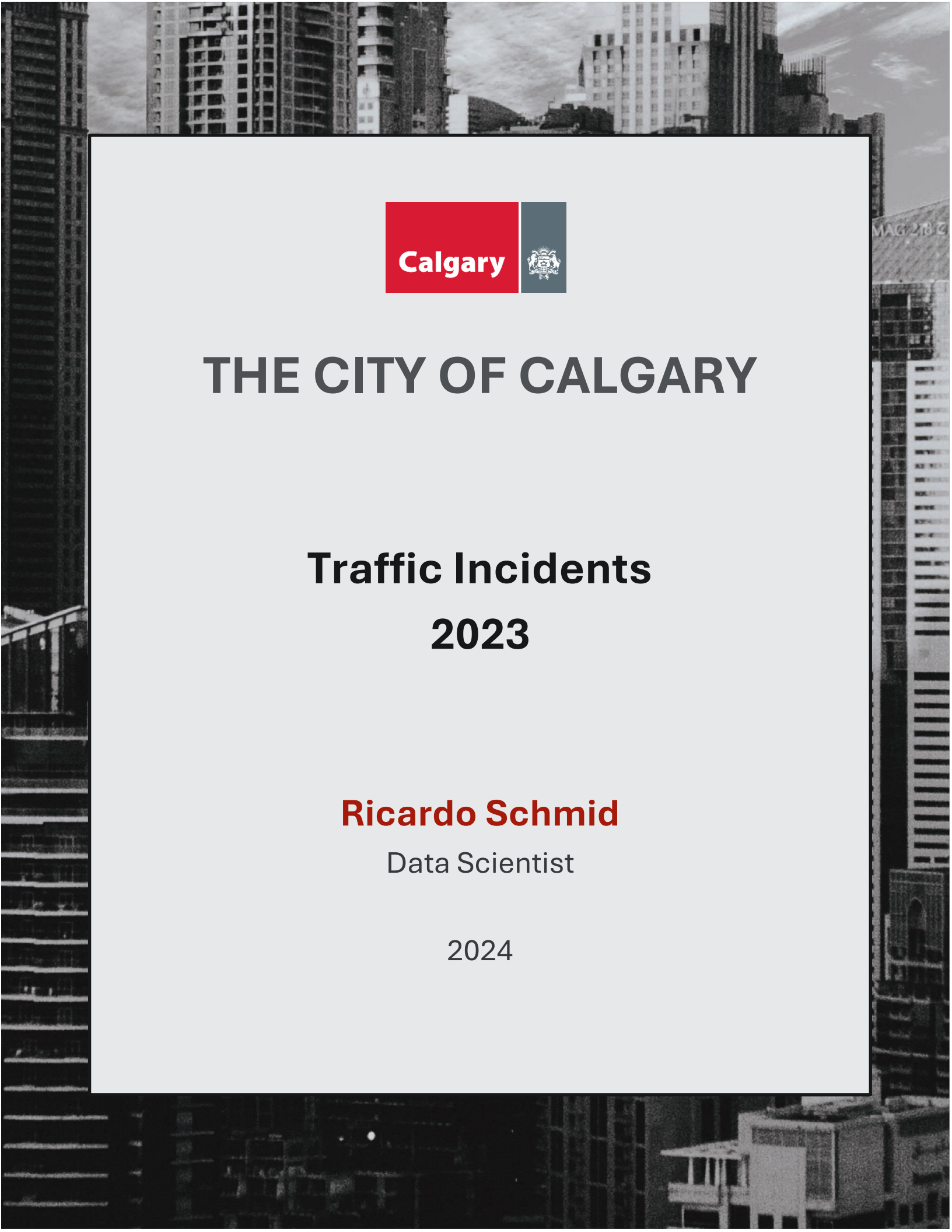

THE CITY OF CALGARY

Traffic Incidents
2023

Ricardo Schmid

Data Scientist

2024

Table of Contents

1. Introduction

1.1. Purpose

Calgary is experiencing rapid growth, leading to an increase in population density and, consequently, a rise in traffic incidents across the city is expected. The primary objective of this project is to conduct a comprehensive analysis of traffic incidents in Calgary, enabling policymakers to identify patterns, trends, and critical areas for intervention to reduce the number of incidents.

1.2. Objectives

This project covers the following analysis:

1. Determine the times (hours) when traffic incidents are most frequent.
2. Identify regions with the highest number of traffic incidents.
3. Identify trends in traffic incidents over time (daily, weekly, monthly, yearly).
4. Assess how different weather conditions (e.g., lower temperatures) affect the number of incidents.

