CSC316: Process Book

Project Title: Crime in Toronto: A Data-Driven Analysis for Public Safety
Team Name: Safe Streets Squad
Team Members: Ifaz Alam, Meet Patel, Aaditya Mandal
Emails: ifaz.alam@mail.utoronto.ca, meetv.patel@mail.utoronto.ca, aaditya.mandal@mail.utoronto.ca

Project Abstract

Crime rates in Toronto have been rising, making public safety a growing concern for residents, students, and policymakers. By using official government sources, we aim to create interactive visualizations that reveal crime hotspots, seasonal trends, and socio-economic correlations such as income levels, neighbourhood safety and unemployment rates across Toronto.

Background and Motivation:
As lifelong residents and students preparing to enter the housing market, we want to better understand crime patterns and the factors influencing neighbourhood safety. Although crime data is accessible online, it is often presented in a raw format which makes it difficult to arrive at a meaningful conclusion quickly. Therefore we want to construct an interactive, data-driven tool that allows users to explore crime trends and neighborhood safety across Toronto. A key motivation of our project is the development of a Crime Safety Index, which will help assess which neighbourhoods are safer, particularly for activities such as walking alone at night. Using official government sources like the Toronto Open Data Portal and the Toronto Police Service Public Safety Data Portal, we hope to analyze information such as crime distribution, police response times, and the frequency of incidents at different times of the day.

Related Work
Our team was not inspired to pursue this topic after witnessing any form of external papers, websites, or visualizations. Instead, we decided to pursue this topic out of curiosity as we have been lifelong residents of Toronto and would like to learn more about our city.

Audience and Questions
This resource could help residents make informed decisions about their living and travel options, including adjusting commute routes based on neighbourhood safety scores, while also providing insights for policymakers and urban planners to address public safety concerns.
A list of concrete research questions we have are as follows:

- Which Toronto neighborhoods experience the highest and lowest crime rate?
    - Any clusters of high-crime areas?
    - Correlations between crime and economic factors or population density?
        - Economic factors such as unemployment, low income.
- What are the most common types of crime and do they vary by area?
    - Urban vs Suburban areas
- What do the response times look like policing varying by neighborhoods?
    - (Is there bias in poorer neighborhoods vs richer ones etc.)
- Which areas are safer (in terms of a score index) to walk at night alone?

- Is there a seasonal structure of crime in Toronto? (Are certain crimes more likely to occur in certain seasons, frequency of crimes etc.)
    - This can be extended to what times do crimes frequently occur (Evenings or morning, or afternoon)

Dataset

Reliable data will be collected primarily through the Toronto Open Data Portal and the Toronto Police Service Public Safety Data Portal. These portals are directly operated by Toronto's own municipal government, the City of Toronto. Furthermore, they provide ready-to-go data with the ability to download in various formats such as CSV or Excel files.

The links to the specified data portals are as follows:

- Toronto Open Data - https://www.toronto.ca/city-government/data-research-maps/open-data/
- Toronto Police Service: Public Safety Data Portal - https://data.torontopolice.on.ca/

Data Cleanup

We may be interested in sticking to a consistent data filetype such as CSV. It is not expected that substantial data cleanup will be required because the datasets are provided as ready-to-go CSV or Excel files. Data headers in the files and the portal itself make it simple to identify what the values in each column represent. If the amount of columns happens to be too large then we may drop unnecessary columns in the interest of saving space for when our solution is hosted online. We may also drop unnecessary rows that are identified as "outliers". For this kind of data cleanup step, we could look towards using the Python programming language to automate this process since Python contains many useful libraries for data processing / manipulation (e.g. Pandas). Once we have the cleaned and processed data, we plan to derive quantities pertaining to a crime such as: date, time, location, type of crime, classification, severity, contributing factors and more.

CSC316 Group

Project Team Agreement

Communication:

- Discord will be used as the primary communication platform for the group
- Weekly meetings will be held either Saturday or Sunday on Discord. The group will decide ahead of time on which of the two days and what time the meeting will take place
- Attendance at the weekly meetings is mandatory unless a valid reason is provided. Each member will provide an update on the progress they have made as well as share any relevant information
- In the case of an emergency or similar circumstances, you must notify the group as soon as possible

Organization:

- Each individual will work in their own respective branches. Ensure that you work on the latest updated main
- Main branch is protected from direct commits
- Naming convention: feature/description or fix/description
- Commit messages should be short and concise
- The task board will be separated into 4 columns: To Do, In Progress, Review, Completed. Individuals are responsible for actively updating it as work progresses

Quality Standards and Expectations:

- Members will actively contribute to all aspects of the project. Individual features or tasks will be assigned equally all technical and non-technical requirements
- Each member is responsible for completing their appointed work. It is to be completed in a timely manner such that the other group members have time to review it.
- Detailed inline comments and documentation are required for blocks of code. Team members should be able to understand the code in case it needs to be reviewed, updated, or corrected
- Successful deployment is required before merging with the main branch. Errors or failed tests must not be present

Signed: Aaditya Mandal (Team Leader), Ifaz Alam, Meet Patel **Date**: Feb 6, 2025

Week 5 | Map

1. **Who is your audience? Come up with at least three options and pick one target audience.**

   Possible options for our intended audience include the general public of civilians/residents, policymakers, international visitors, or newcomers to the city. For the purpose of this project, we have selected the general public as the target audience for our visualizations.

2. **Describe your target audience in more detail. What do they know? What are their interests? What visualization literacy do they have? At what level of detail will you present information to them?**
   a. Our target audience involved the ordinary civilian/resident - especially newcomers and students (not familiar with the city)
   b. **What they know:** The audience may know the basic geography of where they live (mainly in areas where they live, study or work), however, if they decide to explore other parts of Toronto, they may be unaware of the crime trends and safety statistics in the different neighborhoods. They might hear about crime from news reports or personal experiences, but not from structured data-driven analysis.
   c. **Interests:** Finding safe areas for living, working, and commuting, along with exploring the city whilst avoiding high-crime areas. Making informed decisions about their commute choices.
   d. **Visual literacy**: We expect that our audience will have average visual literacy skills. However, they may not be able to comprehend more advanced visualizations suited for experts in the field of criminology. Our plans to ensure the audience can understand the visualizations is described below:
   e. **Data Presentation:** Our visualizations have to be intuitive and interactive, avoiding overly complex statistical graphs; using clear labels, tooltips, color-coded maps for easier interpretation. There should be enough detail to be able to understand the trends but not too much as to overwhelm the audience (using simplified charts and interactive filters).

Week 5 | Questions

3. **What questions about your data will be interesting for your audience? Come up with a list of interesting questions that your audience may have about your data. The more, the better, but your team should come up with at least ten questions.**

   1. Which Toronto neighborhoods experience the highest and lowest crime rate?

   2. Are there any clusters of high-crime areas? Can we make a heatmap of this?

3. Correlations between crime and economic factors or population density?

4. What are the most common types of crime and do they vary by area? (Urban vs Suburban areas)

5. What do the response times look like policing varying by neighborhoods? (Is there bias in poorer neighborhoods vs richer ones etc?)

6. Which areas are safer (in terms of a score index) to walk at night alone?

7. Is there a seasonal structure of crime in Toronto? (Are certain crimes more likely to occur in certain seasons, frequency of crimes etc.)

8. Is there a certain time group that the crimes commonly occur? (Evenings or morning, or afternoon)), or how has the frequency of crimes regardless of timing changed over the years?

9. Has crime in Toronto increased or decreased over the past 5-10 years?

10. Do certain events (eg. holidays, festivals) lead to a spike in crime (assault, hate crime, etc)?

11. What areas of Toronto frequently experience hate crimes? What reasoning bias serves as the motivation behind those crimes?

12. Which Toronto neighborhoods are the safest to walk alone at night ranked on an Index Safety Score?

13. How does the distribution of crime types differ between urban and suburban areas in Toronto?

14. How many people are actually arrested each year?

4. **What data do you have? Look at it in Excel or Google spreadsheet and briefly describe each attribute and its data type (categorical, ordinal, or quantitative) in your process book. It's OK if you are unsure about the data type for some attributes - you can simply describe them (e.g., geographic location).**

   Many of the datasets that are listed in our individual visualizations have common data attributes. Some are written below. The prefix OCC_ most likely means occurrence and refers to the occurrence of the event being recorded.

   EVENT_UNIQUE_ID - Categorical, uniquely identifies each event.

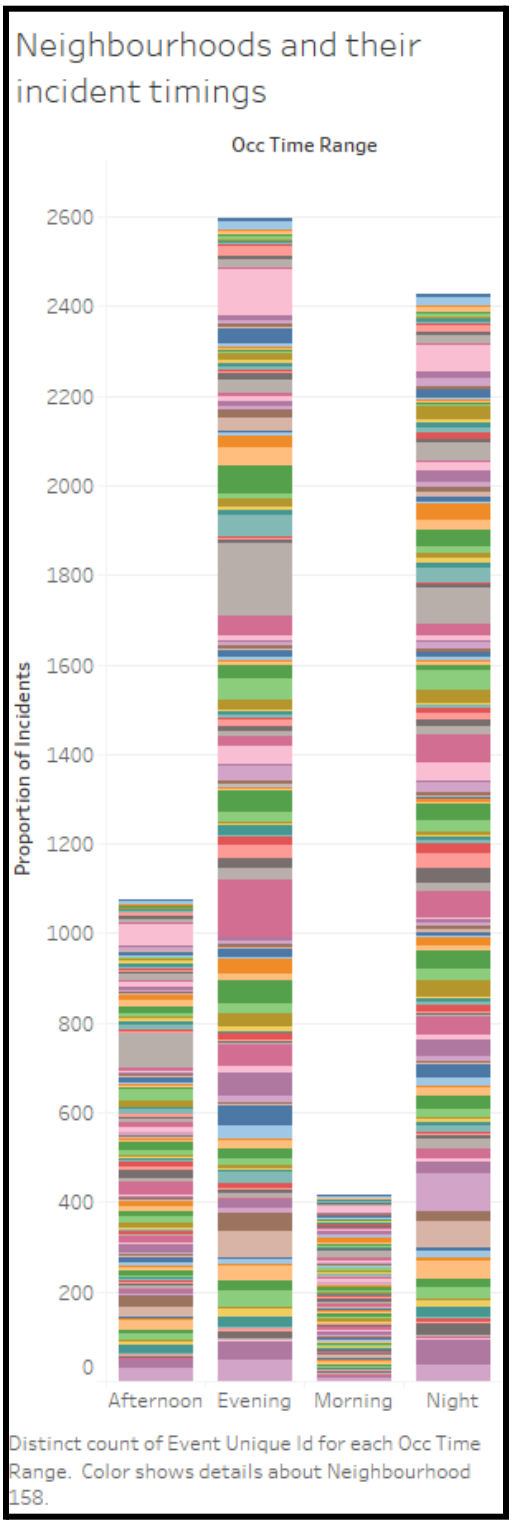
OCC_DATE - Ordinal, specifies the exact date and time when the event occurred (including both date and time).
OCC_YEAR - Quantitative, represents the year the event took place.
OCC_MONTH - Categorical, represents the month the event occurred (e.g., January, February).
OCC_DOW - Categorical, specifies the day of the week the event occurred (e.g., Monday, Tuesday).
OCC_DOY - Quantitative, represents the day of the year the event occurred (1-365 or 1-366 for leap years).
OCC_DAY - Quantitative, represents the day of the month the event occurred (1-31).
OCC_HOUR - Quantitative, specifies the hour of the day the event occurred (0-23).
OCC_TIME_RANGE - Categorical, indicates the time range during which the event occurred (e.g., Morning, Afternoon, Night).
DIVISION - Categorical, represents the division or district where the event occurred (e.g., D31, D54).
DEATH - Quantitative, indicates the number of deaths resulting from the event
INJURIES - Quantitative, represents the number of injuries reported from the event.
HOOD_158 - Categorical, identifies a neighborhood or area code associated with the event (e.g., 023, 078).
NEIGHBOURHOOD_158 - Categorical, represents a neighborhood or area name/ID where the event occurred (e.g., Regent Park).
HOOD_140 - Categorical, identifies another neighborhood or area code, similar to HOOD_158 but potentially mapped differently.
NEIGHBOURHOOD_140 - Categorical, represents another neighborhood or area name/ID corresponding to HOOD_140.
LONG_WGS84 - Quantitative, represents the longitude in the WGS84 coordinate system.
LAT_WGS84 - Quantitative, represents the latitude in the WGS84 coordinate system.
ETHNICITY_BIAS, RACE_BIAS, GENDER_BIAS etc. - Nominal, represents whether this particular bias was the motivation behind a hate crime

