DOUGLAS COLLEGE

DATA VISUALIZATION – TERM PROJECT

Insights from Reported Crashes in Metro Vancouver for 2018 - 2022: A Visual Analysis Using Tableau

Authors: John Zapana & Jose Cornelio

Student IDs: 300378845 & 300382456

Professor: Rupa Manabala

CSIS 3860 - 070

This Photo bv Unknown Author is licensed under CC BY-NC

April 12, 2024

Table of Contents

1. Introduction

In recent years, road safety has emerged as a critical concern worldwide, with a growing emphasis on understanding and mitigating the factors that contribute to vehicular traffic accidents. Therefore, exhaustive analysis of vehicle accident data is essential to identify trends, patterns, and risk factors, thus allowing the formulation of evidence-based strategies to improve road safety outcomes. In that sense, this project explores the dynamics of vehicular crashes in Metro Vancouver during the period 2018 – 2022 using data from ICBC Reported Crashes and the British Columbia Population.

The data set for this analysis covers a wide variety of data related to reported crashes, including crash severity, location details, vehicle involvement, temporal information, and demographic factors. Through Tableau's powerful visualization tools, raw data is transformed into actionable insights, presenting our findings in a visually compelling and easily interpretable way. Leveraging the capabilities of Tableau encourages data-driven decision-making and drives positive changes in road safety policies.

Having made these observations, the rest of this work is organized as follows. The second section points out the data sources to be used and describes the variables to be used in the analysis of this project. In the third section, the nine analysis questions for this project are listed. The fourth section describes in an orderly manner the steps performed to clean and structure the database. Likewise, the creation of Measures or transformation of Fields to be used in the analysis are explained. The fifth section shows the most relevant visual analysis that responds to each of the questions formulated in the project. Therefore, this section is subdivided into a Time Series Analysis of Crashes, a Geospatial Analysis of Crashes, a Correlation Analysis between the Crashes and Population, a Crash Rates Per Capita Analysis, a Types of Crashes Analysis, a Crash Severity Analysis, a Vehicle and Victim Analysis, a Street Analysis, a Crashes and Victims Forecast Analysis. Finally, in the sixth section, the conclusions that the data allow us to achieve are presented.

2. Dataset Overview

As mentioned above, the main objective of this project is to analyze vehicle crashes in Metropolitan Vancouver during the period 2018 – 2022. In this sense, the availability of data and its quality were considered for the case of Metropolitan Vancouver. Consequently, the data set used for the analysis of this project is: ICBC Reported Crashes and British Columbia Population.

The first database Reported Crashes was obtained from the statistical series of the Insurance Corporation of British Columbia (ICBC). This database was published in April 2023 and covers the period 2018 – 2022 for the province of British Columbia. However, for the purposes of this project, only municipalities that belong to Metro Vancouver will be analyzed. In that sense, below is a table with the main variables that will be used in this analysis.

Crash Count	The distinct number of crashes with the above parameters.
Crash Severity	The level of crash severity: • Casualty Crash: A crash resulting in an injury or fatality. • Property Damage Only Crash: A crash resulting in material damages to property (vehicle or non-vehicle, such as structures) with no injuries or fatalities.
Cross Street Full Name	The full name of the cross-street, thoroughfare, or road infrastructure component where a crash occurred. Applies to crashes at intersections only.
Cyclist Included	Yes if one or more cyclists were involved.
Date Of Loss Year	Year of crash (e.g., 2018).
Day Of Week	Day of the week of the crash (e.g., Monday).
Heavy Vehicle Included	Yes if a vehicle involved had a Gross Vehicle Weight (GVW) of more than 10,900 kg.
Intersection Crash	Yes if the crash occurred at an intersection between roads. Includes interchanges, onramps, and off-ramps.
Latitude	The latitude component is the crash location.
Longitude	The longitude component is the crash location.
Month Of Year	The month of the crash (e.g., January).
Motorcycle Included	Yes if one or more motorcycles were involved. Motorcycles include mopeds, limited-speed motorcycles, scooters, and trikes.

Municipality Name	The municipality in which the crash occurred (e.g., Vancouver), based on claim reports.
Municipality With Boundary	The municipality in which the crash occurred. If a crash occurred on a municipality border, all municipalities are included (e.g., Vancouver & Burnaby)
Parked Vehicle Included	Yes if one or more vehicles involved in the crash were parked at the time.
Pedestrian Included	Yes if one or more pedestrians were involved.
Region	The ICBC region in B.C. in which the crash occurred: • Lower Mainland • North Central • Southern Interior • Vancouver Island • Unknown
Road Location Description	The description of the intersection, road segment or road infrastructure where the crash occurred.
Street Full Name	The name of the street, thoroughfare, or road infrastructure component where the crash occurred.
Total Victims	The total number of victims injured or killed in crashes with the above parameters.

The second database, the British Columbia population, was obtained from the statistical series of The Official website of the Government of British Columbia. Likewise, this database includes the population of all the regions and municipalities of British Columbia for the period 2011 - 2023. However, for the purposes of this project, only the municipalities that belong to Metro Vancouver in the period 2018 - 2022 will be analyzed. Below is a table with the variables that will be used.

Municipality Name	It is the name of the municipality (e.g., Vancouver)
Population	Number of Population
Year	Year

3. Analysis Questions

The objective of this project is to answer the following questions.

i. How has the number of crashes evolved over time?

ii. What municipalities experience the highest number of crashes?

iii. Is there a correlation between the frequency of reported crashes and the population size of municipalities?

iv. How do crash rates per capita differ between municipalities?

v. What are the most frequent types of crashes?

vi. What is the distribution of crash severity levels?

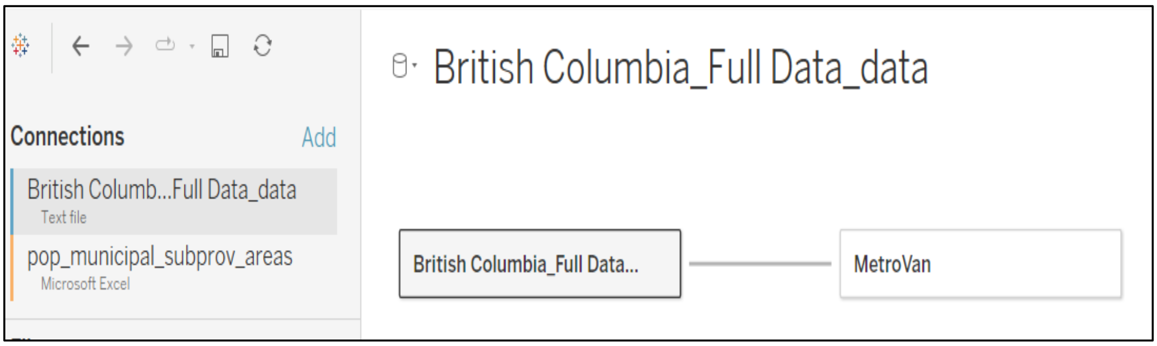
vii. What types of vehicles are most commonly involved in crashes?

viii. Do crashes correlate with specific street locations?

ix. What is the forecast for accidents and victims in the future?

4. Data Preparation

This section describes the steps for preparing the databases before answering the analysis questions. In that sense, it is important to note that in this project the name of the corresponding ICBC Reported Crashes database is ***"British Columbia_Full Data_data.csv"***, while the British Columbia Population database is called ***"pop_municipal_subprov_areas.xlsx"*** *("MetroVan"* sheet).



4.1. Data Cleaning and Standardization

❖ **British Columbia_Full Data_data.csv**

Ensured consistency in the Municipality name format and managed missing or inconsistent data entries in the Municipality name field.

❖ **pop_municipal_subprov_areas.xlsx (MetroVan Sheet)**

Checked for any discrepancies or inconsistencies in the Municipality name field and standardized the format of Municipality names and pivoted columns to make the data set more uniform and to be able to relate to the other dataset.

