Visualising the Victorian road accidents data

Sarath Gopinathan, 30434904

FIT5147 — Data Visualization Report

Tutorial - 08

Tutor names:
Farah Tasnuba Kabir
Aldrich Clarence

1. INTRODUCTION

In 2021, compared to 2020, there was a 10% increase in traffic deaths in Victoria (ABC News, 2022). Despite strict lockdowns throughout the year, there was a surge in the number of fatal accidents. Getting a visualisation of this data can help us identify valid data and help give key insights to try and reduce these accidents in the future. Bureau of Infrastructure and Transport Research Economics collects various data including road accidents data. The road accidents data include latitude, longitude of the accident, accident date and time, region where the crash occurred, Road Management Act classification, alcohol time, accident number, accident status, day of the week, the definition for classifying accident code, if it was a hit and run case, the light condition during the crash, if the police attended the scene of the crash, a number of people involved, the severity of the accident and several other important metrics, data. It collects and records approximately 60 different types of data. The data collected by the BITRE is very feature-rich and makes it a good choice to explore and visualize the crash data with the influencing factors.

This report gives a thorough overview of the application that was developed to aid the general public and emergency medical service providers to understand the trends in road accidents and make predictions. The application's features and more information may be found further down.

1.1 About datasets used in this project

There were multiple datasets that were used in the developmet of this project. The data provides basic details of road traffic crashes in Australia as reported by the police each month to the State and Territory road safety authorities. Since the data is collected manually, there may be situations for erroneous and NULL values being recorded. Fatality data from 1989 to 2013 does not have any data about National Road Type, National LGA Name, SA4 Name and National Remoteness Areas. All of these discrepancies were handled well with a lot of data wrangling and cleaning to interpret the correct information.

• Victorian Road Crash Statistics dataset(2012-2017)

The observations here are made from Bureau of Infrastructure and Transport Research Economics(BITRE) and converted into a csv file that was available in (Kaggle).

Link: https://www.kaggle.com/stevenrferrer/victoria-road-crash-statistics

The dataset is in the form of a csv has data about all the accidents that took place in Victoria, Australia from January 2012 to 2017. This data was combined with the next dataset to form a combined dataset that contains records of all accidents that took place between January 2012 and June 2020.

• Victorian Road Crash Statistics dataset(2014-2020)

The observations here are made from Bureau of Infrastructure and Transport Research Economics(BITRE) and converted into a combined csv file and made available on the official BITRE website. This dataset contains data about all the accidents that took place in Victoria, Australia from January 2014 to June 2020. This was combined with the previous dataset making a combined dataset that contains records of all accidents that

took place between January 2012 and June 2020. It contains a total of 65 columns and 124,326 rows.

Link: https://discover.data.vic.gov.au/dataset/road-crashes-for-five-years-victoria

This (Bureau of Infrastructure and Transport Research Economics) contains data about latitude, longitude of the accident, accident date and time, region where the crash occurred, Road Management Act classification, alcohol time, accident number, accident status, day of the week, the definition for classifying accident code, if it was a hit and run case, the light condition during the crash, if the police attended the scene of the crash, a number of people involved, the severity of the accident and several other important metrics, data. It contains approximately 60 different types of data.

Australian fatal road crash dataset

The dataset is in the form of a csv and has data about all fatal accidents that took place in Australia from January 1989 to January 2022. It has 23 columns and 53,232 rows.

Link: https://www.bitre.gov.au/statistics/safety/fatal_road_crash_database

This (Bureau of Infrastructure and Transport Research Economics) was obtained firectly from the BITRE official website.

Seasons dataset

The seasons (Five design sheet) (Bureau of Meteorology) is not available anywhere in the form of a CSV and so a python script is written and executed to scrape the season's data from the BOM website and it is written into a new CSV file. It contains 2 columns and 12 rows.

Link: http://www.bom.gov.au/climate/glossary/seasons.shtml

• Australian States Location

To obtain the Australian State latitude and longitude values, Google's Geocoding API was used and a python script was written to retrieve the data in JSON format after hitting the API. This was then converted into a CSV file for ease of use in the development application. This csv contains 3 columns and 8 rows.

The following is the link to the API's documentation.