Week 5 | Ifaz - Individual Visualizations

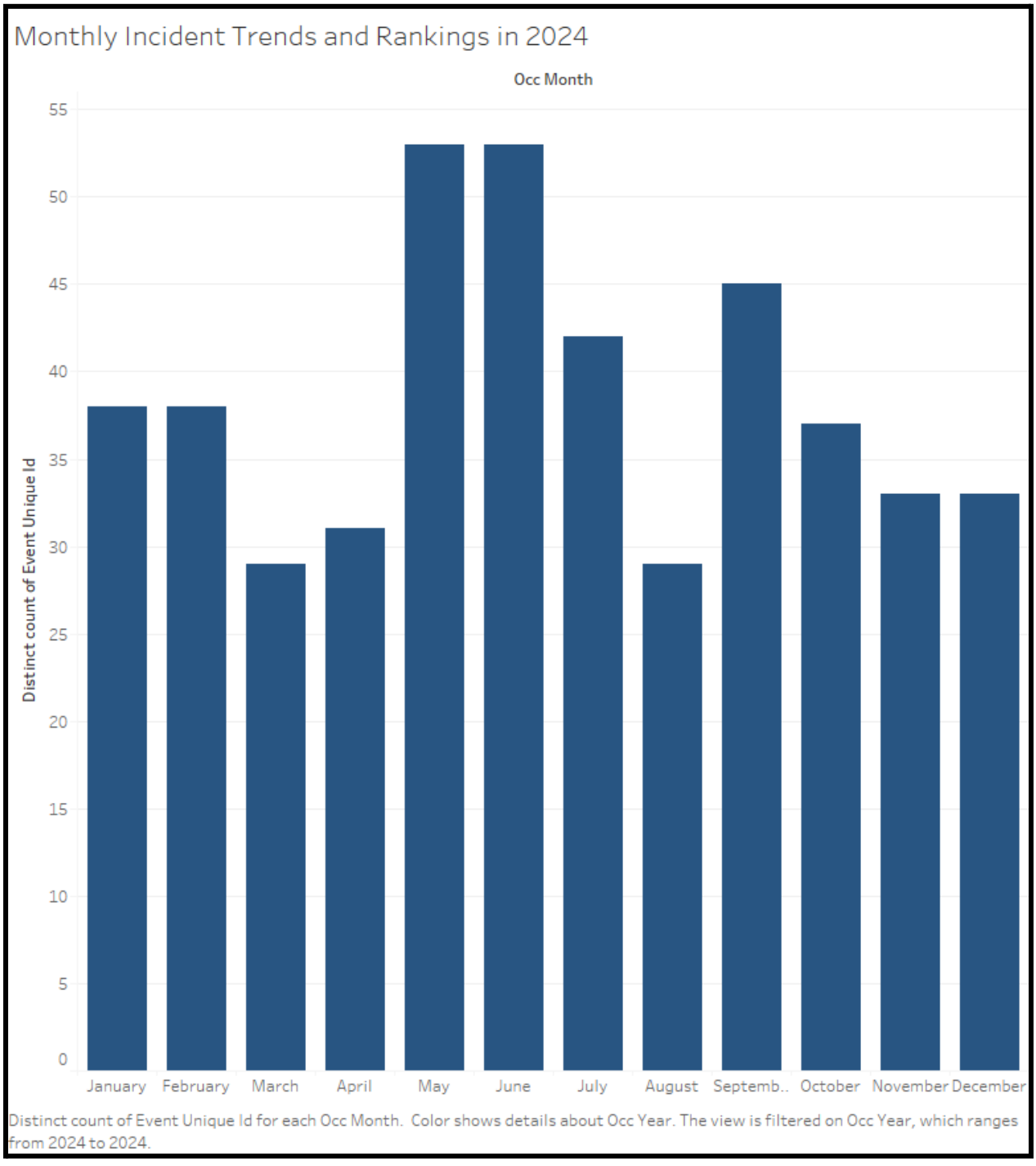
https://data.torontopolice.on.ca/datasets/TorontoPS::shooting-and-firearm-discharges-open-data/about

Dataset:

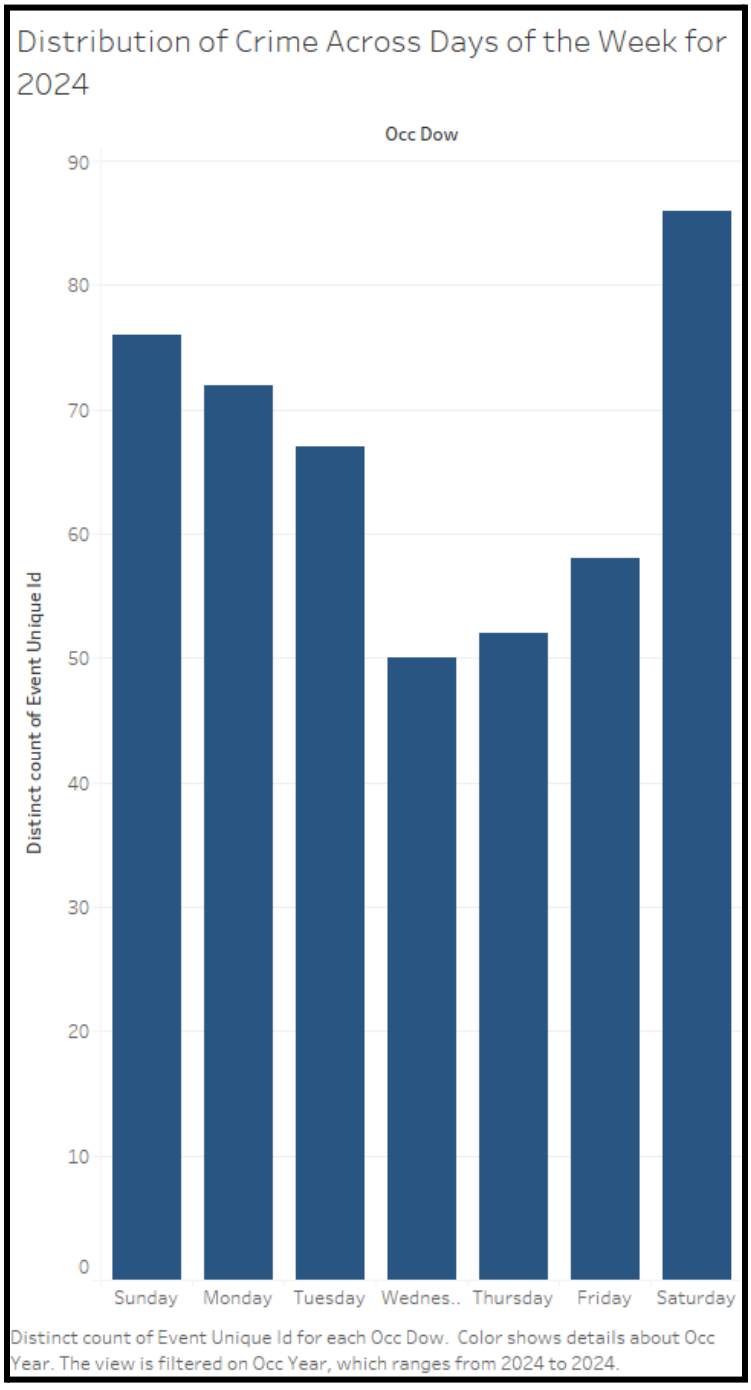
Visualization #1:



Visualization #2



Visualization #3:



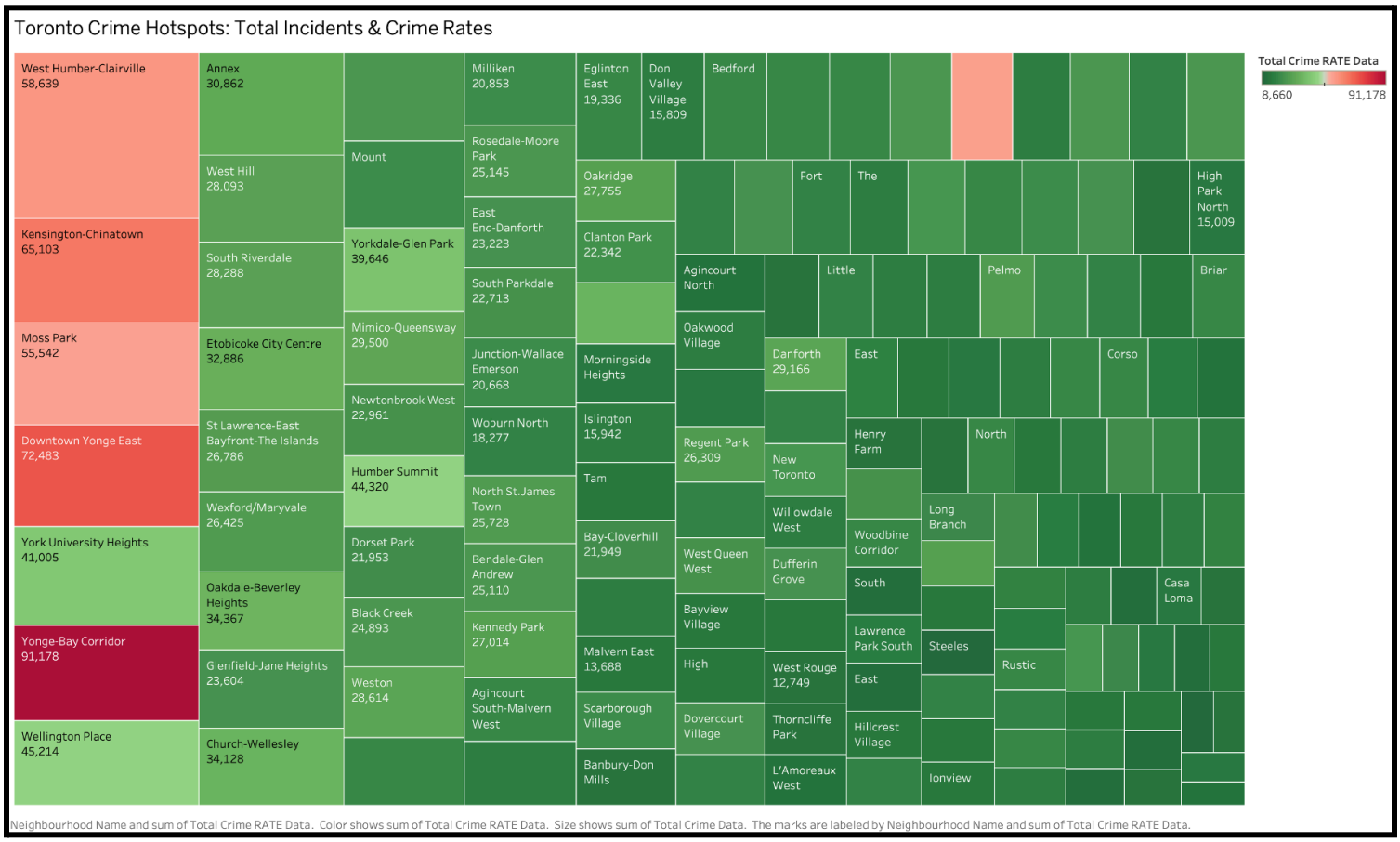
Brief Discussion: The first two questions that I answered in Tableau stay consistent with the questions that we came up as a team, because these are natural and common questions. This means that even if we did not have a discussion about our initial questions, I would have constructed the visualizations for these questions anyway. The final question that I answered slightly differed from the initial questions that we came up with but on the other hand I can also

argue that it is a subcategory for one of our questions which was “Is there a certain time group that the crimes commonly occur?”. This is also a natural question to have.

Week 5 | Aaditya - Individual Visualizations

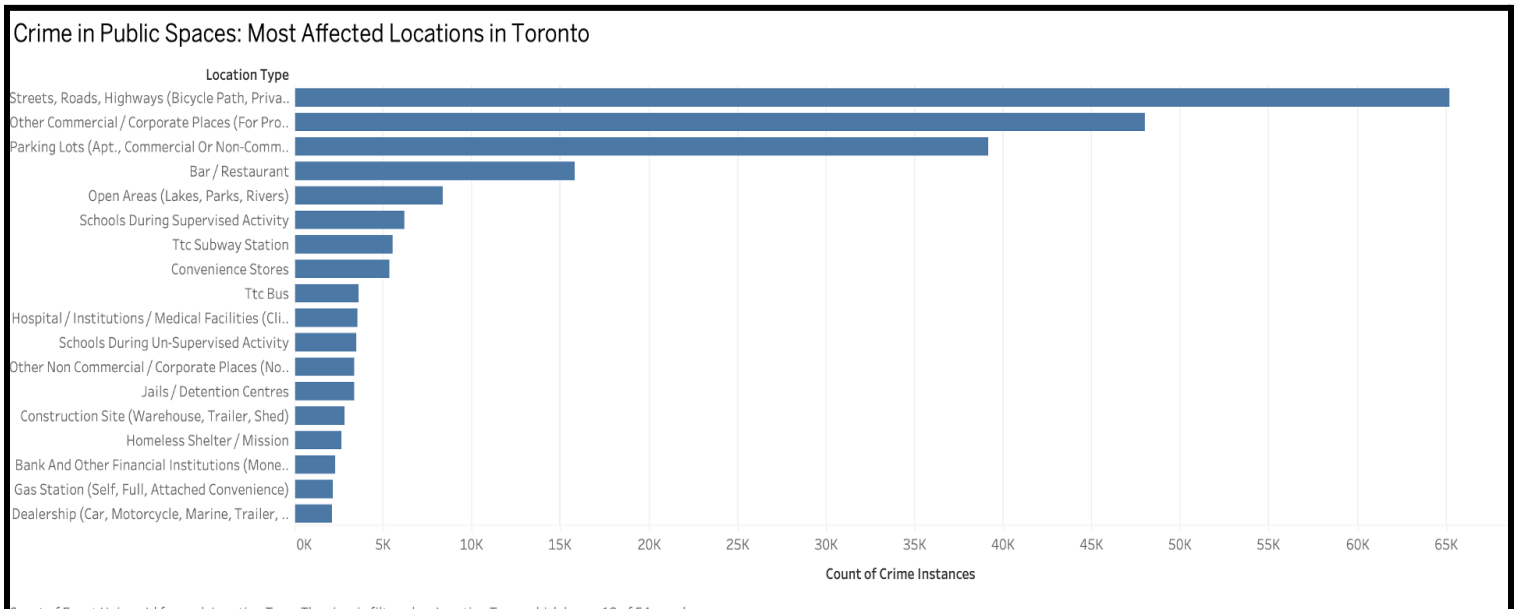
Visualization #1: Toronto Crime Hotspots: Total Incidents & Crime Rates from 2014-2024

Dataset: https://data.torontopolice.on.ca/datasets/TorontoPS::neighbourhood-crime-rates-open-data/about

