EdX + University of Toronto School of Continuing Studies
Data Bootcamp

Project 1:
Toronto Major Crime Indicators
Result Write-up

Group 2:
Fanny Sigouin
Merve Celme
Tania Barrera

October 10th, 2023

Major findings

In general there has been an overall increase in crime in the city of Toronto year to year, although the pandemic years saw a significant reduction, the counts are back to pre-pandemic levels. Auto Theft did not see a reduction over the pandemic years, and it is strongly increasing as year increases. On the other hand, Robbery was already decreasing before the pandemic and it seemed to accelerate its decrease over 2020 and 2021. When we look at where crime is occurring, crimes in transit have been increasing since 2014 and crime in educational settings have decreased in recent years.

The number of crimes that occur daily changes according to season, with the colder months seeing lower daily crime and warmer months seeing higher daily crime. Crime is also significantly higher on the weekend, especially on Friday. When looking at how many crimes occur each hour, although there is variation by category, the general pattern follows a 12-hour cycle, with the peak being at 6pm and the lowest point being at 6am.

From 2014 to 2022, the average day difference between the offence date and the reported date has decreased over 9 days to less than 3 days. Offences occurring in January are consistently reported much later than offences occurring in other months. The period from June to September also sees an increase in the average day difference. In the last four years, there has been a 75% decrease in the day difference between the offence date and the report date.

The 10 neighbourhoods with the most offences over the 8-year period experienced a 12% decrease in the two years following the pandemic, but offences have been on the rise again since. Assault stands out as the main offence across those neighbourhoods, making up over 50% of offences. However, auto theft has been on the rise in recent years. Offences in these neighbourhoods mainly occur in commercial businesses and outdoor areas. The 10 neighbourhoods with the least offences have been on a strong upward trend since 2021, with an 18% increase in offences from 2021 to 2022. Assault remains the top offence in those neighbourhoods, but auto theft and break & enter make up a more significant percentage than in the top neighbourhoods (16% and 23% vs 12% and 18%). Offences in these neighbourhoods mainly occur in houses and apartments.

When looking at year-over-year (YoY) changes in neighbourhoods, we identified that Forest Hill South had the greatest YoY increase as offences rose by 94% from 2016 to 2017. This was mainly due to a 100% increase in assaults and 120% increase in break and enter. On the other hand, Highland Creek experienced the greatest YoY decrease from 2019 to 2020, with a decrease in offences of 52%. Offences across the board were reduced, but the main contributing factors were a reduction of 84% in robbery, 60% in break and enter, and 50% in assault.

Assault stands out as having a very strong positive correlation with outside premises (Pearson r = 0.95). As well, auto theft has a very strong positive correlation with houses (Pearson r = 0.9). Through a correlation matrix, we also identified that auto theft and theft over $5,000 are

positively correlated (Pearson r = 0.74), while assault and auto theft are negatively correlated (Pearson r = 0.49).

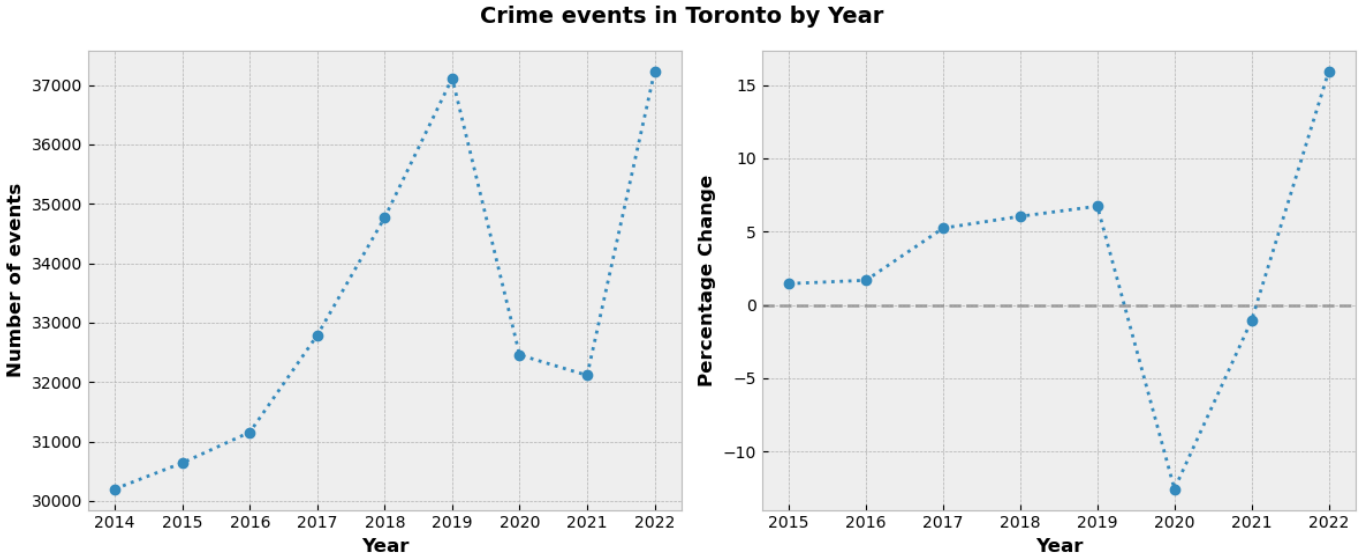
Question 1: Year to year, has there been an overall increase or reduction in crime? Has a certain type of offence seen a significant decrease or increase?

To answer this question, we counted the number of crime occurrences per year, both in total, by crime category and by premises type (i.e. where the offences happened). In addition, for the total occurrences we calculated the percentage change year over year, to visualise how this change has increased or decreased over time. We also calculated the correlation coefficient of these counts with year, to see the relationship between increasing the year and how the counts change.

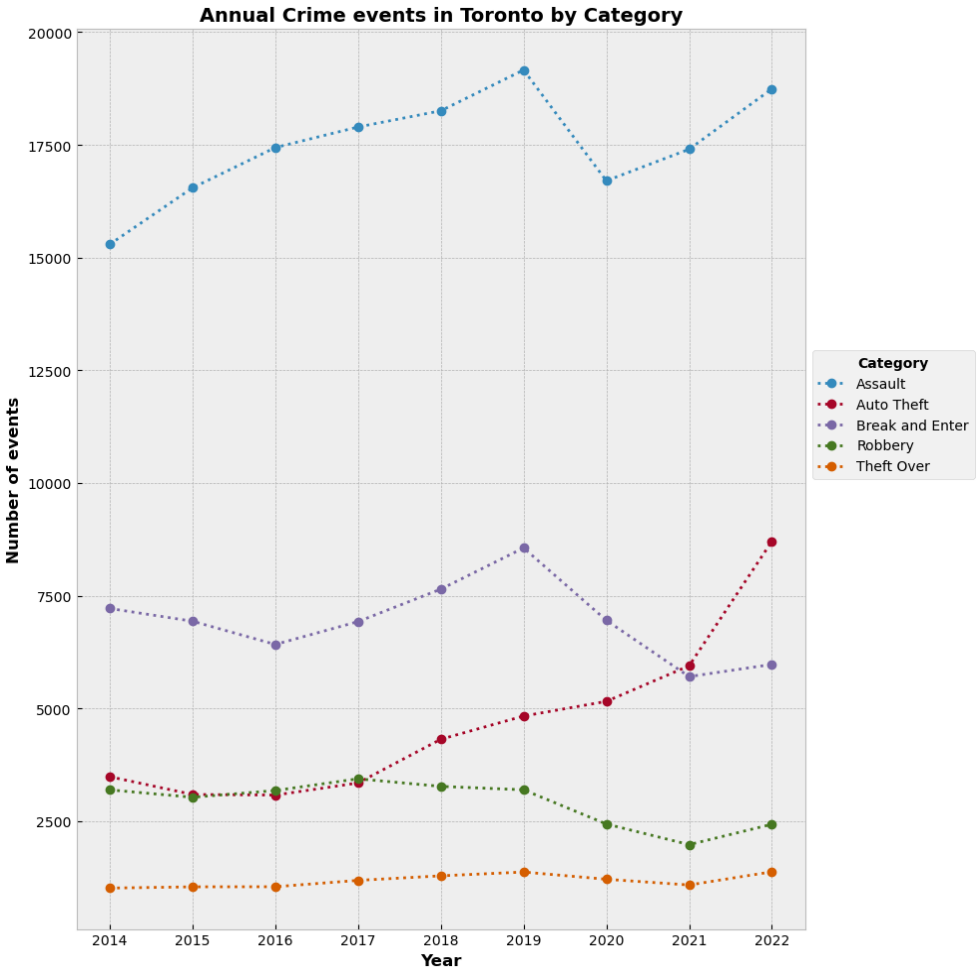
Total crime occurrences

In general, crime events were increasing close to linearly in the years between 2014 and 2016, evidenced by the flat line between 2015 and 2016 that shows that the percentage change year over year remained constant. However, from 2017 to 2019 crime occurrences were increasing exponentially, as we can see both in the counts and in the linear increase for that period in the percentage change plot. The COVID-19 pandemic interrupted this trend and brought a sharp decline in crime occurrences for the year of 2020, a number that was even lower in 2021. The last year in our data shows a great increase compared to the previous year, but the count is just slightly higher than the 2019 number.