Link: https://developers.google.com/maps/documentation/geocoding/overview

1.2 Intended audience

This application is mainly targeted towards the general public and emergency medical service providers and is designed to benefit from historical data gathered across a number of Australian locations. This application has abundant details about major road accidents data with intrinsic details about the seasons in which the accidents took place, age group of the people who met with accidents, road geometry of the accident and so on. This provides a good understanding of the road accidents data and to identify which locations had the highest number of accidents and what were the age groups of the people, speed zones in which the accidents took place. This data can also be used to build a model to predict if accidents will

take place given the season, vehicle type, and location. This can be used to alert the general public and emergency medical service providers. The ultimate goal is to find trends and patterns in the data and help prevent accidents with the understanding of the data.

1.3 Narrative and Message conveyed

- Initially, the users are displayed an overview of how to use the application and this provides a clear understanding of the application to the users.
- General public and emergency medical service providers are the targeted audience and this application provides more understanding on the areas in Victoria with accidents and different categories in which the accidents took place location, road type, speed limit, road geometry and so on.
- The markers on the map are based on accident counts. The more the count, the greater the size. The size of the markers will clearly distinguish the number of accidents on a particular site and this will help convey the message to the users in a simple but effective manner.
- Users can get an understanding of the top sites where the accidents took place and also view their respective locations or routes through which they travel regularly to get an understanding of their location.
- The accidents can also be checked with respect to the 4 Australian seasons and how the climate, seasons affect the number of accidents.
- Overall, the fatal Victorian crash data can be compared with all of the other states fatal crash data by simply checking out the Fatal crash data tab.
- Users can also explore granual information on a specific location by clicking or choosing points on the map and line graph. The data changes according to the user's choice and the accident distribution across the week is displayed.
- An animated bar plot displays the 8 most accident-prone zones.
- Interesting trends are mentioned in each tab to showcase the data in an interesting manner.
- The conclusions tab has a few metrics that are also animated and this attracts the users to it and the data will be conveyed effectively in an attractive manner.

2. DESIGN

To understand the strategy to follow when visualising the findings and presenting them appropriately to the target audience, decisions on the type of data visualisations that give the insights successfully must be made. Other important considerations include whether technologies, such as Tableau, R Shiny, or D3, should be used to produce unique visualisations. It is also vital to identify design structures that best inform and test the practicality of concepts. To respond to these inquiries, The five-sheet design process was used to formalise drawing and create a wireframe like mockup so that we could explore their ideas without worrying about technology concerns.

2.1 Description of five design sheet

- On the brainstorming sheet, at least 10-20 ideas are drawn down, regardless of viability. The relevancy and repetition of these concepts are then filtered. The next step is to categorise the thoughts and check for any missing concepts or facts. Then, to generate more complete solutions, merge and improve the categories, and assess whether ideas compliment one another or whether merging designs might result in superior thoughts. Question your ideas and choose the three most successful ones to apply on the following sheets.
- Alternative design sheets 2,3, and 4 are made to delve deeper into the three
 brainstormed concepts. Metadata such as title, author, and date are included in these
 sheets. Then define the layout and operations/functions that are relevant to the users,
 as well as the design's focus. Finally, the discussion section shows the benefits,
 drawbacks, and practicality.
- The final realisation sheet contains all of the detailed concepts used in the application.
 This document discusses design functionality and interactions, as well as design
 patterns, algorithms, math computations, software requirements, cost, and time
 estimates.

2.2 Summary of the five design sheet used in this project

This part delves into the (Five design sheet) to gain a better knowledge of the features, interactions, and probable design structure that aided in the development of my ideas. Below is further information on the specific sheets.

• Design Sheet 1 – Brainstorming (Refer Appendix)

- ➤ A total of nine distinct visualisation concepts were found, and seven of them were chosen for further consideration based on relevance and feasibility.
- The ideas where shortlisted based on the questions including overall accident areas, top 8 accident-prone locations and other features in the data.
- ➤ Possible user interactions for different aspects including accidents in different locations, states, selecting the location, classifying them based on season, age group, speed zones and so on.
- ➤ Leaflet view with heatmaps and markers, bar view, interactive tree maps, density plot, and word cloud were used as examples of interactive visualisation concepts that best enlighten the target audience and successfully provide the insights visual representations.