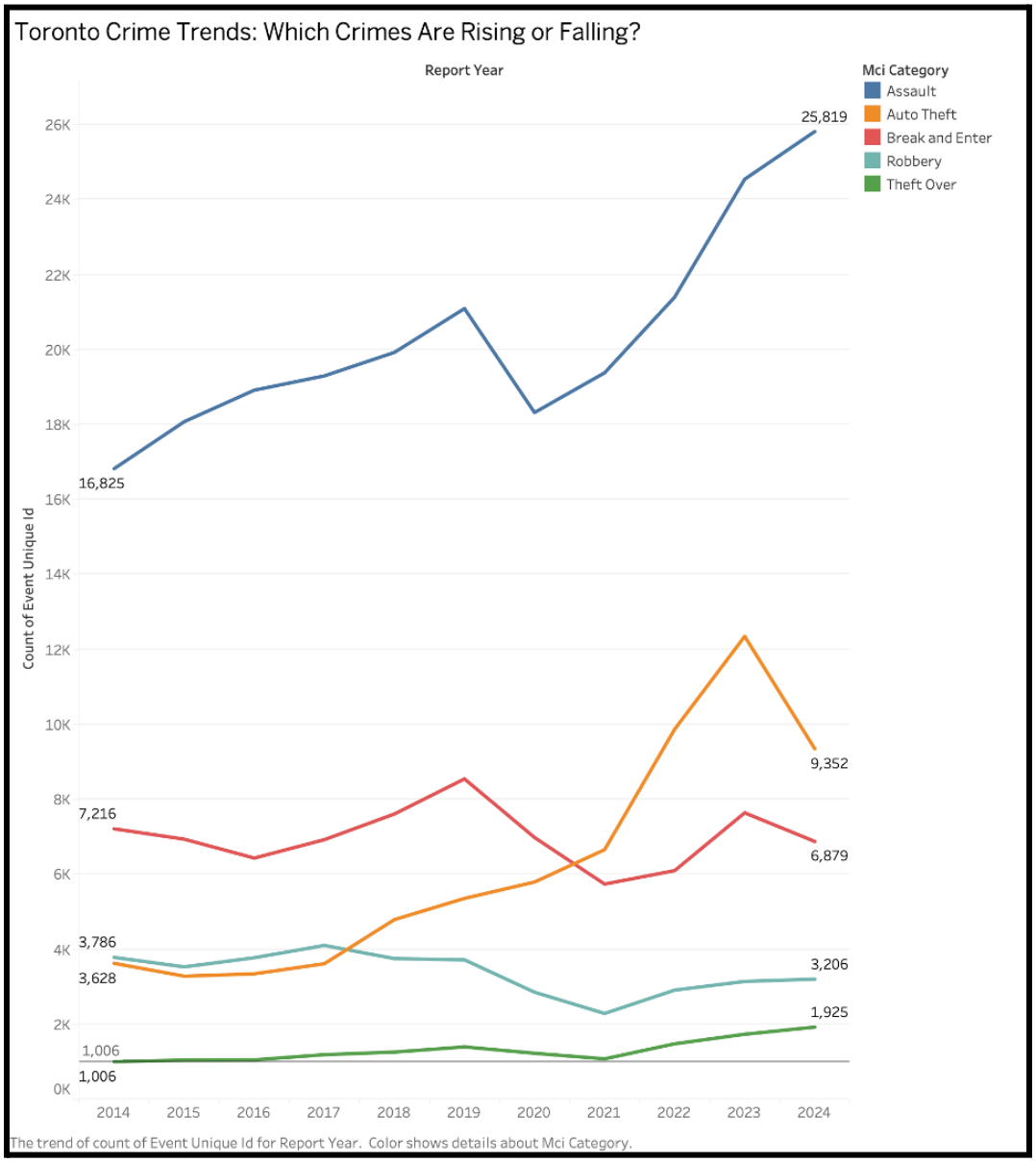


Visualization #2: Crime Distribution Across Public Spaces in Toronto

Dataset: https://data.torontopolice.on.ca/datasets/TorontoPS::major-crime-indicators-open-data/about



Visualization #3: Change in Different types of Crime in Toronto from 2014 to 2024

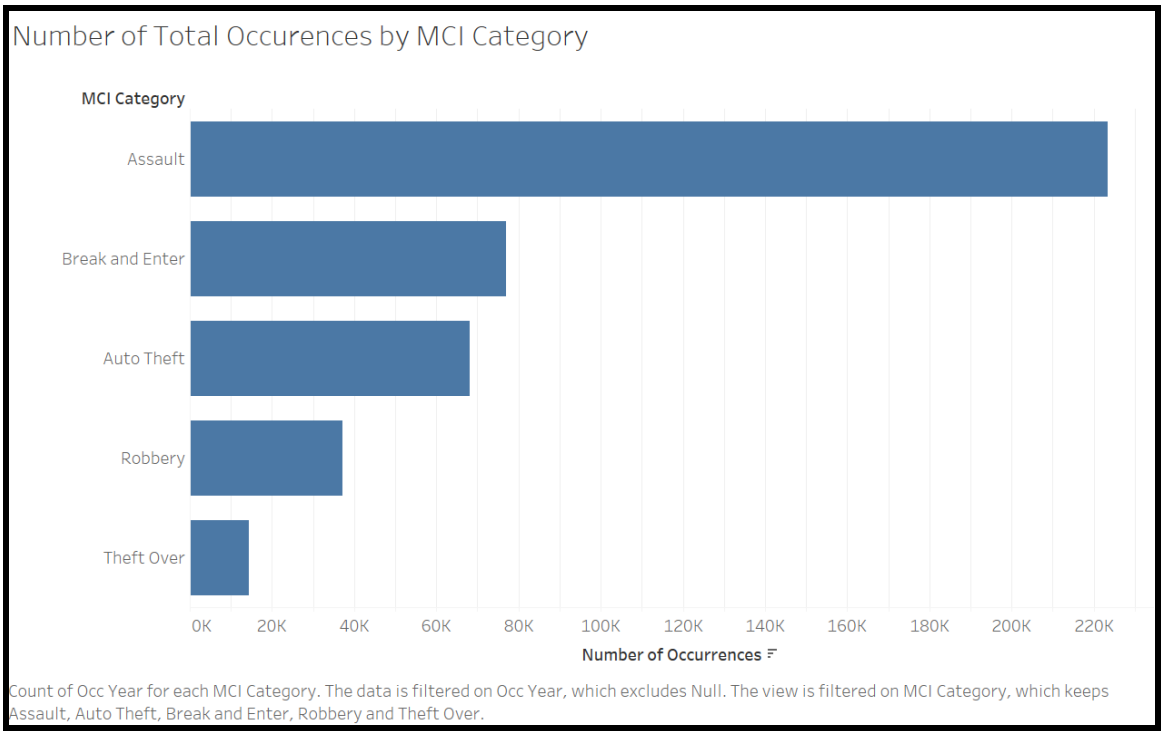


Dataset: https://data.torontopolice.on.ca/datasets/TorontoPS::major-crime-indicators-open-data/about

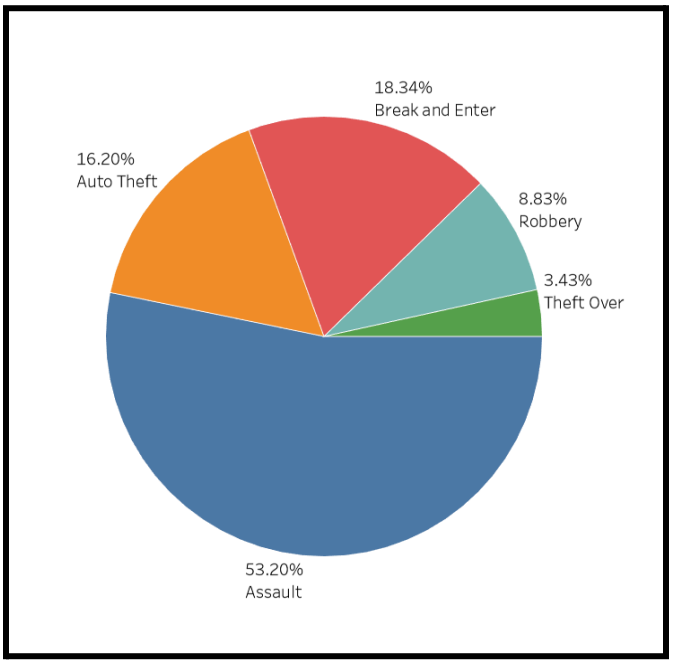
Brief Discussion: These three visualizations closely align with several of the initial questions we outlined as a team. The crime hotspots treemap directly answers which neighborhoods have the highest to lowest crime rates (Q1), along with the line graph of crime trends over time directly addressing the trend of the different types of crimes in Toronto (Q9). The bar chart of crime in public spaces closely aligns with analyzing most affected locations in Toronto where crime has occurred (Q4); Overall, these are very generic questions that are required as a baseline visualizations for our more relational questions which would require combining multiple dataset to get different related fields.The questions we ended up answering are still highly relevant to our target audience and provide insights that align with their concerns about safety in Toronto.

Week 5 | Meet - Individual Visualizations

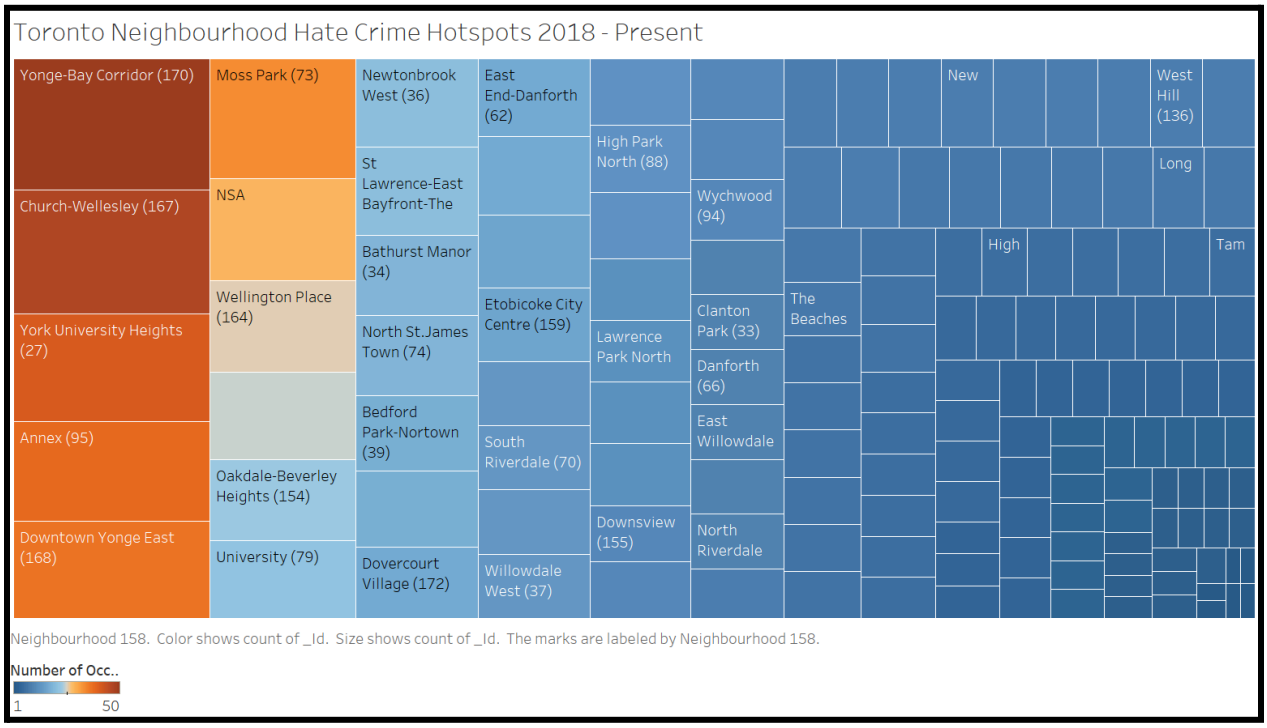
Visual 1 Dataset:
https://data.torontopolice.on.ca/datasets/TorontoPS::major-crime-indicators-open-data/about



Visual 2 Dataset:
https://data.torontopolice.on.ca/datasets/TorontoPS::major-crime-indicators-open-data/about
Title: Percentage of Total Occurrences by MCI Category



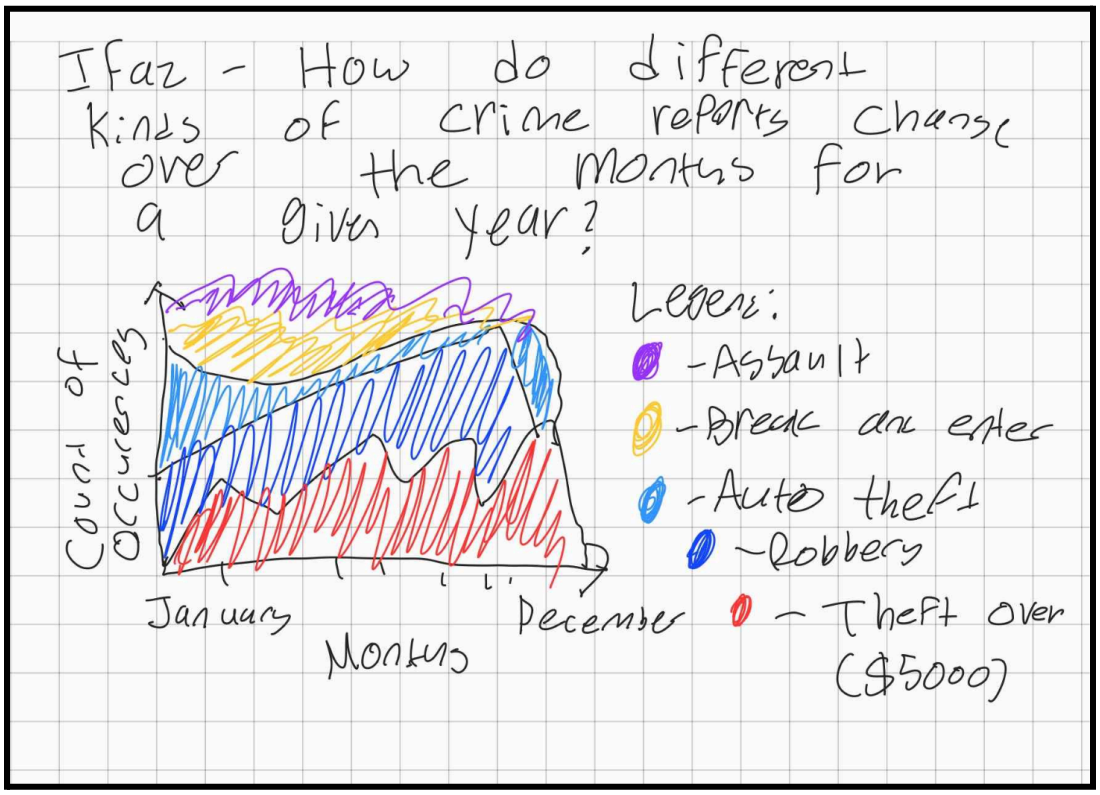
Visual 3 Dataset: https://open.toronto.ca/dataset/hate-crimes-open-data/



