SafeNav - Ottawa's Traffic Collision Data Analysis

Prepared by:

Arit Akpan

Esin Bilgin Savkli

Eduardo Manotas

Julia Saavedra

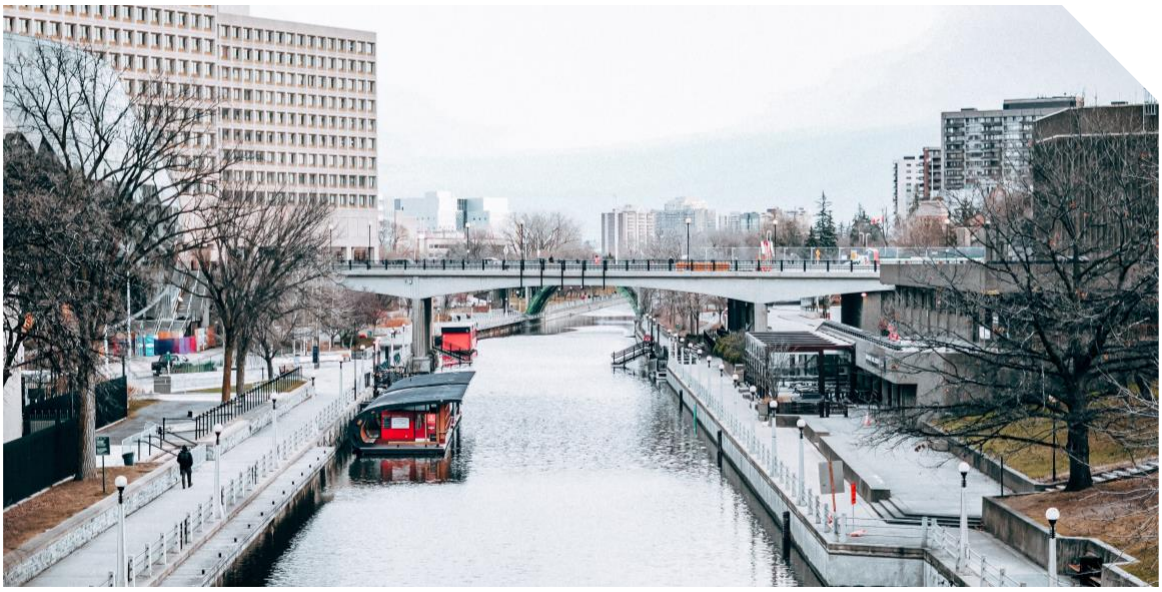

Image from: https://unsplash.com/

Algonquin College

Data Visualization Topics

Siju Philip

December 9, 2023

Content

Executive Summary

The SafeNav project is a comprehensive initiative focused on analyzing Ottawa's traffic collision data to enhance safety strategies and gain a profound understanding of factors impacting road incidents. The project's objective is to leverage temporal analysis for proactive interventions in traffic safety measures, aiming to reduce accidents and make well-informed decisions.

The project began with two fundamental assumptions: the accuracy and reliability of data provided by the City of Ottawa and the potential of temporal analysis to reveal hidden patterns within traffic collision data.

A diverse array of research methods, including data cleaning, feature engineering, and predictive modelling with Python and machine learning, were employed to analyze Ottawa's traffic collision data thoroughly. The project team, consisting of four members, worked collaboratively to ensure the project's success.

Data Analysis

From 2017 to 2022, Ottawa witnessed 71,040 reported accidents, resulting in 13,290 injuries and 140 fatalities. The analysis revealed temporal trends, univariate findings, bivariate insights, and multivariate factors impacting accident counts. Unexpectedly, positive environmental conditions and traffic signals in place emerged as key influencers.

Various visualization techniques, including smart narrative, maps, card visualizations, slicers, tree maps, key influencers, line charts, clustered bar charts, and stacked column charts, were employed to present the data comprehensively.

