Executive Summary

The below report analyses car crash data in NSW inorder to inform Transport for NSW's strategies for reducing car crashes and their severity. The report focused on three research questions examining how external factors e.g. speed limits and weather affected car crash severity, timing and seasonal patterns of car crashes as well as exploring the extent of geographical influences on the frequency of car crashes. Some key insights included how higher speed limits result in high fatality rates, accident injuries and fatalities have dropped over the five year period, surprisingly daytime accidents are more fatal than night time accidents and the region of South Western Sydney has a higher accident rate compared to the rest of NSW. Finally after using data visualisation and predictive modelling to backup these insights we recommended focusing on enhancing daytime driving education, managing speed limits, and implementing safety measures in high-incident areas like Bankstown and Liverpool.

Introduction

The purpose of this report is to analyse car crash statistics in New South Wales (NSW) between 2017-21 and provide recommendations to Transport NSW to reduce car crashes and its severity.

The current state of NSW roads is that a person is killed or hospitalised due to a car crash every 46 minutes, while car crashes cause the economy $9 billion annually*(Transport for NSW 2026 Road Safety Action Plan pg 10).*

As a result, the NSW government has implemented their 2026 Road Safety Action Plan which has goals to halve deaths and reduce serious injuries from car crashes by 30% of

2018-22 levels by 2030. They also have an optimistic goal of achieving 0 fatalities and serious injuries on roads by 2050 *(Transport for NSW 2026 Road Safety Action Plan)*

Our team developed three pivotal research questions:

1. ***How do external factors such as weather conditions and speed limits influence the severity of car crashes in NSW?***

2. ***How can we identify timing and seasonal patterns that can be observed in car crashes between 2017-2021 in NSW***

3. ***To what extent does geographical location influence the frequency of car crashes?***

To investigate these research questions we leveraged sophisticated data visualisation tools such as Pandas, Matplotlib, and Seaborn to portray trends and patterns.

Q1: **How do external factors such as weather conditions and speed limits influence the severity of car crashes in NSW?**

It is anticipated that external factors such as weather conditions and speed limits affect the severity of car crashes in NSW and thus several different models were created to explore the significance of these factors.

The impact of weather conditions was assessed by distributing the number of fatalities in car crashes during different weather conditions.

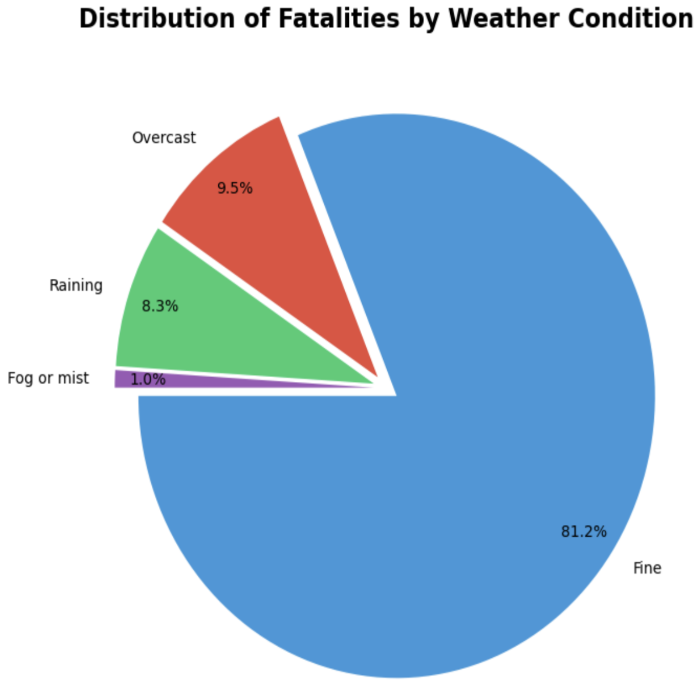


Figure 1

The pie chart in Figure 1 was created to understand how weather conditions impact the severity of car crashes. Traditionally, poor weather conditions are expected to be more dangerous for driving, however the data indicates that the majority of crashes occur in fine weather conditions. This is evident as 81.2% of crashes are in fine weather compared to 8.3% and 1% of crashes occurring during raining and foggy weather respectively.

Therefore, this may be attributed to higher volumes of cars driving during fine weather conditions or complacent driving which increases the likelihood of crashes occurring.