The correlation coefficient between total annual crime occurrences and year is 0.68, showing a moderate correlation that was probably weakened by the two years of great decrease, 2020 and 2021.



By crime category



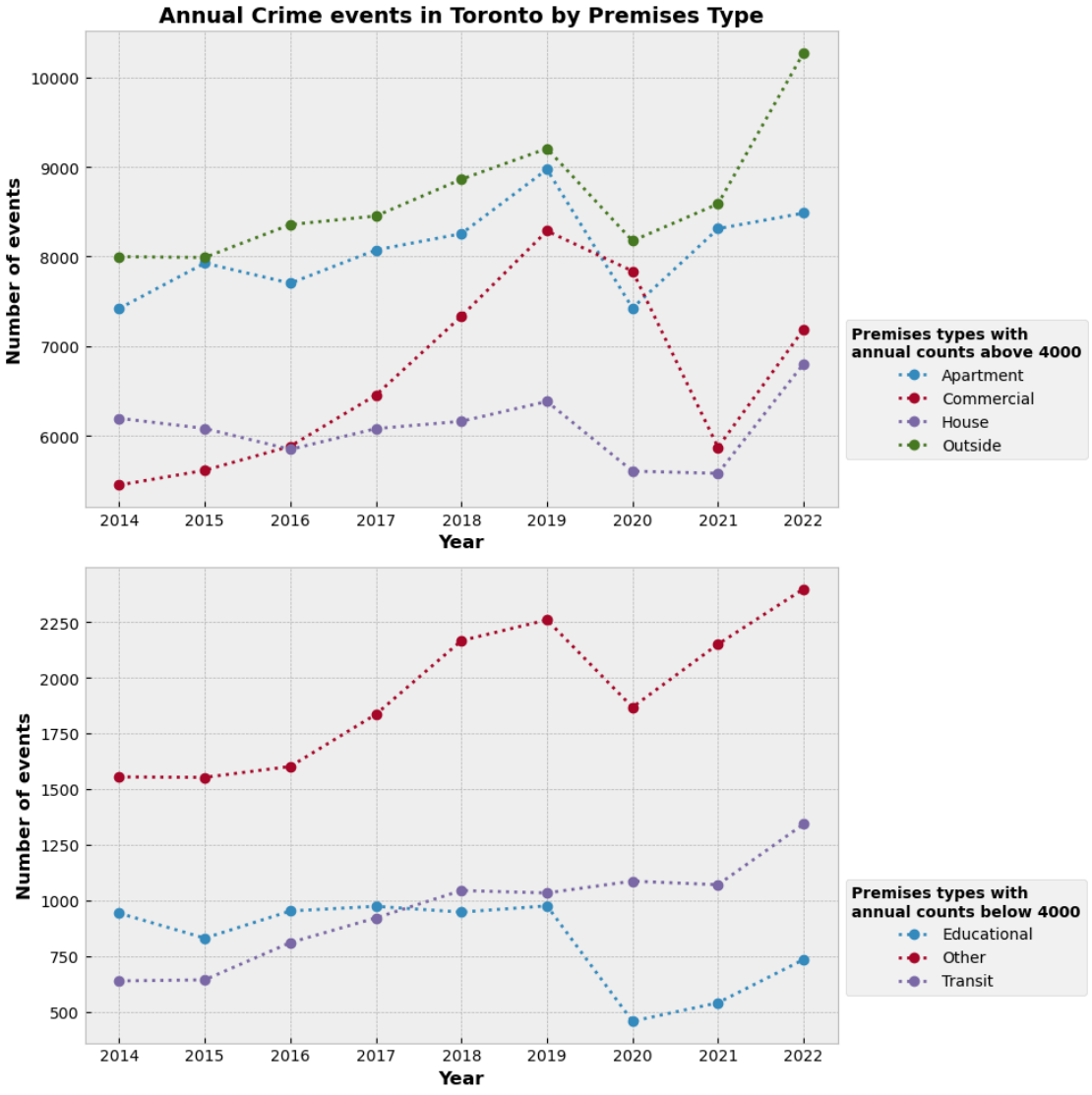
When grouping crime counts by Major Crime Indicator category, we see that different categories follow different patterns over the year. Assault is the most common crime event in Toronto, while Theft Over 5000 CAD (Theft Over hereafter) is the least common.

Auto Theft count has a strong positive correlation with year, 0.88, putting it in first place for how it increases as year increases. This can be observed in the plot of annual crime events by category, where we see that Auto Theft numbers have increased every year since 2014, even during the pandemic, with the highest jump being recent, between 2021 and 2022.

Assault and Theft Over counts both have positive moderate correlations with year (0.62 and 0.67, respectively). Assault has increased every year since 2014, with the exception of a decrease in 2020 that brought it close to 2015 levels. Since then, Assault has continued to increase over 2021 and 2022, with the later year approaching the levels that were seen right before the pandemic (i.e. 2019). A very similar pattern can be seen for Theft Over, although because the counts are much smaller it is not as noticeable in the plot.

Break and Enter count has a weak negative correlation with year (-0.32), while Robbery count has a strong negative correlation with year (-0.73), which means that as year increases, the number of events per year decreases. The annual count of break and enter events was declining from 2014 to 2016, then increasing up to 2019 and then sharply decreasing for 2020 and 2021. The 2022 break and enter count is slightly higher than 2022, but the increase is not at the same level as the 2016 to 2019 period. The number of annual robbery events was slightly decreasing year by year from 2014 to 2019. In 2020, robbery had a larger decrease. Although it declined again in 2021, we can see that the 2022 level increased compared to the year before, reaching similar levels to 2020.

By premises type

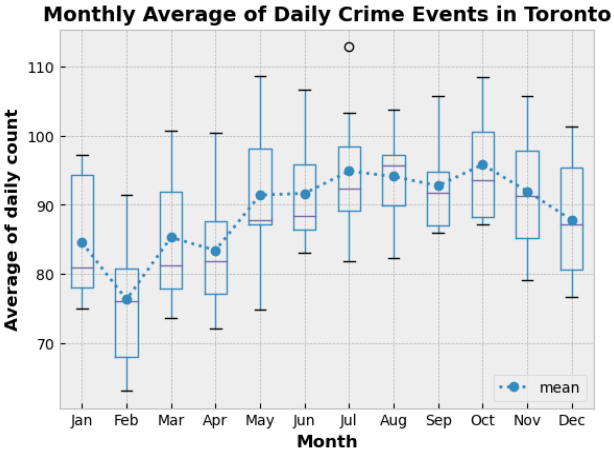


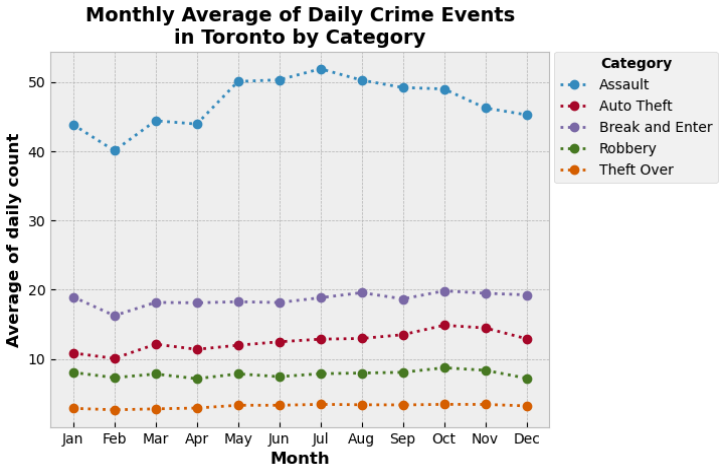
The visualisation by premises type was divided between those types with counts above 4000 and those with counts below 4000, so the patterns could be observed more clearly.

