INFO5060: Data Analytics and Business Intelligence

Dashboard Solution II

Tutorial 02 Group 05
Vinit Iyer SID 520356283
Ananya Vadhera SID 520600278
Reenal Pereira SID 520112117
Harshitha Balakumar SID 520578140
Sanjukta Gain SID 520372582

July 15, 2023

THE UNIVERSITY OF SYDNEY

Contents

1 Acknowledgement

I would like to acknowledge and express my gratitude to Professor Nasim Ahmed for their invaluable guidance and expertise throughout the duration of this project. Additionally, I extend my sincere appreciation to my tutor Eman Sayed for their continuous support and insightful feedback that greatly contributed to the success of this work.

We would like to thank Team 02-04 for their valuable contributions to the dashboard report. Their insights and feedback were extremely helpful in ensuring that the report was accurate, insightful, and user-friendly. Specifically, we appreciated their well-formatted report which gave us insights on how to improve the part as well as their tips on how to make the report more accessible to a wider audience. Their feedback helped us to create a report that is both informative and engaging.

2 Introduction

2.1 About us as a Group

We are a dynamic and cohesive group brought together by our shared passion for Data Analytics and past experiences as a team. Having worked together before as a team in another unit, we have great synergy with each other. Comprising individuals from diverse backgrounds, experiences, and expertise, we collaborate to achieve our common goals and make a positive impact in our respective fields. We firmly believe that by working together, we can achieve far more than what we could individually. Our collaborative efforts foster an environment of mutual respect, open communication, and creative thinking. As we move forward, we remain dedicated to making a difference and achieving our goals. We aim to be the best team in this unit and be the top scorers.

2.2 About the Assignment

In our Dashboard Solution I assignment, our group will tackle the challenge of formulating a compelling scenario and creating a Business Intelligence (BI) dashboard to support it. With a focus on data visualization and analysis, we will design an intuitive dashboard and provide a comprehensive report highlighting its functions and capabilities.

2.3 Introduction

This project aims to develop a Business Intelligence (BI) dashboard for the Transportation Safety and Infrastructure Improvement Agency (TSIIA). TSIIA is a hypothetical organization that combines transportation safety officials, urban planners, and engineers. Operating in the government sector, TSIIA is committed to enhancing road safety and improving transportation infrastructure in a bustling metro city.

The target audience for the dashboard is the internal stakeholders of TSIIA, including professionals and decision-makers from various departments within the organization. These departments encompass data analysis, safety interventions, infrastructure improvement, policy and advocacy, and community engagement. Each department plays a crucial role in improving road safety and ensuring efficient transportation infrastructure.

For the target audience, value is derived from the ability to access and analyse comprehensive data on road accidents and associated factors. The dashboard should provide a user-friendly interface, interactive visualizations, and relevant metrics to make informed decisions, prioritize interventions, allocate resources effectively, and advocate for policies that enhance road safety and improve transportation infrastructure in the metro city.

In the subsequent sections of this report, we will outline the functionalities of the dashboard, detailing the key metrics, visualizations, and interactivity that will empower the target audience within TSIIA to make data-driven decisions and drive positive changes in transportation safety and infrastructure.

3 Scenario

The Transportation Safety and Infrastructure Improvement Agency (TSIIA) is an organization dedicated to enhancing road safety and improving transportation infrastructure in a bustling metro city. With increasing traffic congestion and a growing population, the city has experienced a rise in accidents and concerns regarding road safety. TSIIA has been tasked with analysing accident data, identifying high-risk areas, and implementing targeted interventions to reduce accidents and enhance overall transportation safety.

3.1 Audience and Roles

The primary audience for our assignment is the Transportation Safety and Infrastructure Improvement Agency (TSIIA) and its various divisions. This includes executive leadership, data scientists, statisticians, analysts, safety intervention specialists, transportation engineers, planners, and policy advocates. They are responsible for analysing accident data, developing targeted interventions, implementing infrastructure improvements, and advocating for evidence-based transportation safety policies. Our audience relies on data-driven insights and effective communication to make informed decisions and drive positive change in road safety and transportation infrastructure within the country of New Zealand. The following is a short description of the organizational structure of TSIIA:

- Executive Leadership: The Executive Director of TSIIA, supported by Deputy Directors, sets the strategic vision for the agency.
- Data Analysis and Research Division: TSIIA's data scientists, statisticians, and analysts work diligently to collect, clean, and analyse accident data from various sources.
- Safety Interventions Division: The Safety Interventions Division collaborates closely with the Data Analysis and Research Division to translate insights into action. They develop targeted interventions, programs, and policies based on data-driven findings.
- Infrastructure Improvement Division: TSIIA's team of transportation engineers and planners from the Infrastructure Improvement Division uses the insights provided by the data analysis to design and implement infrastructure improvements.

- Policy and Advocacy Division: This division engages with government officials, policymakers, and stakeholders to develop evidence-based transportation safety policies.

Through this integrated approach, TSIIA aims to make a significant impact on road safety in the metro city. The data-driven insights generated by the Data Analysis and Research Division guide the targeted interventions developed by the Safety Interventions Division. The Infrastructure Improvement Division implements infrastructure enhancements, while the Policy and Advocacy Division ensures the adoption of evidence-based policies. By effectively leveraging their organizational structure and expertise, TSIIA collaborates with key stakeholders, utilizes data-driven decision-making, and implements a comprehensive approach to improve road safety and enhance transportation infrastructure in the metro city, making it a safer place for its residents and commuters.

3.2 Dataset

The Crash Analysis System is a database that captures and analyses detailed information about traffic crashes. It is commonly used by transportation agencies, law enforcement agencies, researchers, and other stakeholders to study crash patterns, identify high-risk areas, and develop strategies to improve road safety. CAS data typically includes various parameters related to crashes, such as the date, time, and location of the crash, the types of vehicles involved, weather conditions, road characteristics, contributing factors, severity of injuries, and more. By analysing this data, agencies can gain insights into the causes and consequences of crashes, identify trends and patterns, and make data-driven decisions to enhance transportation safety.

The Crash Analysis System plays a crucial role in understanding and addressing road safety issues, enabling agencies to prioritize resources, implement targeted interventions, and evaluate the effectiveness of safety programs and policies. For this assignment, we were provided with this dataset consolidated with multiple more datasets. This dataset was provided to us after cleaning and no further cleaning has been done as it wasn't required.

While the data received was sufficiently clean, we did do some data transformations which are mentioned below:

- We created a Crash Severity Filter which can be used to filter the different values of crash severity and be used with parameters that have been developed.
- We created a group for Junction Types wherein Crossroads, Default, Driveway, and End of Road were grouped under Straight Roads. T Junction and Y Junction were grouped under Junctions. Multileg and Roundabouts were kept as is.
- We created a group for vehicle usage categories as well due to the high number of vehicle types. There were 4 groups created which are 4-Wheeler Cars, Heavy Vehicles, 2 Wheelers, and Others.
- Multiple calculated fields were created to aid the parameter creation for the dashboard and for the predictive questions as well.
- 3 parameters were created, namely Current Year, Previous Year, and Select Crash Severity.

3.3 Refinements based on Feedback

Suggestion	Action
Use more complex questions for the descriptive questions to align with what you are discussing in its visualization and description. Use more visualisations that represent your descriptive questions better.	Descriptive Questions were made more complicated by adding more factors and new visualizations were introduced. A new descriptive question was also added in place of a simpler one.
The report is well written. Use larger sizes for the font of the figures, such as figures 6 and 7.	The visuals have been modified as per the suggestion.
Lack of clarity in audience and roles.	Compilation of audience and roles as suggested and further clarification for the same.
Lack of detailed information on data.	Addition of all data transformations and addition of information.
The dashboard is too complicated and filled.	Single dashboard has been divided into 3 dashboards to relay the same amount of information.
Dashboard has a lot of functionalities and there is no information on how to use it.	Addition of Dashboard overview and functionalities in the report.
Lack of cumulative total for all the years	Unfortunately this feature could not be added as we were pressed for time and required a lot of complicated calculated fields.

Table 1: Refinements based on Feedback

4 Dashboards

The TSIIA BI dashboard is a comprehensive tool designed to support the Transportation Safety and Infrastructure Improvement Agency in their mission to enhance road safety and improve transportation infrastructure in a bustling metro city. The dashboard serves as a centralized platform for internal stakeholders within TSIIA, including professionals and decision-makers from various departments.

Visually, the dashboard features a user-friendly interface with a clean and intuitive design. It employs a modern and professional look, with a layout that allows users to easily navigate and access the information they need. The dashboard's design is optimized to present data and visualizations in a clear and concise manner, enabling stakeholders to quickly grasp important insights.

The functionality of the dashboard revolves around providing access to comprehensive data on road accidents and associated factors. It enables users to analyze and explore this data through interactive visualizations, empowering them to make informed decisions and prioritize interventions effectively. The dashboard incorporates various key metrics, charts, and graphs that highlight important trends and patterns in road safety and transportation infrastructure.

Users can interact with the visualizations to drill down into specific data points, filter information based on various parameters, and customize views according to their needs. This interactivity allows stakeholders to gain deeper insights and extract meaningful conclusions from the data. Additionally, the dashboard provides the ability to generate reports and export data for further analysis or sharing with external stakeholders.

The TSIIA BI dashboard caters to the specific roles and responsibilities within the organization. For example, data scientists, statisticians, and analysts in the Data Analysis and Research Division can utilize the dashboard to perform in-depth data exploration, conduct advanced analytics, and generate data-driven reports. The Safety Interventions Division can leverage the dashboard to identify high-risk areas, track the effectiveness of interventions, and evaluate the impact of their programs. Similarly, the Infrastructure Improvement Division can use the dashboard to assess the state of transportation infrastructure, monitor ongoing projects, and identify areas for improvement. The Policy and Advocacy Division can rely on the dashboard to gather evidence-based insights, support their policy development efforts, and communicate the importance of transportation safety to external stakeholders.

Overall, the TSIIA BI dashboard provides a powerful tool for the internal stakeholders of TSIIA to access, analyze, and leverage comprehensive data on road safety and transportation infrastructure. By combining a user-friendly interface, interactive visualizations, and relevant metrics, the dashboard enables data-driven decision-making, effective resource allocation, and evidence-based policy advocacy, ultimately driving positive changes in transportation safety and infrastructure within the metro city.

4.1 Dashboard 1

Figure 1: Collective dashboard of descriptive questions 1 to 3

This comprehensive dashboard provides key insights into the distribution of crash severity, vehicle types, and the occurrence of accidents based on days, months, gender, and holiday status in New Zealand.

Let's explore the different components of the dashboard to understand the information it offers.

Starting on the left-hand side, the Key Performance Indicators (KPIs) provide an overview of the total number of accidents and the distribution of crash severity and year-on-year rate over a 5-year period (2016-2020). This gives stakeholders an understanding of the overall accident trends and severity levels in New Zealand.

Moving to the top figure, the dashboard highlights the relationship between vehicle types and the occurrence of casualties. It focuses on specific categories such as two-wheelers, four-wheelers, heavy-duty vehicles, and others. By examining this chart, stakeholders can identify which vehicle types are most commonly involved in accidents and assess their impact on casualty rates.

The highlighted table in the dashboard presents the count of accidents categorized by days and months. This information allows stakeholders to analyze the distribution of accidents throughout the year and identify any patterns or trends based on specific days or months.

The stacked bar graph provides a visual representation of the division of accidents by gender and holiday status. This graph helps stakeholders understand the gender distribution of accidents and how it may vary during holidays compared to non-holiday periods. It provides insights into whether there are any gender disparities in accident occurrence and how holidays may impact accident rates.

The interactive feature of the dashboard enhances the user experience by allowing the audience to explore the data in more depth. Users can click on specific values, choose different crash severity levels, and select specific years for comparison or vehicle type. For example, by selecting a Tuesday in September, the dashboard will update to display information relevant to that specific value, such as the incidence of accidents for males and females and the comparison between weekdays and weekends.

By leveraging the interactive functionality of the dashboard, stakeholders can gain a deeper understanding of accident patterns and trends based on crash severity, vehicle types, days, months, gender, and holiday status. This information can inform decision-making processes, interventions, and road safety strategies to reduce accidents and improve overall safety on New Zealand roads.

4.2 Dashboard 2

This interactive dashboard provides a comprehensive analysis of crash counts in New Zealand from 2016 to 2020, categorized by road type and natural light conditions. The data is presented in a stacked bar graph format, where each bar represents a specific road type, and the segments within the bar correspond to different natural light conditions.

There are 14 road type categories, including access rural, access urban, arterial rural, arterial urban, footpath, foot track, major rural, major urban, medium rural, medium, minor rural medium urban, minor rural, minor urban, and motorway. Each category is colour-coded for easy visualising natural light, with the bright sun represented by blue, dark by orange, overcast by red, and teal for twilight.