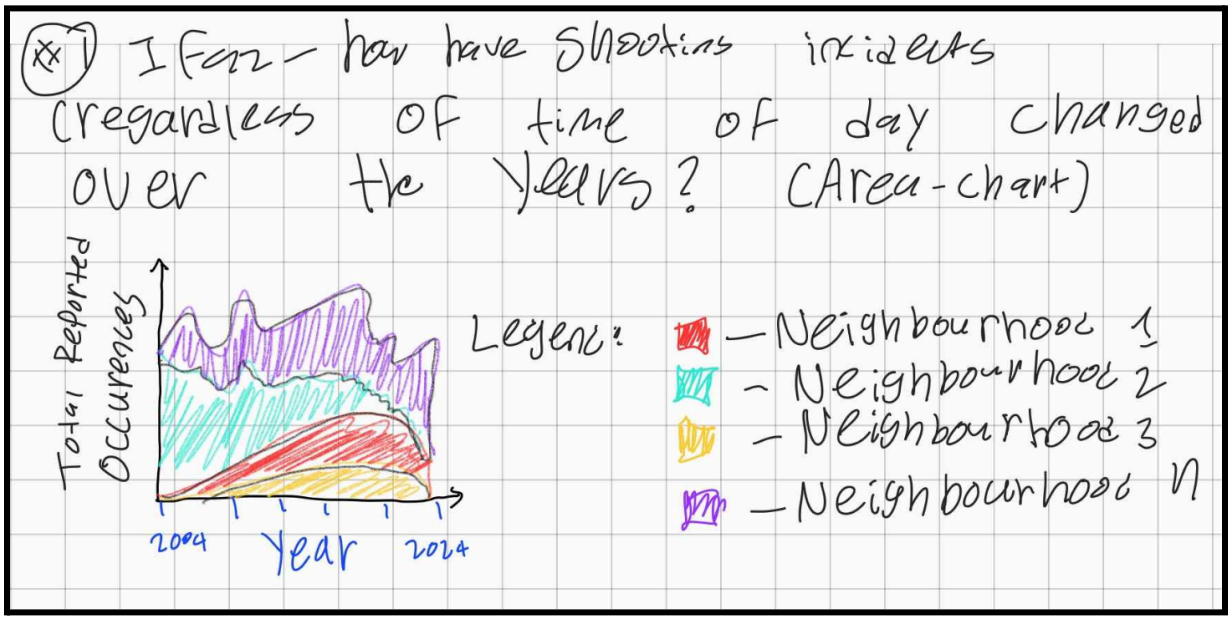
Brief Discussion: The first two visualizations I created, expressed in 2 different ways, briefly answer Question 4 which pertains to breaking down crime incidents into major categories. This is a general question that most people, irrespective of knowledge or experience, would come up with when the topic of crime is being discussed. My third visualization focuses more specifically on addressing Question 11. However, it does touch upon the broad aspect of Question 2. The treemap shows the number of hate crimes that occurred in every Toronto neighbourhood since data collection began in 2018.

Week 6 | Sketch Step

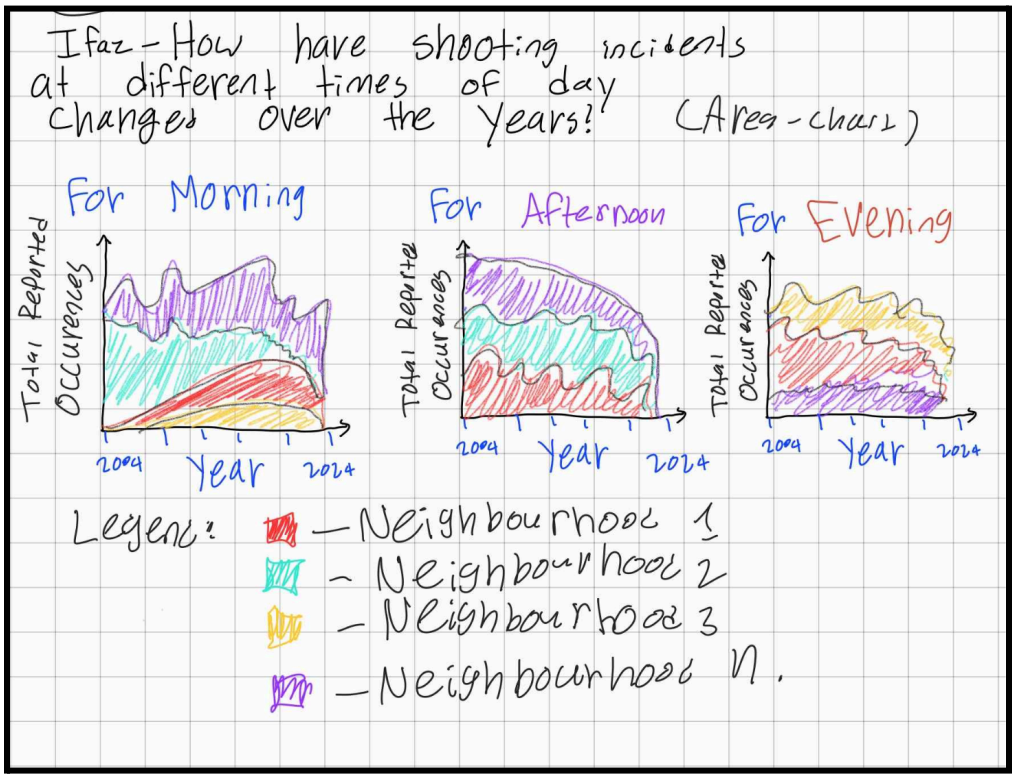
Sketch ID: 1 - Ifaz | Q7 | ([Major Crime Indicators Open Data - Toronto Police Service Public Safety Data Portal](#))



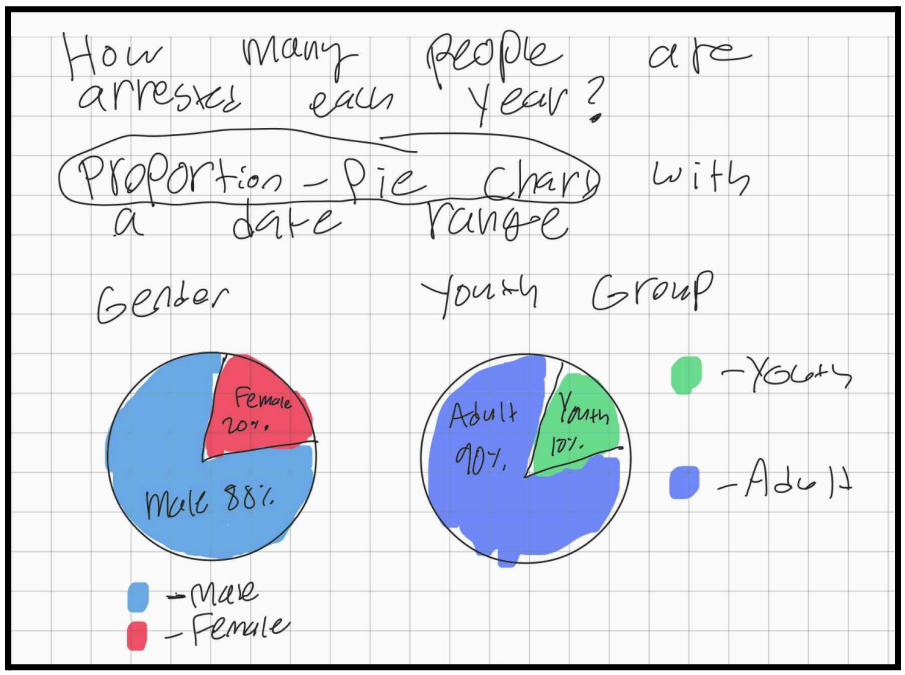
Sketch ID: 2 - Ifaz | Q8 | [(Shooting and Firearm Discharges Open Data)](#)



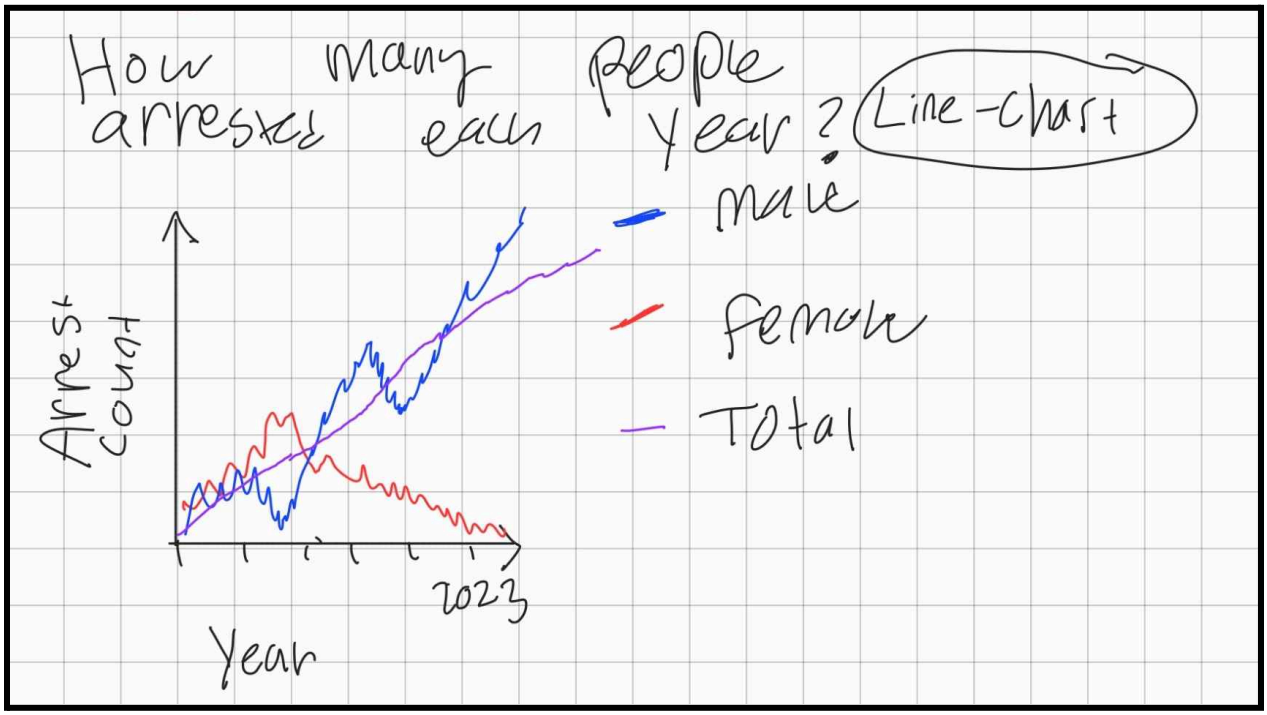
Sketch ID: 3 - Ifaz | Q8 | [(Shooting and Firearm Discharges Open Data)]



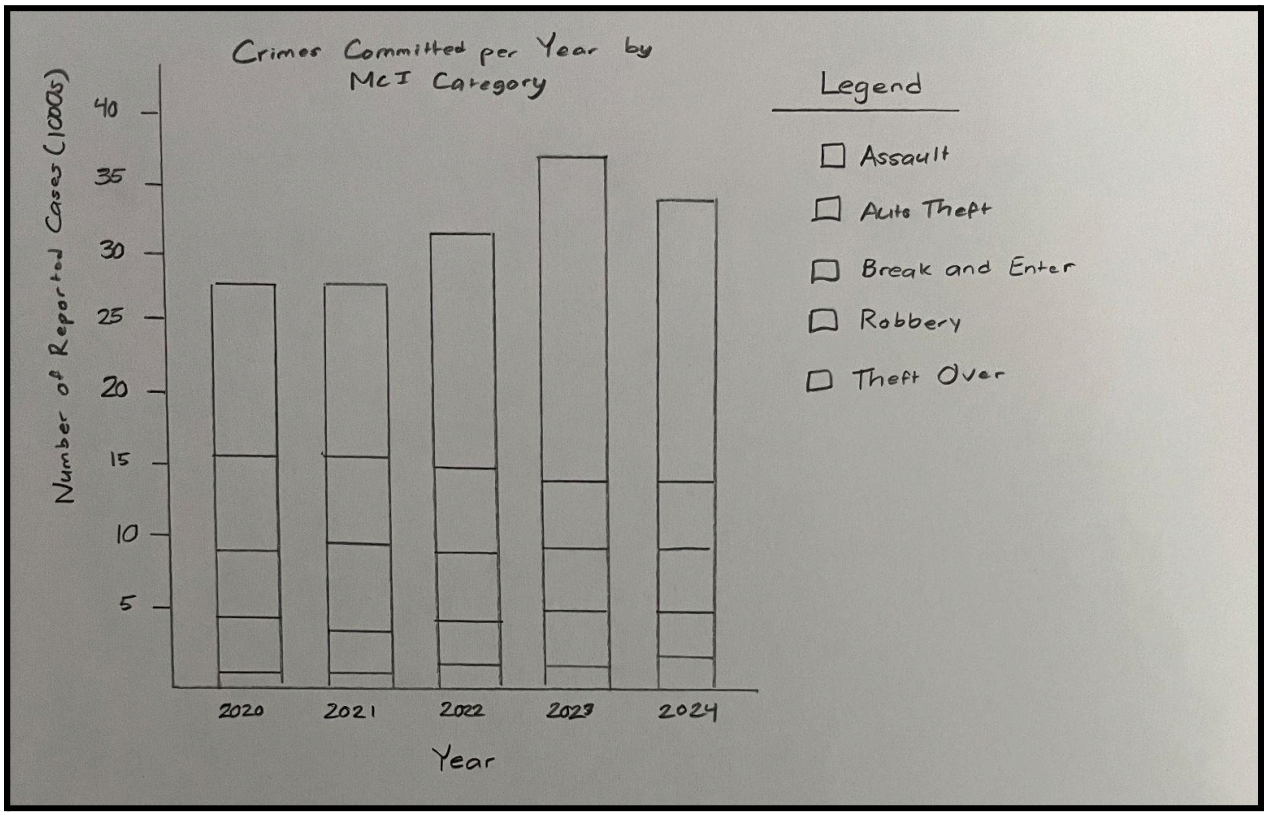
Sketch ID: 4 - Ifaz | Q14 | [(Arrested and Charged Persons (ASR-ENF-TBL-001))]