Conclusion & Recommendations

The analysis provides critical insights into road safety in Ottawa, challenging preconceived notions about accident occurrences. The findings suggest the need for targeted safety measures during winter, Fridays, and afternoons.

A Power BI dashboard was developed to consult the observations, allowing filtering by main categories. Recommendations focus on environmental conditions, seasons, and days of the week with the highest accident counts. The dashboard serves as a tool for informed decision-making to enhance road safety in the city.

In conclusion, the SafeNav project's comprehensive analysis of Ottawa's traffic collision data not only sheds light on key insights and unexpected patterns but also provides actionable recommendations for improving road safety. The findings underscore the importance of targeted interventions during specific environmental conditions, seasons, and days of the week.

Background

Project Scope and Purpose

The SafeNav project is a comprehensive initiative aimed at analyzing Ottawa's traffic collision data to uncover patterns, enhance safety strategies, and gain a profound understanding of factors impacting road incidents. The project's scope encompasses the exploration of temporal patterns and the identification of high-risk zones for proactive interventions in traffic safety measures. The objective is to leverage temporal analysis to reduce accidents and make well-informed decisions regarding traffic safety measures.

Initial Assumptions

The project began with two fundamental assumptions are the data provided by the City of Ottawa is accurate and reliable and temporal analysis can reveal hidden patterns and insights within traffic collision data.

Research Methods and Tools

Throughout the SafeNav project, a diverse array of research methods was skillfully employed to thoroughly analyze Ottawa's traffic collision data. The process involved meticulous data cleaning, feature engineering, and the proficient use of Power Query Editor for data transformation [1][2]. To enhance data interpretability, measures were employed to compute aggregated values for data visualizations [3][4]. Additionally, a well-structured star schema was implemented to optimize data modelling, ensuring efficient data relationships. Notably, predictive modelling played a significant role, with Python and machine learning methods applied to anticipate future trends and outcomes based on historical data and aggregate the high-risk zones [5]. These predictive models enriched the analysis, while Power BI and Power Query Editor facilitated data preparation and transformation. Moreover, temporal analysis emerged as a key method, shedding light on patterns and insights within the data. Collectively, these research methods empowered the project team to gain a comprehensive understanding of Ottawa's traffic collision data, informing proactive measures for road safety enhancement.

Project Stakeholders

Esin Bilgin Savkli: Responsible for data preparation and modeling.

Arit Akpan: Focused on data visualization and analysis.

Eduardo Manotas: Responsible for predictive modelling.

Julia Saavedra: Provided recommendations based on data insights.

Data Analysis

Introduction

The city of Ottawa has witnessed a total of 71,040 reported accidents from 2017 to 2022, resulting in 13,290 injuries and 57,746 incidents of property damage. Unfortunately, 140 fatalities occurred during this period. This paper delves into the findings of exploratory data analysis, aiming to shed light on key insights, unexpected patterns, and visualization techniques employed for a comprehensive understanding.

Data Exploration

The analysis begins with an overview of accident counts over the years. In 2019, Ottawa experienced a peak with 16,437 accidents, a stark contrast to 2022, which recorded the lowest count at 5,937. Univariate analysis indicates that vehicle accidents outnumber bicycles and motorcycles. Temporal trends reveal a notable decrease from 2019 to 2020, with 10,049 accidents reported. Property damage is the most common classification, occurring in 57,749 accidents, followed by non-fatal injuries (13,150) and fatal injuries (140).