4.2. Connection of Fields

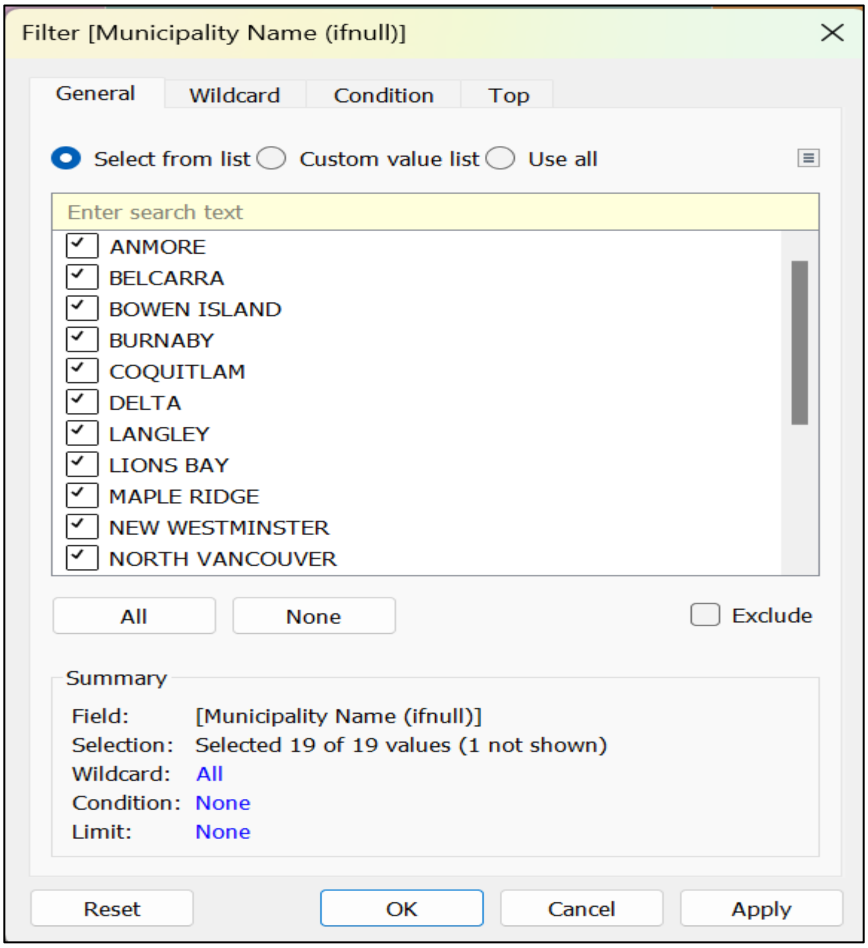
Municipality names were identified as the ***common key*** for establishing connections between the datasets.

4.3. Cardinality Considerations

We used the many-to-many relationship between the datasets, implying that one municipality from the British Columbia dataset may correspond to multiple entries in the population dataset, and vice versa.

4.4. Filters

In our project, we specifically filtered the dataset to focus solely on municipalities within this region. By narrowing our scope to Metro Vancouver cities, we aimed to provide a more targeted assessment relevant to our project's objectives. This filtering ensured that our insights and conclusions were directly applicable to the areas of interest, omitting data from municipalities outside of Metro Vancouver.



4.5. Calculated Fields

Here is an explanation of the calculated field created for this project.

❖ **Merged Dates**

- Role: Discrete Dimension
- Type: Calculated Field
- Formula: DATEPARSE ('yyyy-MMM', STR ([Date Of Loss Year]) + '-' + STR ([Month Of Year]))

This calculated field is used to merge the year and month columns into a single date field. It takes the year and month of the loss as separate columns, converts them to strings, concatenates them with a hyphen, and then parses them into a date format 'yyyy-MMM' (e.g., '2022-Jan').

4.6. Calculated Measures

The Calculated Measures created for this project are shown below.

❖ **Crashes / AVG(Population) Ratio**

- Role: Continuous Measure
- Type: Calculated Measure
- Formula: (SUM ([Total Crashes]) / AVG ([Population])) * 10

This calculated measure computes the ratio of total crashes to the average population, multiplied by ten for scaling purposes. It divides the sum of total crashes by the average population and then multiplies the result by ten to obtain a more interpretable value.

❖ **Victim / Crashes Ratio**

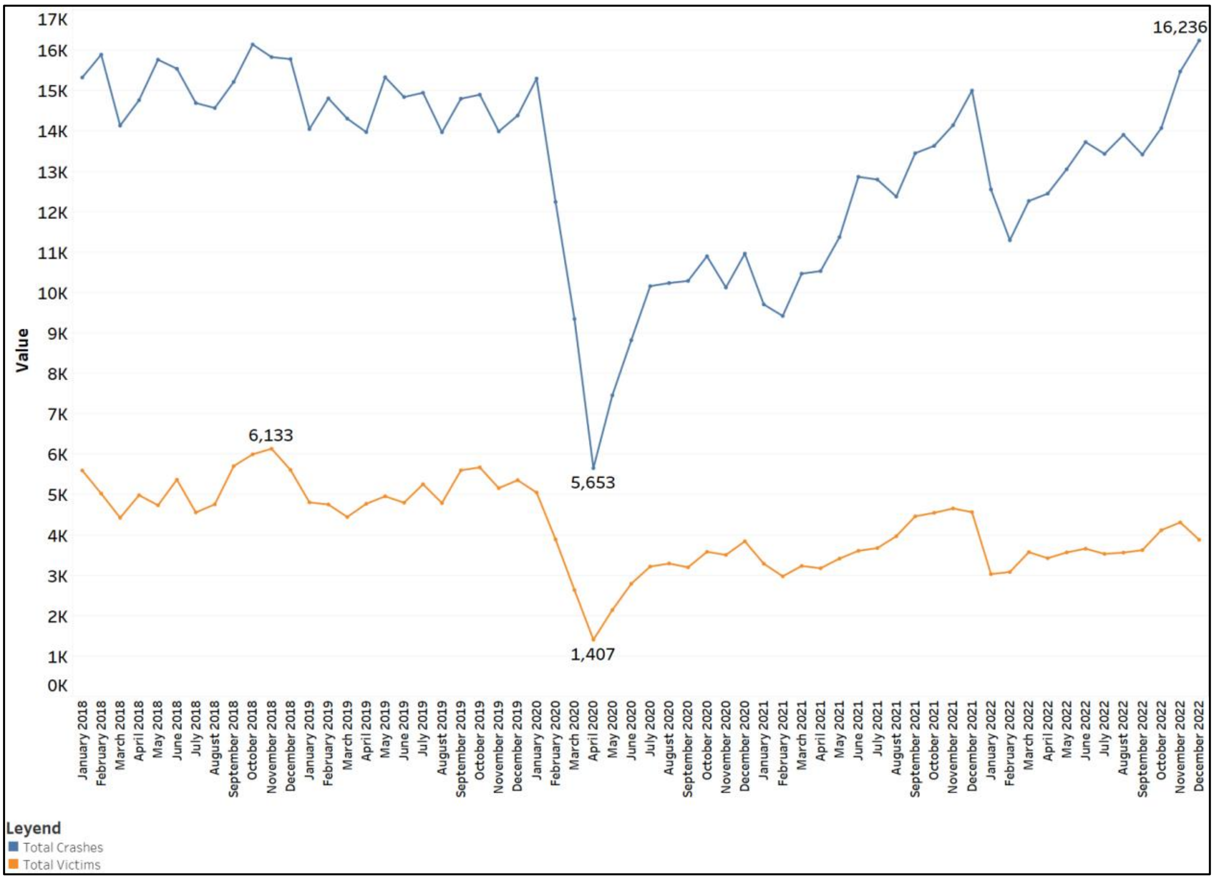
- Role: Continuous Measure
- Type: Calculated Measure
- ❖ Formula: SUM ([Total Victims]) / SUM ([Total Crashes])

This calculated field calculates the ratio of total victims to total crashes. It divides the sum of total victims by the sum of total crashes to determine the average number of victims per crash. This ratio provides insight into the severity of crashes in terms of the number of victims involved.