2. Data Collection

The analysis was based on three datasets:

- **Traffic Incidents**: This dataset includes accident descriptions, dates, times, latitudes, longitudes, and addresses of incidents. It can be accessed from Calgary Traffic Incidents Dataset.
- **Neighborhoods Dataset**: This dataset provides Calgary's neighborhoods' latitude and longitude boundaries (Multipolygon) and sector information. Available at Calgary Community Boundaries Dataset.
- **Hourly Temperature**: This dataset contains hourly temperature records for the period covered by the incident data. It is sourced from Canada Historical Climate Data.

3. Data Preparation

Python was used to obtain the neighborhood and sector information for each traffic incident. The Traffic Incidents table was merged with the Neighborhoods dataset based on the geographic boundaries defined in the Multipolygon column. After the initial merge, some incidents still had missing values for neighborhood and sector information, for these incidents, spatial analysis was used to determine the closest neighborhood and sector based on the incident's latitude and longitude coordinates.

The hourly temperature data was merged with the incident dataset to include only the relevant temperature records based on the date and time of each incident. The left join ensured that all incident records were preserved, even if some did not have corresponding temperature data. A GeoJSON file containing the sector and neighborhood coordinates was created using Python. This file was then converted to TopoJSON format using Mapshaper for visualization and analysis in Power BI.

4. Data Model

In power BI, the dataset was split into three tables to form one fact table and two dimension tables. The diagram on figure 1 below, illustrates the relationship between the Traffic_Incidents, Neighborhood_Lookup, and Calendar_Lookup tables. Traffic_Incidents is the fact table and is connected to the Calendar lookup and Neighborhood_Lookup table by the foreign keys, Date, and Neighborhood ID.

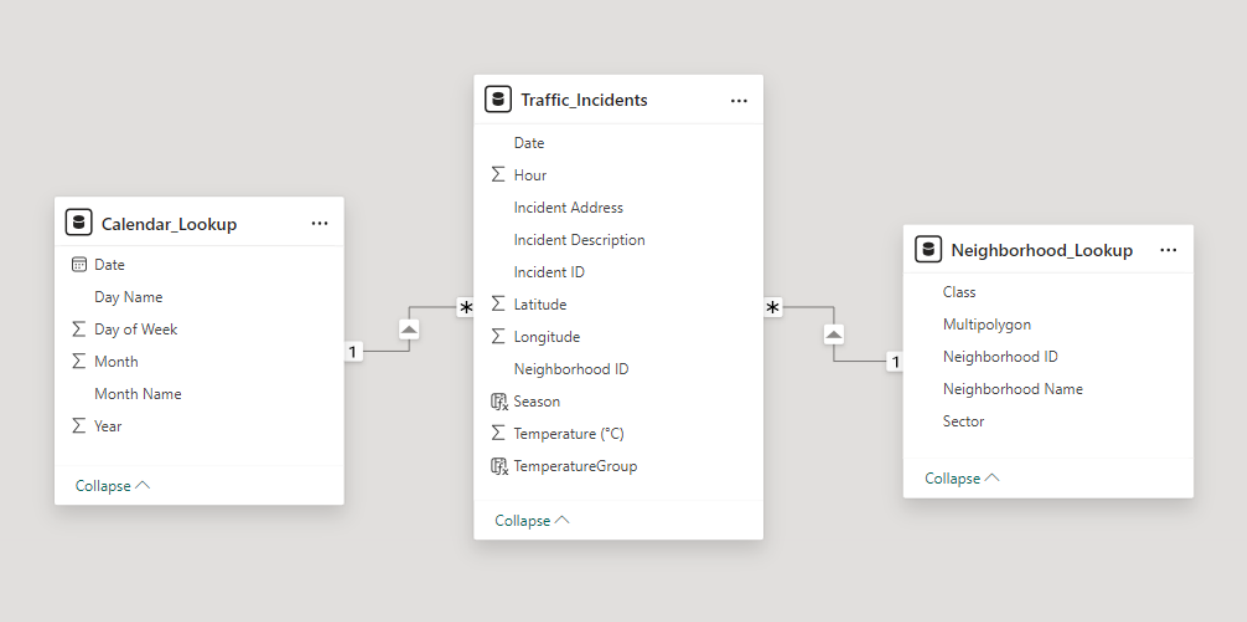

Figure 1 – Data Modeling

5. Analysis

5.1. Overview

Figure 2 illustrates the population growth in Calgary and the percentage increase over the years. Between 2020 and 2023, the population grew by 5%. When comparing the annual growth rates, the population increased by 1.645% from 2019 to 2020, by 2.06% from 2020 to 2021, by 1.31% from 2021 to 2022, and by 1.54% from 2022 to 2023.

