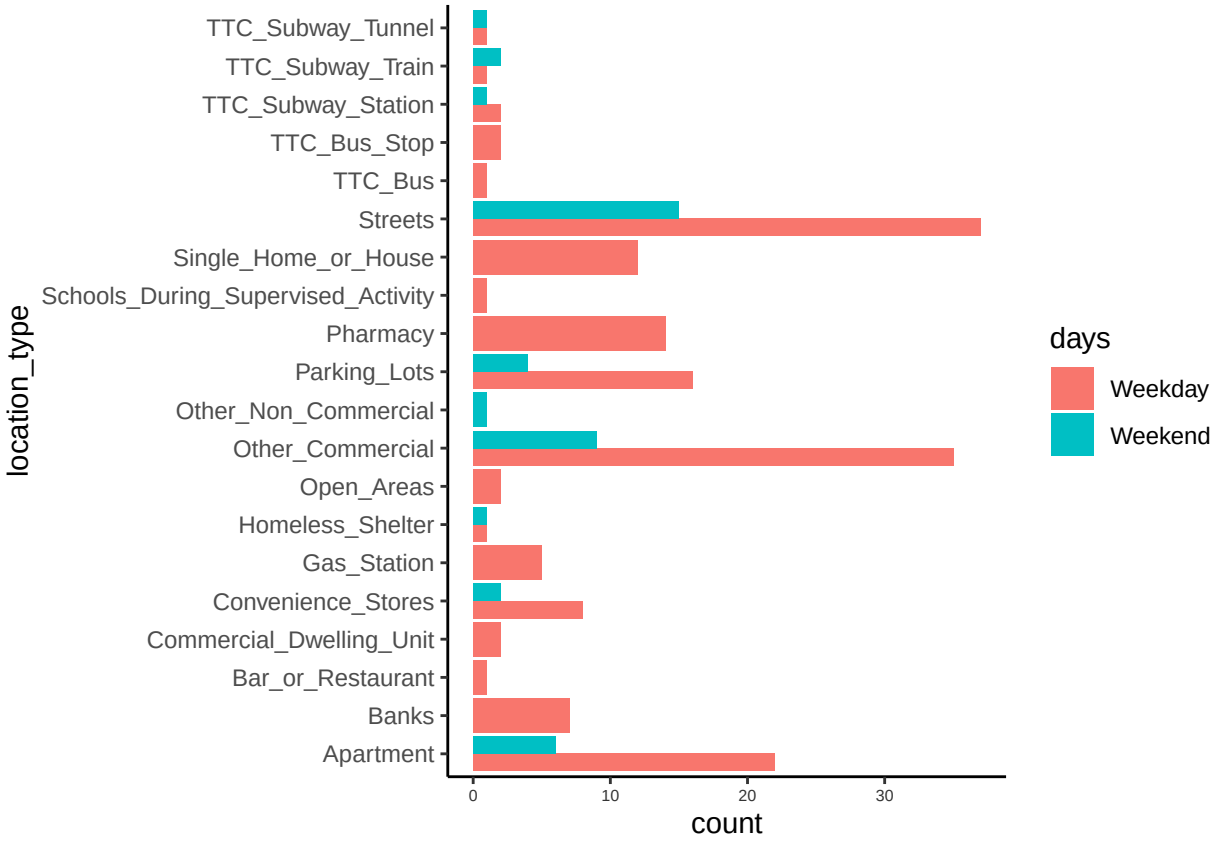

Figure 2: Location type with days filled in

Table 1: Variable Meaning

variable	Meaning
location type	The type of the location that had been robbed
offence	The offence type of this robbery
time	The robbery occurred in morning, afternoon, or night
occurrence hour	The hour of the date robbery occurred
occurrence day	The date of the month robbery occurred (beginning, middle, or end of the month)
premises type	The premises type of the place that robbery occured
occurrence day of week	The day of a week the robbery occured
days	Weekday or weekend the robbery occurred