The ones above 4000 are: Apartment, Commercial, House and Outside. They all have positive correlations with year, although the ones for Apartment and Commercial are moderate (0.52 and 0.59, respectively), the one for House is weak (0.09) and the one for Outside is strong (0.71). In general we see that they were all increasing each year until 2020, when they decreased. Then in 2021 they either started increasing again (Apartment and Outside), remained at the same level as 2020 (House) or decreased further (Commercial). In 2022, all four categories increased.

The categories with counts below 4000 are Educational, Transit and Other. The counts of the Educational category are moderately negatively correlated with year (-0.62), probably due to them being relatively stable until the sharp decrease seen in 2020. The Transit category has a very strong positive correlation with year (0.96) and it increases as year increases. Finally the Other premises type also has a strong positive correlation (0.86), although it did see a decrease for the year of 2020, every other year has seen an increase in the count of crime occurrences.

Question 2: What are the peak times for crime occurrences? Does it change according to season? Is there a concentration of crime around a specific period of the year?

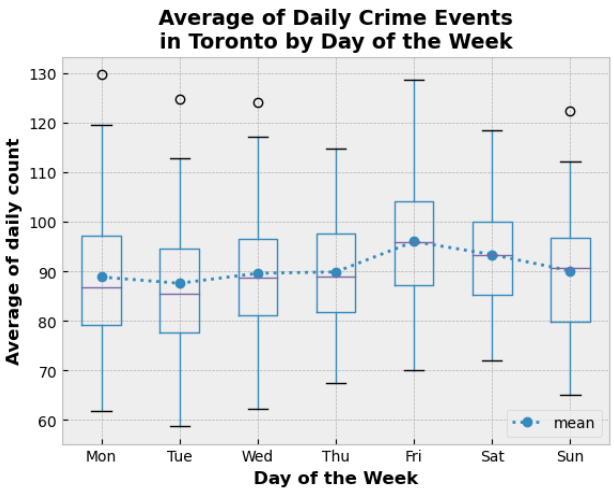


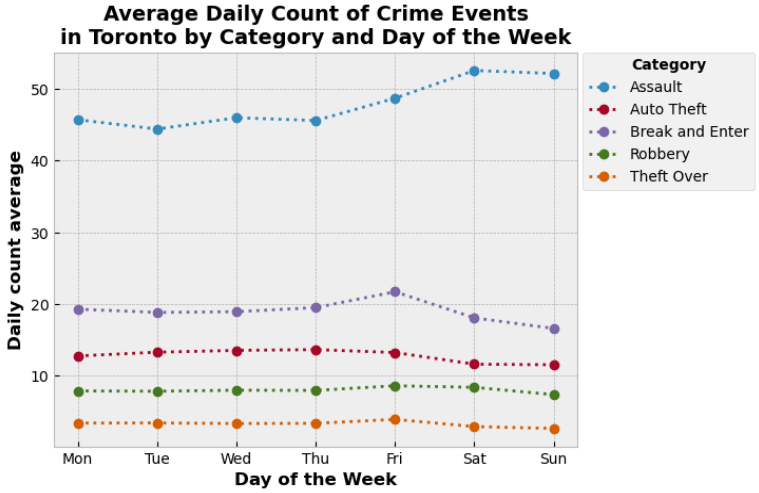


Month and Day of the Week

To answer this second question, we calculated the daily count of crime occurrences for all the dates, and then we grouped these counts by either their month of occurrence or the day of the week. These groups were then used in two separate one-way ANOVAs to determine if month or day of the week have an effect on daily crime counts. Both have a significant effect (Month ($p=9.46 \times 10^{-5}$) and day of the week $p=5.17 \times 10^{-7}$). For the visualisations, we calculated the average for each group and plotted that, along with the boxplots of the individual counts by month and day of the week because it made it clearer to appreciate the pattern. In general, the winter months have a lower daily crime count, compared to warmer months, and crime count is higher on the weekend, mainly on Friday. We also visualised the average daily counts by category. In these visualisations by crime category, it is again evident that assault is the most common type of crime, and it seems that it also has the highest variability by month, probably driving the seasonal changes observed in the total counts. However, in the day of the week plot

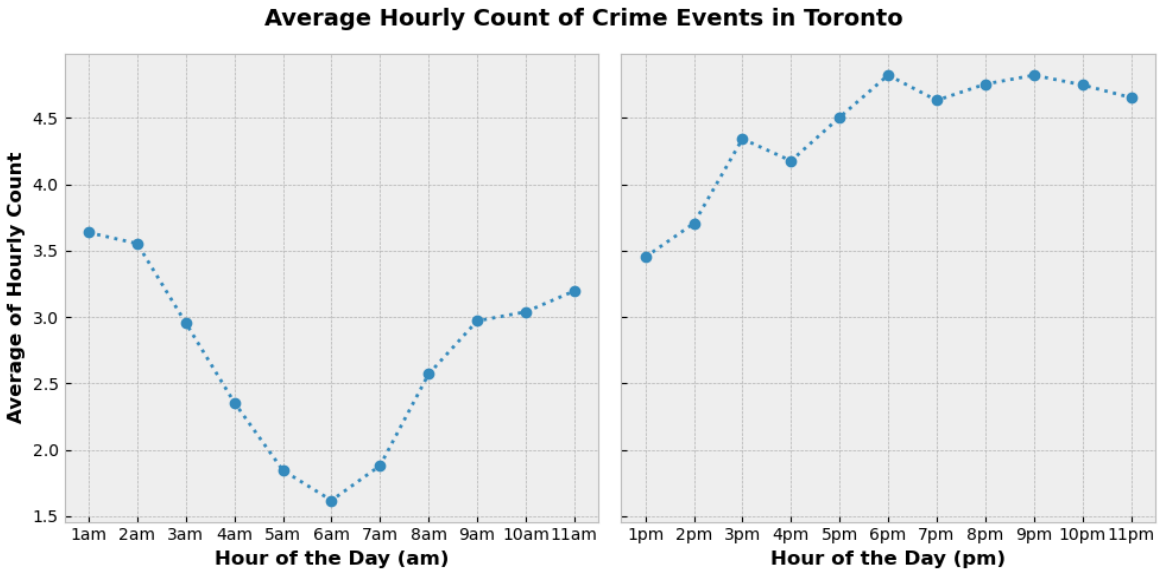
by category, we can see that although Assault has an increase on Friday, it is highest on Saturday and Sundays. The Break and Enter category is the one that has an evident highest count on Fridays, which coupled to the variability seen in Assault, is what results in the highest crime occurrences being on Fridays.