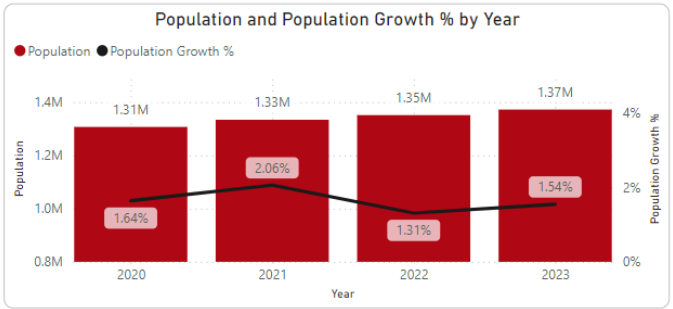

Figure 2 – Population Growth

Figure 3 shows the number of incidents per year. There was a significant increase from 2020 to 2021, which may be related to COVID-19 restrictions and the shift to working from home. The number of incidents grew by 20% from 2020 to 2021, increased by another 8.5% in 2022, and then decreased by 5.2% in 2023.

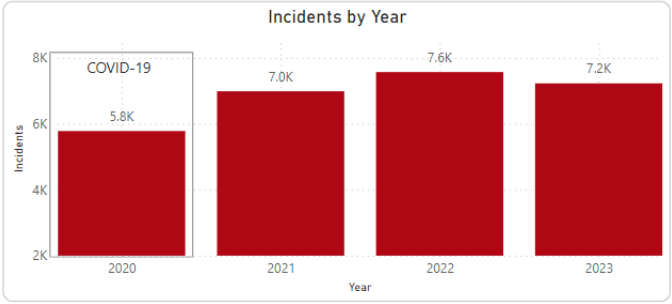

Figure 3 – Number of Incidents Over the Years

Analyzing the relationship between the number of incidents per 1,000 people reveals a pattern similar to the total number of incidents. The rate increased by 18.28% from 2020 to 2021 and by 6.87% from 2021 to 2022, before decreasing by 5.89% from 2022 to 2023. This decline may reflect the effectiveness of actions taken by the city. The decrease occurred not only in the rate per 1,000 people but also in absolute numbers, which is a positive sign.

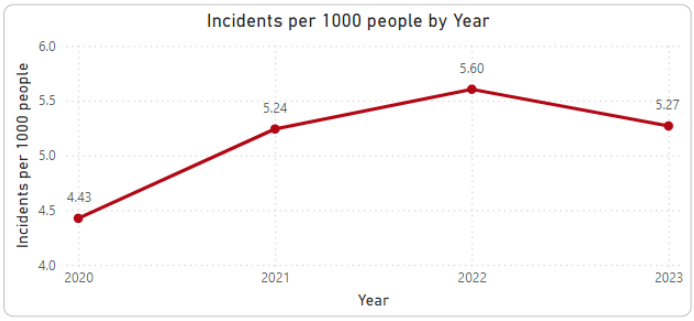

Figure 4 – Incidents Per 1,000 People Over the Years

5.2. Neighborhood mapping

The dashboard in Figure 5 presents a shape map where neighborhoods are colored according to the number of incidents—the darker the color, the higher the number of incidents. It is evident that the Centre sector of the city has the highest number of incidents. Additionally, the horizontal bar chart displays the top five neighborhoods with the highest number of incidents..