Table 2: Summarize Table for Occurrence Day of Week

min	median	max	mean	sd
0	16	23	14.86321	6.238177

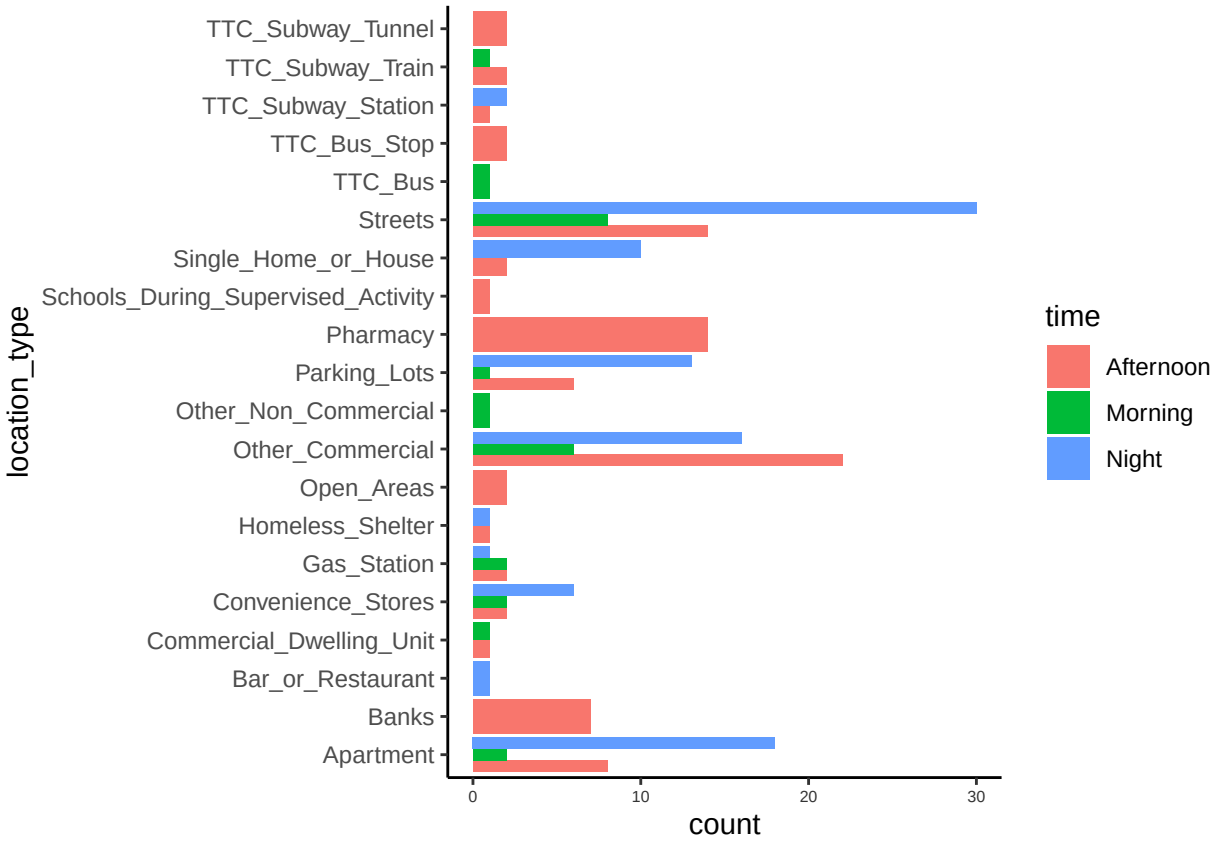

Figure 3: Graph of location type with time filled in

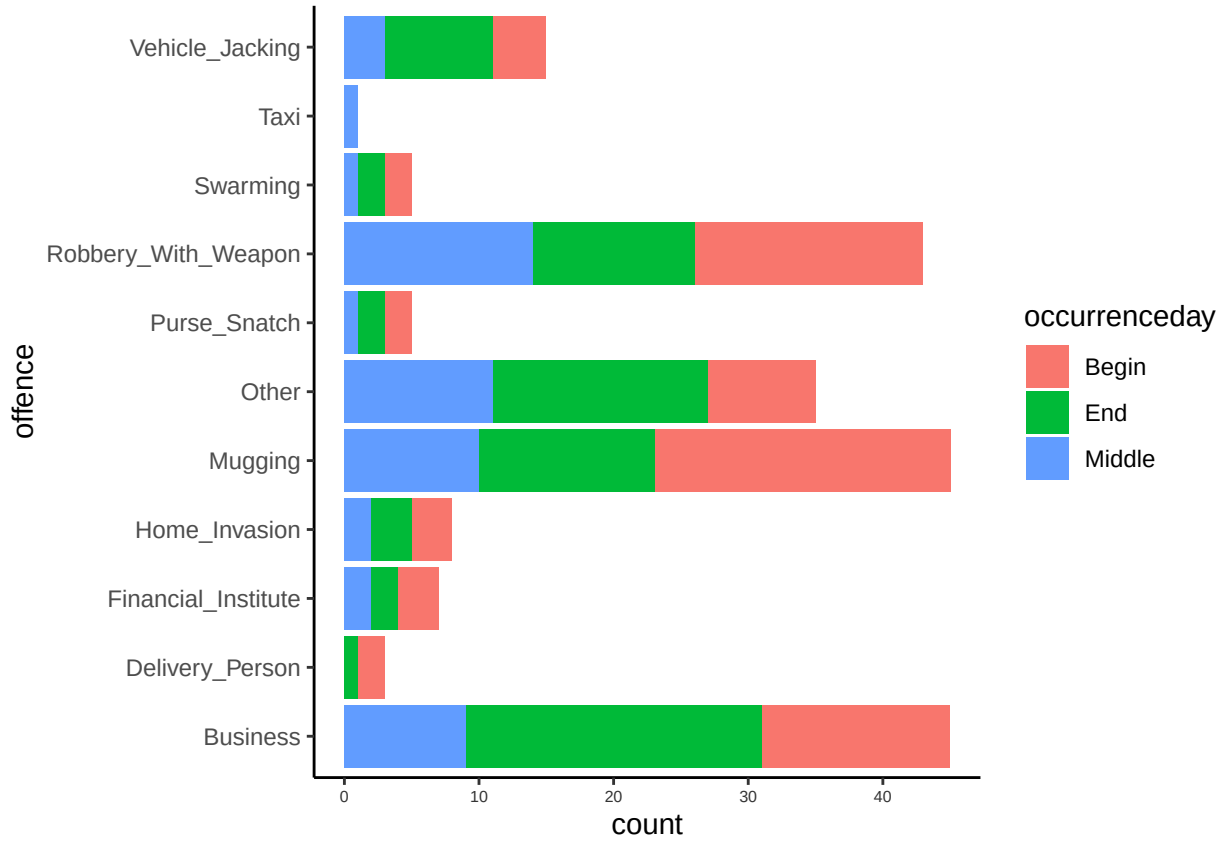

Figure 4: Graph of offence types with occurrence day filled in

3 Method

The model we are building is the logistic regression model. Logistic regression models are often used for predictive analysis and modelling. The reason for choosing this method is that it can help understand the relationship between variables that is independent and dependent, and it also can predict the likelihood of something happening or the likelihood of making a choice. There are some advantages of the logistic regression model, that is this model is not only a classification model, but it also gives probabilities. But, there are still some cons to the logistic regression model. Since the interpretation of the weights is multiplicative, not additive, the interpretation is more difficult.(Molnar 2022)

This model will estimate probabilities with the equation for logistic regression. The model generally looks like represented by $log(\frac{p}{1-p}) = \beta_0 + \beta_1 x$. In this model, "p" means the probability of the thing occurring, "β_0" and "β_1" are the regression coefficients. If the β_1 is positive, it means an increasing x will be interrelated with an increasing p. If the β_1 is negative, it means a decreasing x will be interrelated with a decreasing p. The independent variables can be multiple, which means there can exist $\beta_1, \beta_2, ..., \beta_n$. (Kassambara 2018) The R function "glm()" had been used here for a generalized linear model which is used to compute the logistic regression(Kuhn 2008).