The interactive features of the dashboard allow the audience to explore the data by selecting different years and vehicle types, such as two-wheelers, four-wheelers, heavy-duty vehicles, and others. Additionally,

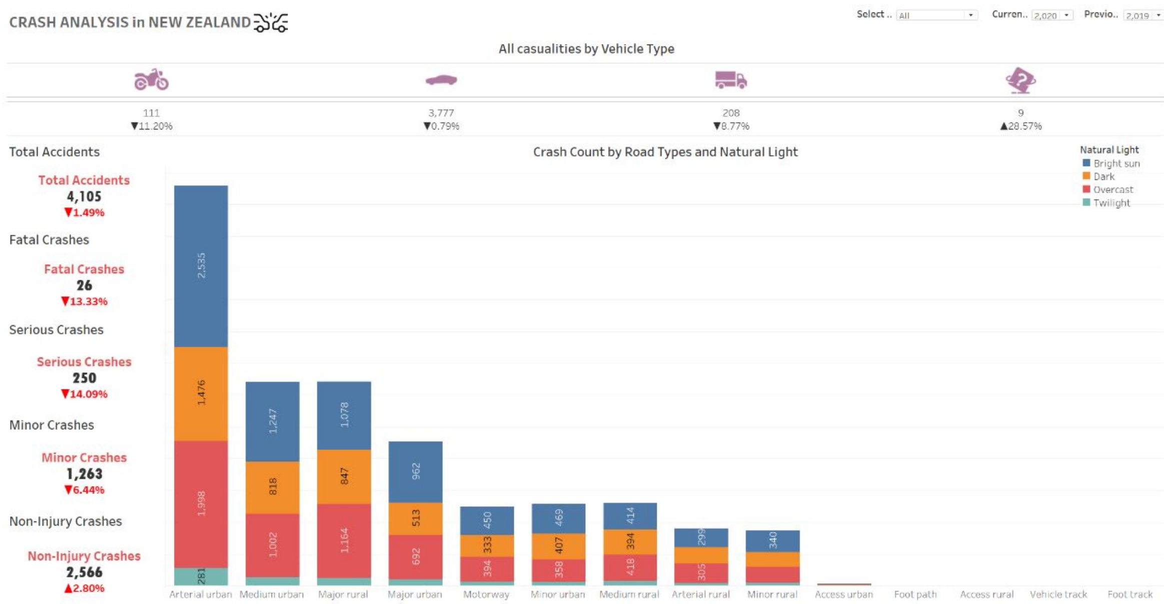

Figure 2: Dashboard for descriptive question 4

the dashboard provides the option to view the data based on crash severity, offering a more detailed understanding of the patterns and trends.

By leveraging the interactive functionality of the dashboard, the audience can gain valuable insights into the crash counts based on road type, natural light conditions, vehicle types, and crash severity. This information can assist in identifying high-risk areas, developing targeted interventions, and formulating effective road safety strategies to mitigate accidents and improve overall safety on New Zealand roads.

4.3 Dashboard 3

The dashboard consists of multiple components. On the left side, there are key performance indicators (KPIs) displaying total accidents and crash severities categorized as fatal, serious, minor, and non-injury crashes. Each KPI shows the number of crashes in that category and the percentage change compared to the previous year.

At the top, there is a band featuring four images representing different vehicle categories, along with numbers indicating the involvement of each vehicle type in accidents. The numbers are accompanied by a percentage indicating the change compared to the previous year.

The dashboard includes a donut chart showing accidents based on different weather conditions, with the total number of accidents displayed in the center. Another donut chart represents accidents based on different junction types.

All the charts are interconnected filters, meaning that clicking on any component filters the others accordingly, allowing for dynamic exploration of the data.

At the top right, there are three parameters: Select Crash Severity, Current Year, and Previous Year.

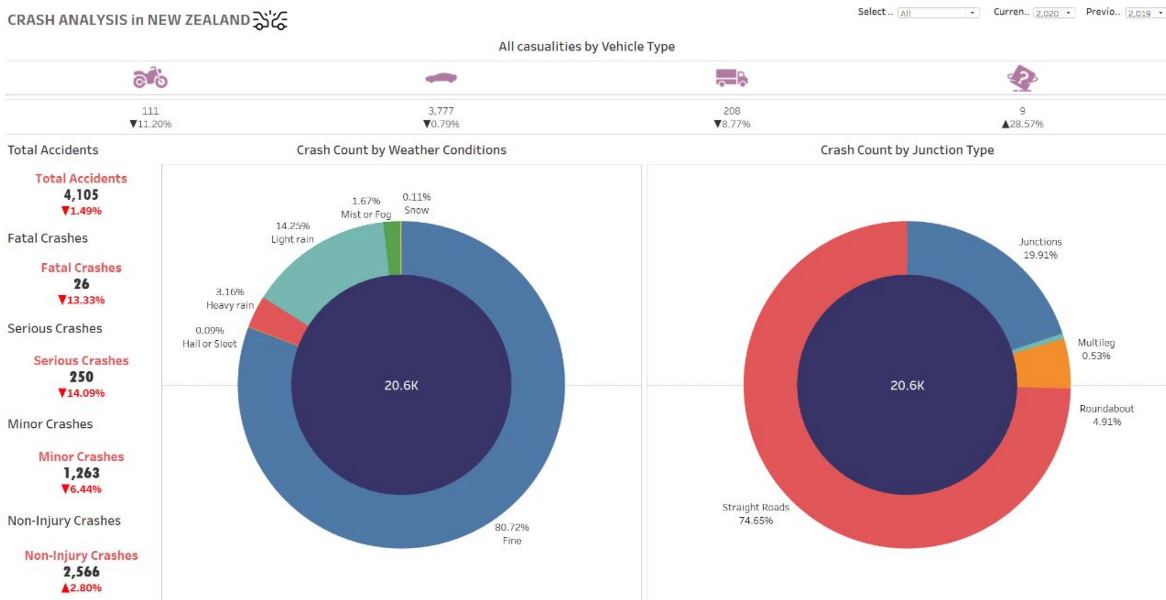

Figure 3: Dashboard for descriptive question 5

These parameters provide options to select the desired crash severity level and compare data between the current year and the previous year.

Overall, the dashboard presents an overview of key accident statistics, crash severities, vehicle involvement, weather conditions, and junction types. The interactive nature of the dashboard allows users to analyze and explore the data by filtering components and selecting specific parameters, facilitating data-driven insights and decision-making.

5 Descriptive Questions

5.1 What are the total number of crashes and their distribution according to the severity in New Zealand?

Description: Our descriptive question aims to examine and understand the total number of accidents and the distribution of crash severity in the recorded accidents over the 5 years between 2016 and 2020 in New Zealand. We look at data collected over the 5-year period to understand patterns and general trends. The data collected is categorized based on the crash severity into 4 categories namely Fatal, Severe, Minor and Non-Injury accidents.

Relevance to the audience: Our collaborative effort in understanding the total number of accidents in New Zealand and their distribution across severity levels assists our audience in multiple ways. Government agencies can assess the effectiveness of current road safety measures, law enforcement can identify areas of concern, transportation planners can prioritize interventions, and policymakers can develop strategies to reduce accidents. This information helps create a safer road environment and protect the lives of people across the country. It is vital for the National Traffic Safety Bureau to analyse and understand this distribution due to the following reasons:

- Highlight any prevalent trends: Understanding the distribution of crash severity over the years will help the bureau identify any existing patterns. It can also help them figure out the direction in which the trend, if any exists, seems to be progressing, which can, in turn, aid them in highlighting any potential factors leading to such crashes.
- Evaluate existing safety measures/interventions and develop any if necessary: They can assess the effectiveness of any existing measures/policies that are in place and develop a few suitable measures if necessary. A few examples of such potential interventions are better traffic regulations, better road structures, etc.
- Allocate resources appropriately: Making sound decisions regarding resources and policies is an imperative activity. Studying information regarding the distribution of crash severity will allow the bureau to make informed decisions in matters of allocating resources. They can focus on high-risk areas and devote more effort and resources there.
- Formulate strategies for prevention: Studying crash severity distribution can help the bureau identify potential causes and factors. This knowledge can be used to develop and curate strategies and plans to mitigate the occurrence of severe crashes.
- Create awareness among people: Campaigns can be held periodically to bring about more awareness among the public regarding safe driving practices. This will greatly reduce the frequency of major and even minor crashes.

Relevant Dataset Parts: To attempt and answer this question, we looked at the count of crashes segregated by the crash severity for each year between 2016 and 2020. The fields used were 'Crash Severity' and Count of crashes by year. The crash severity field has 4 categories namely fatal accidents, serious accidents, minor accidents, and non-injury accidents. For calculating Total Crashes for each year, we have aggregated the counts of 4 types of crashes for each year and used this aggregate as a measure. To determine the change in percentage, we compare the counts of the current and previous years to calculate the change in percentage as an increase or decrease from the previous year.

Description of the dashboard component: In our interactive dashboard, we have created separate worksheets for different crash severities, namely Fatal crashes, Serious crashes, Minor crashes, and non-injury crashes. Each worksheet displays the count of crashes for its category each year. A percentage value is also displayed that shows how much the number of crashes has changed in comparison to a previous/different year.

Sub Questions:

- What is the rate of change in the number of accidents between two consecutive years for each severity category?

  This sub-question aims to analyse the year-on-year changes in the number of accidents for each severity category. By comparing the accident counts from one year to the next, we can identify whether there has been a significant increase or decrease in accidents within each severity level. This information helps us understand the trend and direction of change in accident rates, enabling us to assess the effectiveness of road safety measures and identify areas that require further attention.

- How does the distribution of accidents vary across different road categories and vehicle types for each severity level?

  This sub-question explores the distribution of accidents across road categories and vehicle types within each severity level. By analyzing the data, we can examine whether certain road categories or vehicle types are more prone to specific severity levels of accidents. Understanding these patterns helps in identifying high-risk areas and vehicle types that require targeted interventions and safety measures. It also provides insights into the factors contributing to different severity levels of accidents, allowing for more focused strategies to improve road safety and reduce the occurrence of accidents.

- How has the number of each type of crash (minor, serious, fatal, non-injury) changed year by year?

  This question aims to analyse the variations in the crashes for different crash severities. Answering this question can help us understand whether certain crashes have become less or more prevalent. The knowledge gained can be used to strategize our plan for implementing interventions and measures.

- Which year had the most serious or fatal accidents?

  We notice that the most fatal accidents took place in the year 2019 and the most serious accidents occurred in the year 2017. 2019 displays a record of 201 fatal crashes(an increase of 2.03% from 2018). 2017 shows a record of 1588 serious crashes(an increase of 137.72% from 2016).

Conclusion: Our collaborative effort yielded valuable insights into the total number of accidents occurring in New Zealand within a year, categorized by severity levels. In general, the number of accidents has continually increased across almost all types from the successively between the years 2016 and 2019. Between 2019 and 2020, there is a dip in the numbers. 2019 has recorded the greatest number of fatal, minor, and non-injury accidents. We believe that our work as a team contributes to the assessment of the overall road safety situation in the country, the identification of areas of concern across severity levels, and the development of strategies to reduce accidents. By presenting our findings on an interactive dashboard, we aim to provide an accessible and informative platform for our audience to explore and visualize the distribution and trends of accidents across severity categories. Together, we can make a difference in enhancing road safety and protecting lives.

Dashboard Visualization:

Figure 4: Total number of Accidents and their classification of crash severities.

5.2 How does the type of vehicle contribute to the number of casualties in accidents?

Description: This question seeks to understand the relationship between vehicle type and the occurrence of casualties in accidents. It focuses on distinct vehicle categories - two-wheelers, four-wheelers, heavy-duty vehicles, and others - aiming to reveal any patterns or trends related to vehicle type and accident severity.

Relevance to the audience: The insights from this analysis will assist our target audience, mainly safety intervention specialists and policy advocates, in understanding where to focus their interventions and policy efforts. If certain types of vehicles are more prone to severe accidents, this information can guide the development of vehicle-specific safety measures, driver training programs, and policy changes.

Relevant Dataset Parts: We've utilized specific columns from our dataset to answer this question. This includes the 'Road_usage_type', which classifies the vehicles involved in each accident; we group these into categories such as two-wheelers, four-wheelers, heavy-duty vehicles, and others. We've also employed the 'Crash Severity' and 'Crash Year' columns. Crash Year have been crucial in deriving additional data points like 'Current Year' and 'Year-on-Year Rate of Accidents.' These derived fields were created in Tableau using calculated fields.

Sub Question: Which vehicle type has the highest number of fatal or serious accidents?

Four-wheelers, particularly cars, have the highest number of fatal and serious accidents. This highlights the need for targeted safety measures and interventions to address the risks associated with four-wheeler accidents. Two-wheelers also experience fatal and serious accidents, albeit in lower numbers compared to four-wheelers. Heavy-duty vehicles and others have relatively fewer fatal and serious accidents.