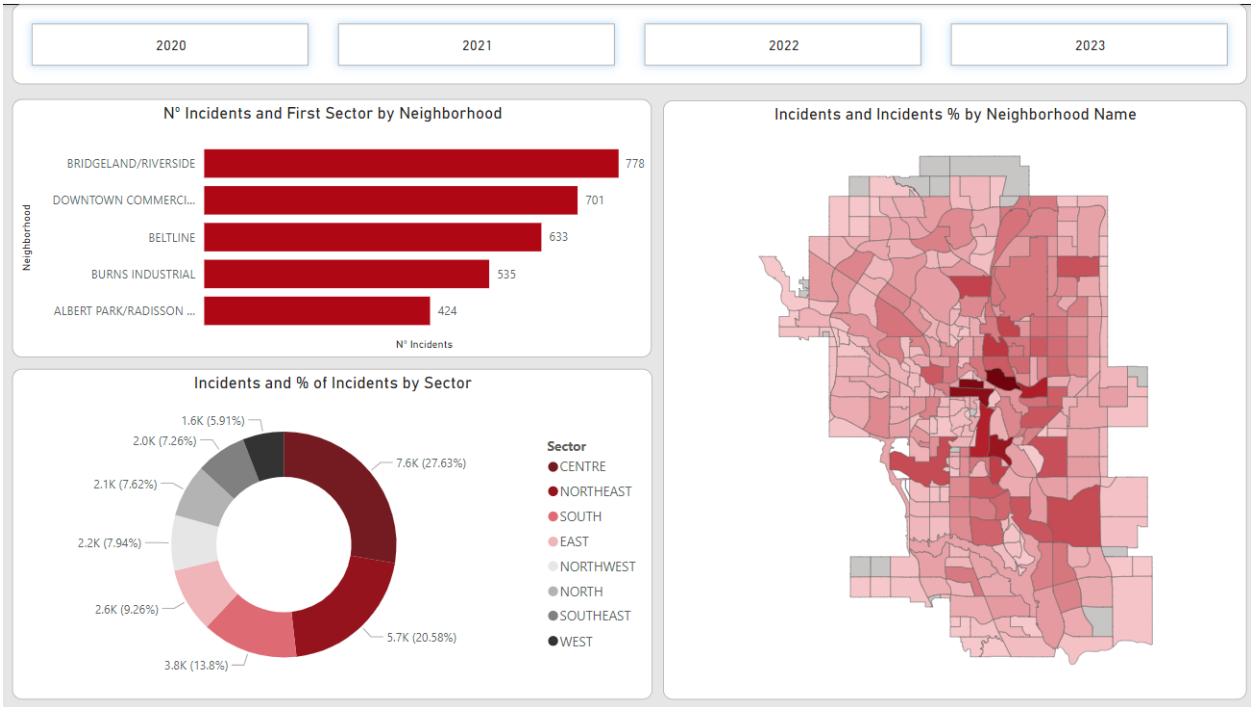

Figure 5 – Neighborhood Map Showing Incident Distribution

The donut chart in Figure 6 shows that, when combined, the Centre and Northeast regions accounted for 46.76% of the total incidents in 2020 and 48.68% in 2023. This indicates a consistent prevalence of higher incident rates in these two sectors of the city throughout the period.

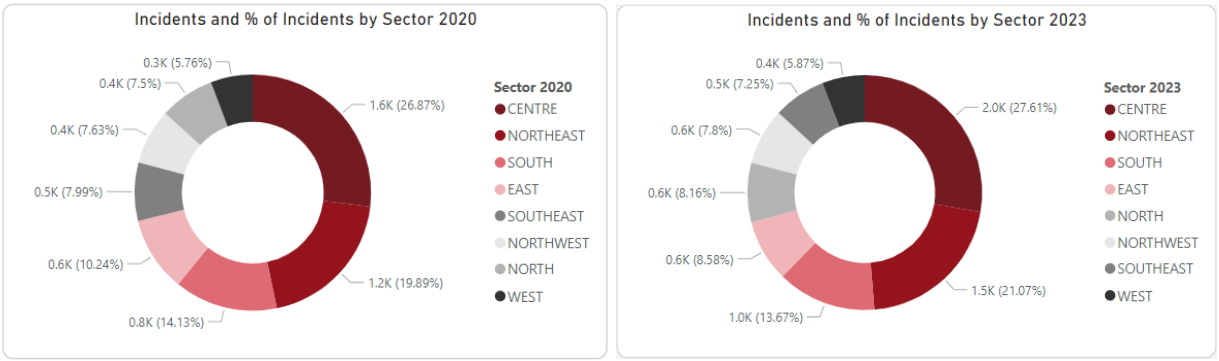

Figure 6 – Comparison of Incidents by Location: 2020 vs. 2023

In Figure 7, the bar chart highlights the top five neighborhoods with the highest number of incidents within the Centre sector. The chart shows that the neighborhoods with the most incidents in the Centre sector are Bridgeland, Downtown, Beltline, Burns Industrial, and Manchester Industrial.