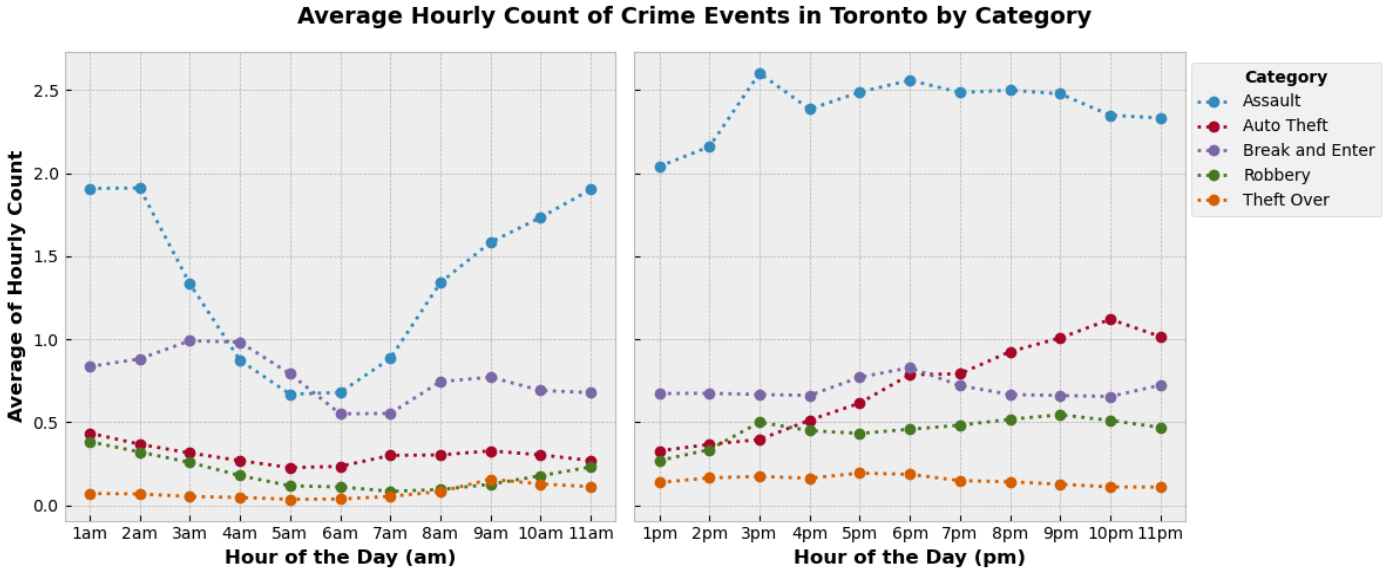




Hour of the Day

We also calculated how many crime events occurred each hour on each date and then calculated the average of this hourly count to visualise the changes over a day. The general pattern follows a 12-hour cycle, with the peak being at 6pm and the lowest point being at 6am. However, when we separate the hourly count by crime category, we see that each category has different fluctuations. Assault is highest at 3pm and remains high until 9pm. Auto Theft peaks at 10pm. Break and Enter peaks at 3-4am, then goes down for 6-7am and remains at a constant mid level during the day with another slight increase at 6pm towards that early morning peak. Robbery is highest in the afternoon and evening, between 3pm and 9pm, and Theft Over is highest during the work hour between 9am and 6pm.



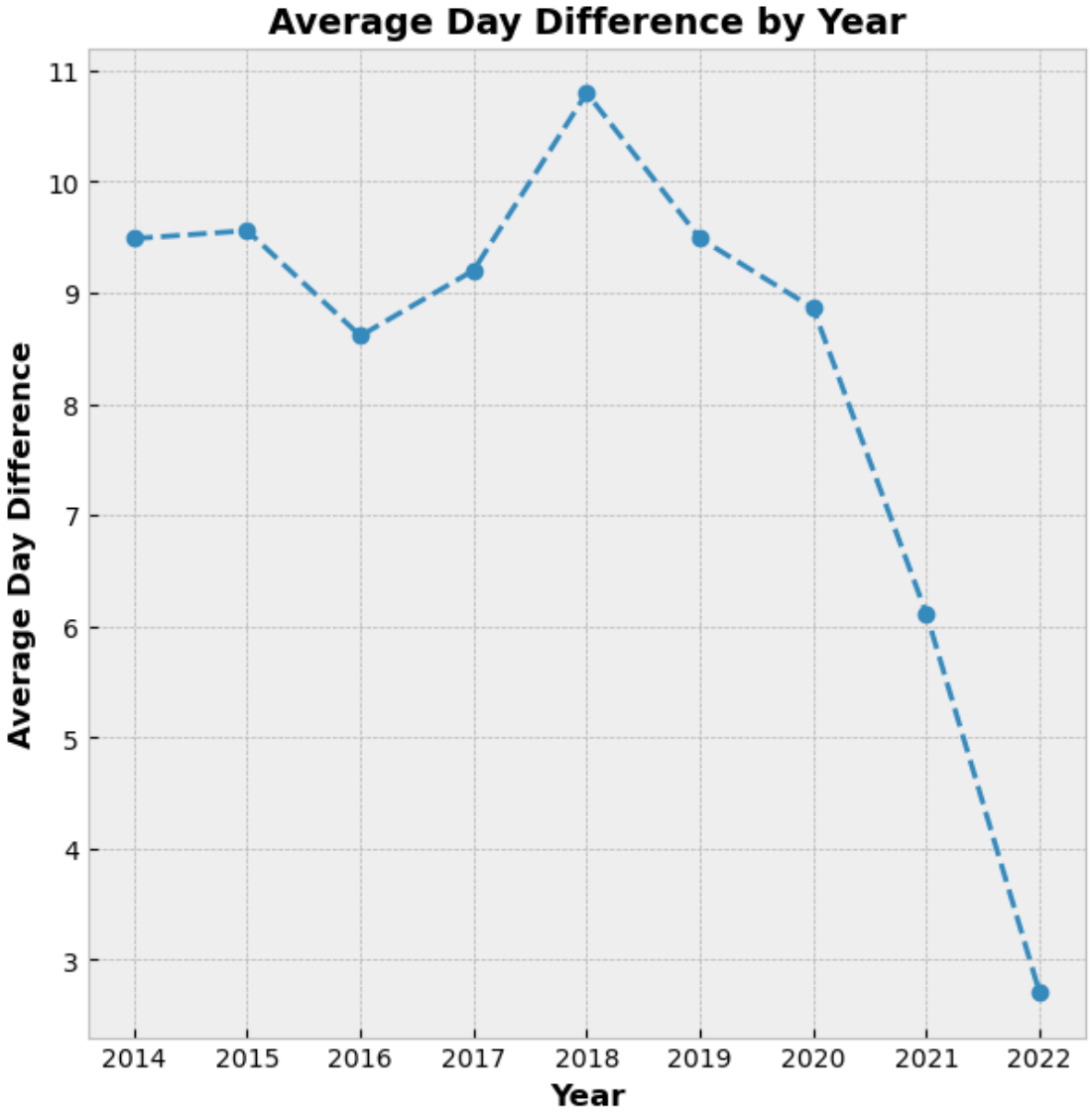


Question 3: On average, how long after the offence date was the offence reported?

We analysed the average day difference between offence dates and reporting dates for 2014 to 2022. We looked into each year as a difference in hours and days and found similar results which show a very high difference in days for January. There is a similar pattern from June to September, where there is an increased day difference in reporting offences each year.

Between 2014 and 2016, there is a decrease, followed by a greater increase between 2016 and 2018. The delay peaks in 2018. Since then, along with the pandemic, the difference between

offence and reporting hits the bottom and reaches an average of 2.7 days whereas it was nearly 11 days in 2018. In the last four years there was a 75 percent decrease.



Question 4: Can we identify trends or patterns in specific neighbourhoods?

As there are over 150 neighbourhoods in the city of Toronto, we narrowed our scope to the 10 neighbourhoods with the most and least offences from 2014 to 2022. We also looked at year-over-year percent changes in offences and analysed trends in the top 10 and bottom 10 neighbourhoods.

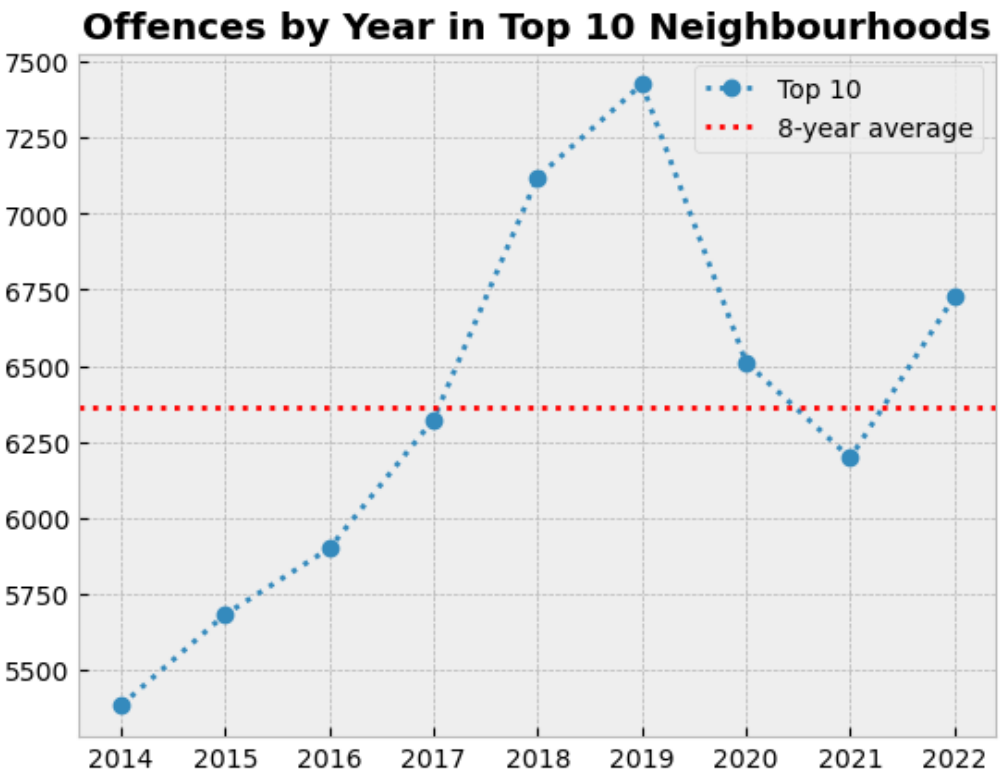
Offences by year

To plot offences by year in the top and bottom neighbourhoods, we separated the top and bottom 10 neighbourhoods from the original dataset and got the sum of their respective offences

each year. We then plotted the offences by year on a line graph, with a horizontal line representing the 8-year average for the neighbourhoods.