Firstly, we set the seeds to 1000 and we initialize a pseudorandom number generator. Then randomly separated into a training dataset and testing dataset, where the test data contains 20% of the data, and the train data contains 80% of the data, and we will build a predictive model from the training dataset. In this case, the dependent variable is if the cases occur on weekdays or weekends, and the independent variable is occurrence hours. This means we are looking at the likelihood of how likely the cases will occur on weekdays or weekends with the different hours.

4 Results

After doing the calculation, the model looks like $log(\frac{p}{1-p}) = -1.176 - 0. - 152 x_{occurrencehour}$. The coefficient estimate of the occurrence hour is -0.152, an increase in an hour will be associated with a decreased probability of the case occurring on weekend. Hence, by the model we know, for an additional unit increase in occurrence hour, the odds of the cases happening on weekends change by 0.86 times. The model it is telling us that there is a negative relationship between the time getting later and the probability the cases will happen on weekend. The classification prediction accuracy is only 0.81 which is pretty good, the misclassification error is only 0.19.

Just like the table [3] shows, the estimate is β_0 and the estimated β coefficients are associated with each predictor variable. The std. error represents the standard error of the coefficient estimates, which is the accuracy of the coefficient. Since the standard error is small, then we are really confident about the estimate. The z value is the z-statistic, it is just the coefficient estimate divided by the standard error. The $\Pr(>|z|)$ is the p-value if it is really small, the estimate is significant.

As the table[3] shows, the standard error represents the accuracy of the coefficients, and since it is small, so we feel confident with the estimate we got. The $Pr(>|z|)$ is the p-value, the smaller the p-value, the more significant we think the estimate is.

As shown in {Graph: 5}, it is a left-skewed histogram, and the highest is about 18. Hence we can say that the hour of a day that the robbery is more likely happen is 18, and robbery is more likely to happen from afternoon to midnight. Also as shown in As it shows in {Graph: 6}, the weekdays are more likely to have robbery occurs, and especially on Thursday, this day has the highest amount of cases occur. These graphs are like the proof of the model we got.

Table 3: Estimate Table

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.1755758	0.4286834	-2.7422938	0.0061012
occurrencehour	-0.0151542	0.0270509	-0.5602125	0.5753345