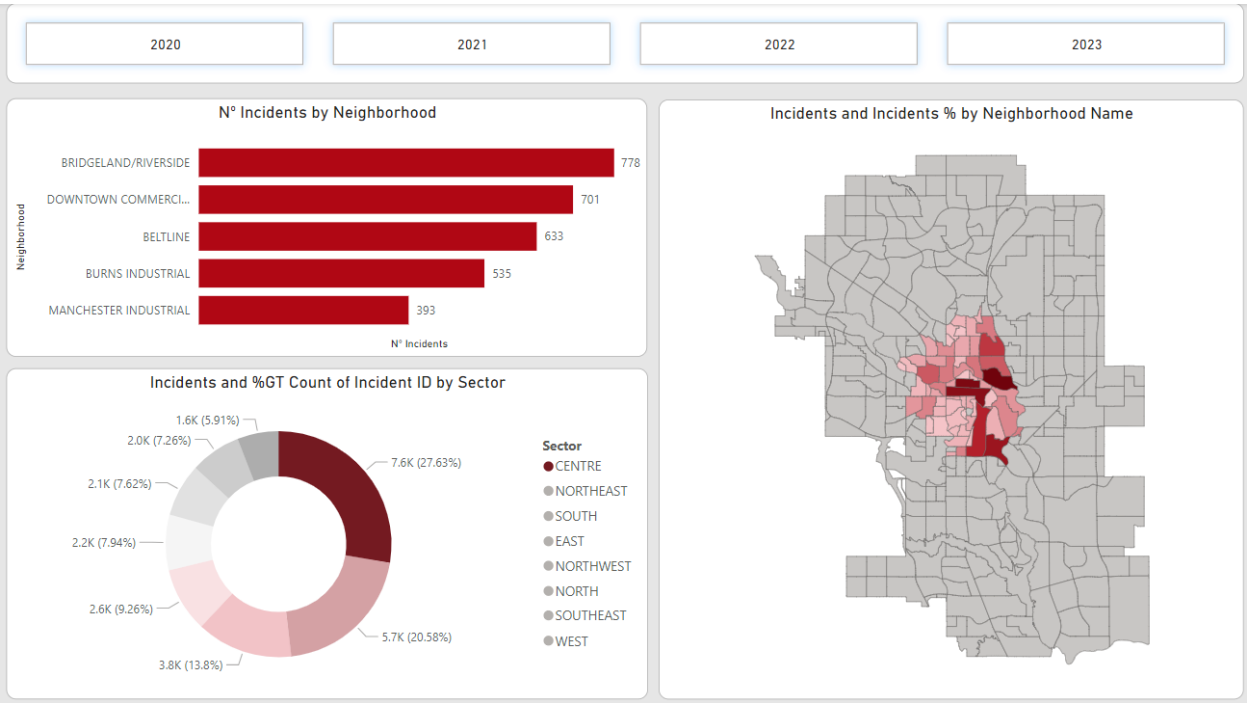

Figure 7 – Incidents in the Centre sector

Figure 8 highlights the darkest neighborhood on the map for the Centre sector, which is Bridgeland, indicating the area with the highest number of incidents. A tooltip is also provided to display both the number of incidents and the percentage of the total for the region.

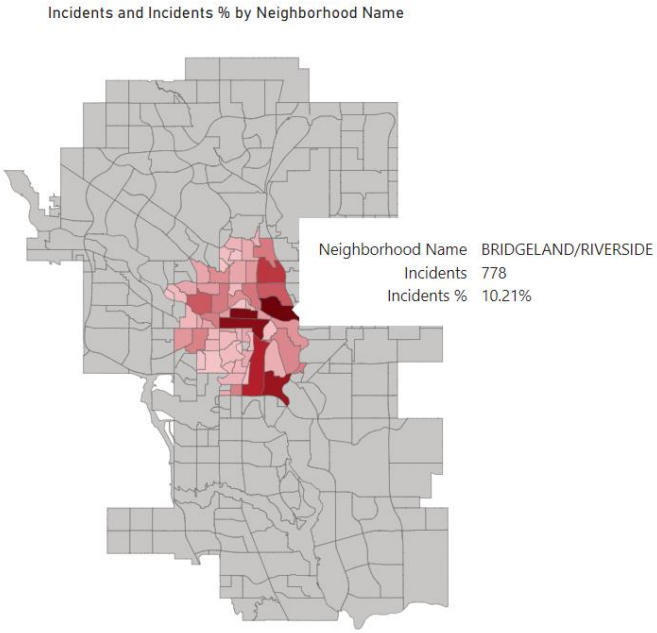

Figure 8 – Highest Incidence in the Centre Sector: Bridgeland

The bar chart in Figure 9 shows the top five neighborhoods in the Northeast region with the highest prevalence of incidents: McCall, Saddle Ridge, Sunridge, Marlborough, and Mauland Heights.