Conclusion: In analyzing the data, it becomes evident that four-wheelers, particularly cars, consistently have the highest number of casualties across all crash categories. This highlights their significant involvement in accidents and underscores the importance of implementing measures to enhance their safety. Furthermore, heavy-duty vehicles also experience a considerable number of casualties, while two-wheelers and other vehicle types have relatively lower casualty rates. It is crucial to address the unique risks associated with each vehicle type in order to mitigate casualties effectively.

Dashboard Visualization:

Figure 5: Vehicle and the corresponding Crash Severities

5.3 What is the distribution of accidents by day and month in the current year, and how does it vary based on gender and the presence of a holiday?

Description: This question aims to analyze the distribution of accidents in the current year, considering the variables of day and month, and further examine how this distribution varies based on gender and the presence of a holiday. The dashboard consists of two key elements: a highlight table and a stacked bar graph. The highlighted table presents the count of accidents categorized by days and months, while the stacked bar graph illustrates the division of accidents by gender and holiday status.

Relevance to the Audience: The distribution of accidents by day, month, gender, and the presence of a holiday is crucial information for the Transportation Safety and Infrastructure Improvement Agency (TSIIA) and its various divisions, as well as for data scientists, statisticians, analysts, safety intervention specialists, transportation engineers, and policy advocates. Understanding the patterns and variations in accidents based on these factors allows stakeholders to identify high-risk periods, gender-specific trends, and the impact of holidays on accident occurrences. This knowledge can inform targeted interventions, resource allocation, policy decisions, and road safety campaigns.

Relevant Dataset Parts: The relevant dataset for this question comprises data on accidents that occurred in the current year. The dataset was constructed using a calculated field derived from the crash year information. Additionally, attributes such as the day of the week, month, gender, and holiday status were derived from the corresponding columns: day of the week, month of year, gender, and holiday.

By utilizing this dataset, we can analyze the distribution of accidents based on various factors, including day and month, gender, and the presence of a holiday. These attributes provide valuable insights into the patterns and variations of accidents, allowing stakeholders to gain a comprehensive understanding of the current year's accident landscape and its associations with specific variables.

Sub-Question: What is the breakdown of accidents by gender in the current year? Analysis of the data reveals a notable variation in the breakdown of accidents by gender during both holiday and non-holiday periods across the years. Generally, there is a higher incidence of accidents involving males during non-holiday periods, while female accidents tend to follow a similar pattern but notably less than males. These gender-specific trends have important implications for stakeholders, particularly the Transportation Safety and Infrastructure Improvement Agency, in their efforts to implement targeted interventions, road safety campaigns, and policies aimed at reducing accidents and promoting road user safety. By considering these trends, stakeholders can tailor their initiatives to address the specific needs and behaviours of different genders, ultimately working towards creating safer road environments for all individuals.

Dashboard Visualization:

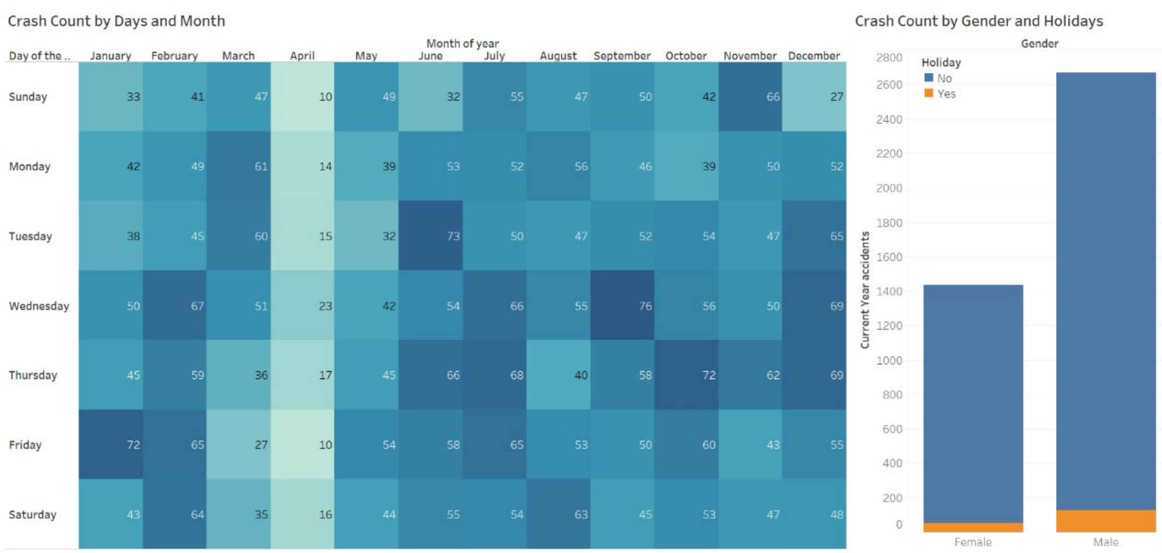

Figure 6: Distribution of accidents over days, months varied by gender and holidays

Conclusion: The analysis of the data reveals several significant trends in the distribution of accidents. It is observed that the months from January to May exhibit a lower number of cases compared to the period from June to October, with the highest number of cases occurring in November and December. Additionally, Fridays consistently have the highest number of accidents among the days of the week.

Furthermore, the data indicate that males have a higher incidence of accidents compared to females. This gender disparity could potentially be attributed to a higher number of male drivers on the road. However, further analysis is needed to explore the underlying factors contributing to this trend.

Moreover, it is noteworthy that the number of accidents during non-holiday periods surpasses those during holiday periods. This could be attributed to various factors, such as the ratio of holiday to non-holiday periods or different driving behaviours during these times.

These findings emphasize the need for the Transportation Safety and Infrastructure Improvement Agency (TSIIA) to develop and implement additional policies, road safety campaigns, and interventions to address the identified patterns of car crashes. By considering these trends and implementing targeted measures, the TSIIA can work towards reducing the number of accidents and promoting road safety in the country.

5.4 How do the frequency and severity of vehicular incidents vary across distinct road categories and natural light?

Description: This question aims to understand the differences in the frequency and severity of vehicular incidents across various road categories and natural light. It focuses on analyzing and comparing the number of incidents and their severity in different types of roads such as urban, rural, arterial, etc and over various types of natural light like bright sun, overcast, dark, and twilight.

Relevance to the Audience: By addressing this question, our target audience, including safety intervention specialists, transportation engineers, and policymakers, can gain valuable insights into the distribution and impact of vehicular incidents across different road categories and under various natural light conditions. This information will enable the development of targeted interventions, infrastructure improvements, and policies that account for both road type and natural lighting factors, effectively reducing incidents and mitigating their severity.

Relevant Dataset Parts: To answer this question, we analyze specific columns from our dataset, namely 'Road category', 'Natural light', and 'Current year' (derived from 'Crash year'). The 'Road category' column classifies the roads into different types such as urban, rural, arterial, and more. The 'Natural light' column classifies into the bright sun, overcast, dark, and twilight. We also consider the 'Crash Severity' parameter to understand the severity level of incidents. By examining these dataset parts, we can create visualizations that allow us to compare and visualize the frequency and severity of vehicular incidents across distinct road categories and natural light.

Sub Question: Which road category and natural light aspect has the highest number of crashes?

By addressing this question, our target audience, including safety intervention specialists, transportation engineers, and policymakers, can gain valuable insights into the distribution and impact of vehicular incidents across different road categories and under various natural light conditions. The data reveals a consistent impact of 'bright sun' across most road categories, suggesting a potential correlation between intense sunlight and the number of crashes. Among the different road categories, the arterial urban road category consistently exhibits the highest number of crashes across all analyzed years, emphasizing the need for focused safety interventions and infrastructure improvements in densely populated urban areas. This comprehensive analysis will aid in the development of targeted measures to reduce accidents, improve road safety, and account for the influence of natural lighting factors, particularly 'bright sun'

Dashboard Visualization:

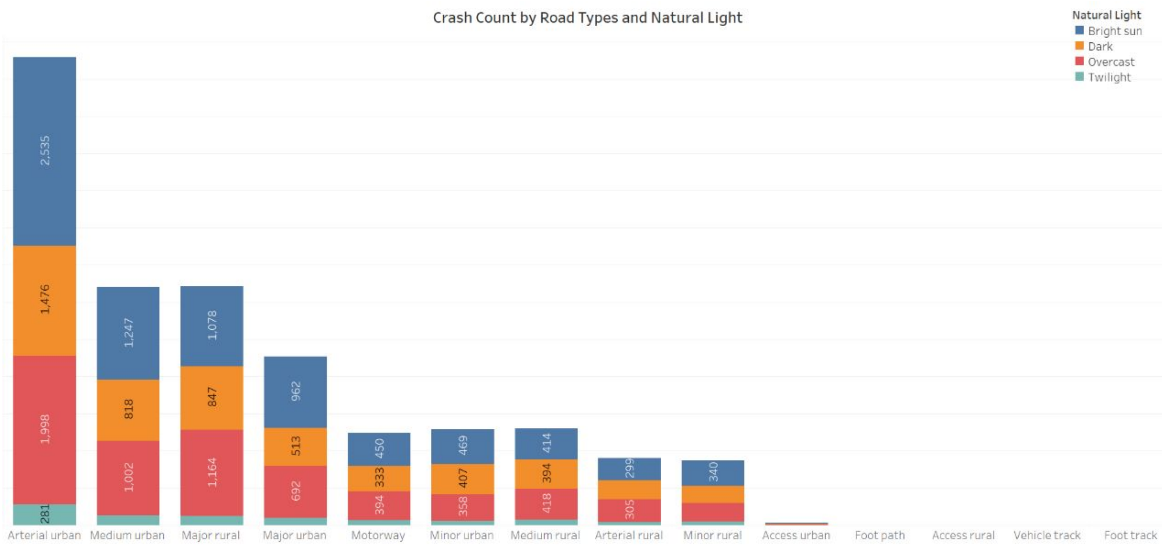

Figure 7: Accidents based on Road types and Natural Light

Conclusion: The analysis of the dataset reveals that the frequency and severity of vehicular incidents vary across distinct road categories and natural light conditions. Arterial urban roads consistently have the highest number of crashes over the years, followed by major urban and medium urban roads. Notably, the aspect of natural light, particularly the bright sun, has a significant impact on car crashes, with dark and overcast conditions also contributing to incidents. In comparison, twilight has a relatively lower contribution to car crashes. Access urban and footpath roads have relatively lower crash numbers. This information highlights the need for targeted interventions and safety measures in the road categories with higher incident rates to improve road safety and reduce the number of accidents.

5.5 What is the distribution of accidents over junction types and weather conditions?

Description: This question aims to analyze the distribution of accidents across various junction types and weather conditions. By utilizing filters such as Crash Severity, Vehicle Type, and Road Category, we can gain insights into how accidents are distributed based on different junction types and weather conditions. Understanding this distribution helps identify high-risk junction types and corresponding weather conditions and enables the development of targeted interventions and safety measures to reduce accidents and improve road safety.

Relevance to the Audience: This analysis provides valuable information to a diverse range of stakeholders, including transportation planners, policymakers, and law enforcement agencies. Understanding the distribution of accidents across junction types and weather conditions helps transportation planners prioritize infrastructure improvements and implement safety measures specific to high-risk junctions. Policymakers can develop regulations and guidelines to enhance safety at these critical points, and law enforcement agencies can focus their efforts on monitoring and enforcing traffic regulations in areas with a higher incidence of accidents.

Description of the Dashboard Component: The interactive dashboard component presents two donut visualization of the distribution of accidents across different junction types and weather conditions. The first donut represents the total number of accidents based on the junction types. The second donut represents the total number of accidents based on the weather conditions. Users can interact with the donut visualizations by selecting filters such as Crash Severity, Vehicle Type, and Road Category to explore the distribution of accidents based on specific criteria. When one junction type is chosen corresponding weather conditions data is highlighted.

Relevant Dataset Parts: The analysis utilizes a dataset that includes information about accidents, including the junction type and weather conditions involved in each accident. By leveraging this dataset and applying appropriate filters, we can categorize accidents based on their respective junction types and weather conditions and determine the distribution across these categories.

Sub Questions:

- Which junction type and weather conditions have the highest number of accidents?

  According to the data analyzed, the junction type with the highest number of accidents is "Straight Roads" with "Fine" weather conditions. Understanding this information is crucial for transportation planners and policymakers to prioritize safety measures and infrastructure improvements for this particular junction type and weather condition, aiming to reduce the number of accidents and enhance road safety in those areas.

- Are there any significant differences in the distribution of accidents across junction types and weather conditions between different crash severity levels?

  Upon analyzing the distribution of accidents across junction types and weather conditions, we observe certain variations between different crash severity levels. For example, "Straight Roads" with exhibits a higher proportion of severe accidents compared to the rest. This suggests that certain junction types may be more prone to more severe accidents than others. Understanding these differences allows transportation planners and policymakers to tailor safety interventions and infrastructure improvements specific to each junction type and crash severity level. By addressing the factors contributing to severe accidents in particular junction types, stakeholders can implement measures to mitigate risks and reduce the severity of accidents at those locations.