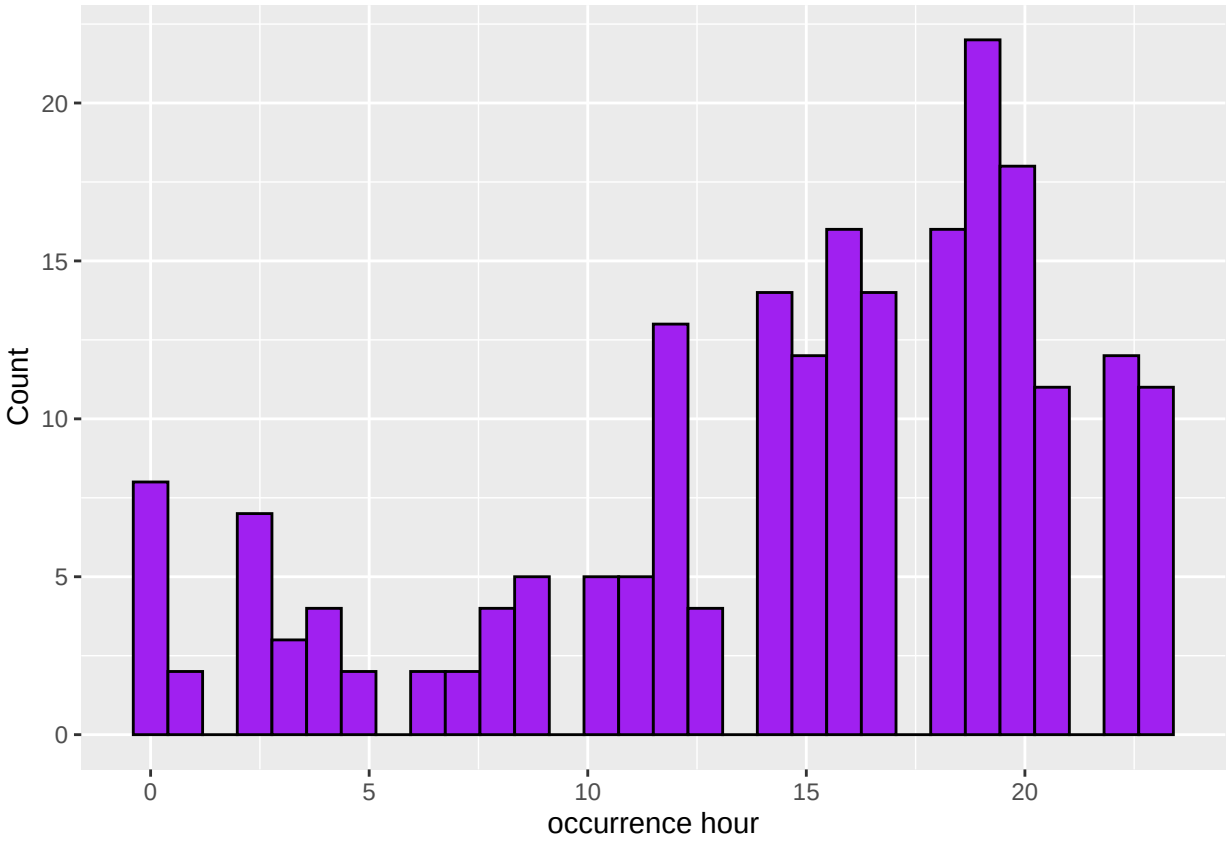

Figure 5: Occurrence hour of each day count graph

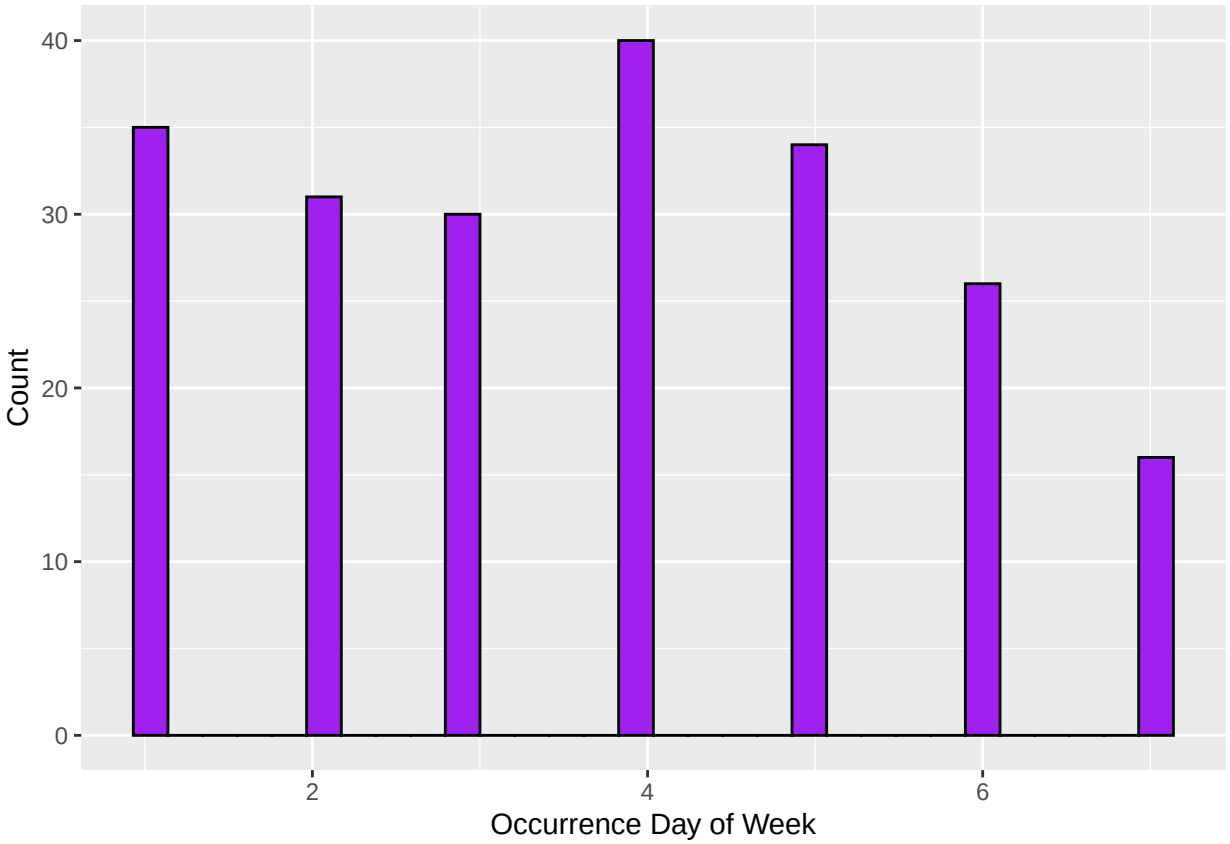

Figure 6: Occurrence day of week count graph

5 Discussion

5.1 About the paper

In this paper, the main purpose is to analyze the robbery cases that happened in Toronto in December 2021. This is important to study since robbery is a major crime, and it happens every day in the whole world, it not only targets people, it also targets businesses. These robbery victims don't just lose property, they were also in danger while they were robbed. We made several graphs and built a model to see where and when is more likely to have robbery cases occur.

From the graphs, we see that the robbery always happens on the weekday afternoon to midnight. In the early morning, there are fewer robbery cases that occur which is understandable by our common sense. Also, the outsides and the commercial places are the places that are easier to get robbed, and the educational place is less likely to get robbed.

From the model we made, $log(\frac{p}{1-p}) = -1.176 - 0. - 152x_{occurrencehour}$, we know that there is a negative relationship between the occurrence hour of the day and if the days of a week are going to be weekends or not. As the graphs had proved to us that in the results show, we can know that the robbery cases are more likely to happen while the hours of the day are growing and while it is on the weekday.

5.2 Connection to our life and the world