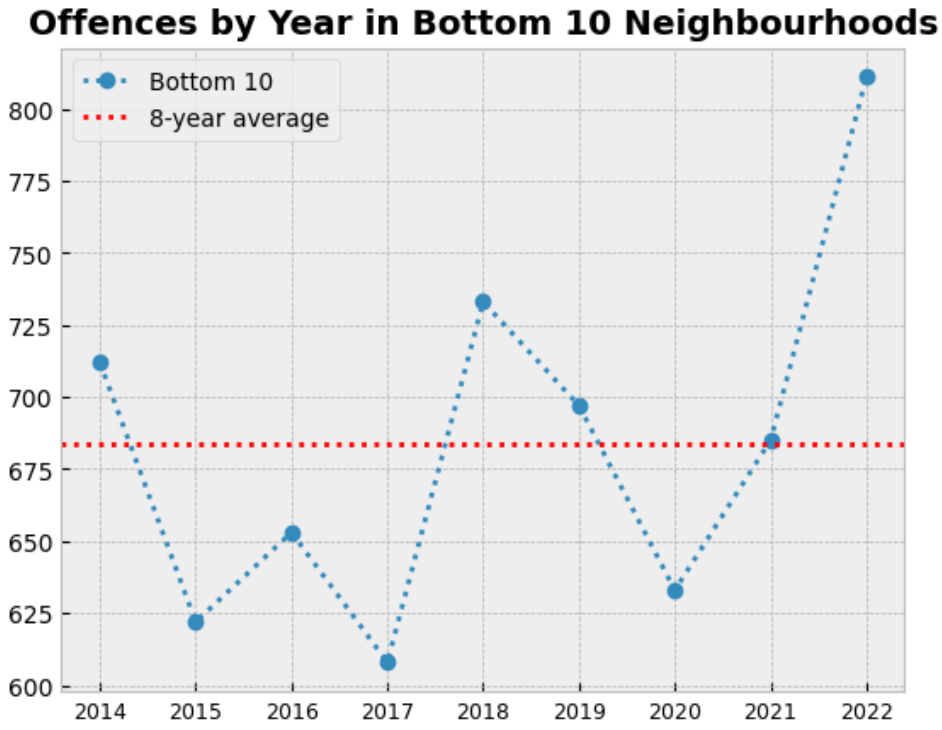
Top 10 neighbourhoods

Offences in the top neighbourhoods were consistently increasing year over year until 2020, where they decreased by 12%. The decline continued in 2021 as offences dipped below their 8-year average. However, offences have been on the rise again since and have increased by 8% from 2021 to 2022, indicating that these neighbourhoods are returning to their previous pre-pandemic trend of year-over-year increases.



Bottom 10 neighbourhoods

Offences in the bottom neighbourhoods were less consistent in the years prior to the pandemic. They fluctuated yearly from 2014 to 2018, experiencing a spike of 20% from 2017 to 2018. After a decline of 13% in the following two years, offences have reached their 8-year peak in 2022 with an 18% increase over 2021.

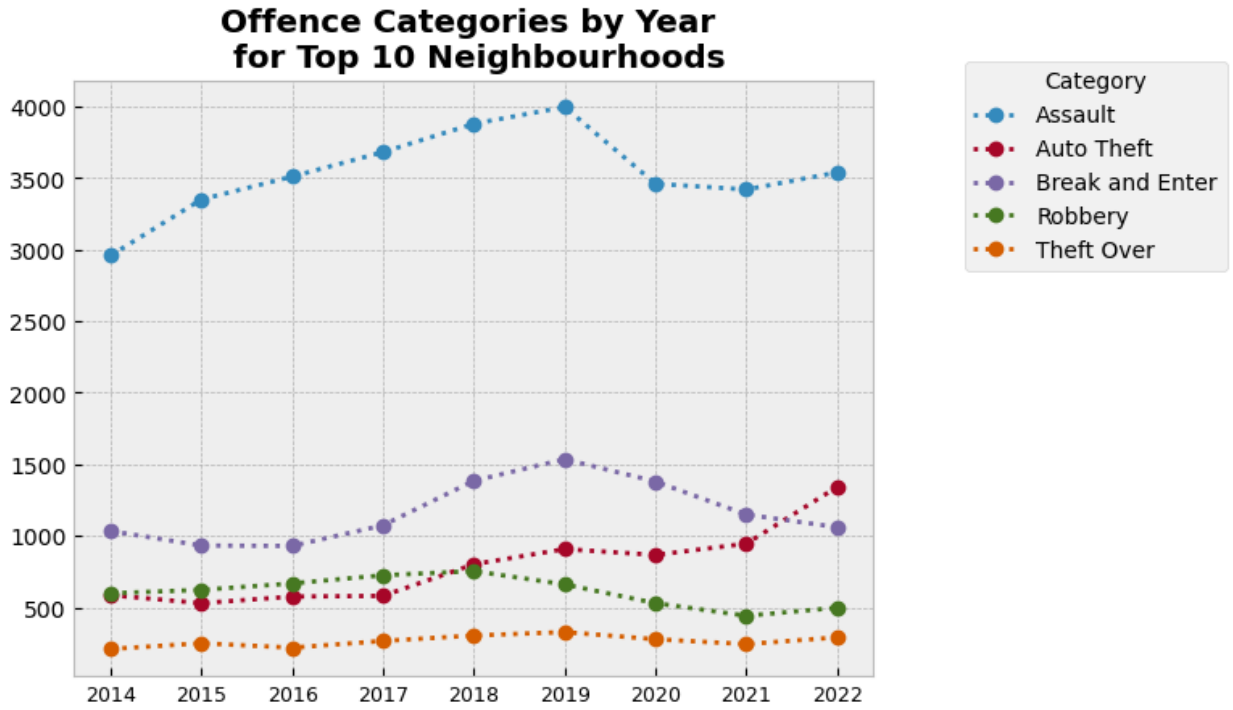


Offence categories by year

Next, we wanted to break down offences by category and year in these top and bottom neighbourhoods, with the goal of identifying trends with specific offences.

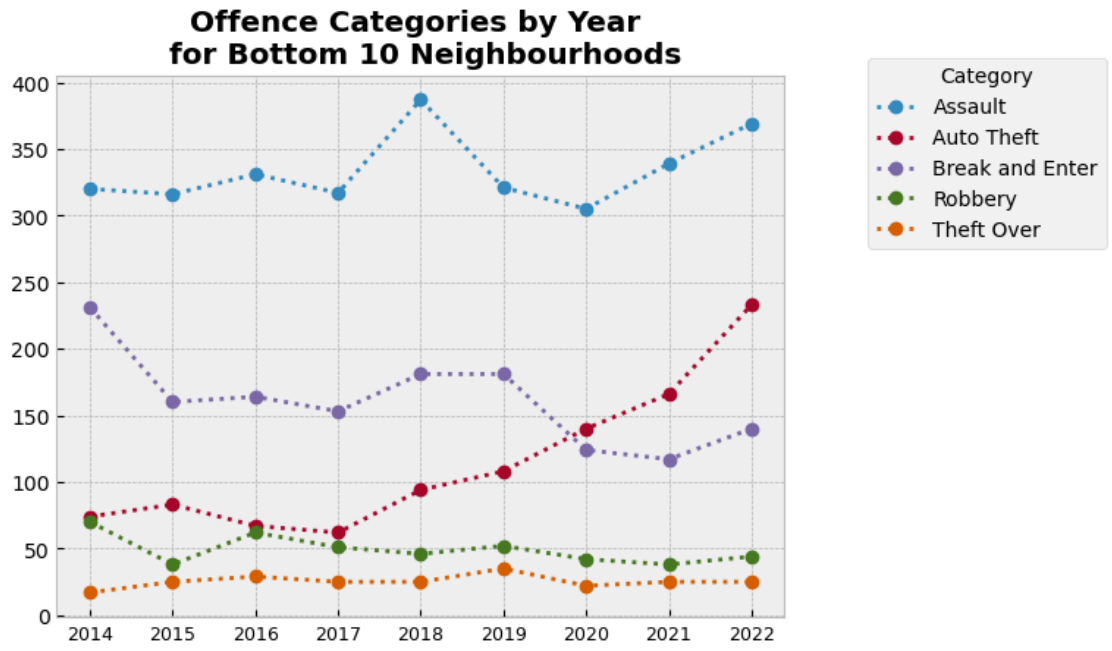
Top 10 neighbourhoods

Assault in the top neighbourhoods, which accounts for 55% of all offences, was on an upward trend until 2019. It has since declined slightly but remains the top offence category. Auto theft has been on the rise while break and enter has been declining in the last two years, but other offences have remained relatively consistent.