5. Exploratory Data Analysis

In order to answer the analysis questions of this project, below are the most relevant graphs made in Tableau and arranged in the same order that the analysis questions were formulated.

5.1. Time Series Analysis of Crashes

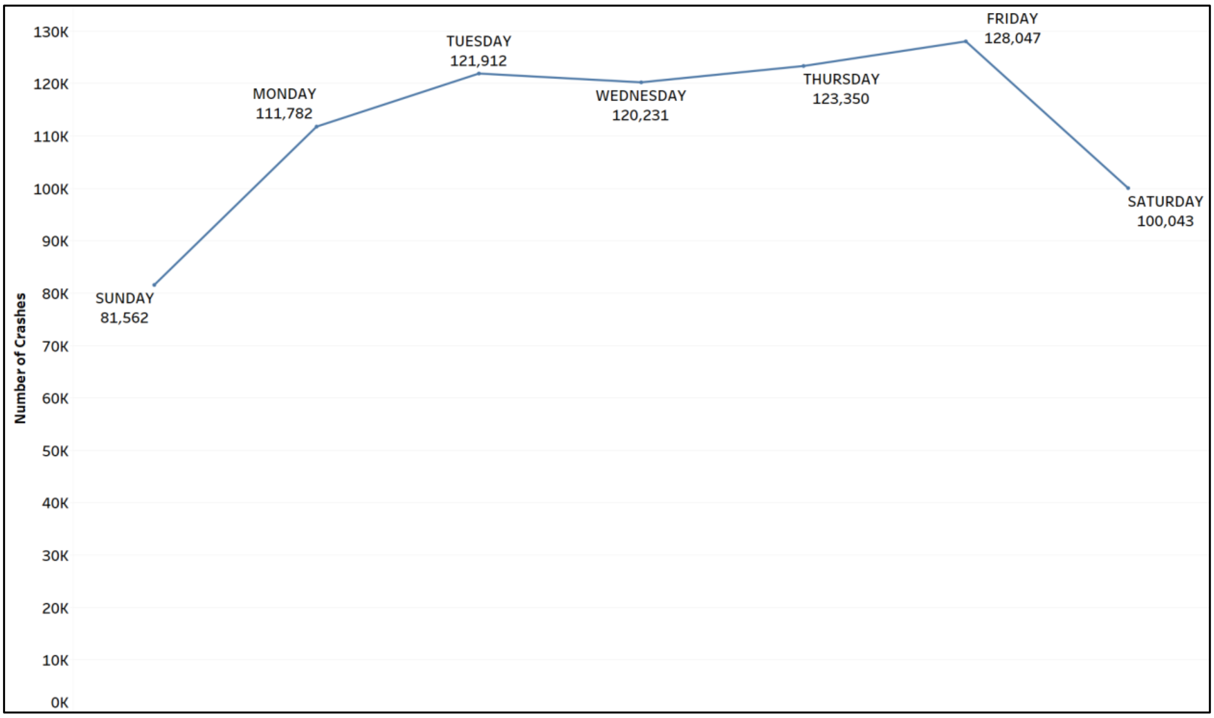


The double-line chart visualizes the trends in Total Crashes and Total Victims over the monthly period from January 2018 to December 2022 in Metro Vancouver. This graphical representation enables viewers to observe the fluctuations in these values over time. Additionally, numerical measures provide insights into the range and sum of total victims and crashes, offering a deeper understanding of the data.

Of particular note is the substantial decrease in the number of accidents and victims from January to April 2020, attributed to the total shutdown of the city due to the

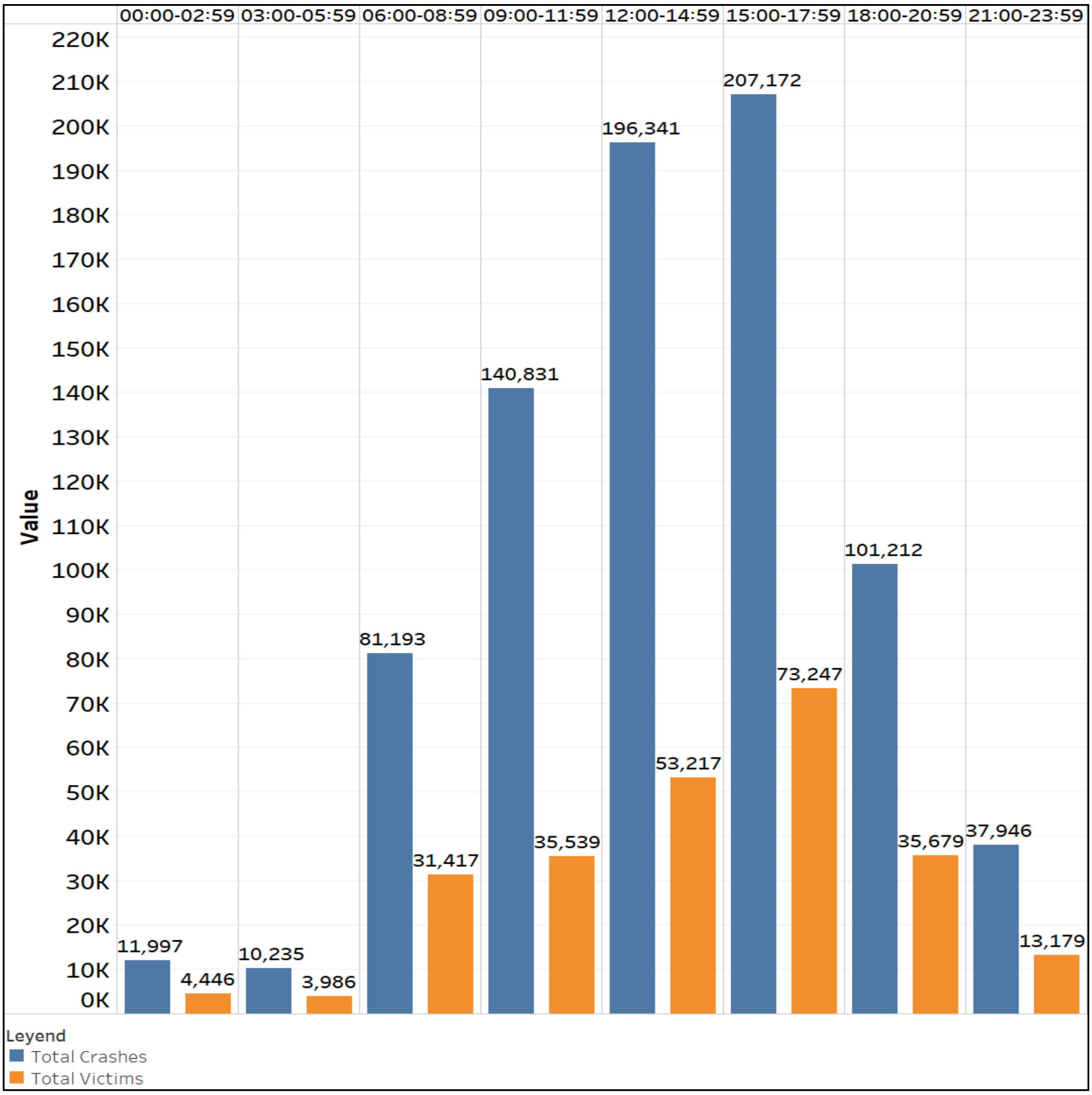
COVID-19 pandemic. This significant drop reflects the widespread impact of the pandemic on daily activities and transportation.