Robbery has always been a very serious thing in the world and it is also a major crime in the world. People lose money, get injured, and even lose their lives because of robberies. From 1976 to 1979, about 10 percent of criminal homicides in the U.S. were classified as robbery murders. In the U.S. in 2019, the average value of property stolen per person in reported robberies was 1,797 U.S dollars. Losses from the robbery were estimated at $482 million. ("Robbery" 2020)

In 2020, the robbery rate per 100,000 people in Canada is 50.65, which is really high. ("Incident-Based Crime Statistics, by Detailed Violations, Canada, Provinces, Territories and Census Metropolitan Areas" 2021) A car robbery occurred on April 24. Although the robbery was unsuccessful, the victim was stabbed three times. Fortunately, it was in front of the Robarts Library, and there were many people nearby, and the victim was rushed to the hospital in time. Something similar happened, but more serious because this time it happened in a place with not many people. It was in the car park of a flat in North York and nobody was there when the victim was beaten. This time it wasn't just property damage, the victim was even beaten with blood, possibly affecting one of his eyes. The first victim was a student at the University of Toronto, and the second was even a friend of my friend. From this, we see that crime is so close to us. It happens every day across the country, and even around the world. So many people are robbed every day, and so many people are injured every day.

5.3 Suggestion based on the result

We can give suggestions about the robbery to two groups of people, the victim or potential victim, and the policies that are catching the criminals for robbery or prevent the robbery happens. Hence in this section, we will write these in separate sections.