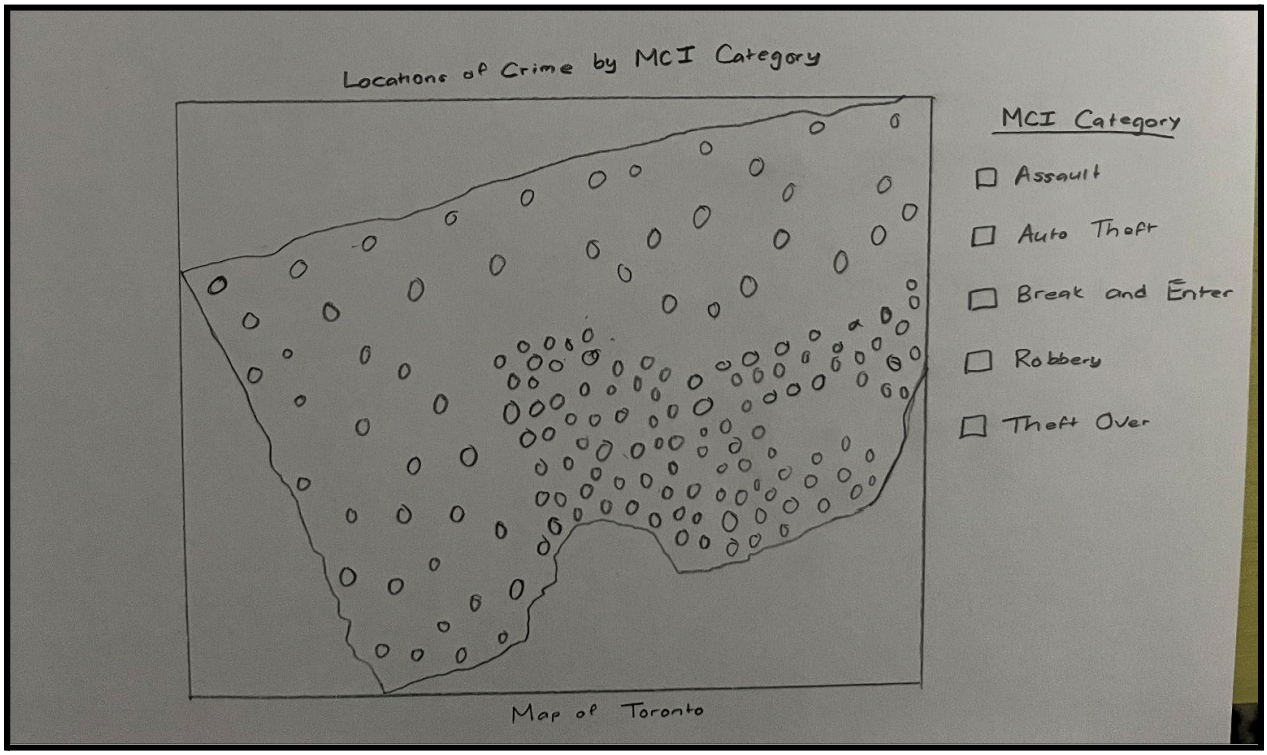
Sketch ID: 5 - Ifaz | Q14 | (Arrested and Charged Persons (ASR-ENF-TBL-001))



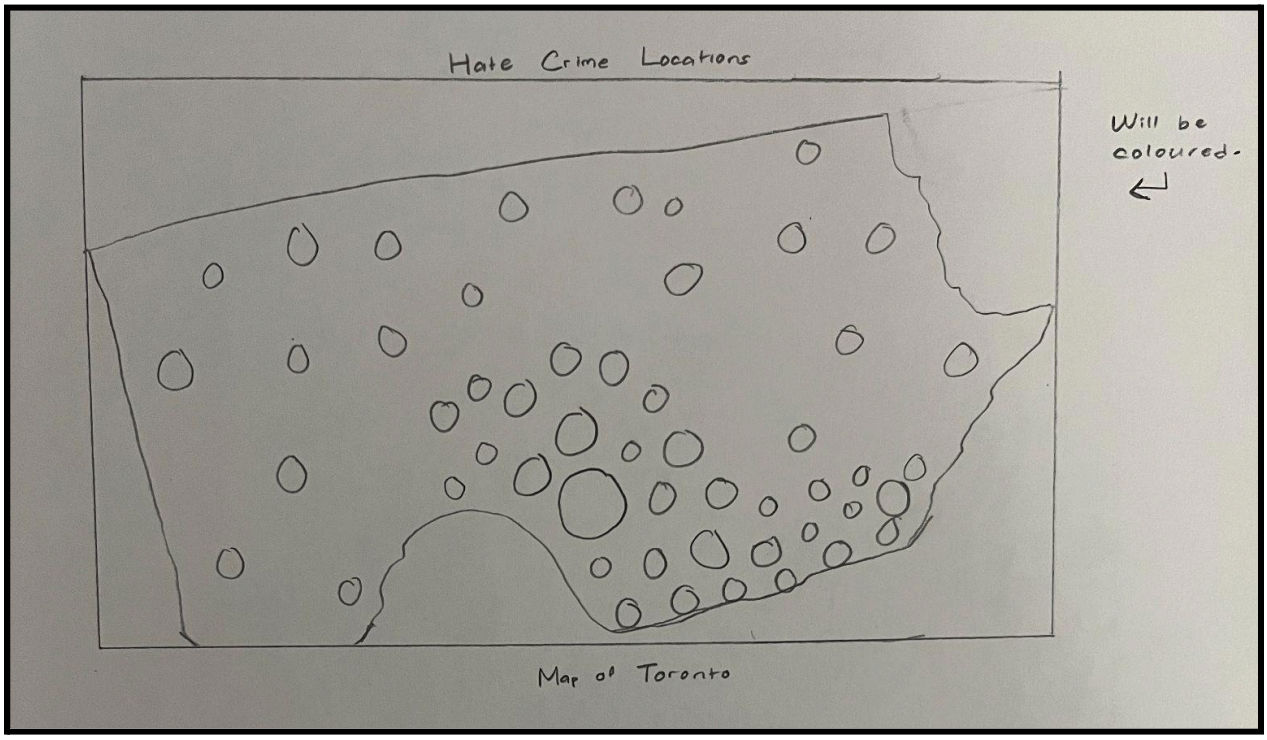
Sketch ID: 6 - Meet | Q4 | (Major Crime Indicators - Toronto Police Service Public Safety Data Portal)



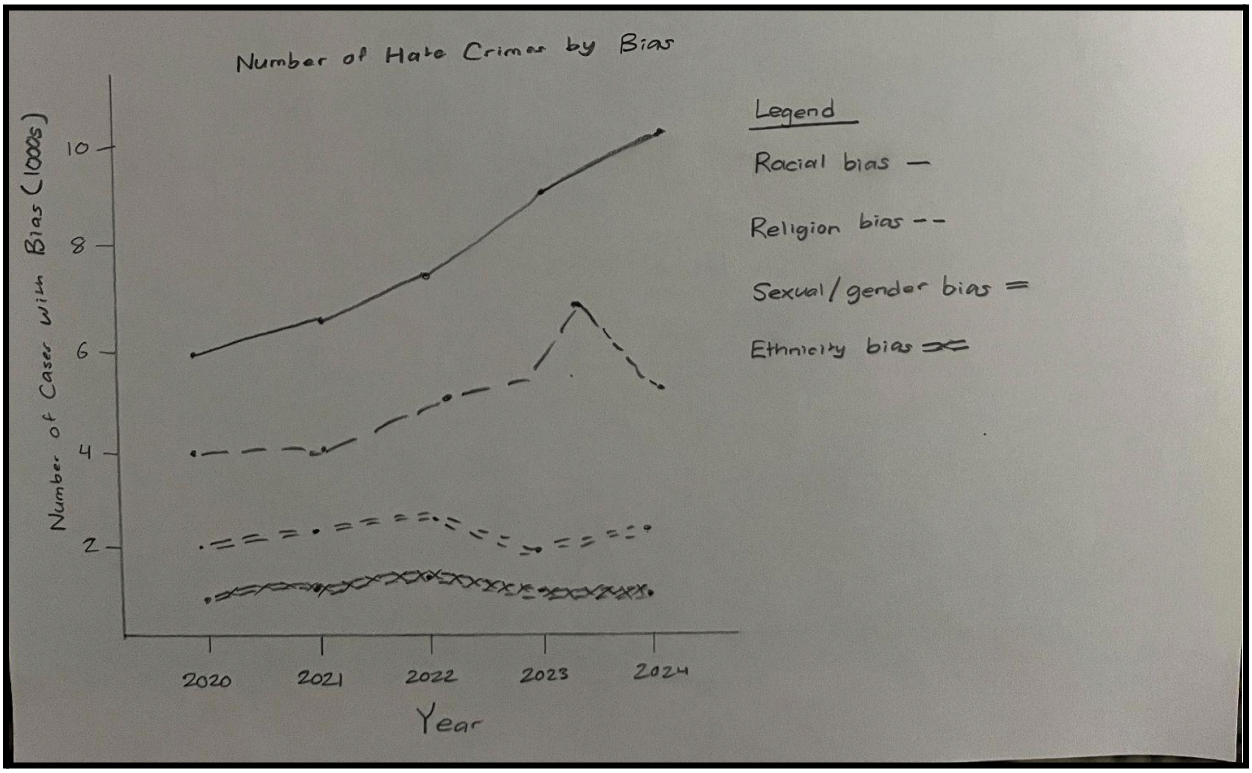
Sketch ID: 7 - Meet | Q4 | (Major Crime Indicators - Toronto Police Service Public Safety Data Portal)



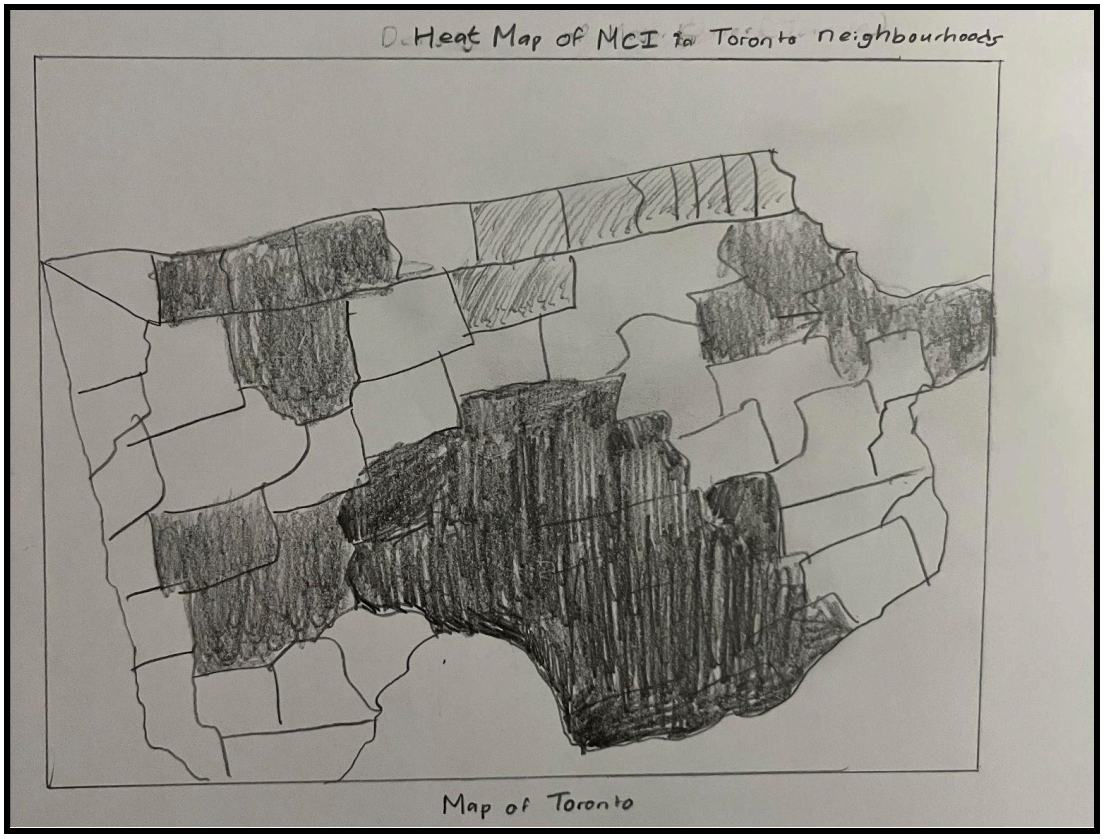
Sketch ID: 8 - Meet | Q11 | (Hate Crime Open Data - Toronto Police Service Data Portal)