• Design Sheet 2 – Layout 1 (Refer Appendix)

- ➤ This layout was structured mainly into 2 views, Leaflet view and types of bar plots animated barplot, normal barplot and circular barplot.
- ➤ The map provides information about the accident areas and it provides information on a tooltip when the user clicks on a marker. The greater the count of the accidents, the larger the size of the marker. The barplot data changes and displays the accidents that took place in the chosen location over the days of the week.
- ➤ The main barplot is an animated barplot that displays the top 8 accident prone locations and the circular barplot displays the distribution of accidents over each season.

There is an overall selection of fatal data and victorian crash data.

• Design Sheet 3 – Layout 2 (Refer Appendix)

- ➤ This layout was structured into 2 views, Treemap and Circle Packer.
- ➤ The treemap and circle packer display the different features of the data. It classifies the data into the age group of the people, vehicle type, road geometry, and speed zone.
- ➤ Users can also choose the day of the week from the slider and the data in both the charts will change accordingly.
- There is an overall selection of fatal data and victorian crash data.

• Design Sheet 4 – Layout 3 (Refer Appendix)

- This layout was structured into 2 views, line chart and bar plot.
- ➤ The line chart is an interactive linechart where the users can click on the points and the bar plot data will change accordingly. The location and count of accidents will change accordingly.
- ➤ Users can also choose the day of the week from the slider and the data in both the charts will change accordingly.
- There is an overall selection of fatal data and victorian crash data.

• Design Sheet 5 – Realisation (Refer Appendix)

- ➤ This sheet combines all of the preceding layouts and includes information on Design, including a sidebar panel, About, Top sites, seasons and conclusion.
- ➤ Overall operations including accidents sites plot, top sites charts, season wise distribution of accidents, circular bar plot, tree map and circlepackers are used here.
- ➤ Users can pick information interactively from the preceding sheets, as previously described.
- ➤ R, Shiny, CSS, HTML, VS CODE, and D3 are all essential tools. This application is expected to take 9 days to complete, with 7 working hours each day.

3. IMPLEMENTATION

- Shiny dashboard, which is part of the R libraries shiny and shinydashboard that enable design interactive web interfaces/applications, was used to build the application. Both libraries have a skeleton for a website structure that aids in the building of applications. Shiny aids with the creation of visual components for web interfaces by utilising built-in methods that are managed by the ui and server objects.
- We may leverage the usual server capabilities like www to add title page pictures and custom icons to aid build run-time UI changes because this application is a webpage.
- The interactive plotting library plotly was used to construct the Bar plots and tree maps that are part of this application. This package allows you to adjust the plot's tooltips, legends, typefaces, and colours, among other things.
- To construct plots, library ggplot2 was used in conjunction with library dplyr ability to edit the dataset and select exact data for the plot. Ggplot2 may also be used in conjunction with the plotly library to create interactive plots.
- Library Leaflet was used to create an interactive geographic plot that displays the dataset's latitude and longitude.

- Libraries shinythemes, shinydashboard, dashboardthemes and shinyWidgets were
 used to build and design the dahboard and modify the colours, themes of the overall
 dashboard.
- Libraries gganimate and gifski were used to create the animated barplots.
- Libraries shinyjs and shinycssloaders were used to enable making design changes using javascript and css.
- Libraries treemapify and packcircles were used to plot treemaps and circle packer plots respectively.

4. USER GUIDE

This section gives you a thorough overview of the shiny application's features and functions. There are a total of 4 tabs in this application that are used to display the data and identify trends. The 4 tabs are – About, Top sites, Seasons and Conclusion. This app must be viewed in full screen window view or on the browser. Below is further information about each tab.

4.1 About tab

The about tab contains a visual showing how the users must use this application. The two major button clicking areas are displayed here. The users are informed to click on the circle parker and the points of the line chart in order to display the other charts. The bar chart, treemap and circle packer data changes with respect to what the user chooses. In the Leaflet map, when the user clicks on a circle marker, the barplot values change and it displays the day of the week distribution of accidents in the chosen site. When the user clicks on a point in the line chart, the treemap and the circle packer data changes according to what the user chooses. The data changes with respect to the season and location the user chooses on the line chart. Figure 1 displays the about tab.