Bottom 10 neighbourhoods

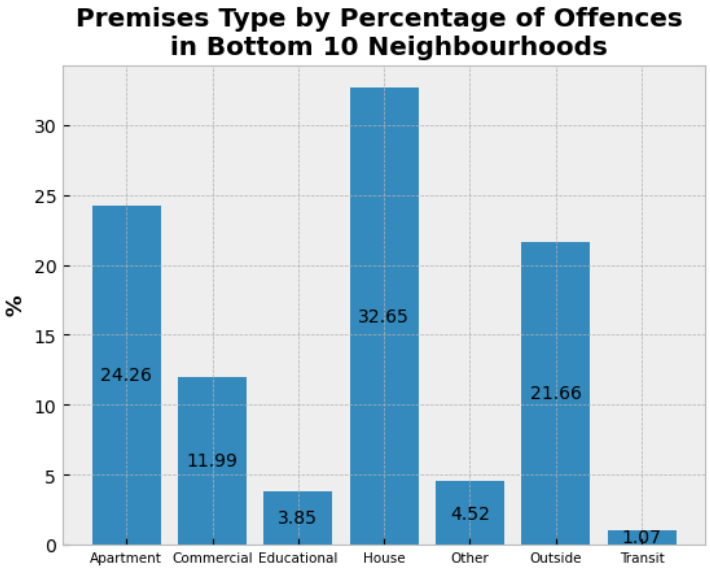
While assault is still the top offence in the bottom neighbourhoods, making up 48% of offences, it has been much less consistent than in the top neighbourhoods. Nonetheless, it has been steadily increasing since 2020. Auto theft has also been on a strong upward trend since 2017, while break and enter has been declining. Robbery and theft over $5,000 have remained consistent with little change year over year.

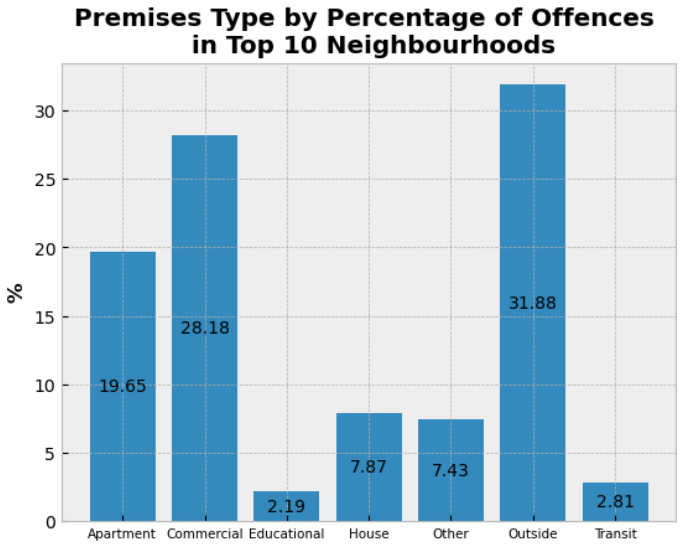


Premises type

Top 10 neighbourhoods

When grouping offences by premises type, we identified that offences in the top neighbourhoods mostly occur outside (32% of offences) and in commercial places (29% of offences). Transit and educational premises make up less than 3% of premises type respectively.





Bottom 10 neighbourhoods

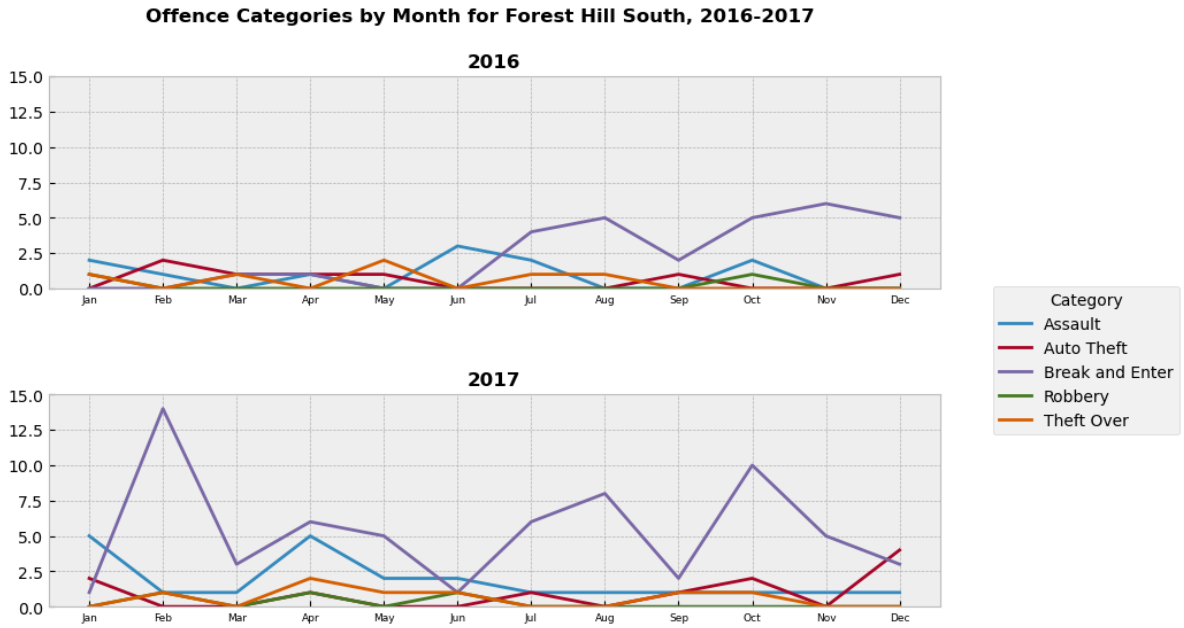
In the bottom neighbourhoods, we identified that offences mostly occur in houses (32.6%) and apartments (24.3%). Offences occurring on educational and transit premises also make up a negligible percentage of offences, with 3.8% and 1% respectively.

Year-over-year changes

In identifying trends and patterns in neighbourhoods, we also wanted to analyse year-over-year changes in neighbourhoods. To do so, we grouped the dataset by neighbourhood and by occurrence year. We could then calculate the percent change from year to year.