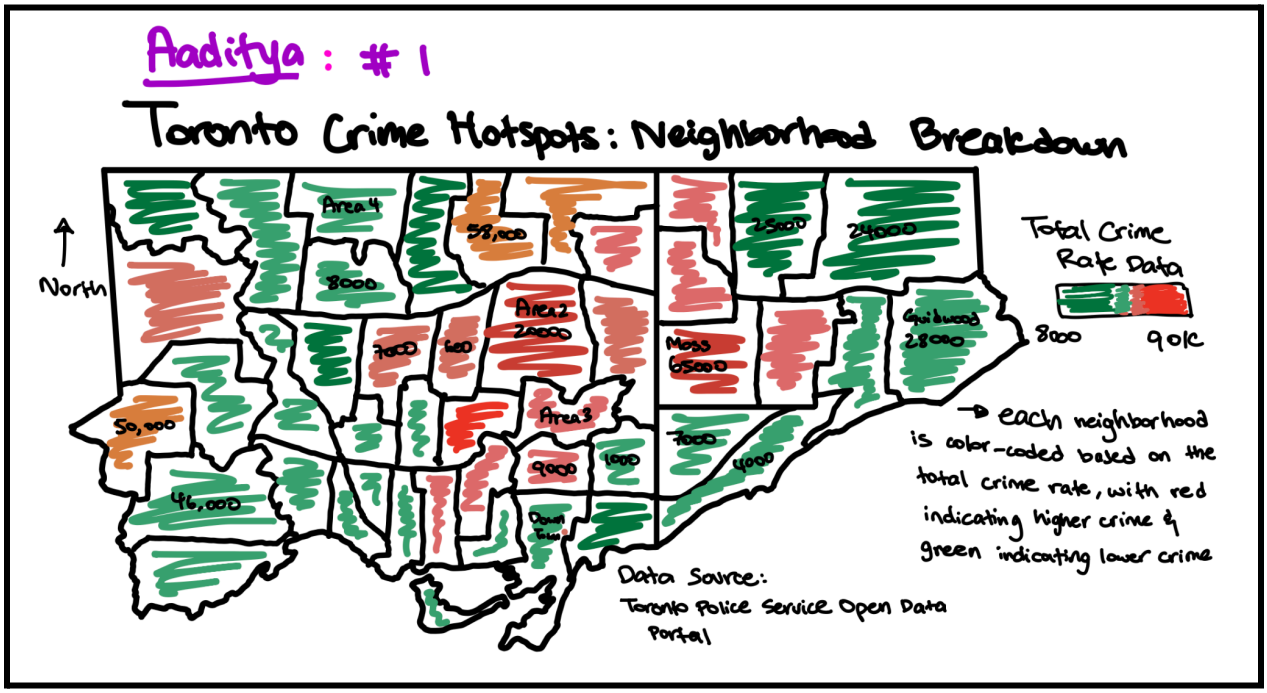
Sketch ID: 9 - Meet | Q11 | (Hate Crime Open Data - Toronto Police Service Data Portal)



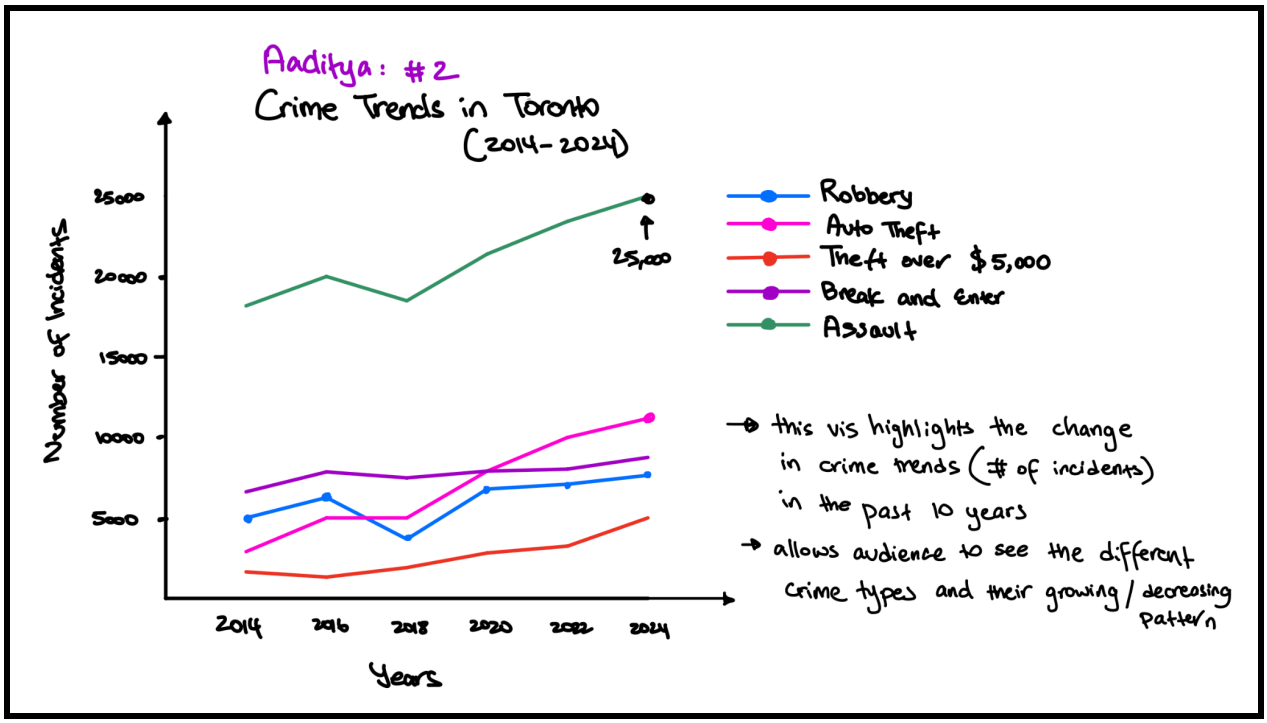
Sketch ID: 10 - Meet | Q1, Q4 | (Major Crime Indicators - Toronto Police Service Data Portal)



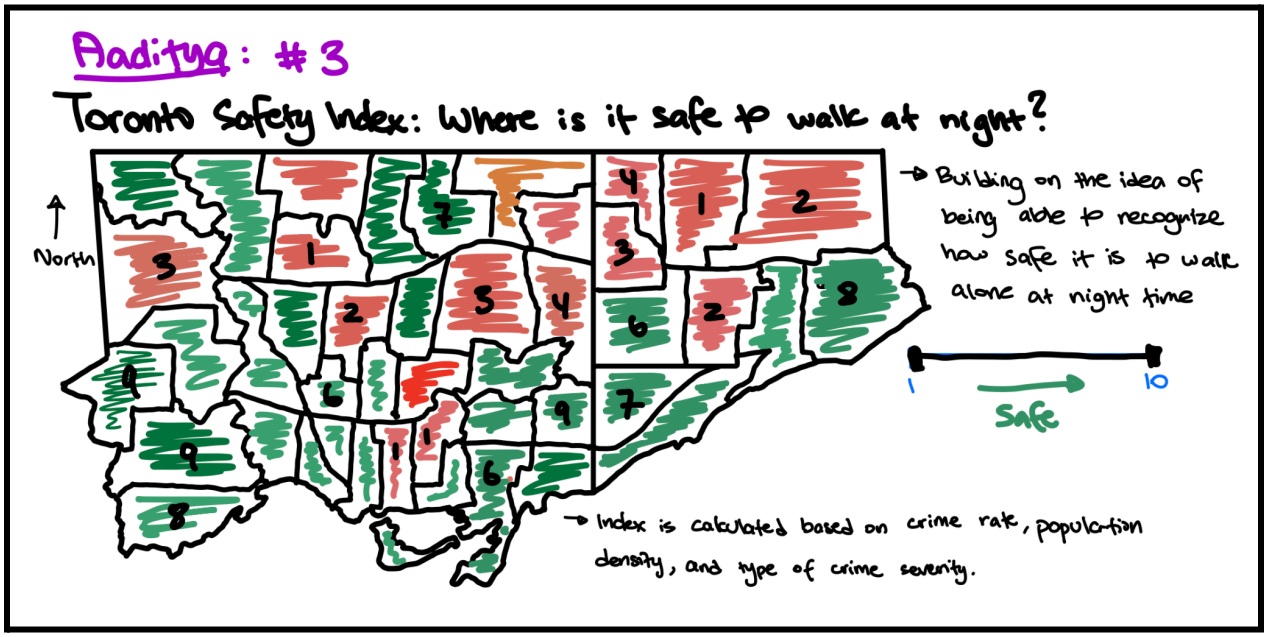
Sketch ID: 11 - Aaditya | Q1 | (Neighbourhood Crime Rates Open Data)



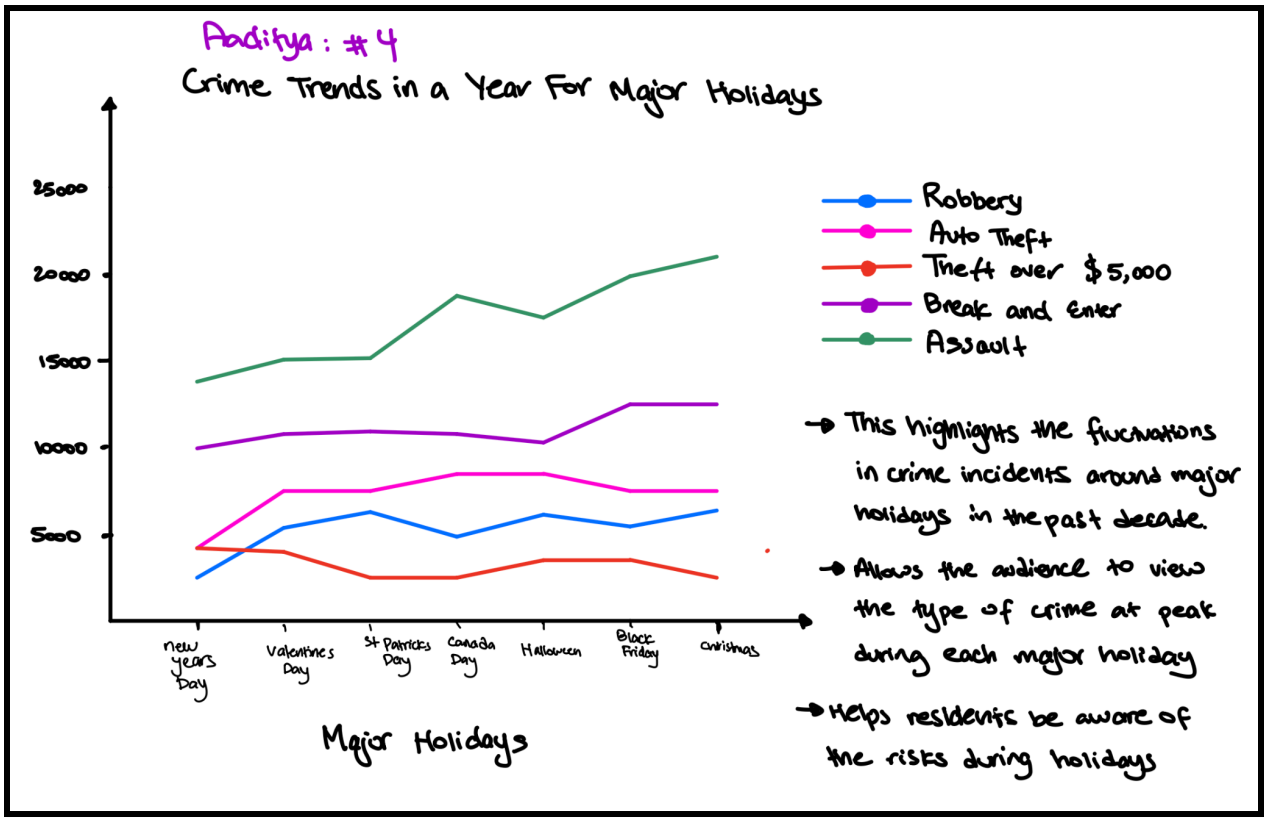
Sketch ID: 12 - Aaditya | Q9 | (Neighbourhood Crime Rates Open Data)