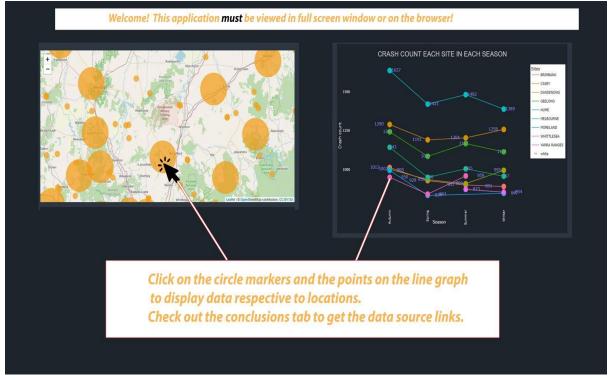


Figure 1 About tab

4.2 Top Sites Tab

The top sites tab focusses mainly on the sites where maximum accidents have occurred. This tab has 2 major options. The users can choose if they want to view data specific to Victoria or they can switch to the all Australian states fatal data. The user can compare the fatal accidents that took place in Victoria to all the other states of Australia. This tab is divided into 4 boxes. The first one is the Leaflet map view. This map view shows all the accident sites. It has circle markers on it to show the accident sites. The bigger the circle marker, the bigger the accident count in that site. The next box has an animated bar chart and it contains the top 8 sites with maximum crash count. The next box has a circular bar chart that displays the spread of the crash data across the 4 seasons in Australia. The same box also has information about interesting insights in the data presented in the current tab. Finally, the last box is another bar chart that displays the distribution of accident count over each day of the week. This chart is directly linked to the Leaflet map. When the user clicks on a marker on the map, the data for that particular site is displayed in this tab.

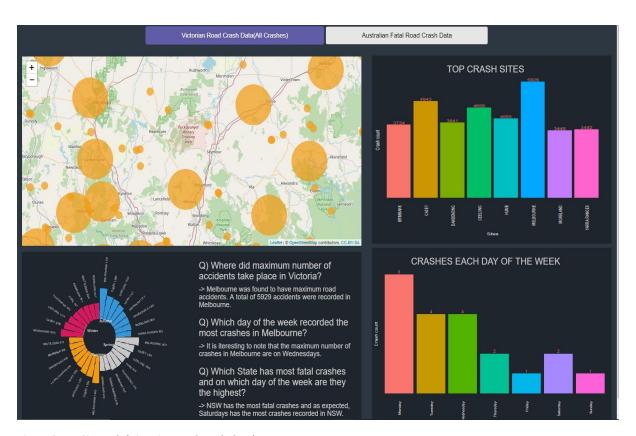


Figure 2 Top Sites tab (Victorian road crash data)

Figure 2 shows the Top Sites tab with the Victorian road crash data option selected by the user. Purple background on the button indicates that it is selected.

Figure 3 shows the Top Sites tab with the Australian Fatal Road crash data option selected by the user. Purple background on the button indicates that it is selected.

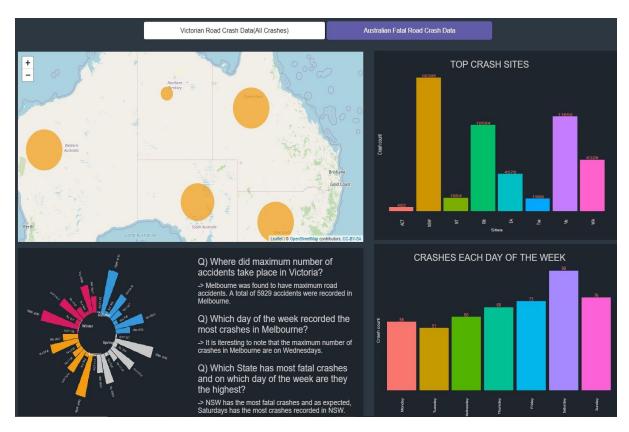


Figure 3 Top Sites Tab (Australian Fatal Road Crash data)

Figure 4 shows the change in the final barchart where the user clicks on NSW and the data on the bar chart changes.