Conclusion: The analysis of the dataset and visualization on the donut dashboard provides insights into the distribution of accidents across various junction types and weather conditions. Stakeholders can utilize this information to make data-driven decisions and implement targeted interventions to improve road safety. By understanding the junction types and weather conditions associated with a higher incidence of accidents, transportation planners can prioritize improvements and safety measures at these critical points. Policymakers can develop regulations and guidelines specifically addressing the challenges associated with different junction types and weather conditions. Overall, this analysis contributes to the reduction of accidents and the enhancement of road safety at various junctions depending on different weather conditions.

Dashboard Visualization:

Figure 8: Distribution of accidents over Junction type and weather conditions.

6 Predictive Question

6.1 Based on past trends, what is the predicted number of accidents in the upcoming year if no new safety measures are implemented?

Description: For our predictive question, we wish to study and analyze what the estimated number of accidents would be in the next year, that is 2021 if no fresh safety measures are implemented. We wish to analyze old data and make estimates about the possible number of accidents in the next year. A linear regression model can help us answer this question. The insights gained from answering this question can help us make informed, sound decisions. The accuracy of this prediction may vary depending on a range of factors, however, this is a worthy analysis.

Relevance to the audience: Predicting the number of accidents in the absence of any new safety measures is vital for the bureau because of the following reasons:

- Planning and allocating resources appropriately: Resources such as personnel, safety aid kits, medical aid, etc. can be taken into account at the very beginning and allocated suitably with the knowledge of possible crashes in the next year. They can also ensure that all the necessary resources are available in adequacy.
- Assess risks: Predicting crashes will allow the bureau to analyze and evaluate potential risks such as road conditions, vehicle conditions, etc. This will further aid them in developing mitigation measures.

- Making decisions: The Bureau for national traffic safety is involved in making a number of policies and decisions regarding the overall road safety of the public. By anticipating a certain number of accidents for the following year, the bureau can effectively evaluate the effectiveness of any current measures that are in place. They can then either modify them or implement fresher measures.
- Educate the public about road safety: The information and knowledge gained by predicting the possible number of accidents can be shared with the public. This will enlighten them to be alert and cautious about their on-road behaviours and activities. This imparting of knowledge can be carried out in various ways, such as campaigns, posters, flyers, etc.
- Study long-term patterns: The predictions made can be studied alongside older, historical data to gain an insight into the overall direction of the trends.

Relevant dataset parts: To answer our predictive question, the parts of the dataset that we have investigated are Date as a measure of year and count of accidents in each year. We have used the calculated field called MonthYear where each year between 2016 and 2021 is split up into different months. Our graph is a linear graph produced by a linear regression model in the Tableau software where the MonthYear is mapped against the number of crashes that took place in that specific time period.

Description of the dashboard component: Our interactive dashboard displays a line graph in which a linear graph is mapped between the fields 'MonthYear' and Count of crashes for each 'Month'. Our chart also displays the trend lines for actual and estimated numbers of accidents. The darker blue on the line graph (non-dashed) represents the actual recorded data mapped against the month. The light blue (non-dashed) portion of the line displays the predicted count of crashes for the year 2021.

Description of the predictive model used: Tableau's 'Analyze' tab offers the 'Predict' feature which allows us to draw out possible estimates for the potential number of accidents in the year 2021. Tableau has employed a linear regression model to try and find a relationship between the number of crashes (dependent variable) and the year of the crash (independent variable). This feature is also known as forecast where historical data is studied to draw inferences about future knowledge. The total number of predicted accidents for the year 2021 as per our model and graph is 26580.

Conclusion: From the graph, we can see that the trend lines for both the predicted and actual values of the count of crashes start at different points but eventually submerge at the top. The estimated/predicted values of the number of crashes for the year 2021 seem to be on par with 2019 (the year with the most crashes) and at some points even higher than 2021. This information is very valuable as the possibilities of dangers can be shared with the public and informed decisions and prevention measures can be put in place.

Dashboard Visualization:

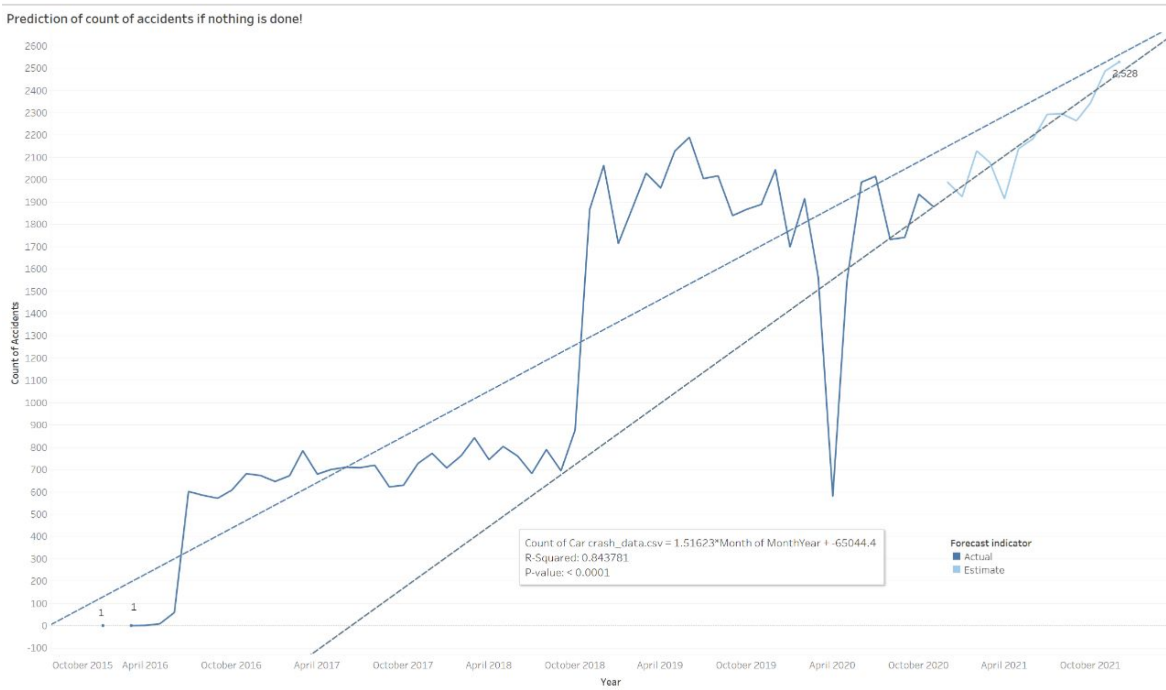

Figure 9: Predicted number of accidents in the upcoming year if no safety measures are taken.

6.2 How does the number of fatal crashes vary over different months and years in New Zealand, and what is the predicted count of fatal crashes based on a linear regression model?

Description: This analysis aims to investigate the variation in the number of fatal crashes over different months and years in New Zealand. Additionally, it includes predictions for the count of fatal crashes in the year 2021 based on a linear regression model. The insights gained from answering this question can help us make informed, sound decisions. The accuracy of this prediction may vary depending on a range of factors, however, this is a worthy analysis.

Relevance to the Audience: The analysis of fatal crash severity and its correlation with months and years holds immense significance for the Transportation Safety and Infrastructure Improvement Agency (TSIIA) and its various divisions in New Zealand. As a government organization, ensuring a minimal number of fatal accidents resulting in loss of life or long-term disabilities is of paramount importance.

- Targeted Interventions: By identifying specific time periods with a higher frequency of fatal crashes, the TSIIA can develop focused interventions tailored to address the challenges presented during those months.
- Infrastructure Improvements: Insights gained from the analysis can guide enhancements in road design, signage, and medical facilities. Additionally, traffic control measures can be reviewed and

improved based on seasonal or yearly trends to enhance road safety.

- Policy Advocacy: The analysis provides a solid foundation for advocating evidence-based transportation safety policies, empowering the TSIIA to present data-driven arguments supporting policy changes or implementations.
- Resource Allocation: Efficient resource allocation is enabled by anticipating crash rates during different time periods. Personnel, emergency services, and resources can be strategically allocated to areas of higher need.
- Public Awareness and Education: The analysis findings facilitate the development of public awareness campaigns and educational initiatives. By informing the public about risks and encouraging responsible behaviour during critical months or years, overall road safety can be improved.

Leveraging insights derived from the analysis of fatal crash severity and its relationship with months and years, the TSIIA and its divisions are empowered to make informed decisions, implement targeted interventions, improve infrastructure, advocate for evidence-based policies, allocate resources effectively, and educate the public. Collectively, these actions contribute to the enhancement of road safety in New Zealand, reducing the occurrence of fatal crashes and minimizing their impact on individuals and communities.

Relevant Dataset Parts: The dataset used for this analysis comprises information on fatal crashes in New Zealand. It includes a calculated field derived from the crash severity data column, representing the severity of the crashes. Additionally, the dataset incorporates a calculated field called month-year, obtained by combining the crash year and month of the year.

To perform the linear regression model and obtain predicted values, we used a predictive modelling function with the target expression as the count of fatal crashes and the predicted expression as the months of the year. This allowed us to establish a relationship between the two variables and generate predictions based on the linear regression analysis.

Description of the dashboard component: The dashboard component consists of a line graph displaying the trends in fatal crash cases over different months and years in New Zealand. The actual recorded data is represented by an orange line, illustrating the count of fatal crashes. Additionally, a blue line represents the results of a linear regression analysis that predicts the number of fatal crashes for the month and year 2021.

Description of the predictive model used: The predictive model employed in this analysis utilizes Tableau's modelling function called MODEL_QUANTILE. The specific configuration of the model includes using a linear regression approach (model=linear) with a quantile value of 0.5. The target expression for the model is the sum of fatal crashes, while the protective expression is the attribute of month and year. This combination of model settings allows for the generation of predictions for the future count of fatal crashes based on the established regression patterns and the quantile specified.

Conclusion: The line chart visualization in Tableau provides valuable insights into the variation of fatal crashes over different months and years in New Zealand. By comparing the actual recorded data with the predicted count of fatal crashes based on the linear regression model, the audience can assess the accuracy of the model and gain important information for decision-making. This analysis can significantly contribute to the efforts of the TSIIA and its divisions in implementing targeted interventions, enhancing transportation infrastructure, and advocating for evidence-based policies to effectively reduce fatal crashes and improve road safety across the country.

Analyzing the predicted visualization, it is evident that the number of fatal crashes showed a consistent increase from 2016 to 2018. However, in late 2019, there was a noticeable drop in the number of fatal crashes, which aligns with the actual recorded data. After the temporary decrease, the fatal crash count began to rise again, exhibiting some fluctuations but without a sustained decrease.

The prediction for the year 2021 indicates an upward slope, suggesting a potential increase in fatal cases if no additional measures are implemented. This finding emphasizes the urgency for the TSIIA to develop new policies and intervention plans to mitigate the risk and address the factors contributing to fatal crashes.

These insights highlight the importance of closely monitoring and addressing the factors contributing to fatal crashes during periods of increasing trends. By utilizing this information, the TSIIA can strategically focus their interventions and initiatives to effectively reduce fatalities on the roads. Additionally, the predicted visualization reinforces the need for ongoing efforts to sustain and further improve road safety measures. Despite some temporary drops, the data suggests that fatal crashes have not shown a continuous decline, underscoring the necessity for continued vigilance and the implementation of evidence-based measures to ensure road safety in New Zealand.

Dashboard Visualization:

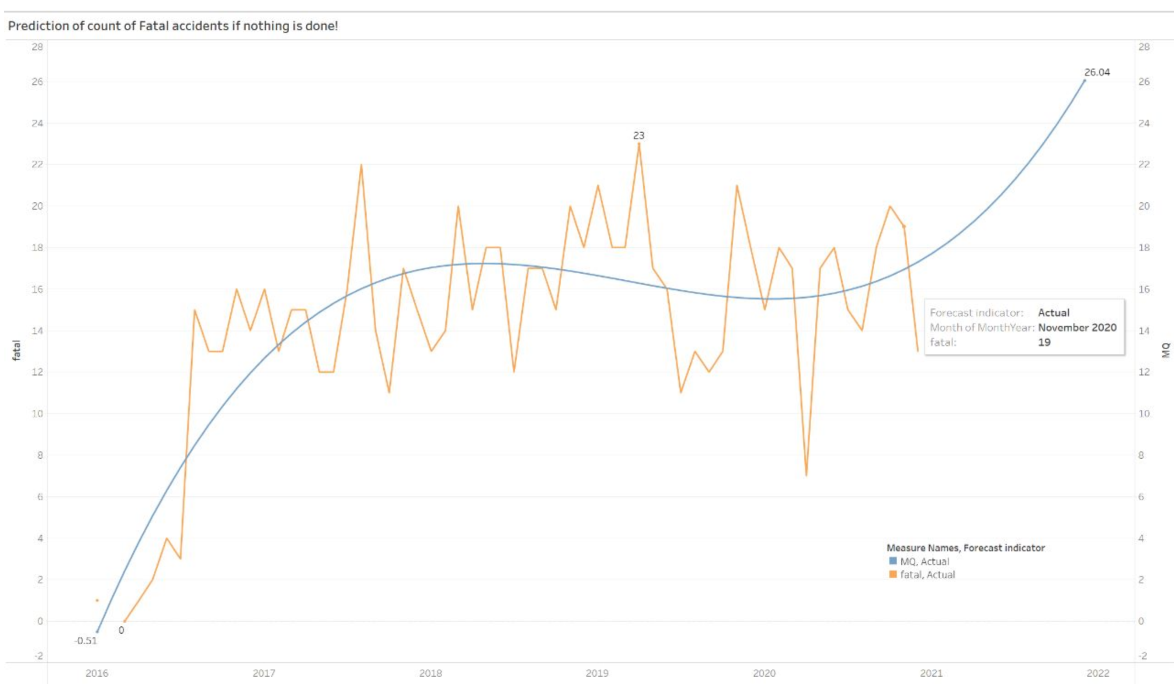

Figure 10: Predictive fatal crashes.