Furthermore, the visualization highlights a historic peak in total accidents, reaching 16,236 in the last month of 2022. Despite this surge in accidents, the number of victims remains relatively steady during this period, indicating potential improvements in safety measures or emergency response systems.



The line chart visualization explores the trends of total crashes across different days of the week, offering insights into how the sum of Total Crashes varies throughout the week. Each line on the chart is labelled with the corresponding day of the week and the sum of Total Crashes, providing clear insights into the data.

Observing the chart, it is apparent that Friday stands out with the highest number of occurrences, totalling 12,047 crashes. Conversely, Sunday exhibits the lowest number of crashes with 8,562 accidents. These trends offer valuable insights into the weekly patterns of crash occurrences, highlighting Friday as a day with increased risk on the roads, potentially due to factors such as increased traffic volume or heightened activity levels.



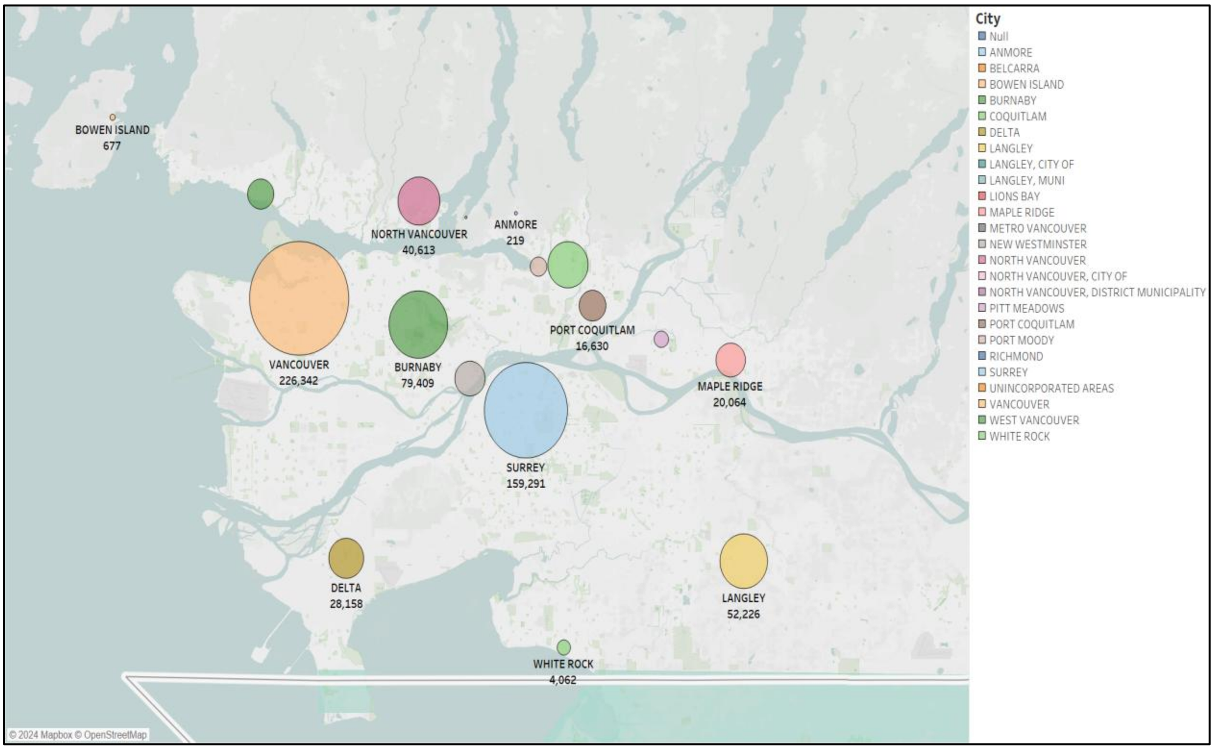
The "Time Category" visualization provides a detailed overview of crash accidents and associated victims throughout the day. Using stacked bar charts, it illustrates total crashes and victims across various time intervals, enabling insights into temporal trends.

Each bar represents a specific time category, aiding in understanding crash occurrences and victim counts at different times. Peaks during rush hours or late-night hours may indicate high-risk periods. Comparing victim proportions to crashes offers insights into accident severity.

Notably, the 15:00-17:59 timeframe records the highest number of accidents and victims, indicating heightened risk. This highlights the need for targeted interventions during these hours to enhance road safety.

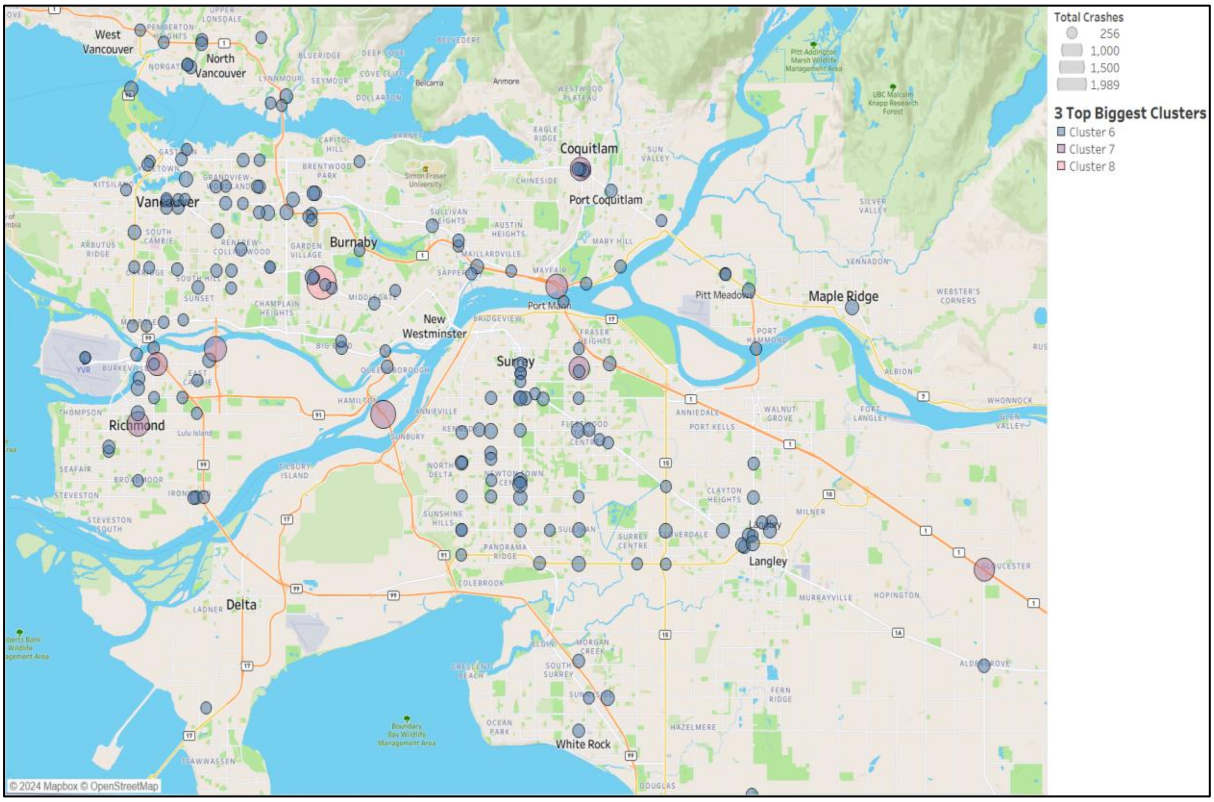
5.2. Geospatial Analysis of Crashes

The "Geo (City-General)" visualization is a powerful tool for understanding and analyzing crash data within a specific region, making it an invaluable asset for a Crash Accident Report project in a data visualization course. Utilizing a map-based format, the visualization presents crash data across various cities, offering insights into the distribution and severity of accidents.