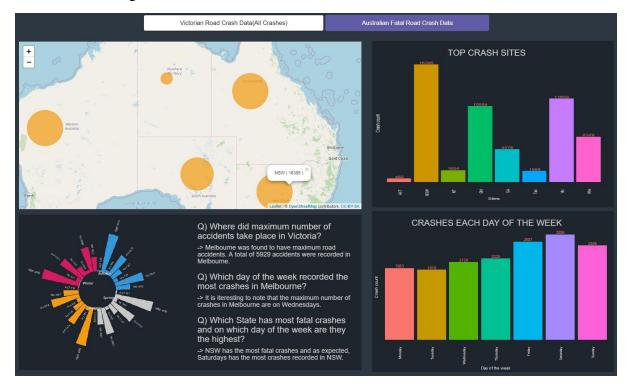


Figure 4 Change in bar chart when user clicks on the circle marker

Figure 5 and 6 displays the circular barcharts. The display the distribution of the crashes in the top 8 locations when on the Victorian crash data tab and the 8 states when on tha Australian fatal road crash data tab respectively.

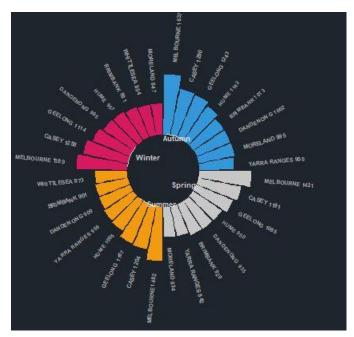


Figure 5 Crash distribution across seasons in the top 8 sites



Figure 6 Crash distribution across seasons in the 8 states

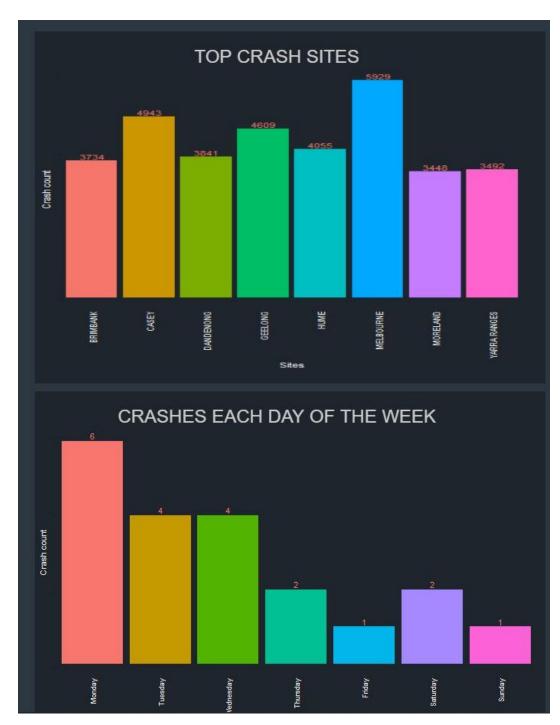


Figure 7 shows the data about the top 8 crash sites in Victoria. This is an animated chart and it does not change. The chart below that displays the spread of the crash data through each day of the week. This chart changes when the user chooses a site on

Figure 7 Top Sites and day of the week crash data of the selected site

the leaflet map. This interaction will help keep the audience's attention. The data also changes to the state fatal crash data when the user clicks on the Australian fatal crash data. Both the charts change the data accordingly. Since there were some null values noted in the Victorian crash dataset, the null values along with the ones categorized as unknown are clubbed together as an unknown category.

4.3 Seasons Tab

The seasons tab mainly contains data about the crash data that is distributed across the seasons. It focuses on the distribution of both the Victorian road crash data and the Australian road crash data across the seasons. This tab is also divided into 4 boxes. The first one is an interactive line chart that displays the top 8 sites and the 8 states data on how they move from season to season. It displays the changes in count of the crashes visibly and it can be clearly observed. The next box is an information box which answers interesting questions about the crash data that is spread across the seasons. It answers questions regarding both the Victorian and Australian fatal crash data. The next box contains a circle packer and it drills down the seasons data and displays the distribution of crashes through different speed zones. The size of the circle indicates the count of the crashes. The bigger the circle, the more the crashes in that speed zone. There are labels on the circles to show the speed zone and the crash count respectively. Overall information about the speedzones is also mentioned below the circle packer plot. The final box contains a treemap. This shows the distribution of crashes across road geometry and age group. For the Victorian crash data, it displays the distribution of road geometry and for the Australian fatal crash data, it displays the distribution of the age group of people involved in the fatal accidents. Overall information about the road geometry and age group is mentioned below the treemap plot. Key insights are mentioned here. Both these plots are directly linked to the line chart. The plots data changes when the user clicks on a point on the line chart.