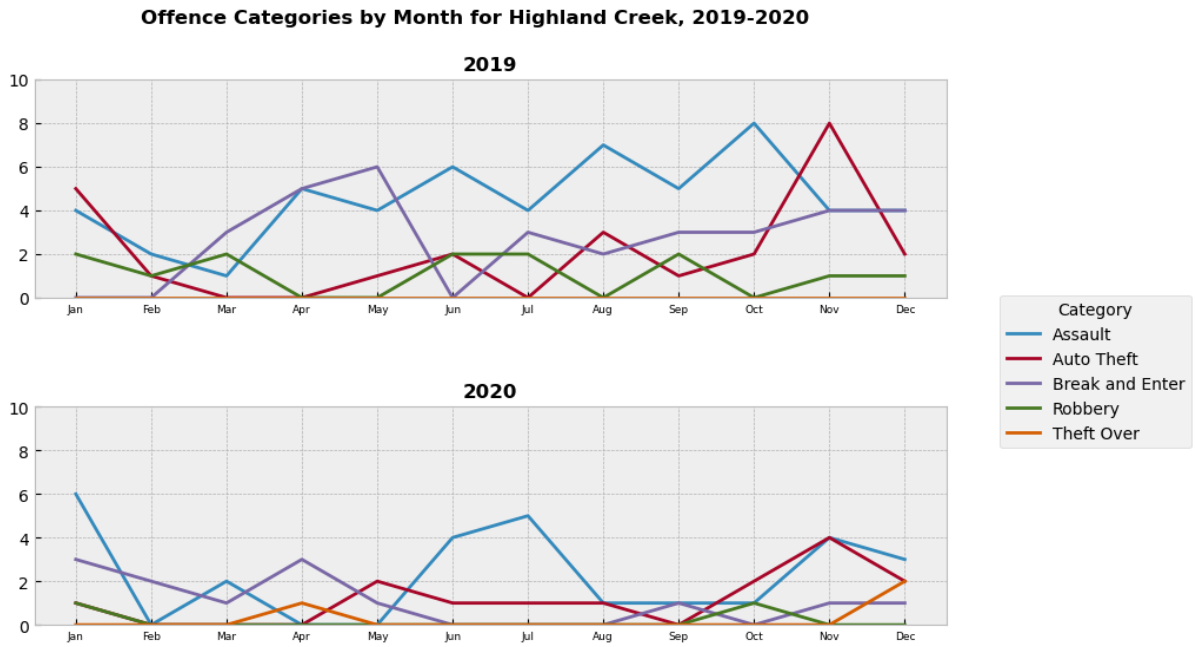
Greatest increase

Forest Hill South experienced the greatest year-over-year increase from 2016 to 2017, with an increase of 94%. When comparing offence categories by month in those two years, we can identify that the main contributing factors were an increase of 120% in break & enter and an increase of 100% in assault.



Greatest decrease

Highland Creek had the greatest decrease year-over-year from 2019 to 2020, with a decrease of 52%. Contributing factors were a 60% reduction in break & enter, a 50% reduction in assault, a 44% reduction in auto theft, and an 84% reduction in robbery.

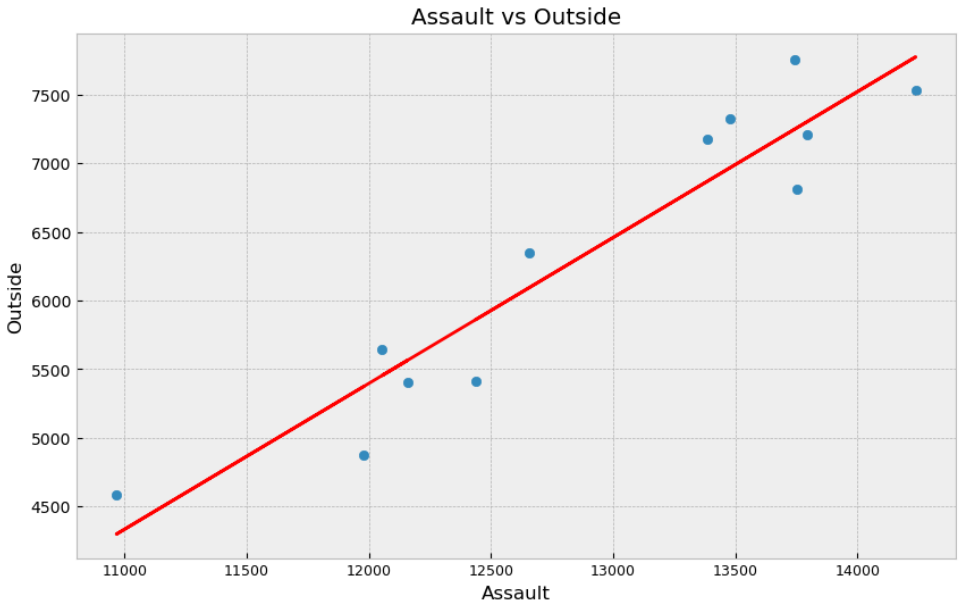


Question 5: Is there a correlation between certain types of offences? Is there a correlation between the type of offence and the location?

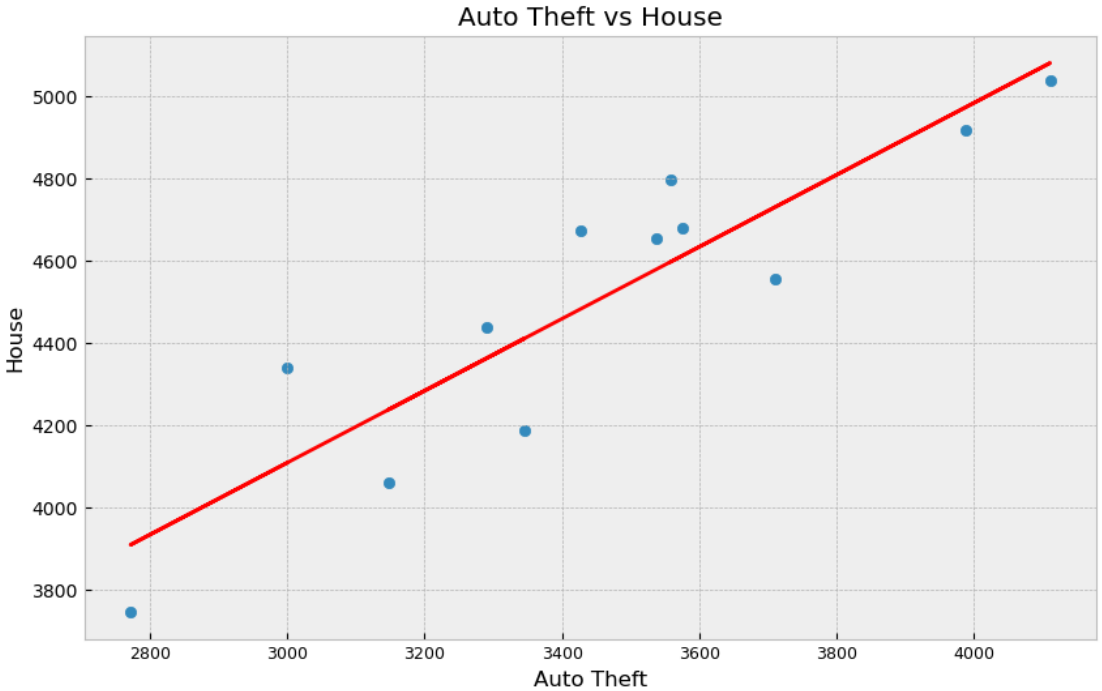
Premises type

To identify any potential correlation between premises and offences, we plotted the count of offences occurring in a premises type and the count of one offence category in a scatter plot. We also plotted the linear regression of these two factors and calculated the Pearson r.

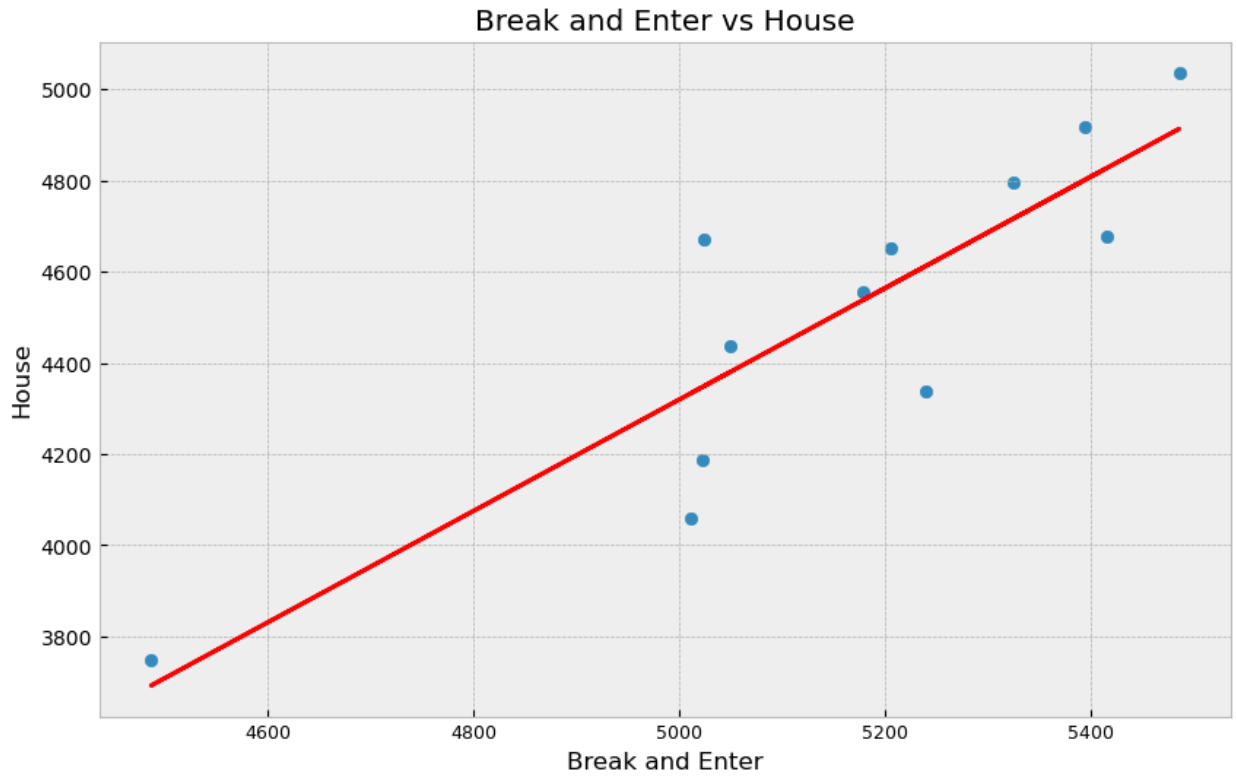
Assault has a very strong positive correlation (Pearson r = 0.94) with outside premises (parking lots, streets/roads/highways, open areas, etc.).