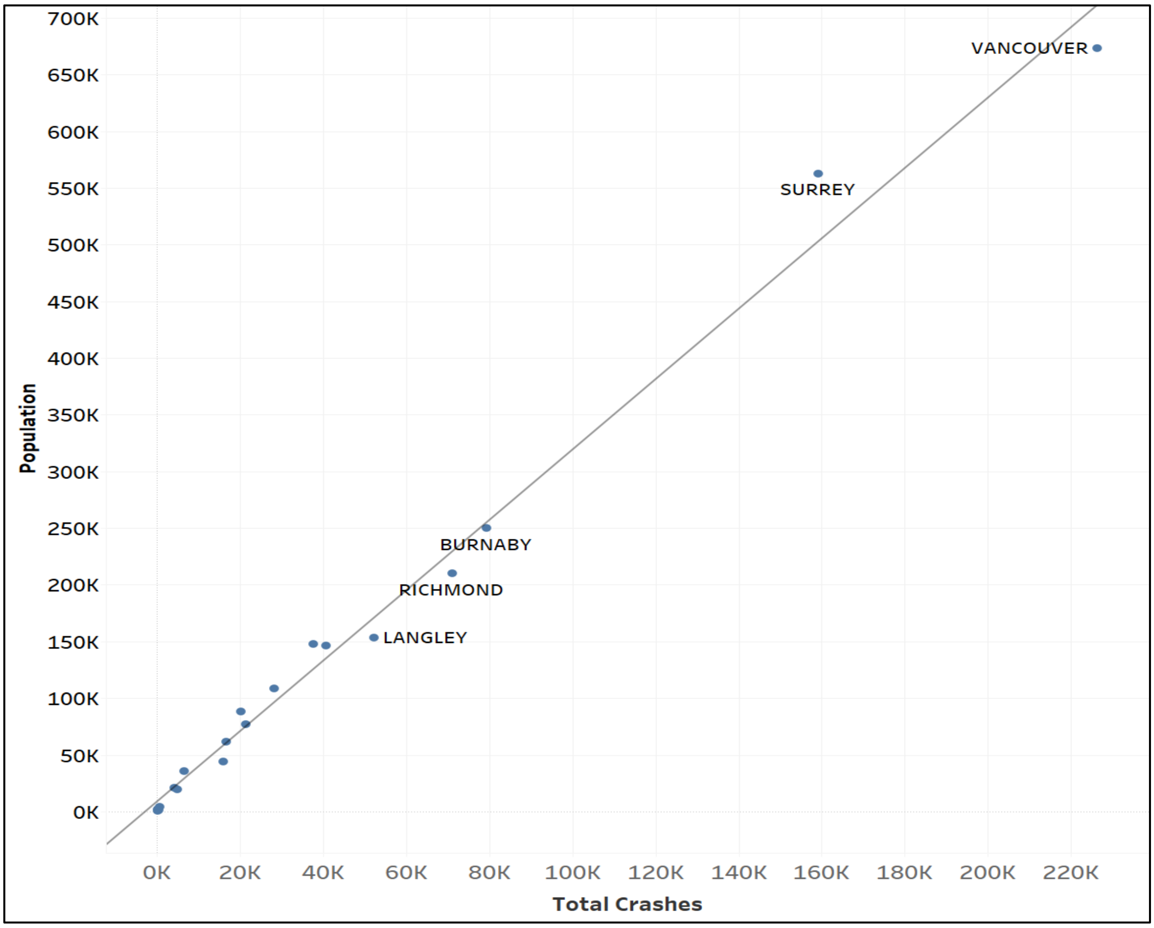
In this visualization, each circle on the map represents a city, with the size of the circle corresponding to the total number of crashes recorded in that location. This allows for a quick visual assessment of which cities experience higher crash frequencies. Additionally, the circles are color-coded based on different cities inside Metro Vancouver, aiding in the identification and differentiation of locations.

For instance, areas with higher housing density might exhibit more traffic congestion, potentially leading to a higher likelihood of accidents.



These clusters provide insights into the grouping of data points based on their total crash counts, indicating diverse levels of crash severity or frequency within each cluster. For instance, Cluster 8 has the highest sum of total crashes, suggesting areas with a significant concentration of accidents. In contrast, Cluster 6 has a much lower sum of total crashes, indicating areas with fewer accidents. Analyzing the characteristics of each cluster can help identify patterns and hotspots, facilitating targeted interventions for road safety improvement.

5.3. Correlation Analysis between Population and the Crashes



The scatterplot shows a positive correlation between population and total crashes, indicating that municipalities with higher populations tend to have more crashes. This is supported by the fact that Vancouver, with the highest population, also has the highest number of total crashes.

Surrey, with a slightly lower population than Vancouver, also exhibits a considerable number of total crashes, indicating it is another high-risk area for traffic incidents.

Burnaby and Richmond, with lower populations compared to Vancouver and Surrey, still have substantial numbers of total crashes. While not as high as the former two municipalities, these areas still experience significant traffic incidents, which may necessitate further investigation into the underlying factors contributing to crashes in these regions.