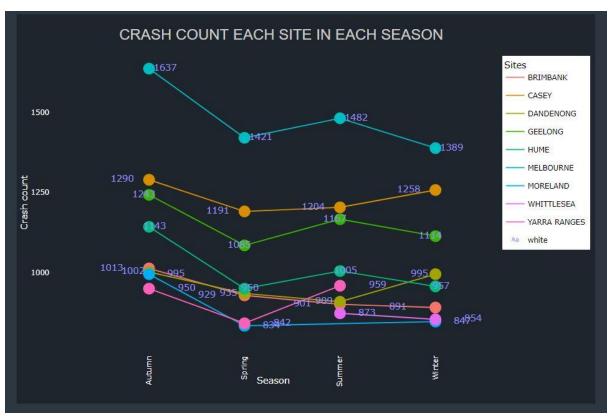


Figure 8 Crash distribution of the top 8 sites in Victoria each season

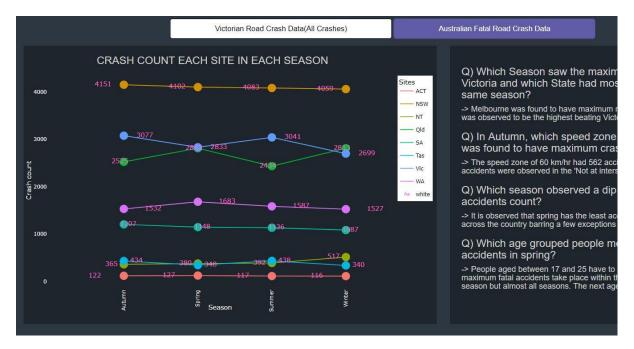


Figure 9 Fatal crash count distribution of the 8 states in each season

Figure 8 and 9 display the line chart that shows the top 8 sites and 8 states distribution of crashes each season respectively.

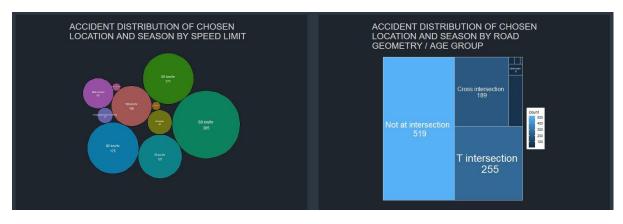


Figure 10 Circle packer and Treemap showing the distribution of speed zones and road geometry

Figures 10 and 11 display the circle packer and treemap respectively. Clearly, the data has changed when the user clicks on different points on the line chart. Also, when the user clicks on Victorian crash data, the treemap shows the road geometry distribution of data and when the user clicks on the Australian fatal road crash data, the treemap displays the distribution across the age groups of people who met with fatal accidents. The are also key insights about the overall data that is mentioned below the two plots respectively.

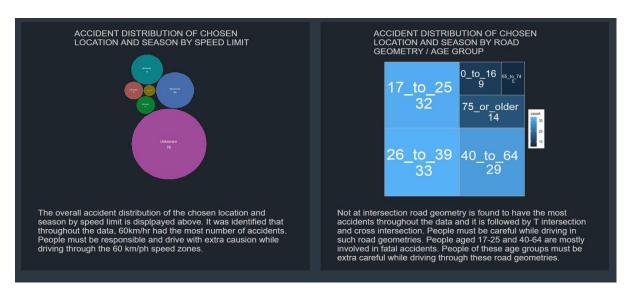


Figure 11 Circle packer and Treemap showing the distribution of speed zones and age group



Since the dataset is big and it takes time for the initial calculations, a simple loading bar is added to each of the plots so as to keep the user engaged and not lose focus. The loading bar animation provides an attractive

Figure 12 Loading bar animation

look and at the same time makes sure the users don't lose focus. Figure 12 displays the loading bar animation.

4.4 Conclusions Tab

This is the final tab of the application and it acts as a conclusion to the application. This tab contains some of the important metrics and displays them to the audience in an engaging manner. This tab has 3 main segments to it. The first segment is the one where the metrics are displayed. These are the 8 most important metrics and these metric gives an overall idea

about the crash data and helps provide an overall idea of the crash data. This is very useful to the general public and the emergency medical service providers.