6.3 What are the observed trends of car crashes based on historical data for weekdays and weekends, and what can these trends indicate for 2021?

Description: This question aims at analyzing crash data that will uncover any patterns or trends that highlight the likelihood of increased or decreased car crash incidents during weekends compared to weekdays. Understanding these variations can inform targeted safety measures and interventions to mitigate risks and enhance road safety on specific days. By analyzing crash data, this investigation aims to uncover any patterns or trends that highlight the likelihood of increased or decreased car crash incidents during weekends compared to weekdays. Understanding these variations can inform targeted safety measures and interventions to mitigate risks and enhance road safety on specific days.

Relevance to the audience: This analysis holds significant relevance for traffic safety analysts, road planners, law enforcement agencies, and policymakers. Understanding the expected difference in car crash occurrence between weekends and weekdays provides valuable insights into the underlying factors and potential risks associated with specific days of the week. This information can aid in the development and implementation of targeted road safety measures, such as increased enforcement or awareness campaigns during high-risk periods. By identifying patterns and variations in crash occurrences, decision-makers can proactively allocate resources, enhance traffic management strategies, and implement preventive measures to reduce the overall number of car crashes and improve road safety outcomes. More workforce could be added during high-risk days to prevent crashes and maintain high alerts.

Relevant dataset parts: The dataset incorporates calculated fields called month-year, obtained by combining the crash year and month of the year, sum(no) which aggregates all the crashes that happened during weekdays and sum(yes) aggregates all the crashes that happen over weekends. We used a linear regression model to generate predicted values. By utilizing a predictive modelling function, we established a relationship between the count of crashes over weekdays and weekends as the target expression and the year as the predicted expression. This enabled us to analyze the variables and generate predictions through linear regression analysis.

Description of the dashboard component: The dashboard component consists of two visualizations:

- The first line graph visualization presents the actual number of crashes that occurred over weekdays, while the predicted number of crashes based on the linear regression model is represented by the green line. We employed a linear regression model to generate predictions without regularization. The model_quantile field predicts the number of crashes at a specific quantile, such as the 50th percentile, indicating the expected crash count in 50% of cases. By comparing the actual and predicted crash counts using linear regression, the line graph aids in identifying which part of the week that are most susceptible to crashes. We observe an exciting negative trend in the prediction during weekdays, suggesting that there is likely no need to allocate additional workforce.
- The second line graph visualization presents the actual number of crashes that occurred over weekends, while the predicted number of crashes based on the linear regression model is represented by the green line. We employed a linear regression model to generate predictions without regularization. Again we see an exciting negative trend in the prediction during weekdays.

Although there are negative trends for future crashes, the estimated number of crashes during weekdays is 700 points, compared to 350 points on weekends. This indicates the need for greater focus and safety programs during weekdays rather than weekends. A potential reason for this disparity could be the mandatory travel during weekdays, whether for work or school. It is unlikely to find people at home during weekdays, unlike at weekends when they tend to engage in household chores and spend time resting at home.

Description of the predictive model used: The analysis utilizes Tableau's MODEL_QUANTILE function for predictive modelling. The model is configured with a linear regression approach and a quantile value of the 50th percentile. The target expression focuses on the sum of crashes over weekdays and weekends, while the predictive expression considers the attribute of crash year. These settings enable the generation of predictions for future car crash counts specifically for weekdays and weekends.

Conclusion: In conclusion, our predictive analysis reveals a notable difference in car crash occurrences between weekends and weekdays. The model suggests that weekdays are expected to have a higher number of car crashes compared to weekends. This insight underscores the need for targeted road safety measures and interventions, particularly during weekdays, to mitigate risks and reduce the overall number of car crashes. By focusing on specific days of the week, stakeholders can implement effective strategies to enhance road safety and minimize the occurrence of car crashes.

Dashboard Visualization:

Figure 11: Predictive count of accidents during the weekdays and weekends if no measures are taken.

6.4 What is the number of accidents that would have alcohol and drug involvement in the year 2021 based on data from 2016 to 2020?

Description: We wish to analyse the provided data and understand it in order to draw up estimates for the number of accidents that could take place in the following year, that is 2021, in which alcohol and/or drugs are involved. We will thoroughly study past data and use tableau's model functions to draw predictive values. Studying this question will provide us with an abundant understanding of how alcohol and drug usage impacts public driving. The bureau can use this insight to curate measures that will be effective. The accuracy of the model may vary while drawing predictive values depending on a range of other factors. However, this is a question worth studying and answering.

Relevance to the audience: Predicting the number of accidents in the year 2021 based on whether or not alcohol or drugs are involved is going to be very beneficial to the bureau for the following reasons:

- Curate specific interventions: Understanding the magnitude of how many crashes could take place due to the usage of drugs and alcohol in the following year would help the bureau immensely in developing programs that focus on the issue of inebriated and intoxicated driving. These programs could be in the form of prevention drives, campaigns that spread awareness, etc. More checkpoints that ensure sobriety can be implemented on roads heavily prone to crashes.
- Developing and evaluating policies: Existing policies that deal with alcohol and drug usage while driving can be evaluated and assessed by understanding the potential number of crashes that could take place in the next year. The outcome of this evaluation can help the bureau further create more effective, relevant, evidence-based policies that address the issues of drunk and intoxicated driving better.
- Educate the public: The bureau can use the knowledge gained via this analysis to create awareness about the ramifications of drunk and intoxicated driving among the general public. This will encourage the public to abstain from indulging in impaired driving and make responsible choices. The bureau can even collaborate with other organisations such as universities, media companies, etc. to disseminate important information to the public efficiently.
- Evaluate existing measures and curate newer, effective ones: Predicting numbers for the following year can help our audience to evaluate how effective existing measures are. Comparing predicted and actual numbers can help them gain a better understanding of whether the measures in place have had the outcome they wished to see.
- Resource sharing: To combat driving under the influence of alcohol and drugs, our audience can collaborate with other agencies that are also focused on this issue. A few examples of such organisations would be law enforcement, healthcare agencies, etc. Resources would be available in plenty that can be shared.

Relevant dataset parts: To predict the number of accidents that could occur in the year 2021 in which the usage of alcohol and drugs are present, we curated three distinct graphs for better understanding. The first one takes a detailed look at how the presence of alcohol usage (usage suspected-tested, usage suspected-not tested, usage not suspected) impacts the number of accidents. The parts of the dataset that were used in this first chart are the following fields: Month of MonthYear (a calculated field that is parsed from its original format of MMMM-YYYY and presented as a string concatenation of month of the accident's occurrence and the year in which it occurred), count of all the car crash data segregated by the whether or not alcohol usage has been suspected and the data field Alcohol Suspected which categorises data rows into 3 mentioned categories. The second chart takes a look at how drug usage (usage suspected-tested, usage suspected-not tested, usage not suspected) has impacted the number of accidents in the past and used to draw inferences for the following year. The parts of the dataset that were used in this first chart are the following fields: Month of MonthYear (calculated field mentioned above), count of all the car crash data segregated by the whether or not drug usage has been suspected and the data field Drug Suspected which categorises data rows into 3 mentioned categories. The third chart takes a look at the overview of information presented by the first chart. Here we have created a new field named 'Substance Suspected' in which we check each entry of data to see whether either alcohol/drug or both alcohol and drug usage has been suspected. Along with this field, we have again used the fields of Month of MonthYear and the count of car crash data fields (segregated by whether substance, ie alcohol and/or drug usage has been suspected).

Description of the dashboard component: We have created and presented 3 charts to predict the number of accidents that could occur in the year 2021 with the involvement of alcohol and/or drugs. The first one primarily deals with whether alcohol usage has been suspected (tested or not tested) or not. We have graphed out the count of car crash data over different months between 2016 and 2020. The colours of the lines signify whether alcohol usage was suspected or not. The lines that are not curved or extended to a year after 2020 (that is, to 2021) depict the actual, present data. The curved lines that have been extended to the year 2021 denote the prediction that our model quantile prediction model has predicted for us based on past data. The second chart deals with whether drug usage has been suspected (tested or not tested) or not. We have mapped out the count of car crash data over different months between 2016 and 2020. The colours of the lines display whether drug usage was suspected or not. The lines that are not curved or extended to a year after 2020 (that is, to 2021) depict the actual, present data. The curved lines that have been extended to the year 2021 denote the prediction that our model quantile prediction model has predicted for us based on past data. The third chart takes a slightly more holistic look at the data provided. A new data field named 'Substance Suspected' has been created to check whether either Alcohol or Drug usage has been suspected for each row of data. Rows with both alcohol and drug usage suspected have also been marked as Suspected for substances. Here, too, we have mapped out the count of car crash data over different months between 2016 and 2020. The colours of the lines demonstrate the presence/absence of suspicion for the usage of alcohol/drugs during the time of the crash. Here, too, the lines that have not been extended up to 2021 denote the actual available data. The curved lines that have been extended up to 2021 denote the predicted values that our model quantile prediction model has predicted for us having analysed available, historic data.

Description of the predictive model used: For the predictive model, we have employed the model-quantile function offered by Tableau. This function takes a model, a quantile, a target expression and a predictor expression as its inputs. The target expression sets the limits for the probable range between which a numeric value is returned. In our figure, we have set the quantile to be 0.5 which means the median will be predicted. The count of car crash data for whether or not alcohol usage has been suspected (graph1), whether or not drug usage has been suspected(graph2) and whether either alcohol or drugs or both have been suspected(graph3) is the target expression that will be predicted. The predictor expression that is used is the Month of MonthYear (converted to attribute). The default model used by model-quantile is the linear regression model, we have employed this model.

Dashboard Visualization:

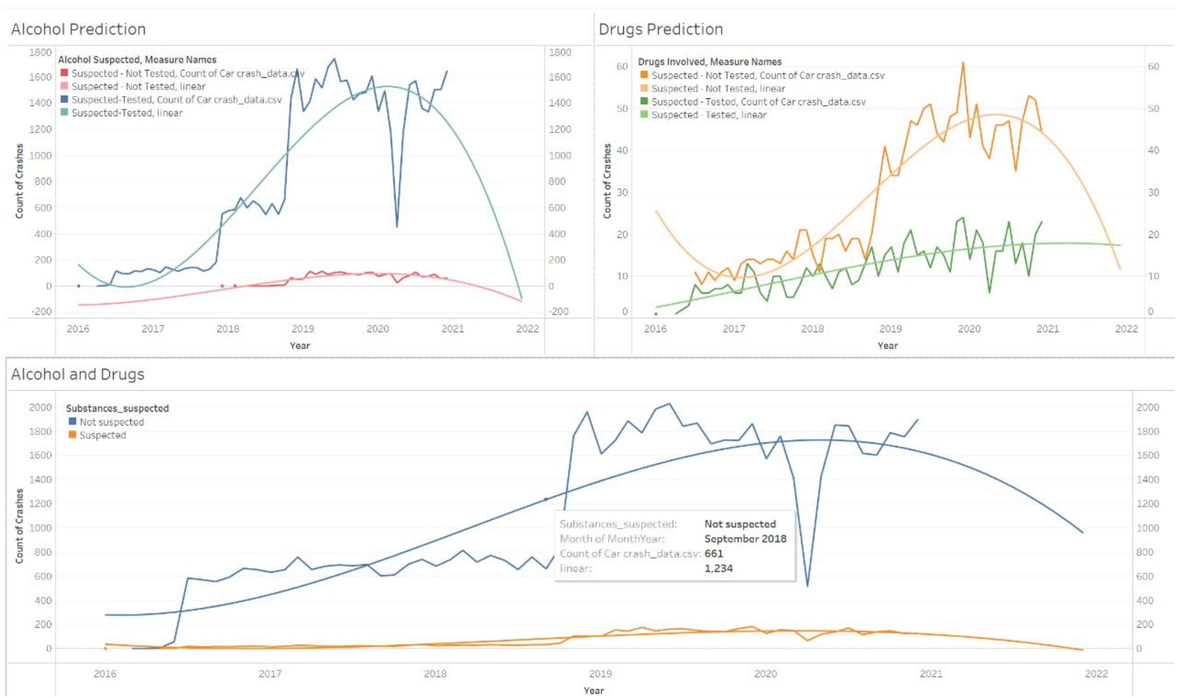

Figure 12: Predictive count of accidents due to substance abuse.