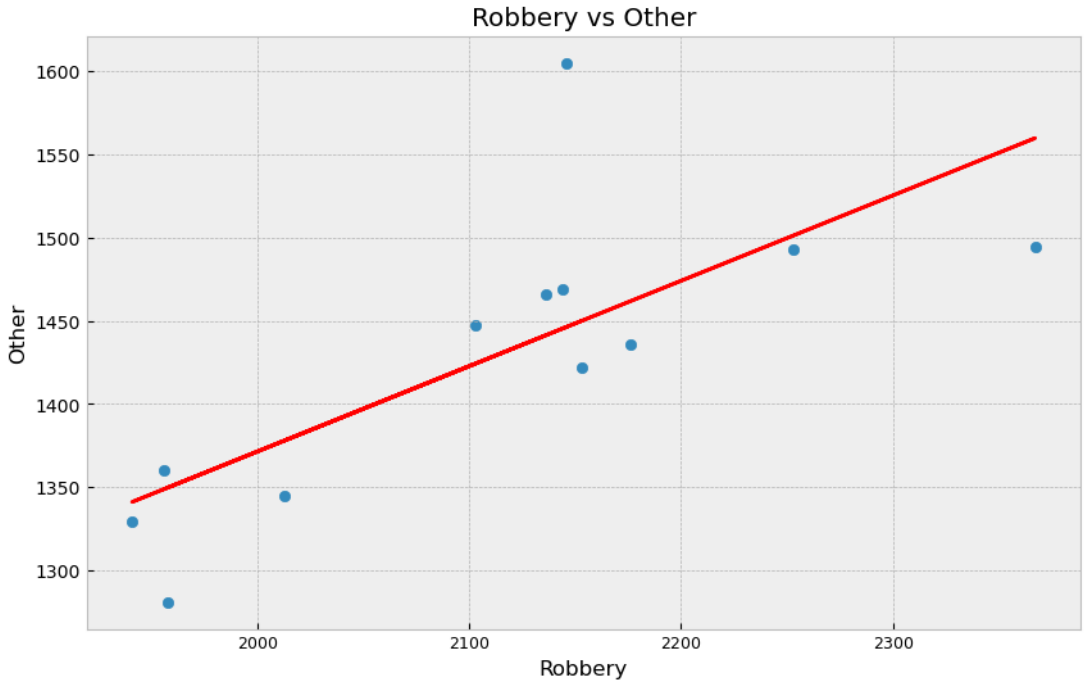
Auto theft has a very strong positive correlation (Pearson r = 0.90) with houses.



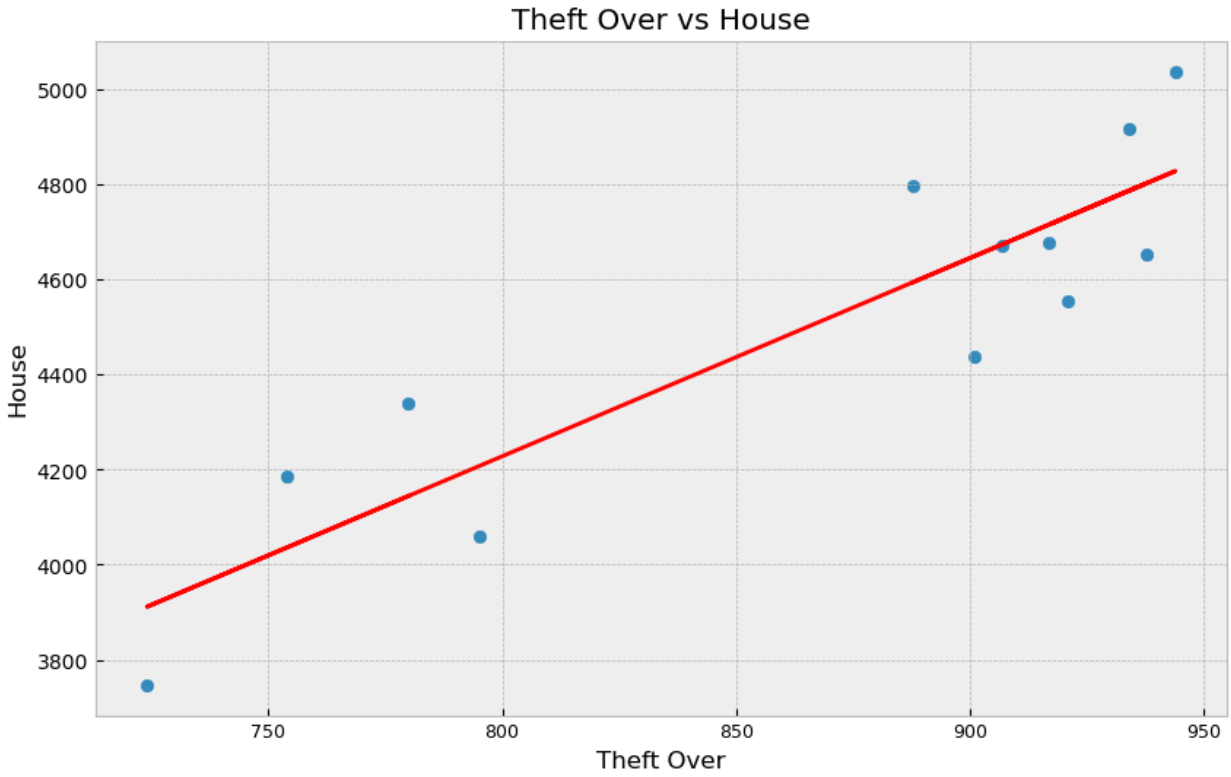
Break and enter has a strong positive correlation (Pearson r = 0.87) with houses.



Robbery has a positive correlation (Pearson r = 0.74) with other premises (religious and medical facilities, social services, etc.).



Theft over $5,000 has a very strong correlation (Pearson r = 0.89) with houses.



Offence categories

From the data on various neighbourhoods, we can observe different patterns in the types of offences. For instance, West Humber-Clairville has a notably high number of auto thefts at 3,466, but also has a significant number of assaults at 2,311. Moss Park stands out with a high count of assaults at 4,468, yet its auto thefts are much lower at 303.

The Downtown Yonge East and Yonge-Bay Corridor areas show high assault numbers, almost paralleling Moss Park, but with fewer auto thefts. Interestingly, York University Heights and Kensington-Chinatown present a balanced spread across different crime types.

We can infer that certain neighbourhoods do seem to have dominant crime types. This might suggest that local factors or conditions play a role in influencing the prevalence of certain offences.

In our analysis, we employed a correlation matrix to examine relationships between different crime types. The matrix allows us to see how the occurrence of one crime might relate to another. For instance, there's a strong positive correlation of 0.74 between Auto Theft and Theft

Over, suggesting areas with more auto thefts tend to also see more of the latter. Conversely, Assault and Auto Theft have a moderate negative correlation of -0.49, indicating that areas with higher assaults might have fewer auto thefts.

For question 5, we did a cross correlation for top 10 neighbourhoods where the offence rate is highest. Among nine neighbourhoods assault is the overall most seen offence category but the neighbourhood, west humber-clairville, who has the highest number shows a different pattern. It suffers from auto theft the most and assault comes the second.
Using a correlation matrix we can see a strong link between auto theft and theft over where the Pearson value is 0.74. we also noticed a negative relationship (-0.49) between Assault and Auto Theft.