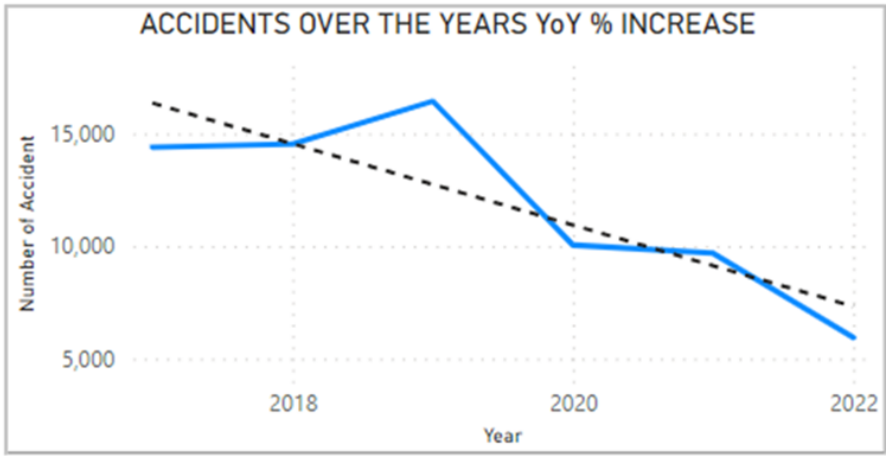

Figure 1. Temporal Trends in Traffic Accidents (2017-2022)

Bivariate Analysis

Minor injuries consistently dominate across the years, with winter accounting for 31.79% of accidents and 35.68% happening in the afternoon. Unexpectedly, Friday emerges as the day with the highest accident count (12,331), representing 17.36% of all accidents.

Multivariate Analysis

Several unexpected findings emerge from the analysis showing that the highest accident counts are associated with traffic signals in place; and positive environmental conditions like clear weather and broad daylight, challenging common perceptions. Although counterintuitive, these findings suggest other indicators like the human factor as a contributory factor to accident counts.

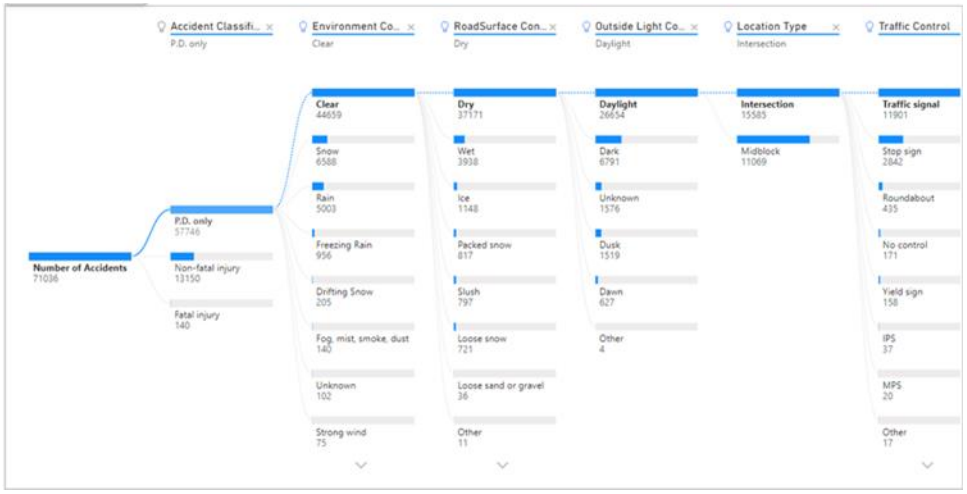

Figure 2. Positive Key Influencers

Visualization Techniques Implemented

Various visualization techniques were employed to present the data comprehensively:

- Smart Narrative: Offers a summary of key insights.
- Maps: Highlights accident hotspots, aiding in spatial analysis.
- Card Visualizations: Displays total accident counts for impactful communication.
- Slicers: Allows users to drill through the data on the dashboard for a clearer picture.
- Tree Map: Provides a breakdown of occurrences, identifying potential influencers.
- Key Influencer: Identifies variables impacting accident counts.
- Line Charts: Illustrates temporal trends in accident counts over the years.
- Clustered Bar Charts: Visualizes comparisons between different categories.
- Stacked Column Chart: Presents data in a stacked format for better understanding.

Risk Clustering

This step was leveraged using 2 relevant Python libraries for Geo clustering: K means from Scikit learn and geo pandas, the process involved dividing the city into 2413 grids that were placed into 5 different clusters based on 2 features, the number of accidents and the number of injuries during the examination time from 2017 to 2022.