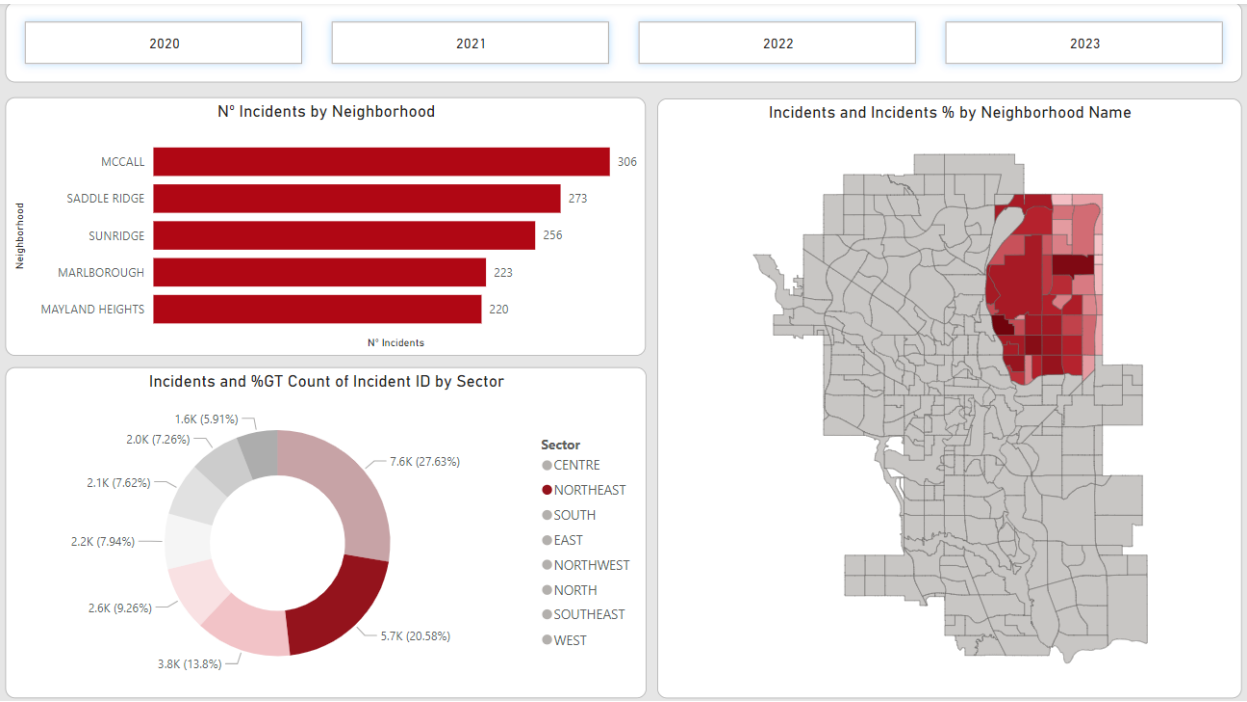

Figure 9 – Incidents in the Northeast Sector

Figure 10 shows McCall as the darkest point on the map for the Northeast sector, indicating the highest incident rate in that area.

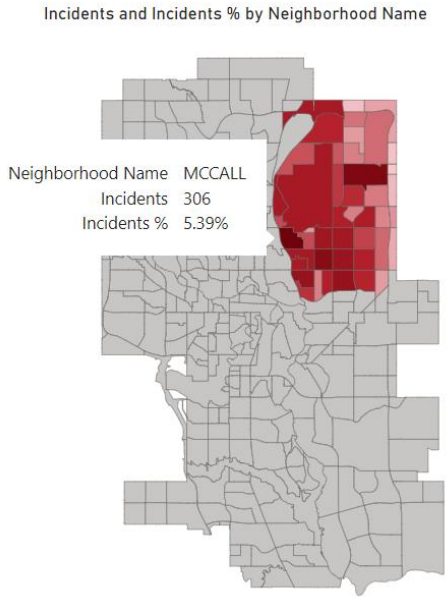

Figure 10 – Highest Incidence in the Northeast Sector: Mccall

Finally, the analysis of incident locations reveals that the distribution of incidents in 2020 is very similar to that in 2023, as shown on Figure 11. This indicates that the areas with high incident rates

in 2020 remain largely unchanged, suggesting that the actions taken to reduce incidents in these areas have not been effective.