5.3.1 Suggestions for the polices

The first is to start with suggestions to the police. In this part, there are not a lot of suggestions we can give since we do not have the power to predict the future to see where the robbery will happen. But, from the previous part, commercial places such as banks and restaurants have always been places with a high probability of being robbed, also the open areas are also more likely to get robbed. I think the police should patrol these areas more often. In addition, the most common time period for robberies is from afternoon to midnight on weekdays. So, so for the police near the area, I would advise the police to increase their attention.

5.3.2 Suggestions for other people

We can give more advice to people that may be the victim. First of all, be wary of people who look suspicious to you, especially on weekday afternoons when you're in an open space or in a commercial area. Secondly, please keep calm if you are robbed, especially when the robber has a weapon in his hand, please cooperate with his request, which can reduce the chance of being injured.

In the process of being robbed, remember some of the robber's characteristics that will help the police find the robber and pay back your losses. When the robber leaves, be sure to go to a safe place as soon as possible and call 911 or go to the nearest police station. Life is always more important than property, so make sure your own safety is higher than your property when being robbed. ("What to Do During a Robbery," n.d.)

5.4 Weaknesses and next steps

Firstly, in order to protect the privacy of the victims, the cases had been moved to the nearest road intersection. This means the data may not be so accurate since it had been moved. There may be several cases that happen in different areas but close by was shown in this dataset as the same place. This gives us a low accuracy when we are looking at the region of the case that occurred, the neighbourhood may not reflect the exact number of cases that were reported. but this is understandable since the privacy of the victim is really important.

Second, the dataset does not contain events that are considered unfounded. According to the definition given by Statistics Canada, this means that a reported crime that did not occur or was attempted will not show up in the dataset. This is also a weakness of our chosen dataset, as the tried cases are also present. Although the robbery was unsuccessful in this place, that doesn't mean this place and this time aren't a chance for another robbery to happen.

Thirdly, the weakness of the report is the variables we choose. Since most of them are categorical variables, it is really hard to make the model. Also, since all of the variables were recorded as a fact, there is not really a dependent or independent variable, in this case, hence the model is even worse to make.

In the next step, since it is not possible for us to change the dataset. The only thing we can change is the variables we choose. There should be more numerical variables we choose to help our analysis, not only the count of these categorical variables.

Appendix

.1 Datasheet

Motivation

1. *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.*
   - This dataset was created for public understanding, use and application of police information. These datasets are subject to the Open Government Licence. All data are opened unless they are exempt for legal, privacy, security, confidentiality or commercially-sensitive reasons.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
   - The Toronto Police Service published this dataset.
3. *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.*
   - The government funded the creation since the Toronto Police Service is like one of the representation of the government.
4. *Any other comments?*
   - No

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*
   - The instances represent the documents. All the cases were reported as robbery.
2. *How many instances are there in total (of each type, if appropriate)?*
   - There is 11 kind of offences in total.
3. *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 (for example, 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 (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*
   - Yes. It contains all possible instances that can be shown to the public. It cannot represent a larger geographic coverage since other places may not be the same.
4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*
   - It includes the cases that were reported and confirmed including variables like occurrence day, month, year, reported day, month, year, neighbourhood, etc.
5. *Is there a label or target associated with each instance? If so, please provide a description.*
   - Each instance is about the robbery.
6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*
   - No, but in order to protect privacy, all the locations have been deliberately offset to the nearest road intersection.
7. *Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.*
   - No.
8. *Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.*
   - No.
9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*
   - In the dataset, the cases of robbery that were unsuccessful were not recorded.
10. *Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example,*

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 (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, 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.

- The dataset is self-contained since it is from the Toronto Police Service, but it also has data from other third parties such as 911 dispatch but we will not be given these data.