Cluster	# of Grid	% Grid	Avg Total Accidents	% Total Accidents
Critical	9	0,4%	770,1	9,8%
High	38	1,6%	341,8	18,3%
Moderate	118	4,9%	176,8	29,4%
Low	252	10,4%	68,2	24,2%
Rare	1996	82,7%	6,6	18,4%
Grand Total	**2413**	**100,0%**	**29,4**	**100,0%**

Finally, We can see that the layout of geographical clusters seems to match the pattern of the grids. Despite only making up 6% of the total grids, the areas identified as Critical, High, and Moderate clusters account for 60% of all accidents. This suggests that a small proportion of the grid areas are responsible for a large percentage of accidents.

Conclusion & Recommendations

The analysis of traffic accidents in Ottawa from 2017 to 2022 provides valuable information that can guide the initiatives to be implemented regarding road safety in the city. As a tool to consult these observations, a dashboard was developed in Power BI that allows filtering by the main categories identified or combining them if necessary [6]. Of all the recommendations that can be found on the dashboard, 3 stand out, which correspond to the categories with the highest values of total accidents, these are:

Recommendations – Environment

In this category, most accidents occur in conditions such as clear and snow. The main recommendations are:

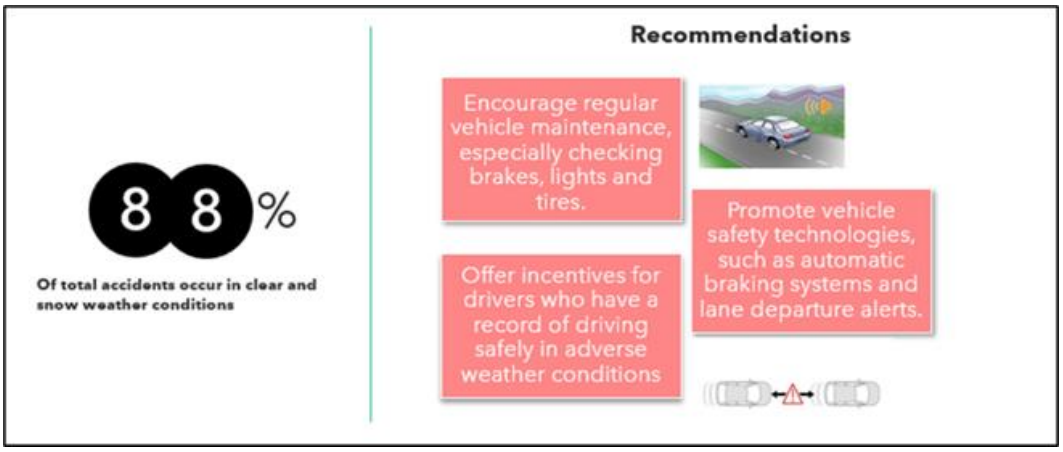

Figure 3. Recommendations – Environment

Recommendations – Season

For this category, it is observed that the greatest number of accidents occur in winter and fall. The recommendations are:

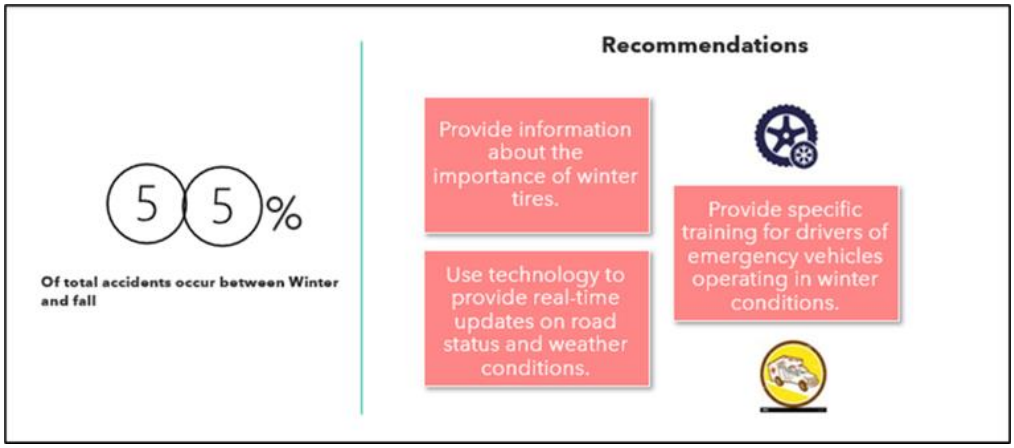

Figure 4. Recommendations - Season

Recommendations – Day of Week

The days on which the most accidents occur were identified, as well as the areas where they occur. Recommendations are:

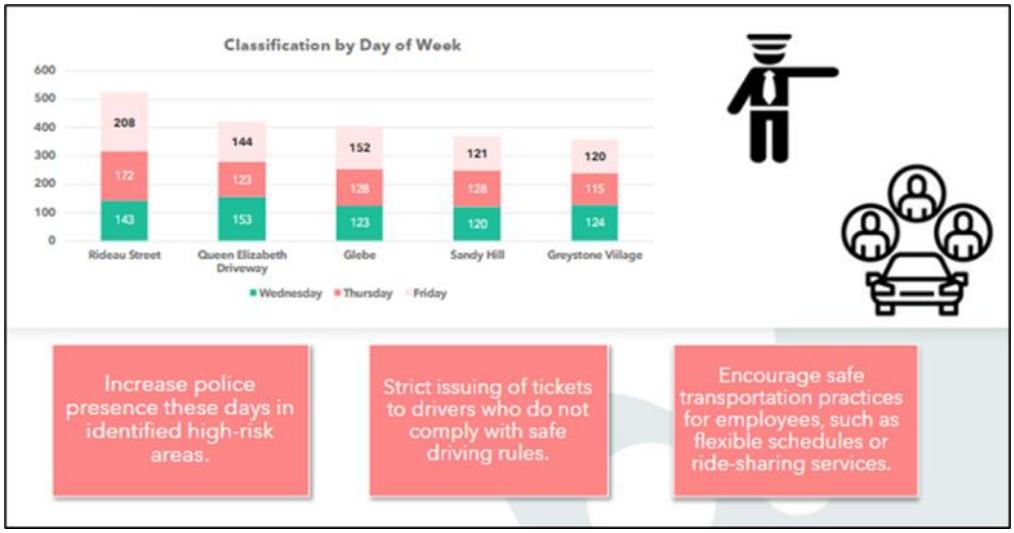

Figure 5.

References

[1] Microsoft, “Modeling view in Power BI Desktop,” 11-Oct-2023. [Online]. Available: https://learn.microsoft.com/en-us/power-bi/transform-model/desktop-modeling-view. [Accessed: 12-Nov-2023].

[2] Microsoft, “DAX function reference,” 20-Oct-2023. [Online]. Available: https://learn.microsoft.com/en-us/dax/dax-function-reference. [Accessed: 12-Nov-2023].

[3] Microsoft, “Power BI visualization: Smart narrative,” 22-Feb-2023. [Online]. Available: https://learn.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-smart-narrative. [Accessed: 16-Nov-2023].

[4] Curbal, “Create ELEGANT LISTS using DAX and the smart narrative visual in Power BI,” YouTube, 08-Mar-2022. [Video]. Available: https://www.youtube.com/watch?v=6L1YmLjmzXA. [Accessed: 16-Nov-2023].

[5] P. Sharma, “Comprehensive Guide to K-Means Clustering,” Analytics Vidhya, 03-Nov-2023. [Online]. Available: https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/. [Accessed: 14-Nov-2023].

[6] City of Ottawa, “Road Safety Action Plan,” 2023. [Online]. Available: https://ottawa.ca/en/parking-roads-and-travel/road-safety/road-safety-action-plan#section-96234999-1bf4-4485-845f-f3bda6f44e9d. [Accessed: 20-Nov-2023].