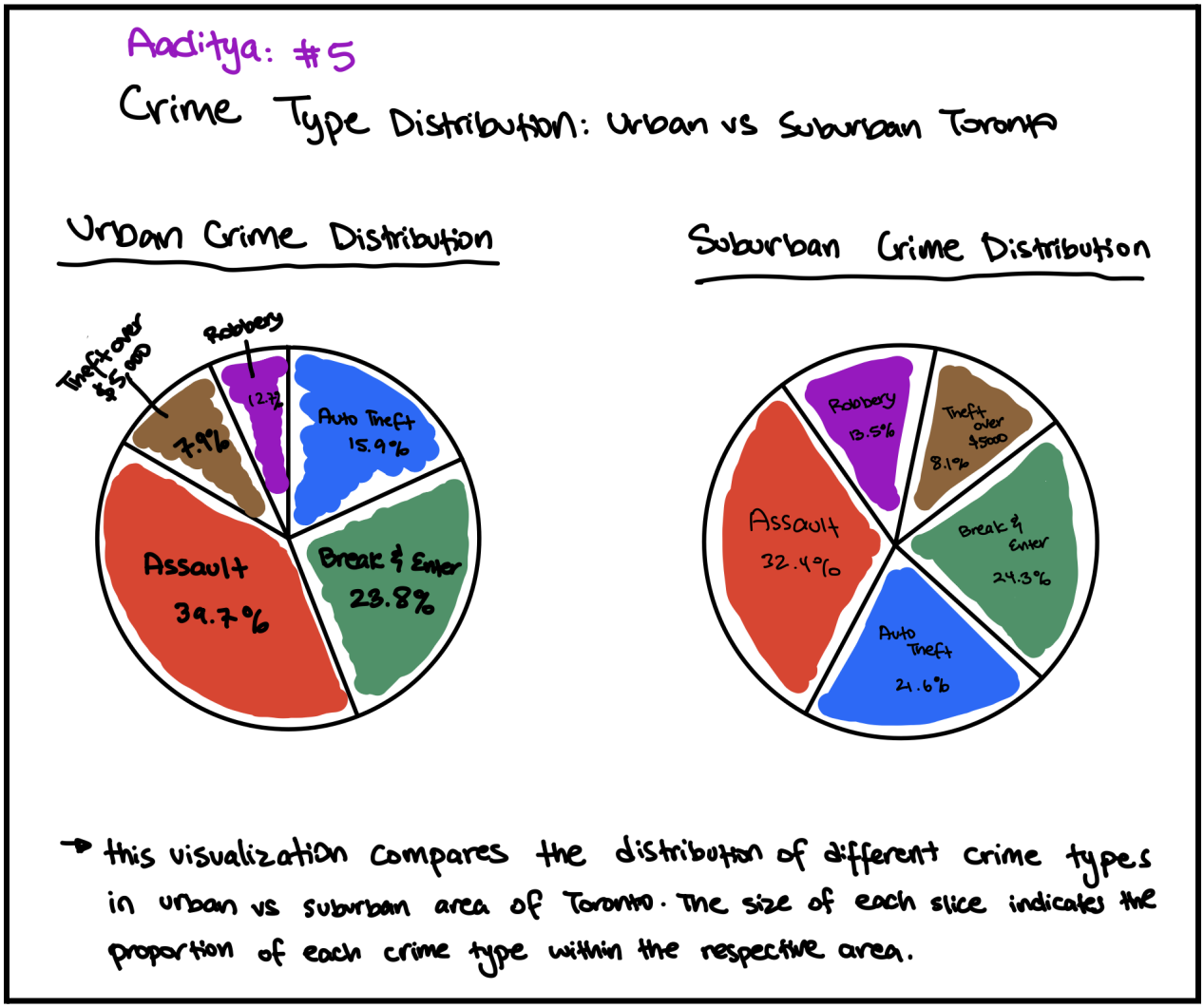
Sketch ID: 13 - Aaditya | Q12 | (Neighbourhood Crime Rates Open Data)



Sketch ID: 14 - Aaditya | Q10 | (Major Crime Indicators Open Data)



Sketch ID: 15 - Aaditya | Q13 | (Neighbourhood Crime Rates Open Data)



Datasets [Google Drive Folder]: Cleaned Datasets (To Use)

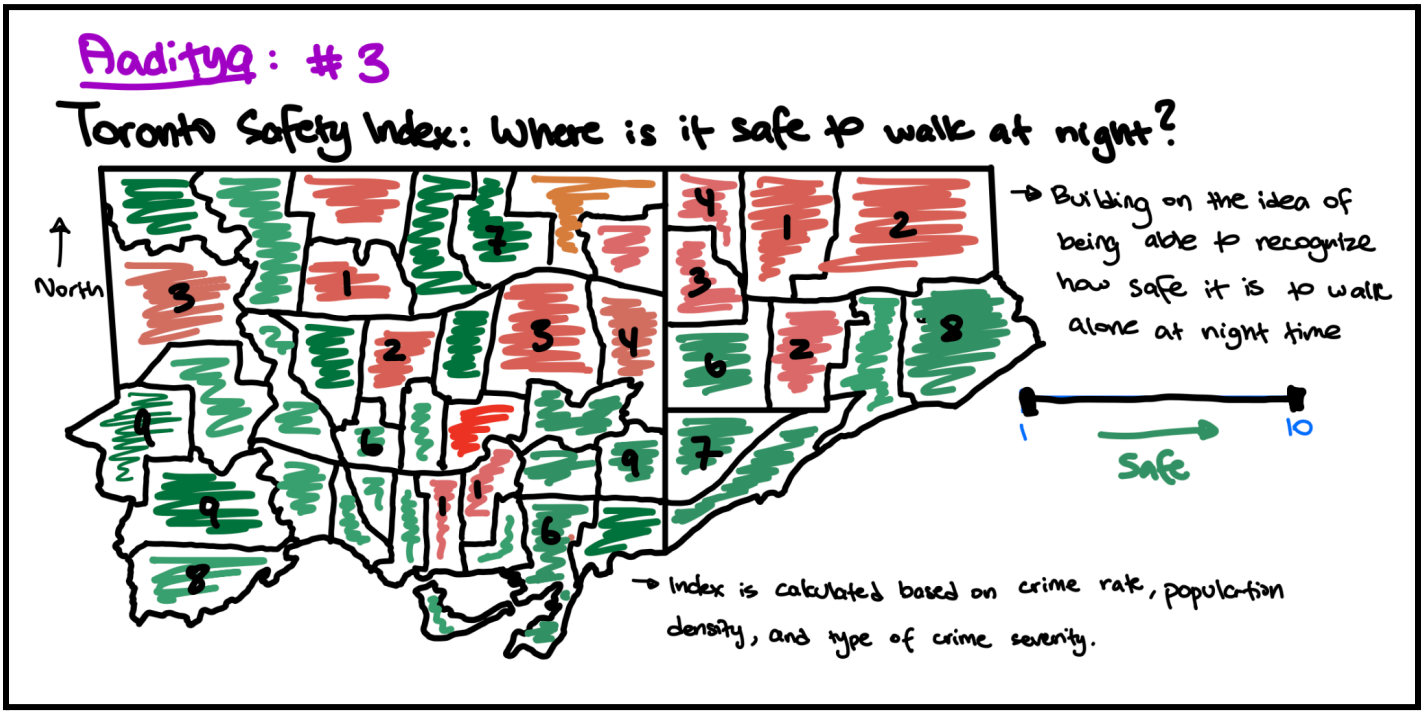
Week 6 | Decide Step

Sketch ID	Question ID	Author	Total Number of Votes
1	7	IA	0
2	8	IA	1
3	8	IA	1
4	14	IA	1
5	14	IA	0
6	4	MP	0
7	4	MP	3
8	11	MP	0
9	11	MP	0
10	1, 4	MP	0
11	1	AM	0
12	9	AM	3
13	12	AM	3
14	10	AM	0
15	13	AM	3

Final Selections:

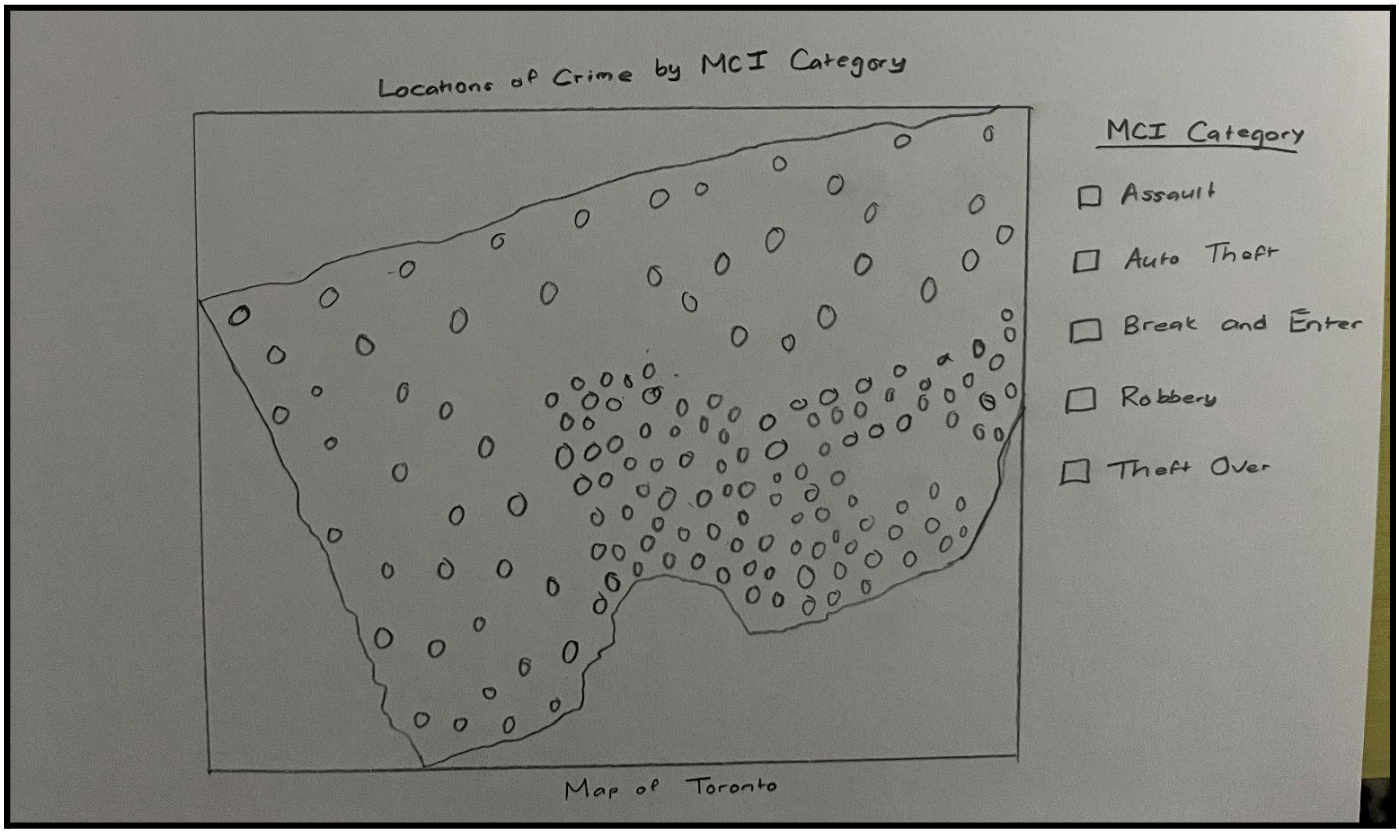
Sketch ID: 13

- Q12) Which Toronto neighborhoods are the safest to walk alone at night ranked on an Index Safety Score?
- (Neighbourhood Crime Rates Open Data - Toronto Police Service Public Safety Data Portal)



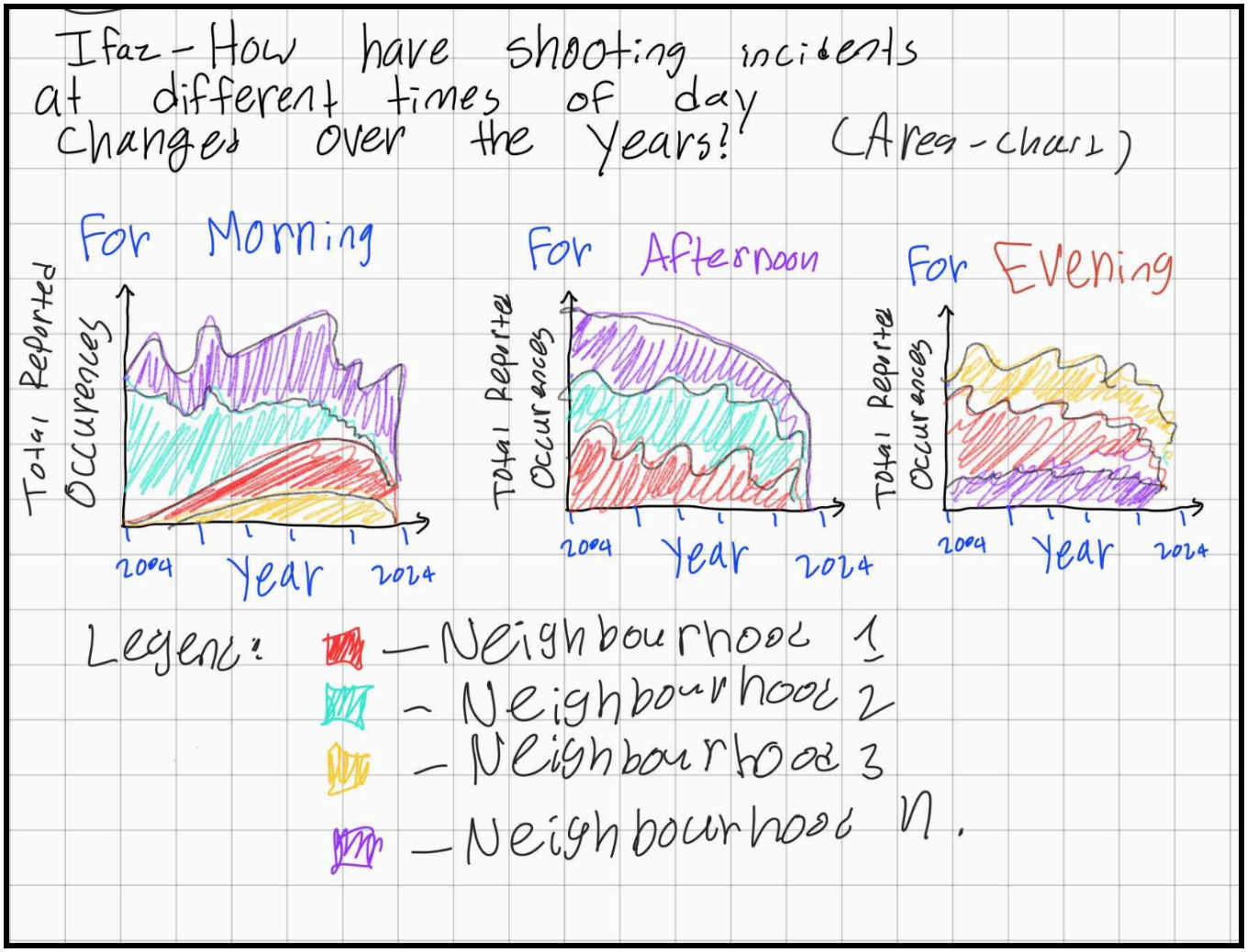
Sketch ID: 7

- Q4) What are the most common types of crime and do they vary by area? (Urban vs Suburban areas)
- (Major Crime Indicators Open Data - Toronto Police Service Public Safety Data Portal)