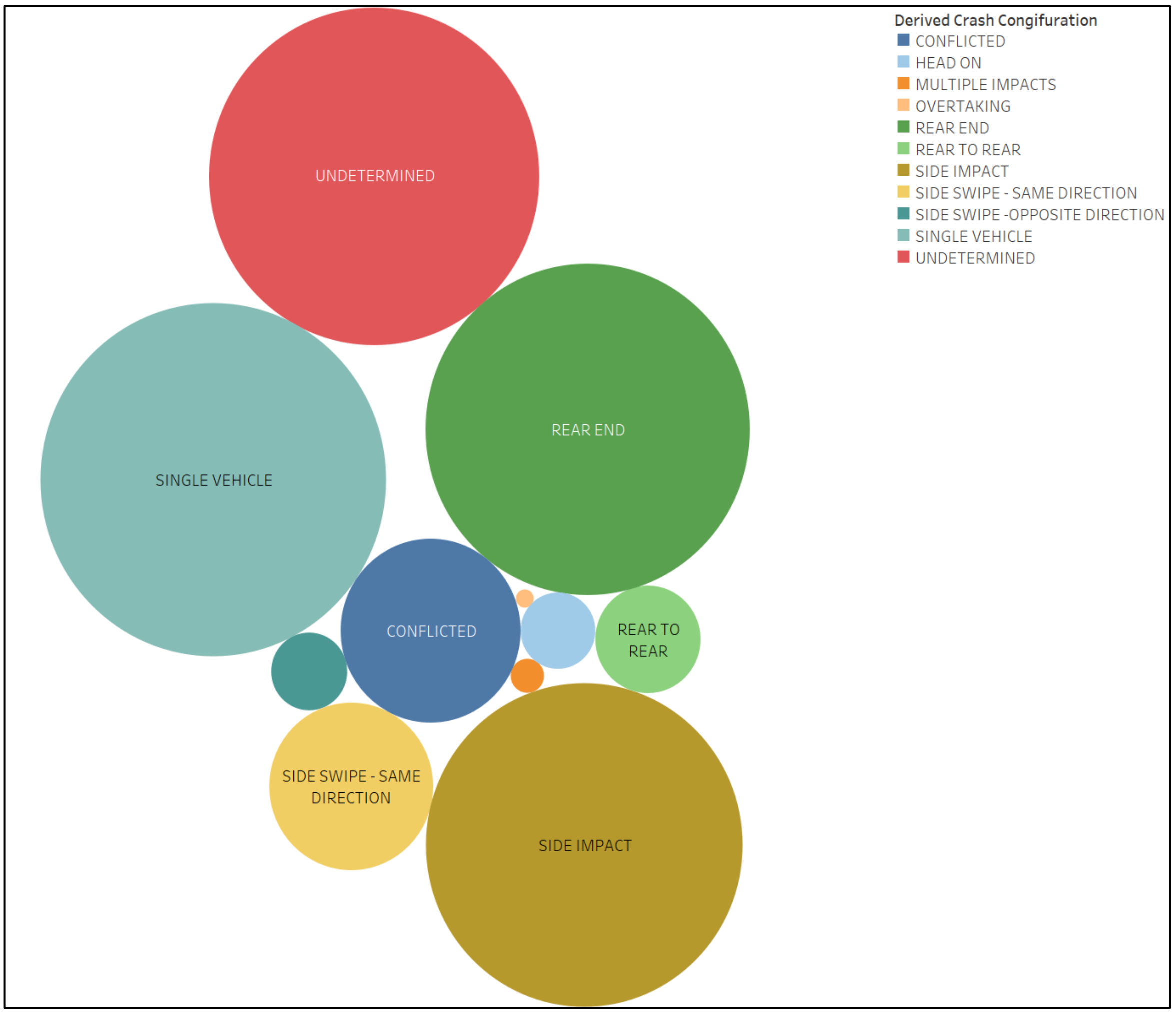
5.4. Crash Rates Per Capita Analysis

From the results of the tree map visualization, we can observe the crash-to-population ratios for several municipalities in Metro Vancouver.

- **West Vancouver**: With a ratio of 3.55, it indicates that for every 10 persons in West Vancouver, approximately 3.55 ~ 4 of them have been involved in a car accident.
- **Burnaby**: The ratio here is 3.06 ~ 4, implying that for every 10 persons in Burnaby, around 3.06 of them have experienced a car accident.
- **Richmond**: This municipality has a ratio of 3.28 ~ 4, suggesting that for every 10 persons in Richmond, approximately 3.28 of them have encountered a car accident.
- **North Vancouver**: With a ratio of 2.70 ~ 3, it signifies that for every 10 persons in North Vancouver, roughly 2.70 of them have been in a car accident.

Burnaby, Richmond, and West Vancouver have higher crash-to-population ratios compared to North Vancouver, indicating a higher frequency of car accidents per capita in these areas. West Vancouver has the highest ratio among the listed municipalities, suggesting a comparatively higher risk of car accidents for its population.

5.5. Types of Crashes Analysis



The "Derived Crash Type" visualization categorizes crash configurations into distinct types, among these, single-vehicle accidents, rear-end collisions, and side impacts stand out as the most prevalent types of crashes.

Insights gleaned from this breakdown reveal significant patterns:

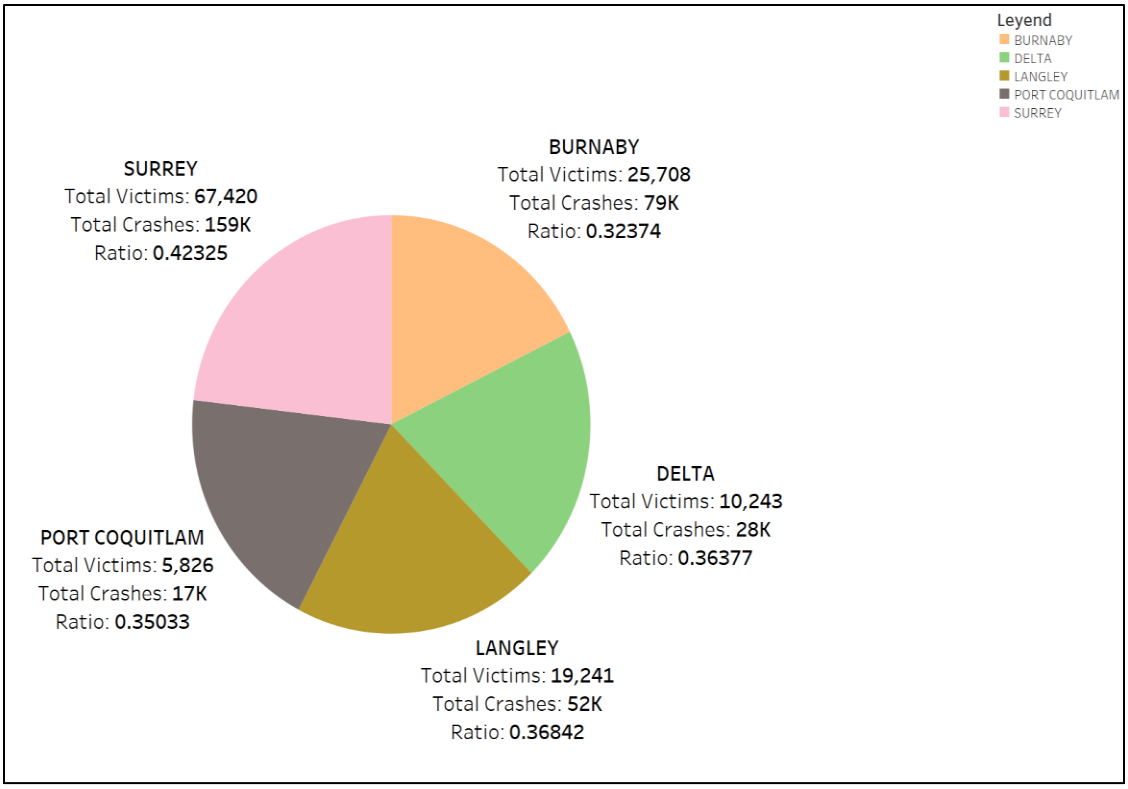
a) **<u>Prevalence of Single-Vehicle Accidents</u>**: Single-vehicle accidents are notable for their frequency, suggesting they are a common occurrence on the roadways. This may indicate issues such as loss of control, road hazards, or driver error. Addressing these factors could help reduce the occurrence of single-vehicle accidents and improve overall road safety.

b) **<u>High Incidence of Rear-End Collisions</u>**: Rear-end collisions also emerge as a prominent crash type. This trend might be attributed to factors like tailgating, distracted driving, or sudden stops. Implementing measures such as increasing awareness of safe following distances or introducing automated braking systems could mitigate the occurrence of rear-end collisions.

c) **<u>Significance of Side Impacts</u>**: Side impacts represent another substantial portion of the total crashes. These accidents often result from failure to yield, running red lights, or improper lane changes. Enhancing intersection safety, promoting defensive driving techniques, and improving vehicle visibility could help reduce the frequency of side impacts.

By understanding the prevalence and characteristics of each crash type, ICBC can develop targeted strategies to address specific issues and enhance overall road safety.

5.6. Crash Severity Analysis



In analyzing the Crash Severity Pie Chart for municipalities within Metro Vancouver, several insights emerge. Burnaby, while presenting a lower victim/crash ratio compared to others, still registers a considerable number of victims per crash, with 25,708 victims from 79,409 crashes. Delta and Langley show moderate victim/crash ratios, indicating a substantial impact on victims per crash despite having fewer total crashes than some other municipalities. Delta recorded 10,243 victims from 28,158 crashes, while Langley saw 19,241 victims from 52,226 crashes. Port Coquitlam, with a lower victim/crash ratio, still demonstrates a notable impact on victims, recording 5,826 victims from 16,630 crashes.

However, Surrey emerges with the highest victim/crash ratio among the municipalities, indicating a substantial impact on victims per crash. Surrey recorded 67,420 victims from 159,291 crashes, showcasing the severity of crashes in the area. Overall, these findings underscore the significance of implementing safety measures and traffic regulations across Metro Vancouver to mitigate the severity of crashes and minimize harm to victims.