11. *Does the dataset contain data that might be considered confidential (for example, 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.*
    - No.
12. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*
    - No.
13. *Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*
    - No.
14. *Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.*
    - No.
15. *Does the dataset contain data that might be considered sensitive in any way (for example, 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.*
    - No.
16. *Any other comments?*
    - No

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, 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.*
    - Since this dataset is from the Toronto Police Service, hence the data are all got from the reported crimes which are just like directly observed and the police even went there to confirm.
2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?*
    - The 911 phone calls were used while the victims report data,
3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*
    - N/A
4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*
    - The polices, the 911 dispatchers, etc.. They worked in different ways to record and notice the data.
5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.*
    - It starts in 2014 the data had been recorded and published on the website. And the data were collected every day.
6. *Were any ethical review processes conducted (for example, 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.

- The information published on the website is already in a way to protect privacy by relocating the place to the nearest street intersection.

7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*
   - I collect the data from the Toronto Police Service open data.
8. *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.*
   - I think when the victims report the cases, they know the cases will be recorded.
9. *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.*
   - Still, the victims should agree with data collection since they reported the cases.
10. *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).*
    - No.
11. *Has an analysis of the potential impact of the dataset and its use on data subjects (for example, 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.*
    - No.
12. *Any other comments?*
    - No.

Preprocessing/cleaning/labeling

1. *Was any preprocessing/cleaning/labeling of the data done (for example, 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.*
   - No.
2. *Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.*
   - N/A
3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*
   - N/A
4. *Any other comments?*
   - No.

Uses

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

- It may be since it is open data and everyone can use it.

2. *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.*
   - No.
3. *What (other) tasks could the dataset be used for?*
   - To analyze, to check other cases, to see if the neighbourhood you are living in is safe or not, etc,
4. *Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact 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 (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to*

mitigate these risks or harms?
   - No.
5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*
   - Any tasks that do not follow the Open Government Licence.
6. *Any other comments?*
   - No

Distribution

1. *Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.*
   - No.
2. *How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*
   - The dataset is distributed on the website, but no DOI as I can find.
3. *When will the dataset be distributed?*
   - It is distributed the beginning of the year.
4. *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.*
   - No.
5. *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.*
   - No.
6. *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.*
   - No.
7. *Any other comments?*
   - No

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
   - The Toronto Police Service.
2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*
   - The address is Access & Privacy Section, Information Access Records Management Services, Operational Support Services, HQ – 4th Floor, 40 College Street, Toronto, ON, M5G2J3.
3. *Is there an erratum? If so, please provide a link or other access point.*
   - No.
4. *Will the dataset is updated (for example, 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 (for example, mailing list, GitHub)?*
   - Yes, it will be updated every year by the Toronto Police Service. People can find it on their website.
5. *If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, 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.*
   - In order to use this information, people agree to follow Open Government Licence. This includes to acknowledge the source of the information.
6. *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.*
   - Yes, the dataset just got updated on April 4th, 2022.

7. *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.*
    - It is not really possible to extend the dataset, the only way to do so is you report a real robbery case.
8. *Any other comments?*
    - No

References

"Incident-Based Crime Statistics, by Detailed Violations, Canada, Provinces, Territories and Census Metropolitan Areas." 2021. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3510017701&pickMembers%5B0%5D=1.1&pickMembers%5B1%5D=2.35&cubeTimeFrame.startYear=2020&cubeTimeFrame.endYear=2020&referencePeriods=20200101%2C20200101.

Kassambara. 2018. "Logistic Regression Essentials in R." *Articles - Classification Methods Essentials*. http://www.sthda.com/english/articles/36-classification-methods-essentials/151-logistic-regression-essentials-in-r/.

Kuhn, Max. 2008. "Building Predictive Models in R Using the Caret Package." *Journal of Statistical Software, Articles* 28 (5): 1–26. https://doi.org/10.18637/jss.v028.i05.

Molnar, Christoph. 2022. "Interpretable Machine Learning." https://christophm.github.io/interpretable-ml-book/logistic.html.

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

Robbery. 2022. Toronto Police Service. https://data.torontopolice.on.ca/datasets/TorontoPS::robbery-1/about.

"Robbery." 2020. https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/topic-pages/robbery.

"What to Do During a Robbery." n.d. https://dps.usc.edu/safety-tips/suspicious-activity/robbery/.

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.