Conclusion: This dashboard aims to empower our audience with the knowledge required to understand how alcohol and drug usage affects the frequency and occurrence of road accidents. In thoroughly studying and comprehending this dashboard, we believe that our audience will be able to make decisions that are evidence-based, curate intervention plans that are effective, allot resources effectively and spread awareness about our findings among the public. The primary functionalities of our dashboard include predicting the number of accidents that could take place in the year 2021 that could be driven by driving under the influence of alcohol and/or drug usage. Past data has been used by the predictive model to come up with numbers for the following year. These predictions will allow the TSIIA to understand periods of the year when drunk/intoxicated driving are high and also take measures to alleviate the possibility of a high number of accidents occurring. A few vital rules and laws can also be enforced that will keep the public in check while on the road.

The TSIIA can also work with other agencies with the same drive to mitigate drunk/intoxicated driving to bring about positive changes and improve overall safety while driving. We also believe that the holistic overview that we have provided will aid the TSIIA in enhancing road safety, creating safety measures and interventions that will ultimately make the metro city a much safer place to drive in.

6.5 Predict the number of accidents on the basis of natural light conditions

Description: We wish to study and analyse the estimated number of accidents that could occur in the next year, that is 2021, if no new interventions are implemented. This will be carried out by thoroughly examining past data that will allow us to draw predictive values for the potential number of accidents in the following year. Answering this question will provide us with a lot of insight on what measures can be taken and what decisions can be made. Despite how the accuracy of the answer to this question may vary depending on a lot of factors, this, we believe, is a question worth analysing.

Relevance to the audience: Predicting the number of possible accidents in the year 2021 based on the light conditions is extremely relevant to the bureau due to a few reasons such as:

- This would allow them to carry out an appropriate risk assessment: the different risks associated with different times of the day depending on the levels of light present can be studied and assessed using a range of suitable measures. Periods that display higher risks and greater chances of crash occurrences can be studied thoroughly and apt measures can be put in place to mitigate such risks.
- Decisions on measures to be taken: Answering this question would expose valuable information regarding times of the day when there is a greater probability of crashes occurring. This information can be used by the bureau to curate measures and enforce rules/laws to reduce the chances of such accidents taking place.
- Allocate resources appropriately: This is a major step in ensuring the measures decided upon are implemented to satisfaction. Analysing the relationship between light conditions and the magnitude of crashes would allow the bureau to allocate equipment, personnel and other resources to areas of concern. This will ensure that all the necessary supplies are available in sufficiency.
- Infrastructure design and road plans: This is a vital measure that has the potential to avoid the occurrence of numerous crashes. A few examples of measures that can be taken are improving the lighting systems in susceptible areas, using markers that improve visibility, etc. Such measures and plans can be drawn up only if the correlation between the number of crashes and light conditions is studied.
- Make decisions rooted in data: Studying the data, especially the relationship between crash frequency and light conditions will help the bureau make data-driven decisions that are more likely to be effective and not arbitrary.
- Curate awareness drives: Being aware of how many accidents occur in the presence/absence of different lighting levels can help the bureau pass on awareness about this to the general public. The public can be empowered to be cautious and safer during different periods of the day. This is likely to decrease the number of accidents in the following years greatly.

Relevant dataset parts: To predict the number of accidents that could occur in the year 2021, we curated two distinct graphs. The first one takes a detailed look at all the levels of natural light that are present, namely Bright sun, twilight, dark and overcast. The parts of the dataset that were used in this first chart are the following fields: Month of MonthYear (this is a calculated field that is parsed from its original format of MMMM-YYYY and presented as a string concatenation of month of the accident's occurrence and the year in which it occurred), count of all the car crash data segregated by the level of natural light present when the accident took place, and the levels of natural light present(field name: Natural Light). The second chart takes a look at the overview of the information presented by the first chart. Here we have created a new field named 'Light/Dark' in which we have classified the values of 'bright sun' and 'overcast' as 'Light' and the values of 'Dark' and 'Twilight' as 'Dark'. Once again, the fields of Month of MonthYear and the count of car crash data fields (segregated by whether they belong to Light or Dark) are used.

Description of the dashboard component: We curated 2 charts to predict the number of accidents that could occur in the year 2021. While the first one takes a deeper look at the intricate information that is how many accidents occur when all levels of natural light are present namely Bright sun, twilight, dark and overcast. We have graphed out the count of car crash data over different months between 2016 and 2020. The colours of the lines signify the type of natural light that was present during the time of the accident. The lines that are not curved or extended to a year after 2020 (that is, to 2021) depict the actual, present data. The curved lines that have been extended to the year 2021 denote the prediction that our model quantile prediction model has predicted for us based on past data. The second chart takes a slightly more encompassing look at the data provided. The values 'Bright sun' and 'Overcast' have been segregated as 'Light' and 'Dark' and 'Twilight' have been assigned as 'Dark'. Here, too, we have mapped out the count of car crash data over different months between 2016 and 2020. The colours of the lines demonstrate the presence/absence of light during the time of the crash. Here, too, the lines that have not been extended up to 2021 denote the actual available data. The curved lines that have been extended up to 2021 denote the predicted values that our model quantile prediction model has predicted for us having analysed available, historic data.

Description of the predictive model used: For the predictive model, we have employed the model-quantile function offered by Tableau. This function takes a model, a quantile, a target expression and a predictor expression as its inputs. The target expression sets the limits for the probable range between which a numeric value is returned. In our figure, we have set the quantile to be 0.5 which means the median will be predicted. The count of car crash data for each level of natural light (graph1) and light/dark(graph2) is the target expression that will be predicted. The predictor expression that is used is the Month of MonthYear (converted to attribute). The default model used by model-quantile is the linear regression model, we have employed this model.

Conclusion: By providing comprehensive data access and analysis on road accidents and associated factors, the dashboard empowers the target audience to make informed decisions, prioritize interventions, allocate resources effectively, and advocate for evidence-based policies. The key functionalities of the dashboard include predicting the number of accidents based on natural light conditions, which is a critical factor in risk assessment, decision-making, resource allocation, infrastructure design, and public awareness campaigns. The dashboard utilizes past data to predict the number of accidents that could occur in the year 2021, based on different levels of natural light. These predictions enable TSIIA to anticipate high-risk periods and take appropriate measures to mitigate risks, enforce rules and laws, allocate resources effectively, design infrastructure improvements, and curate awareness drives. By leveraging data-driven insights and effective communication, TSIIA aims to drive positive change in road safety and transportation infrastructure in the metro city. In summary, the BI dashboard developed for TSIIA equips the organization with the tools and information necessary to make data-driven decisions, prioritize interventions, and advocate for evidence-based policies. By collaborating across departments and utilizing the integrated approach outlined in the report, TSIIA can leverage their expertise to enhance road safety, reduce accidents, and improve transportation infrastructure, ultimately making the metro city a safer place for residents and commuters.

Dashboard Visualization:

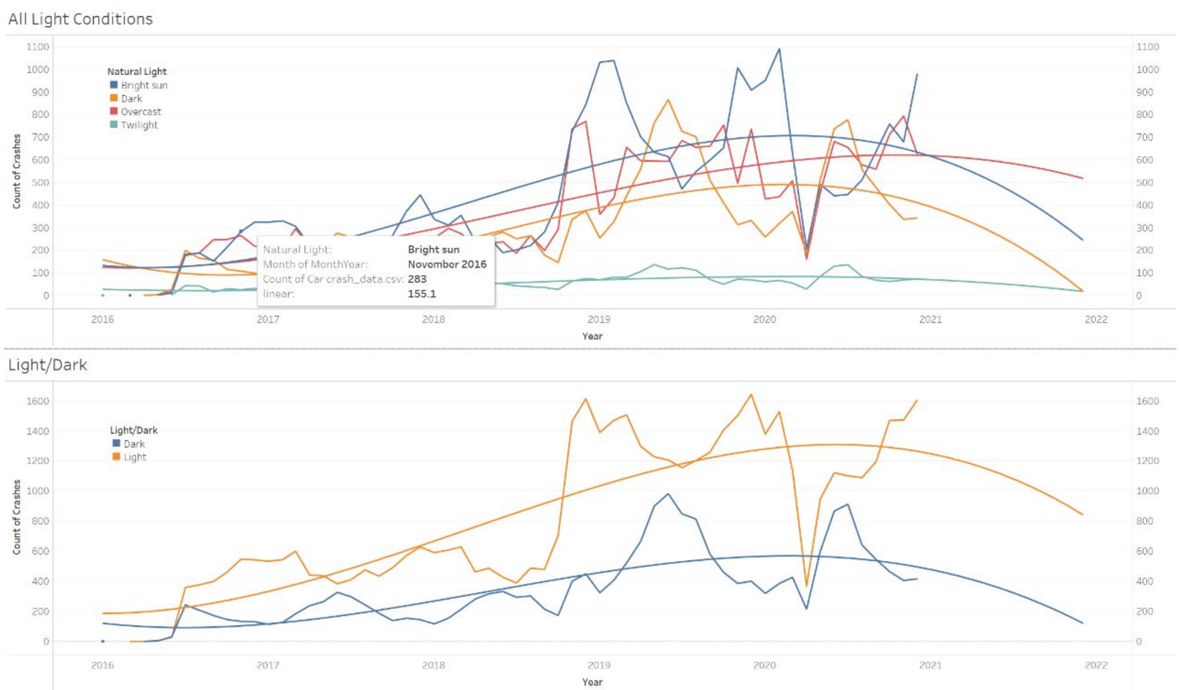

Figure 13: Predictive count of accidents based on natural light condition.

7 Dashboard Story

Once the dataset is analysed, we need to try to understand the data. We need to ask questions to describe the data and these descriptive questions are what makes the first part of the dashboard story. Only through asking questions can we make sense of the data.

Based on the dataset our first main concern is the total number of accidents in a particular year, or which vehicle type would be the one involved in crashes the most. These questions are answered through the first slide Descriptive Questions 1-3. This slide also answers the question of which days of the week might we see the most traffic accidents. Which month? Who is responsible for it? Is it a male driver or a female driver? Furthermore, did the crash happen during a holiday or a regular day? Such questions are skillfully answered through the dashboard presented in slide 1.

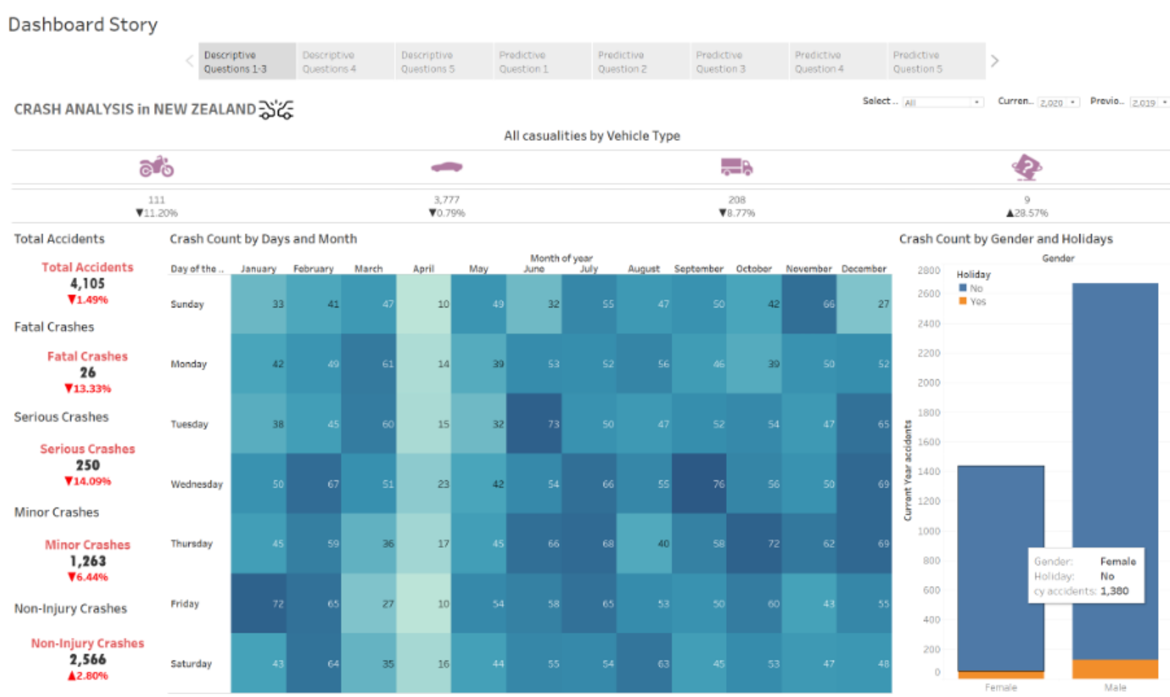

Figure 14: Slide 1: Descriptive Question 1-3

This is usually followed by the question of what the external influences on the crashes are. Through our analysis, we deduced that the type of road and natural light were some of the most important external influences. Which type of road did the accident take place on? How many accidents took place over here? What were the natural light conditions then? These questions are answered through Slide 2: Descriptive Question 4 as shown below:

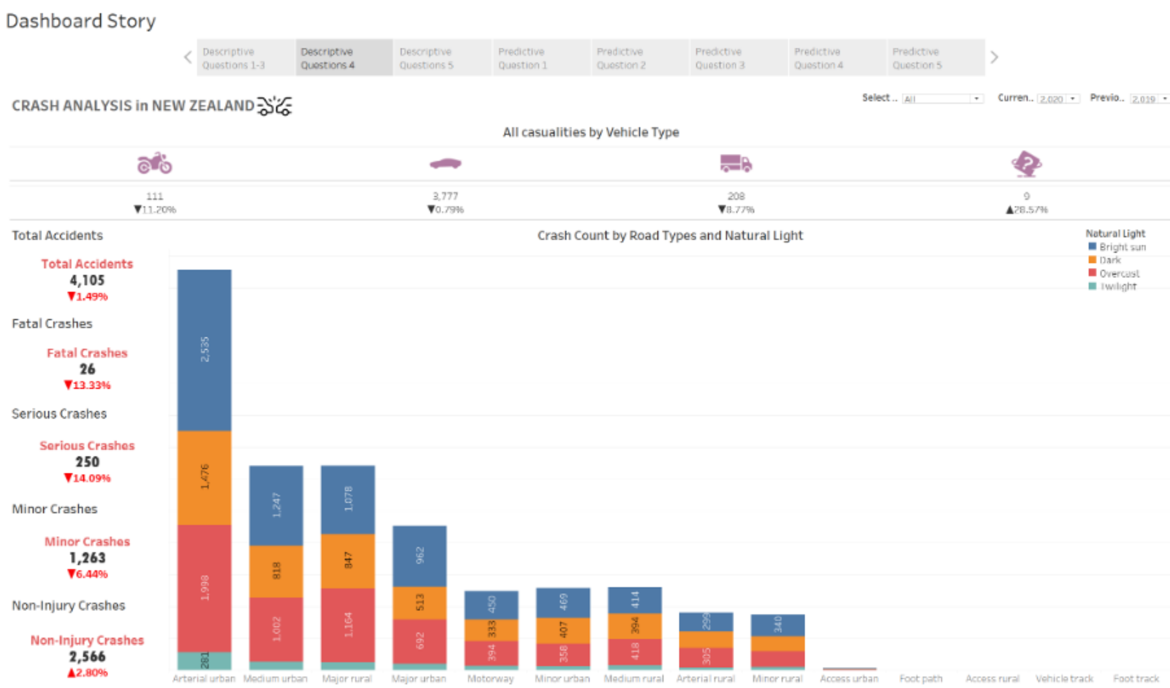

Figure 15: Slide 2: Descriptive Question 4

Finally, considering that most of our data is explored, we can not forget another important criterion in vehicular accidents, Weather conditions. Where did the accident take place? Was it a straight road? Junction? What were the weather conditions? Was it a fine weather? Snowing? Raining? These questions are answered by Slide 3: Descriptive Question 5 as shown below:

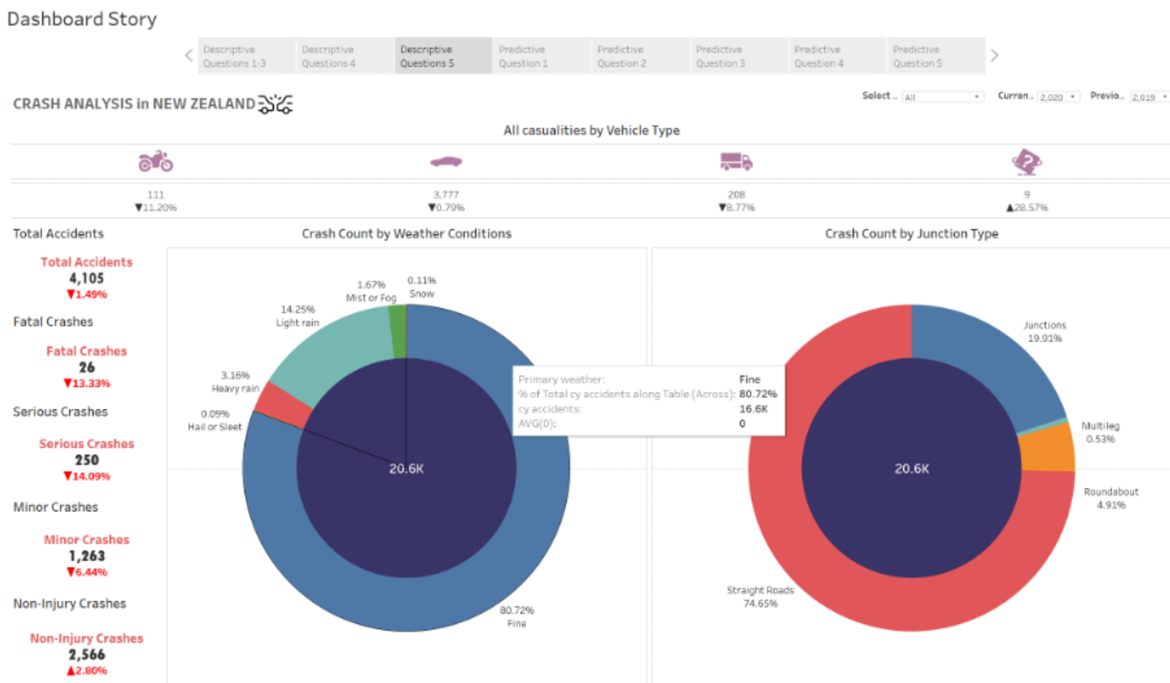

Figure 16: Slide 3: Descriptive Question 5

We have gained a brief understanding of the various factors associated with car crashes in New Zealand. We know about external influences such as weather conditions and natural light conditions. We know about the statistics of the crashes, the severities, where it has taken place and in general most of the dataset. Now, knowing these things, if we do not do anything if our organization does not take any steps, we can expect a cumulative of 26000 crashes in 2021 as seen in Slide 4: Predictive Question 1 which gives an estimate of the number of crashes if we do not take some steps to reduce or raise awareness.

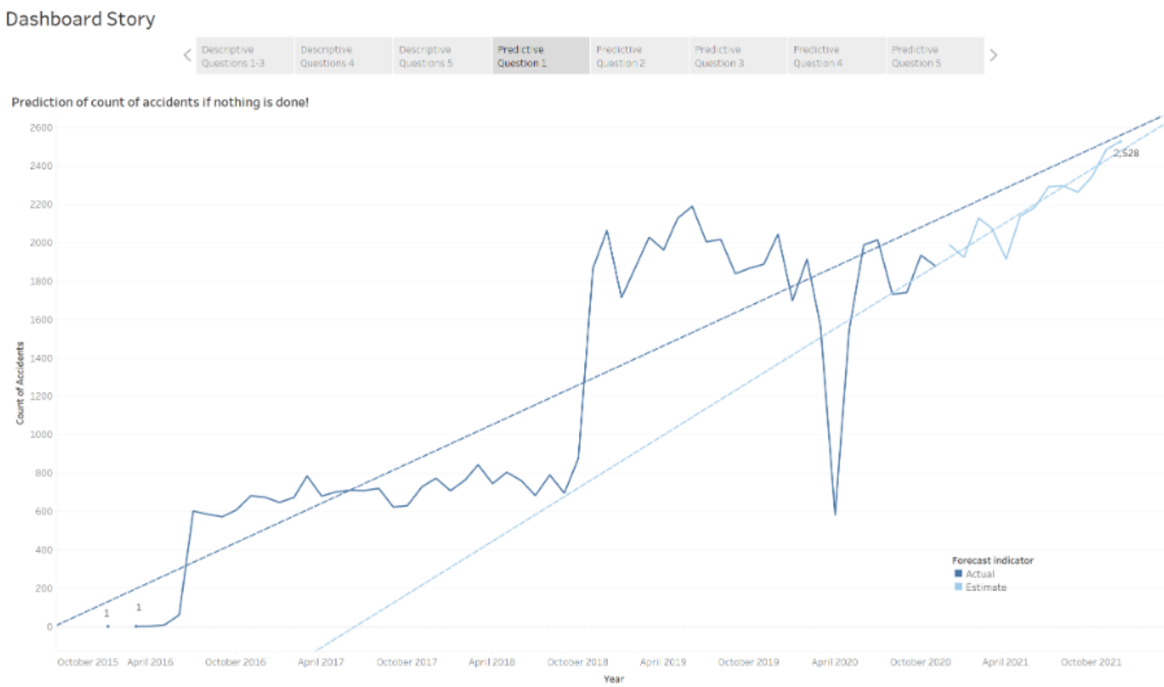

Figure 17: Slide 4: Predictive Question 1

Those were just the prediction for the total number of crashes. What about the fatalities? Those crashes where we would lose our people? The prediction of fatal crashes in Slide 5: Predictive Question 2 shall shock you:

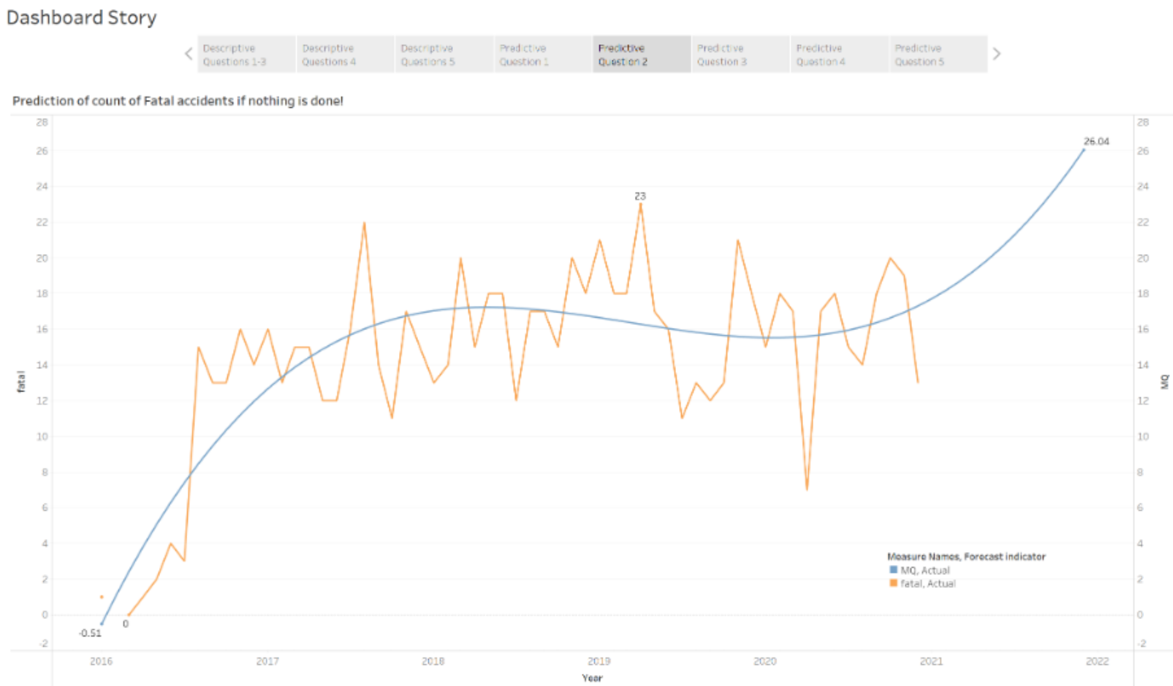

Figure 18: Slide 5: Predictive Question 2

We remember the count of accidents from the first dashboard which showed us some valuable statistics of the count over the different days of the week over different months in the selected year. Based on that, we created a prediction for the count of accidents during weekdays and weekends if nothing is done and this was the result of our prediction:

Figure 19: Slide 6: Predictive Question 3

We are sure that there might be discussions regarding us discussing the external factors influencing car crashes, but what about internal factors? What about the driver's negligence or crimes? This is answered by our specifically curated dashboard of alcohol and drug abuse. The people involved in car crashes who were suspected of substance abuse and their predicted growth rate are as seen:

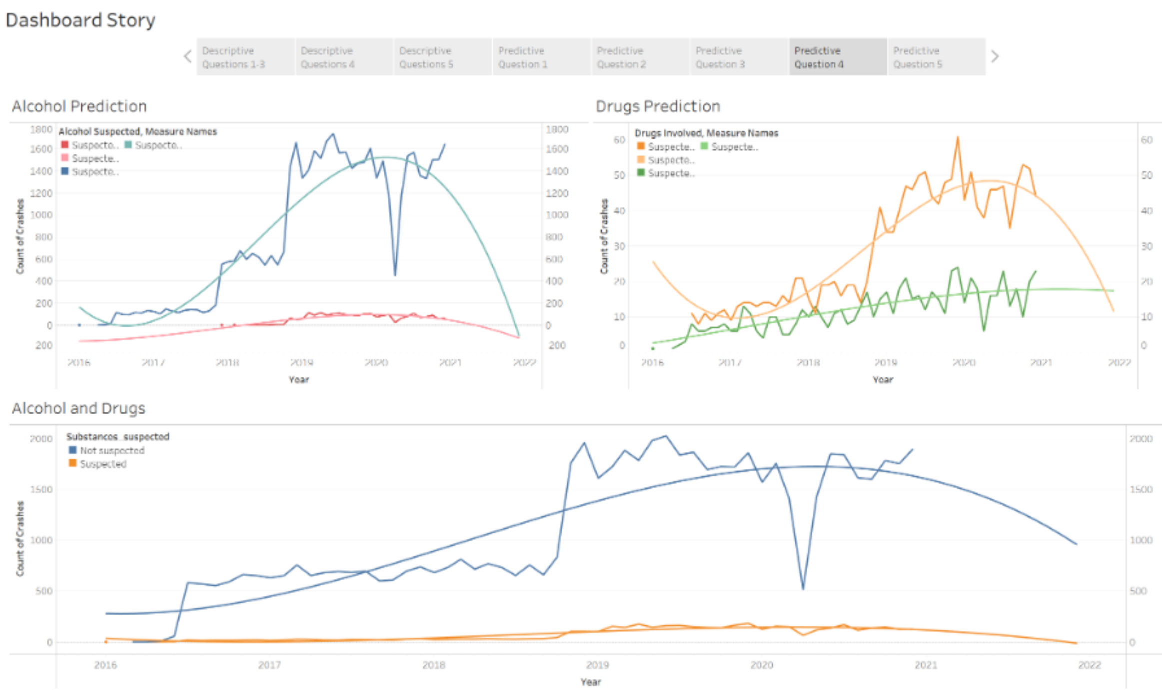

Figure 20: Slide 7: Predictive Question 4

Having spoken about substance abuse as well, we would finally speak about the external factors which can affect car crashes such as the weather and light conditions. The predictive analytics conducted revealed the following dashboard:

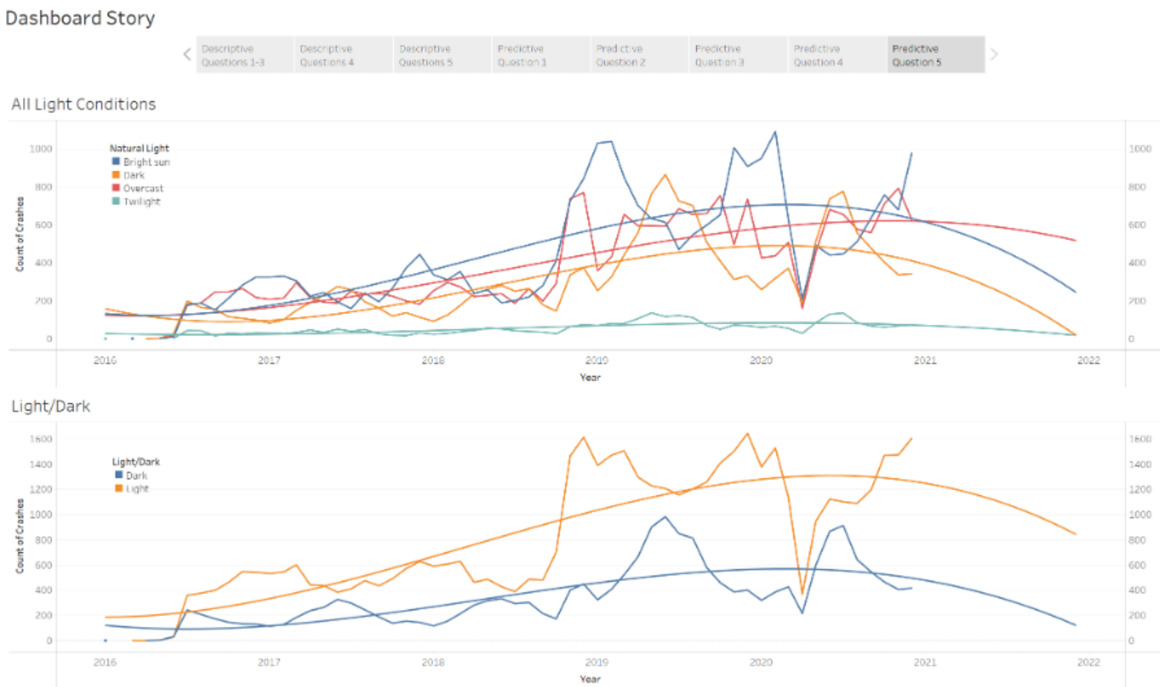

Figure 21: Slide 8: Predictive Question 5

8 Evaluation

Total number of accidents and their Crash Severity: This dashboard provides an overview of the total number of accidents and their severity. It helps TSIIA's stakeholders understand the overall magnitude of the problem and identify areas that require immediate attention. This information can be used by executive leadership to set strategic goals and allocate resources accordingly. Example scenario: The executive director of TSIIA needs to present an annual report to the government highlighting the severity of road accidents in the metro city. The dashboard can support this decision-making by providing clear visualizations of the total number of accidents and their crash severity over time. This enables the executive director to emphasize the urgency of implementing safety interventions and infrastructure improvements.

Potential barriers to use: Stakeholders might face challenges in interpreting crash severity levels if they are not adequately defined or explained within the dashboard. Additionally, if the dashboard lacks user-friendly features or is not easily accessible, it might hinder adoption and engagement.

Vehicle type and Crash Severities: This dashboard focuses on the relationship between vehicle types and crash severities. It helps TSIIA understand which types of vehicles are involved in accidents and the corresponding severity levels. This information can guide the development of targeted interventions and policies to address specific vehicle-related safety issues. Example scenario: The safety interventions division of TSIIA needs to prioritize their efforts in reducing accidents involving motorcycles. By using this dashboard, they can identify the proportion of accidents and their severity specifically related to motorcycles. Armed with this data, they can design awareness campaigns, implement safety training programs, and propose policy changes targeting motorcycle riders.

Potential barriers to use: If the dashboard does not provide a clear breakdown of vehicle types or if the data is incomplete or inconsistent, it may hinder accurate analysis and decision-making. Additionally, if the dashboard lacks interactive features that allow users to drill down into specific vehicle types, it might limit the usefulness and adoption of the dashboard.

Distribution of accidents over days months, varied by gender and holidays: This dashboard explores the distribution of accidents over days and months, considering gender and holidays. It provides insights into temporal patterns and identifies any variations related to gender or holidays. This information can help TSIIA plan interventions, allocate resources during peak periods, and identify potential gender-specific safety concerns. Example scenario: The policy and advocacy division of TSIIA is preparing a proposal to introduce traffic management measures during the holiday season. By utilizing this dashboard, they can analyze historical accident data to determine peak periods and assess if there are any gender-specific differences in accident rates during those times. This analysis will enable them to propose targeted policies and safety campaigns tailored to holiday periods and specific genders.

Potential barriers to use: If the dashboard does not present the temporal patterns clearly or fails to provide gender-specific breakdowns, it may limit the effectiveness of decision-making. Additionally, if the dashboard lacks features that allow users to compare different time periods or filter data based on specific holidays, it may hinder the adoption and utilization of the dashboard.

Accidents based on Road types and natural light: This dashboard focuses on accidents categorized by road types and natural light conditions. It helps TSIIA understand the relationship between road characteristics, lighting conditions, and accident occurrences. This information can guide infrastructure improvement decisions, such as road design, lighting enhancements, and traffic management strategies. Example scenario: The infrastructure improvement division of TSIIA is tasked with prioritizing road safety upgrades in the metro city. By using this dashboard, they can identify the road types and lighting conditions where accidents are most prevalent. Armed with this data, they can allocate resources to improve high-risk areas, enhance lighting infrastructure, and implement traffic calming measures accordingly.

Potential barriers to use: If the dashboard does not provide a clear classification of road types or if the lighting conditions are not accurately captured, it may hinder effective decision-making. Furthermore, if the dashboard lacks interactivity to explore specific road segments or lighting conditions, it might limit the dashboard's usefulness and adoption.

Distribution of accidents over Junction type and weather conditions: This dashboard analyzes the distribution of accidents based on junction types and weather conditions. It helps TSIIA understand the relationship between specific junction types, weather conditions, and accident occurrences. This information can guide decision-making related to junction design, traffic signal optimization, and weather-responsive safety measures. Example scenario: The transportation engineers within the infrastructure improvement division are planning a major junction redesign project. By using this dashboard, they can assess the accident distribution across different junction types and identify weather conditions that contribute to accidents. This analysis will inform their design decisions, such as incorporating better signage, optimizing signal timings, and implementing weather-responsive safety measures at high-risk junctions.

Potential barriers to use: If the dashboard does not provide clear categorization of junction types or lacks accurate weather condition data, it may limit the effectiveness of decision-making. Additionally, if the dashboard does not offer interactive features to compare junction types or explore accidents under different weather conditions, it might hinder adoption and engagement.

Predictive charts of what would happen if we do not do anything: This dashboard provides predictive charts that visualize the potential consequences of inaction in terms of road safety. It helps TSIIA stakeholders understand the projected increase in accidents, injuries, or fatalities if no interventions or improvements are implemented. This information can effectively communicate the urgency and importance of taking action. Example scenario: The executive leadership of TSIIA needs to secure funding for safety interventions from government officials. By utilizing this dashboard, they can present the potential future scenarios in terms of accidents, injuries, or fatalities if no action is taken. This visualization can strongly support their advocacy efforts and highlight the necessity of allocating resources to address road safety issues.

Potential barriers to use: If the dashboard's predictive models are not accurate or fail to consider relevant factors, it may undermine the credibility and usefulness of the dashboard. Additionally, if the dashboard lacks clear explanations or annotations to help stakeholders understand the predictive charts, it might hinder their adoption and trust in the provided insights.

In summary, while the outlined dashboards address key aspects of road safety and transportation infrastructure, their effectiveness in supporting decision-making relies on clear data presentation, user-friendly features, and accurate insights. Potential barriers to adoption include data quality issues, lack of interactivity, limited breakdowns or classifications, and difficulties in interpreting the visualizations. Ensuring these challenges are addressed can enhance the adoption and usability of the dashboards by TSIIA's target audience.

9 Conclusion

In this project, we have developed a comprehensive Business Intelligence (BI) dashboard for the Transportation Safety and Infrastructure Improvement Agency (TSIIA). The dashboard aims to support TSIIA's efforts in enhancing road safety and improving transportation infrastructure in a bustling metro city. It caters to a diverse range of stakeholders within TSIIA, including executive leadership, data scientists, statisticians, analysts, safety intervention specialists, transportation engineers, planners, and policy advocates.

The developed dashboard consists of multiple interactive visualizations that address critical aspects of road safety and transportation infrastructure. These visualizations include total number of accidents and their crash severity, vehicle type and crash severities, distribution of accidents over days months, varied by gender and holidays, accidents based on road types and natural light, distribution of accidents over junction types and weather conditions, and predictive charts illustrating the potential consequences of inaction.

By providing relevant metrics, clear visualizations, and interactivity, the dashboard empowers TSIIA's stakeholders to make data-driven decisions and drive positive changes in transportation safety and infrastructure. It helps them identify high-risk areas, prioritize interventions, allocate resources effectively, advocate for evidence-based policies, and communicate the urgency of action.

The usefulness of the dashboard is evident in its ability to answer critical questions and support decision-making in realistic scenarios. For example, executive leadership can leverage the total number of accidents and crash severity dashboard to advocate for resources and emphasize the need for safety interventions. The safety interventions division can utilize the vehicle type and crash severities dashboard to prioritize efforts in reducing accidents involving specific vehicle types. The infrastructure improvement division can rely on the accidents based on road types and natural light dashboards to allocate resources and plan targeted road safety upgrades. These are just a few examples of how the dashboard's insights can be applied in various decision-making scenarios.

While the developed dashboard has the potential to be highly useful, there are some considerations to maximize its effectiveness. These include ensuring clear classification and breakdowns of data, accurate and complete data collection, user-friendly features, and explanations to aid interpretation. Addressing these factors will enhance the adoption and utilization of the dashboard within TSIIA.

In conclusion, the BI dashboard developed for TSIIA has the potential to greatly contribute to its mission of enhancing road safety and improving transportation infrastructure. By providing comprehensive data analysis, interactive visualizations, and actionable insights, the dashboard equips TSIIA's stakeholders with the tools they need to make informed decisions, prioritize interventions, allocate resources effectively, and advocate for evidence-based policies. With its potential to drive positive changes, the dashboard will play a crucial role in making the metro city safer for its residents and commuters.