5.7. Vehicle and Victim Analysis

Animal Involved?: No Pedestrian Involved?: No Heavy Veh. Involved?: No Cyclist Involved?: No Total Crashes: 723,925	Animal Involved?: Yes Pedestrian Involved?: No Heavy Veh. Involved?: No Cyclist Involved?: No Total Crashes: 3,336	Animal Involved?: No Pedestrian Involved?: No Heavy Veh. Involved?: Yes Cyclist Involved?: No Total Crashes: 40,487	Animal Involved?: Yes Pedestrian Involved?: No Heavy Veh. Involved?: Yes Cyclist Involved?: No Total Crashes: 18
Animal Involved?: No Pedestrian Involved?: Yes Heavy Veh. Involved?: No Cyclist Involved?: No Total Crashes: 9,959	Animal Involved?: Yes Pedestrian Involved?: Yes Heavy Veh. Involved?: No Cyclist Involved?: No Total Crashes: 32	Animal Involved?: No Pedestrian Involved?: Yes Heavy Veh. Involved?: Yes Cyclist Involved?: No Total Crashes: 697	
Animal Involved?: No Pedestrian Involved?: No Heavy Veh. Involved?: No Cyclist Involved?: Yes Total Crashes: 7,973		Animal Involved?: No Pedestrian Involved?: No Heavy Veh. Involved?: Yes Cyclist Involved?: Yes Total Crashes: 150	
Animal Involved?: No Pedestrian Involved?: Yes Heavy Veh. Involved?: No Cyclist Involved?: Yes Total Crashes: 345		Animal Involved?: No Pedestrian Involved?: Yes Heavy Veh. Involved?: Yes Cyclist Involved?: Yes Total Crashes: 5	

The visualization utilized a matrix format to present crash factors, providing a concise breakdown of various scenarios. The most significant outcomes from this matrix analysis are as follows:

❖ **No Involvement of Animals, Pedestrians, Heavy Vehicles, or Cyclists**

- Total Crashes: 723,925
- This portion of the visualization indicates that the crashes involved only common vehicles and no other external factors such as animals, pedestrians, heavy vehicles, or cyclists.

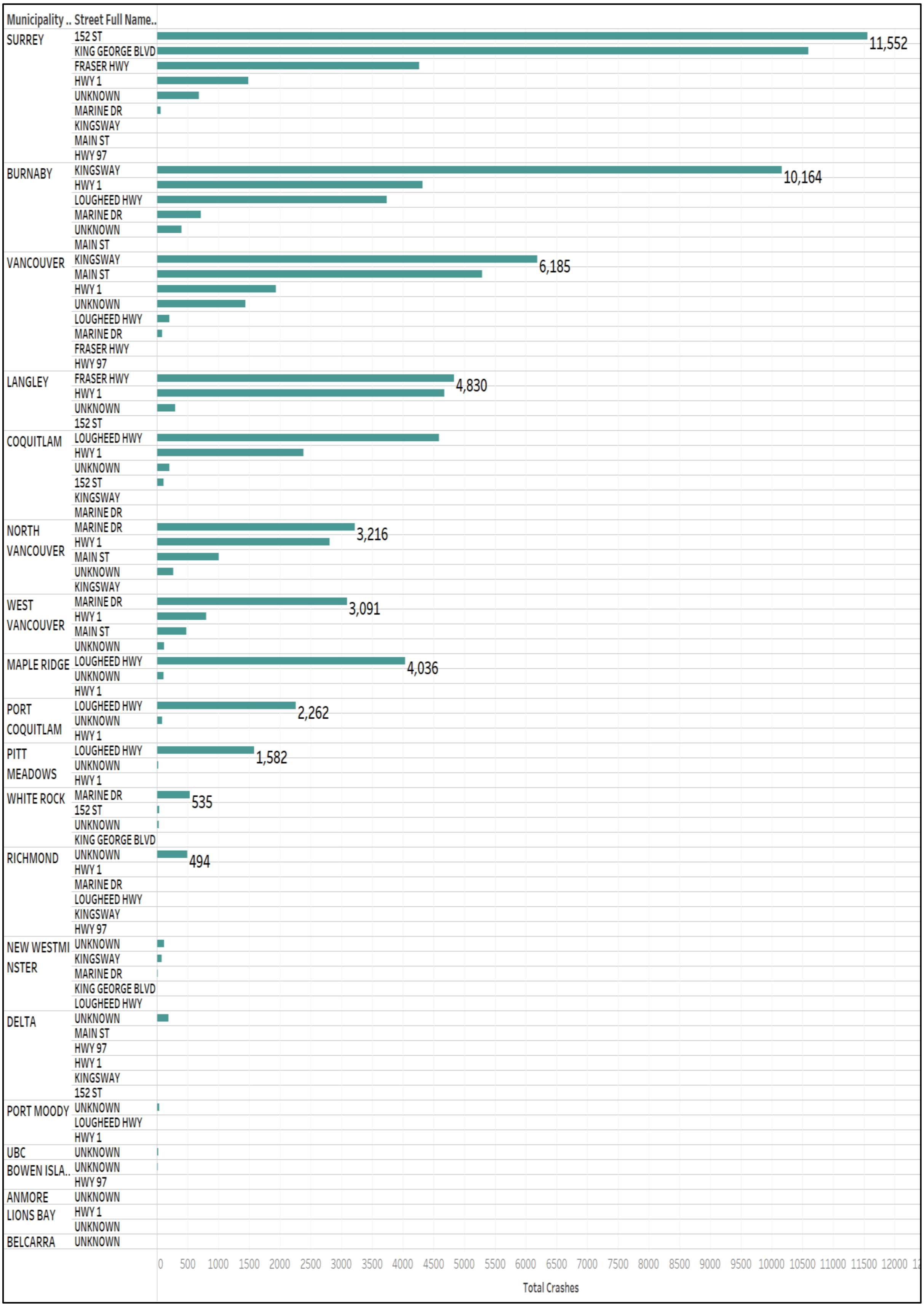
❖ **Involvement of Animals**

- Total Crashes: 3,336
- These crashes involved animals but did not have any pedestrians, heavy vehicles, or cyclists involved.

- ❖ **Involvement of Heavy Vehicles**
  - Total Crashes: 40,487
  - These crashes involved heavy vehicles but did not have any animals, pedestrians, or cyclists involved.
- ❖ **Involvement of Cyclists**
  - Total Crashes: 7,973
  - These crashes involved cyclists but did not have any animals, pedestrians, or heavy vehicles involved.
- ❖ **Involvement of Pedestrians**
  - Total Crashes: 9,959
  - These crashes involved pedestrians but did not have any animals, heavy vehicles, or cyclists involved.

Overall, this matrix-based visualization effectively communicates the involvement of different crash factors in various scenarios, providing insights into the prevalence of crashes solely involving common vehicles versus those involving external factors such as animals, pedestrians, heavy vehicles, or cyclists.