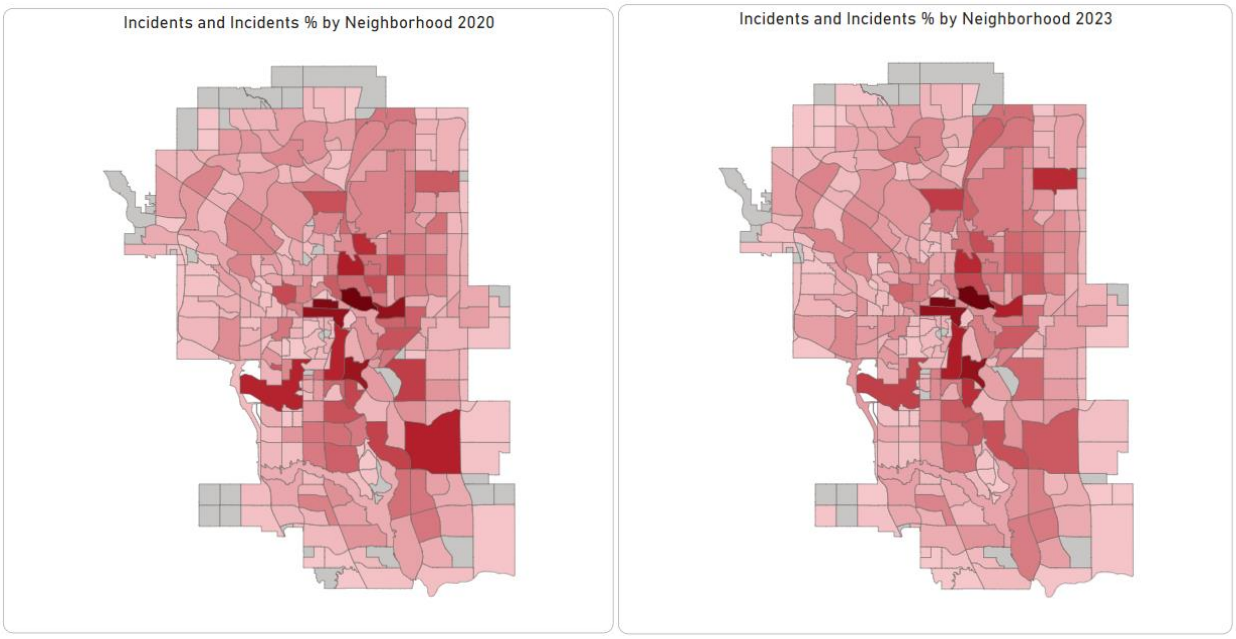

Figure 11 – Comparison of Incidents by Neighborhoods: 2020 vs. 2023

According to the graphs above, the region that requires most of the city attention is the Centre, more specifically, Bridgeland, Downtown, Beltline and Burns Industrial. Also, the Albert Park neighborhood, which is close to downtown, requires some attention too, it comes in 5th overall. As per the regions, Northeast requires attention after the centre.

5.3. Time Analysis

Figure 12, which displays the monthly distribution of incidents, shows that December has the highest number of incidents throughout the period, suggesting a peak during the winter month. Additionally, the hourly graph reveals two significant peaks during rush hours, indicating a notable increase in incidents at these times. This pattern highlights a strong correlation between traffic congestion and the occurrence of incidents. Furthermore, the trend suggests that incidents are more likely to occur on weekdays when traffic volumes are higher due to daily commuting. Thus, addressing traffic congestion during peak times could potentially help reduce the number of incidents.