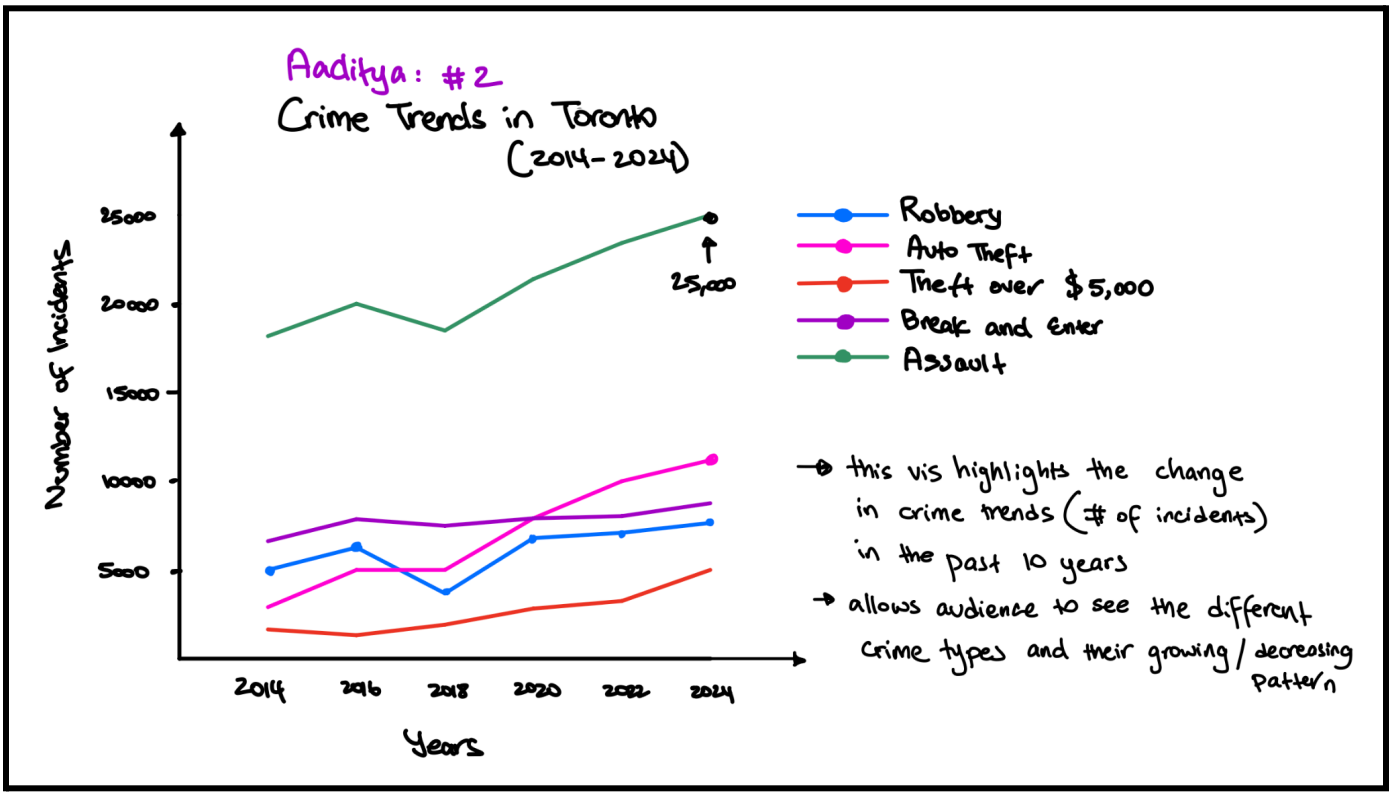
Sketch ID: 3

- Q8) Is there a certain time group that the crimes commonly occur? (Evenings or morning, or afternoon)), or how has the frequency of crimes regardless of timing changed over the years?
- [(Shooting and Firearm Discharges Open Data)]



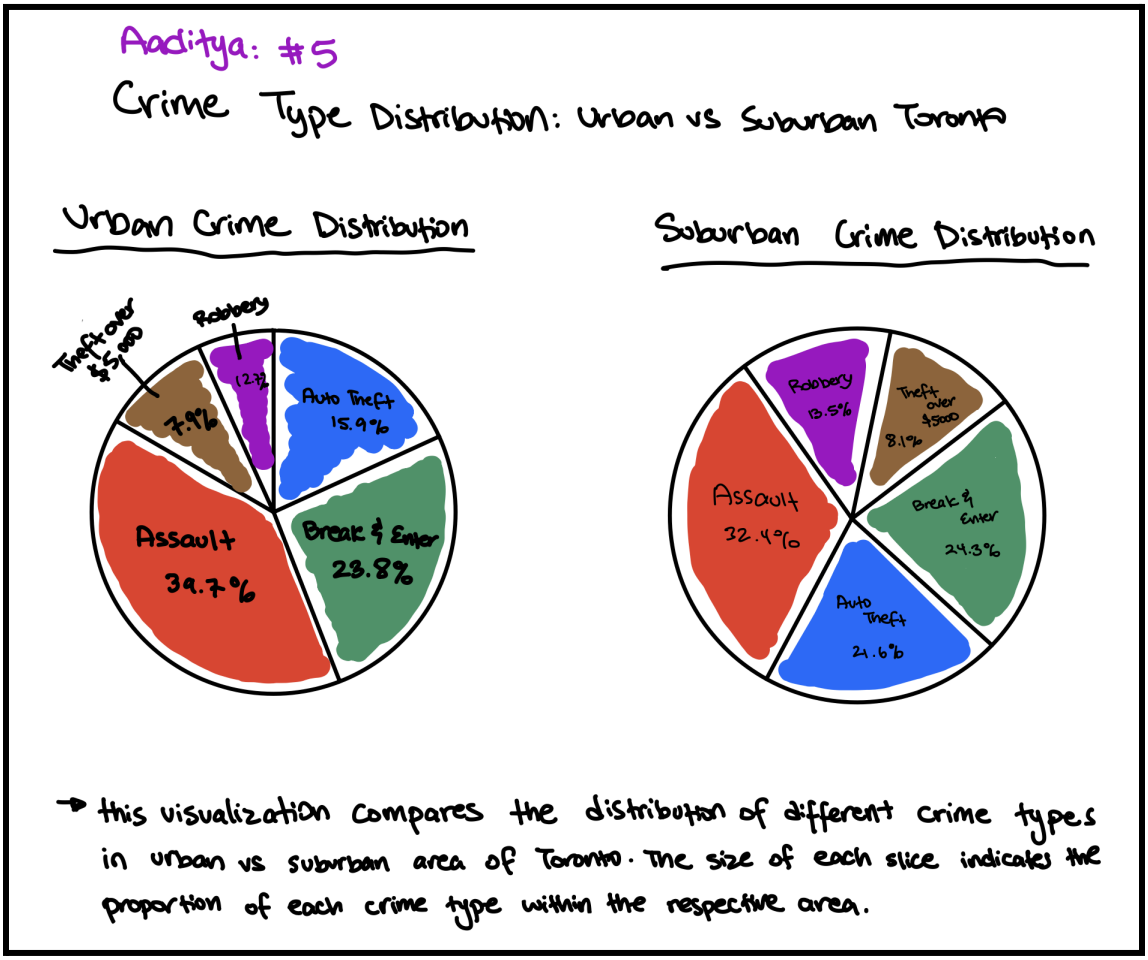
Sketch ID: 12

- Q9) Has crime in Toronto increased or decreased over the past 5-10 years?
- (Neighbourhood Crime Rates Open Data - Toronto Police Service Public Safety Data Portal)



Sketch ID: 15

- Q13) How does the distribution of crime types differ between urban and suburban areas in Toronto? (Neighbourhood Crime Rates Open Data - Toronto Police Service Public Safety Data Portal)



Summary of Selections: The purpose of this project was to disclose the safest and most dangerous locations in the city of Toronto through the analysis of crime statistics. Therefore, our selection of sketches must directly reflect this. Sketch 3 highlights whether gun-related incidents have been on the rise or decreasing over the years for each time of day - this is especially important because one of our main points was how safe it is to walk at night. Furthermore if the user wishes for more information, they also have the opportunity to assess how safe it is to walk in earlier time groups throughout the day too. Sketches 7 and 13 show the locations of major crime incidents (MCIs) and the safety index by neighbourhood respectively. This helps the viewer determine the neighbourhoods that they may want to relocate to or work at in the future. Meanwhile, sketches 12 and 15 present more detailed information that inform the viewer about how the frequency of MCIs have changed over the course of a decade, as well as the difference in the type of crimes occurring in urban and suburban areas. Ultimately, these sketches complement the message we are trying to convey through this project.

Week 6 | Storyboarding Step

Insights:

- Evening to Night Time has been shown to document the majority amount of crime incidents - this pattern holds across various types of crimes, including assault, robbery, and firearm-related incidents; Crime Incidents occurring at night are 2.5 times higher than in the morning (Aaditya)
- The number of shooting reports has risen from approximately 190 in 2004 to 461 in 2024 according to the Firearm and Shooting Discharges Open Data. (Ifaz)
- According to Major Crime Indicators Open Data, occurrences of assault logged from 2014-2024 cumulatively appear to be more frequent on Saturday (3650 reports) and least frequent on a Tuesday (3129 reports). (Ifaz)
- Crime distribution is not uniform in Toronto as certain neighbourhoods experience significantly more crime than others. People living in Downtown Toronto neighbourhoods have a greater likelihood of committing or being a victim of criminal activity. (Meet)
- Urban and suburban crimes vary in the frequency and type of MCIs. People living in suburban communities experience more auto theft and break and enter incidents. Meanwhile, urban areas see more assault and robbery. (Meet)
- Crime trends in the past decade have a tremendous increase per year - with assault cases rising from 16,825 in 2014 to 25,819 in 2024, along with auto theft with 3,786 in 2014 to 9,352 in 2024. (Aaditya)

Main Insight:

- Crime in Toronto is significantly influenced by both time and location, making nighttime travel in certain neighborhoods particularly unsafe. While overall crime rates have risen over the past decade, the risk is highest in the evening and in specific high-crime areas. By analyzing these patterns, we can construct a safety index for each neighborhood, allowing individuals to make informed decisions about when and where to travel alone—ultimately improving personal safety.

Storyboard:

Our storyboard is structured according to the four key elements of a data story:

1. The hook ➜ Introducing the broader issue of crime in Toronto.
2. Rising insights ➜ Exploring trends over time and specific crime distribution.
3. The message ➜ Looking at how the crime trends have been increasing and is based on time and location.
4. The solution ➜ Providing a safety index to inform safer decisions.

Viewing through the Miro Link is recommended.

Miro Link: https://miro.com/app/board/uXjVIYNwsh0=/?share_link_id=106354559355
PDF Format Link: Safe Street Squad - Storyboarding.pdf
Datasets [Google Drive Folder]: Cleaned Datasets (To Use)

Week 7 | Prototype V1

Link to GitHub repository: https://github.com/aadityamandal/safe-street-squad-crime-analysis

Students who worked on the prototype:

- Aaditya Mandal
- Ifaz Alam
- Meet Patel

In accordance to the requirements of completing at least two D3 visualizations, we have implemented the following:

1. A line chart for the Hook section that models Sketch ID 12. (Meet)
2. A line chart for the Rising Insight 1 section that models Sketch ID 3. (Ifaz)
3. Two pie charts for the Message section that model Sketch ID 15. (Aaditya)

All other listed requirements have been met as well.

Week 7 | Prototype V2

New additions:

1. LineChart: (Ifaz)
    a. In Prototype V1 we added filtering based on a date range. We now added further filtering based on category type so the user can choose how many lines they want to see at the given moment. By default we show all the lines. (Ifaz)
    b. Added smooth transitions for the line chart when the date range is modified. Also upon hiding a category, the corresponding line fades out. (Ifaz)
    c. Added circle points to the lines to explicitly indicate a separate datapoint. These circles also have their own individual transitions for when they enter the canvas / exit the canvas. (Ifaz)
    d. Fixed position for the Y-Axis label of the line chart so that it doesn't intersect with the ticks. (Ifaz)
    e. Added flexibility for what type of curve we want to display on the line chart. Some examples are d3.curveLinear (for straight lines) or d3.curveCatmullRom (for curved / rounded lines). (Ifaz)

2. A map chart for the Rising Insight 2 section: (Meet)
    a. Displays the location of MCI crimes on a map of Toronto and its neighbourhoods.
    b. Each MCI crime type is given its own unique color on the map.
    c. Unfortunately, due to Meet's sickness, the option to filter out specific MCI crimes and update the map accordingly has yet to be implemented. This will be completed by the final deadline.
3. Solution Visualization Design + Algorithm (Aaditya)

a. Implemented a dynamic map-based visualization that allows users to enter their starting and destination locations to evaluate safety based on crime data.
b. The route is drawn between the two points with a dashed line, which dynamically adjusts in size based on the zoom level for better visibility.
c. The system calculates a Safety Score based on reported Major Crime Indicators (MCI) along the route.
d. The safety score of walking from Point A to Point B is **weighted based on**:
    i. **Crime severity** (e.g., Assault and Robbery are weighted higher than Theft).
    ii. **Proximity of crimes** to the walking path.
    iii. **Recency of crimes** (Recent crimes have a higher impact on the safety index - 2024 crimes vs 2014 crimes)
    iv. **Route distance** (Longer walks slightly lower the safety score due to increased risk exposure).
e. Users receive a **final safety rating**, along with a recommendation, such as: “Safe to Walk”, “Walk with Caution”, “Better to use another mode of transportation, rather than walking”, “Extremely dangerous to walk, not recommended”
f. This visualization allows users to be able to see which destinations are safe for walking (given the areas that they must walk through)