5.8. Street Analysis



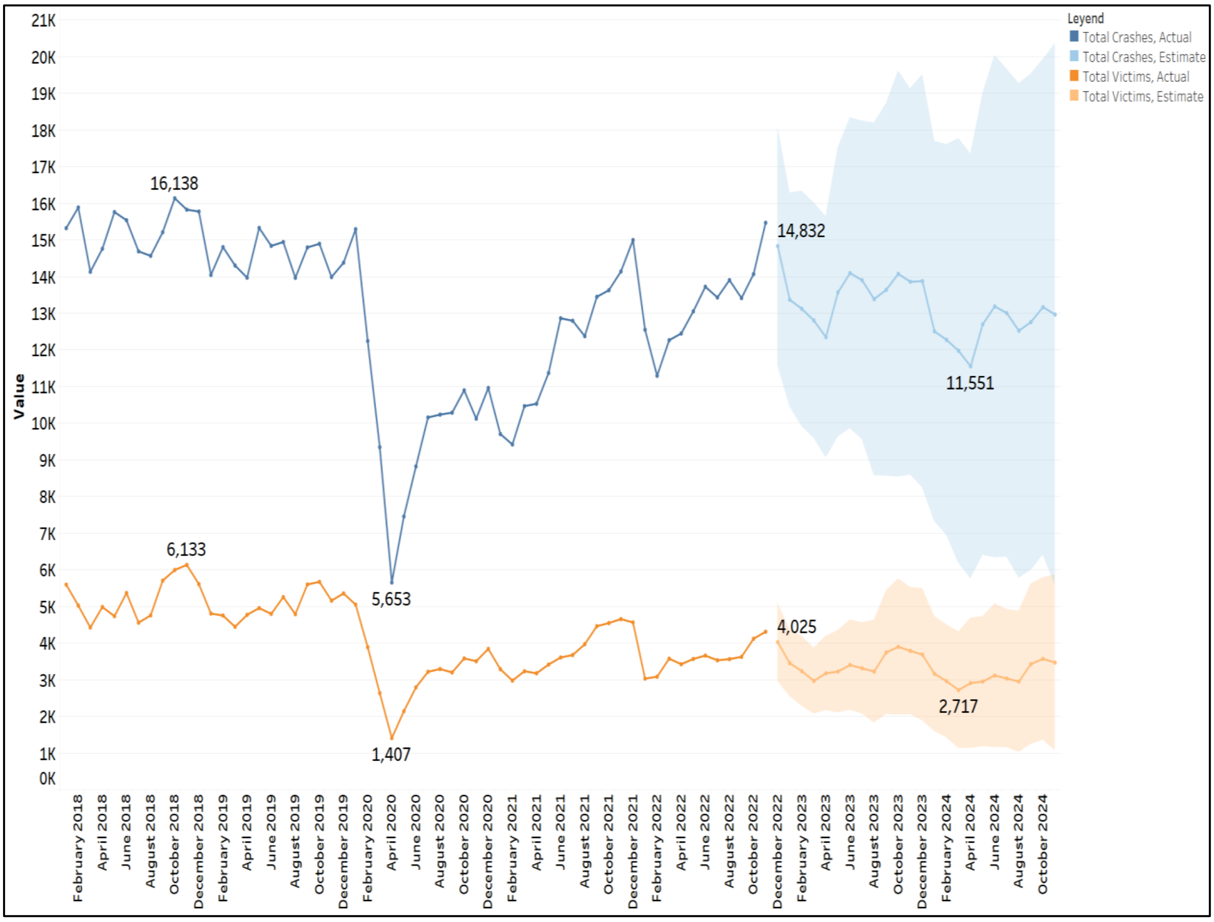
The visualization is a horizontal bar chart displaying streets with the highest number of crashes per city. Here are the top five streets with the highest number of crashes in each city:

1) **Surrey**: 152 ST - 11,552 crashes
2) **Burnaby**: Kingsway - 10,164 crashes
3) **Vancouver**: Kingsway - 6,185 crashes
4) **Langley**: Fraser Highway - 4,830 crashes
5) **Coquitlam**: Lougheed Highway - 4,592 crashes

Insights from the results and visualization:

- **High Crash Frequency**: Streets in Surrey and Burnaby have the highest number of crashes, indicating areas with potentially higher traffic volume or safety issues.
- **Common Street Names**: Interestingly, Kingsway appears in both Burnaby and Vancouver among the top streets with high crash counts. This could suggest that certain roads, despite spanning multiple cities, consistently experience high crash rates, due to factors such as heavy traffic or road design.
- **Suburban Areas**: Streets in Langley and Coquitlam also feature in the top five, suggesting that crashes are not limited to urban centers but also occur in suburban areas.
- **Safety Concerns**: The visualization highlights specific streets within each city that may require further investigation into safety measures or road improvements to reduce the number of crashes.
- **Comparative Analysis**: By visualizing crash data across different cities, authorities can identify patterns and prioritize areas for targeted interventions, such as increased enforcement, road signage improvements, or infrastructure upgrades.

5.9. Crashes and Victims Forecast Analysis



The provided forecast data, coupled with the multiplicative trend and season summary, highlights the anticipated fluctuations in total victims and total crashes over two years:

- **Total Victims Forecast**
  The multiplicative trend and season imply that the trend and seasonal patterns in the number of victims are changing multiplicatively over time. The peak month for total victims occurred in December 2022, with 4,025 victims, followed by a gradual decline to 2,717 victims in March 2024. This suggests a non-linear decrease in the number of victims, influenced by both long-term trends and seasonal variations.

- **Total Crashes Forecast**
  Similarly, the multiplicative trend and season indicate that the trend and seasonal patterns in the number of crashes are changing multiplicatively over time. The peak month for total crashes was December 2022, with 14,832 crashes, followed by a

decrease to 11,551 crashes in April 2024. This implies a non-linear decrease in the number of crashes, affected by both long-term trends and seasonal fluctuations.

In summary, the forecasted trends in both total victims and total crashes exhibit complex, non-linear patterns influenced by multiplicative trends and seasonal variations. Understanding these patterns is crucial to developing effective strategies for mitigating road accidents and enhancing overall road safety.

6. Conclusion

This project analyzes the dynamics of vehicle crashes for the nineteen municipalities of Metro Vancouver during the period 2018 – 2022. For this purpose, nine analysis questions were formulated, which were developed and answered using the powerful visualization tools of Tableau. In that regard, the main results of the analysis are mentioned below.

Firstly, a Time Series Analysis of Crashes was conducted to explain the evolution of the number of crashes over time. The double-line chart used for this analysis does not show a clear trend, but rather a stable behaviour, since at the beginning of 2018 and the end of 2022 the number of crashes was around 16,000. However, the first months of 2020 saw a notable drop in vehicle crashes due to the effect of Covid-19.

Secondly, a Geospatial Analysis of Crashes was conducted to show the municipalities of Metro Vancouver with the highest number of vehicle crashes. Using a Symbol Map for the analysis, it is observed that Vancouver, Surrey, and Burnaby are the municipalities with the highest number of crashes.

Thirdly, the Correlation Analysis between Population and Crashes allows us to know the type of correlation between the number of inhabitants of a municipality and the number of crashes. To do this, a Scatter Plot was used that allows us to observe a positive correlation between both variables, that is, the greater the population in a municipality, the greater the number of crashes in said municipality is likely to be.

Fourthly, the Crash Rates Per Capita Analysis allows us to know the municipalities with the highest crash rate per inhabitant. To do this, a Tree Map was created that shows West Vancouver, Richmond, and Vancouver as the municipalities with the highest crash ratio per inhabitant. In that sense, it is important to mention the atypical case of West Vancouver because it is one of the municipalities with the fewest inhabitants, but it has a remarkably high number of crashes for its population size.

Fifthly, the Crash Severity Analysis shows the ratio of victims per vehicle crash. The Pie Chart used in this analysis shows that Surrey, Langley, and Delta are the municipalities with the highest number of crash victims. Likewise, it is important to mention that Langley and Delta are municipalities with a not-so-high number of accidents, but they have a high number of victims.

Sixth, Street Analysis shows the streets where the greatest number of vehicle crashes occur. Through a Horizontal Bar Chart, it is observed that the streets with the most crashes are 152 St (Surrey), King George Blvd (Surrey), Kingsway (Burnaby) and Main St (Vancouver).

Finally, the results found in the analysis suggest that road safety policies in Metro Vancouver should improve significantly. However, in the context of budgetary restrictions in the public sector, it should focus on specific municipalities such as West Vancouver, Richmond, and Vancouver because they have the highest crash ratios per inhabitant. With equal or greater urgency, it is necessary to implement road safety policies in Surrey, Langley, and Delta due to their high rate of victims. Focusing on the streets with the highest number of accidents.

7. References

ICBC. (2023, April 1). ICBC Reported Crashes. ICBC. https://www.icbc.com/about-icbc/newsroom/Statistics

Province of British Columbia. (2024, February 26). Population estimates. Province of British Columbia. https://www2.gov.bc.ca/gov/content/data/statistics/people-population-community/population/population-estimates