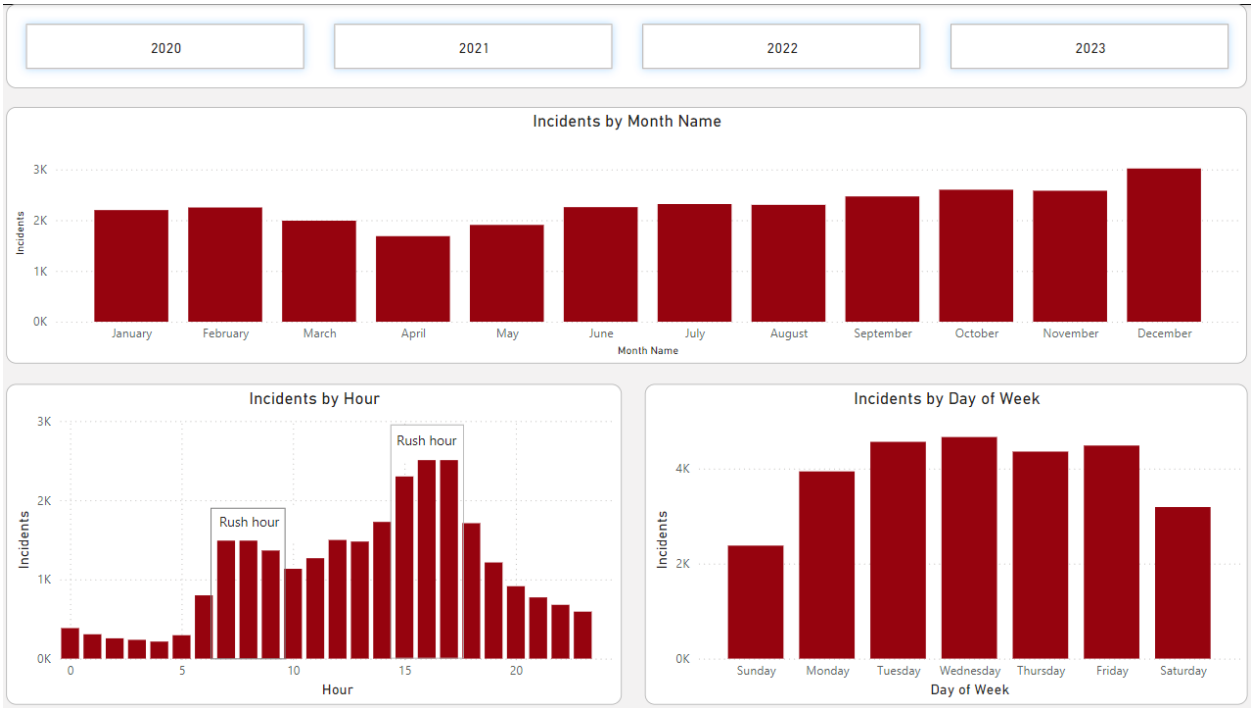

Figure 12 – Month, Hour and day of the week analysis

5.4. Temperature Analysis

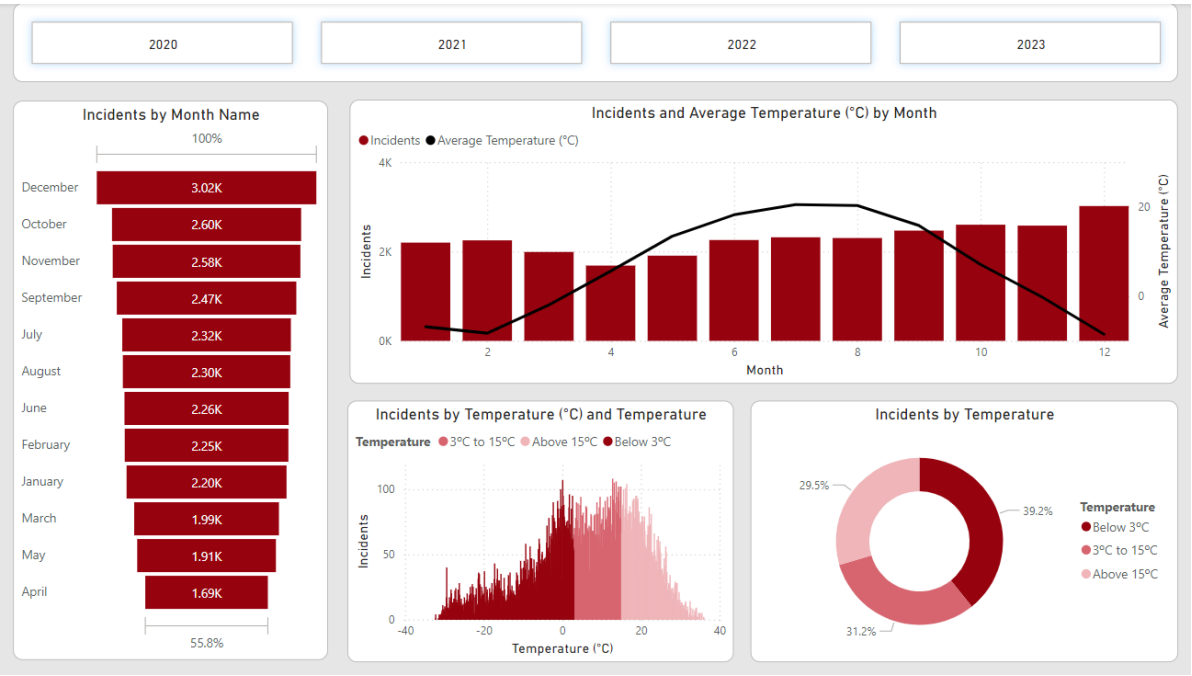

Figure 13 – Temperature x Incidents

Moving to the temperature analysis in Figure 13, the graph showing incidents and average temperature by month reveals an inverse relationship between the number of incidents and

temperature: as the temperature drops, the number of incidents tends to increase. Analysis of the months with the highest number of incidents over the years indicates that the peak months for incidents were cold months, specifically December, October, and November, in that order.

Regarding temperature, incidents were most prevalent when temperatures fell below 3ºC, accounting for almost 39% of the total incidents. This correlation suggests that lower temperatures, which can be associated with snow and ice on the roads, contribute to a higher risk of incidents.

6. Conclusion

In terms of location, it is evident that the same regions consistently rank among the top for the number of incidents. This indicates that the city's efforts to address issues in these areas have been insufficient or ineffective. To improve safety, traffic controls such as speed bumps, traffic circles, and additional signage should be considered.

Regarding time analysis, rush hours are a significant factor contributing to the increased number of incidents, particularly on weekdays. Enhancing public transportation incentives and improving the public transportation system could help reduce traffic congestion and, consequently, the number of incidents.

In terms of climate, lower temperatures are associated with a higher number of incidents. To mitigate this, the frequency of road salting and plowing should be increased to improve road safety during colder months.