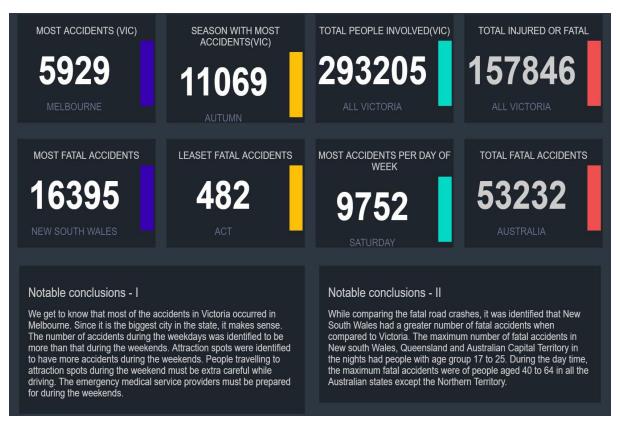


Figure 13 Conclusions tab

The next segment is the notable conclusions segment. This is divided into two and the most notable insights, conclusions are mentioned here. The final segment is the data sources segment where all the data source links are provided. The users can go to the respective data sources to check the data. To keep the users engaged in this tab, all the metrics are animated and a counter is run so that it doesn't blandly display the numbers alone. This will attract the audience. Figure 13 shows the conclusions tab after the animated number counter stops. The final metrics are displayed in Figure 13 along with the notable conclusions segment.

5. CONCLUSION

Overall, to summarize, this application was mainly intended to deliver a trust worthy dashboard of the crash dataset to the target audience – general public and emergency medical service providers. Patterns and trends in the data were observed and were clearly mentioned in the information boxes throughout the application. This application will help keep the emergency medical service providers as they can clearly check when they can expect more accidents to take place and which locations where the accidents will take place and be prepared. A model can also be built on the data to consolidate on such predictions. This application also provides key insights to the general public and they can check their location, regular travel routes to make sure they drive safely and stay alert in the required areas.

Overall, users can use this application and make informed decisions before they take a drive based on these trends and aggregated statistics provided in this application.

There has been a steep learning curve on many elements of Shiny Dashboard, CSS, and other R libraries as a result of building this application. Although there are already interactive plots available to users, several portions, such as the circle packer and treemap, are static due to issues with numerous plots utilising plotly on the same tab.

Due to time constraint, a lot of different features of the dataset could not be explored and converted into trends, plots and charts to display in this application. There are a lot of other features that can be explored and trends can be observed and displayed to the general public. As mentioned earlier, due to the time constraint, it was not feasible to develop them all in this application. Additionally, several graphs might have been created with D3 and then incorporated into R Shiny to provide more interactive and interesting plots.

6. BIBLIOGRAPHY

Bibliography

ABC News. (2022, January 01). Retrieved from ABC News: https://www.abc.net.au/news/2022-01-01/increased-death-toll-on-victorian-roads/100734312

Australian fatal crash metadata. (n.d.). Retrieved from Australian fatal crash metadata.

Bureau of Infrastructure and Transport Research Economics. (n.d.). Retrieved from Bureau of Infrastructure and Transport Research Economics:

https://www.bitre.gov.au/statistics/safety/fatal_road_crash_database

Bureau of Meteorology. (n.d.). Retrieved from Bureau of Meteorology: http://www.bom.gov.au/climate/glossary/seasons.shtml

Five design sheet. (n.d.). Retrieved from https://towardsdatascience.com/five-design-sheet-methodology-approach-to-datavisualisation-603d760f2418

Kaggle. (n.d.). Retrieved from Kaggle: https://www.kaggle.com/stevenrferrer/victoria-road-crash-statistics

VIC Crash Metadata. (n.d.). Retrieved from VIC Crash Metadata: https://data.vicroads.vic.gov.au/metadata/Crashes_Last_Five_Years%20-%20Open%20Data.html

VIC Data. (n.d.). Retrieved from VIC Data: https://discover.data.vic.gov.au/dataset/road-crashes-for-five-years-victoria

7. APPENDIX

To better understand the thought process and design phases made prior to execution, the five design sheets utilised as part of the shiny application are shown below in this section.