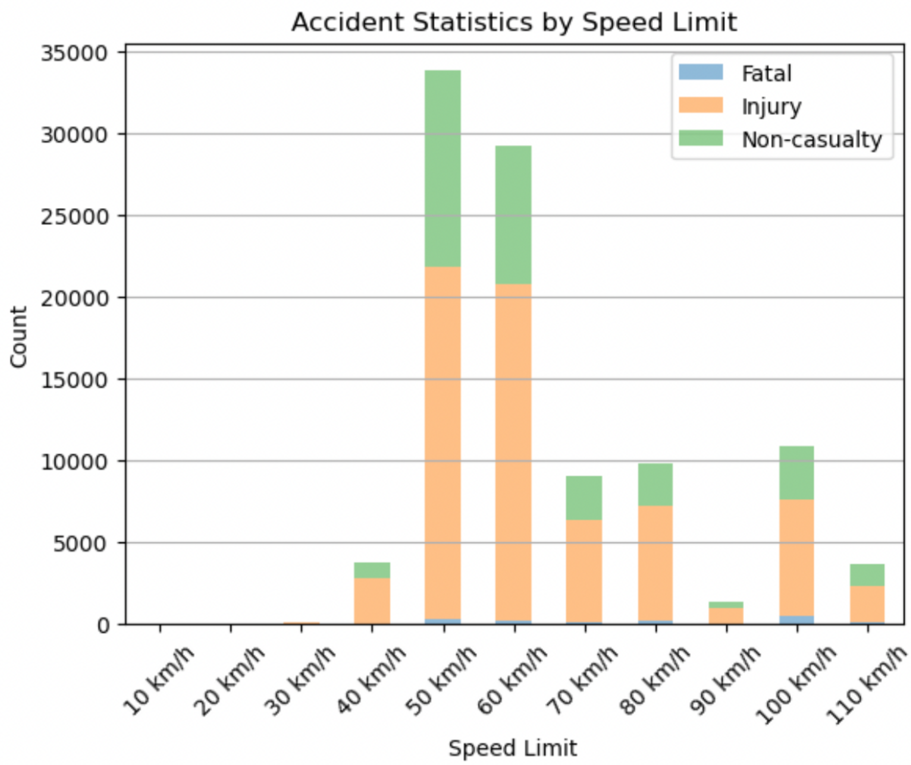


Figure 2

Speeding has been determined as the highest cause of car crash fatalities between 2008-10, accounting for 42% of all fatalities (NSW ROAD SAFETY STRATEGY 2012). Therefore, the varying degrees of car crash severities were analysed across different speed limits in the stacked bar graph in Figure 2.

Figure 2 identifies that the majority of car crashes, 63,000 occur between the 50km/h to 60km/h speed limit range, suggesting that most car crashes occur in metropolitan areas. However, a large proportion of car crashes in the 50km/h to 60km/h speed limit range are non-fatal which reiterated the importance of driving at slower speed limits.

Conversely, the highest fatalities occur at the 100km/h speed limit zones which highlights the substantial impact of high speeds on fatality rates in car crashes.

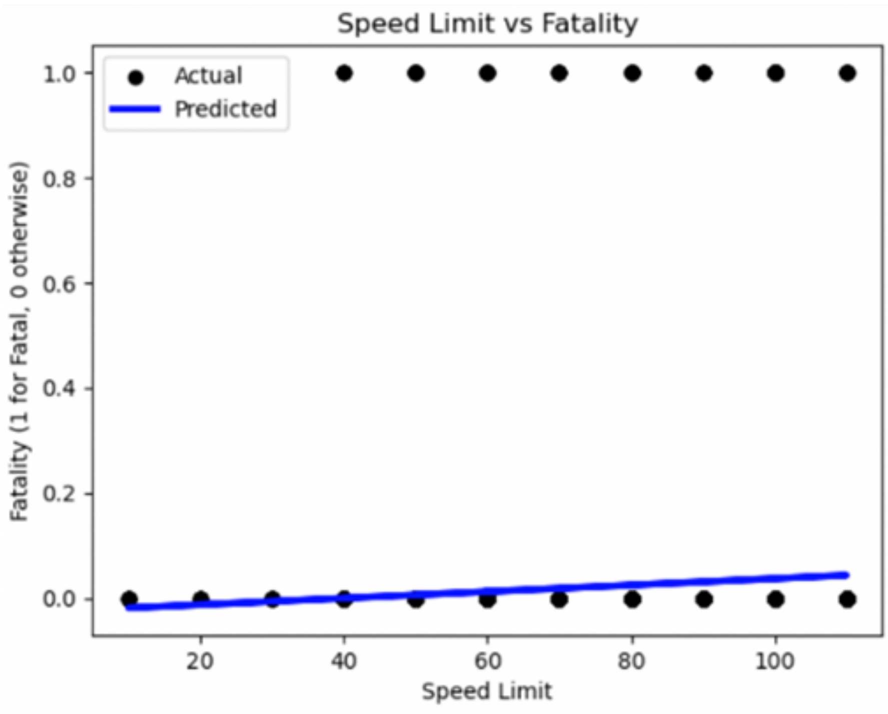


Figure 3

We developed a linear regression to further understand the relationship between speed limits and car crash severity to determine the extent to which these factors correlate with each other.

Figure 3 analyses all crashes across speed limits ranging from 10 km/h to 110 km/h.

Upon completion of this linear regression model, our team determined that it is an unsuitable method of discerning the relationship between speed limit and car crash severity, as it presents a neutral correlation between both variables. While the predicted line of best fit does indicate a trend of higher speed limits resulting in a more likely probability of fatalities from car crashes, developing a supplementary graph was necessary to provide further insight into Research Question 1.

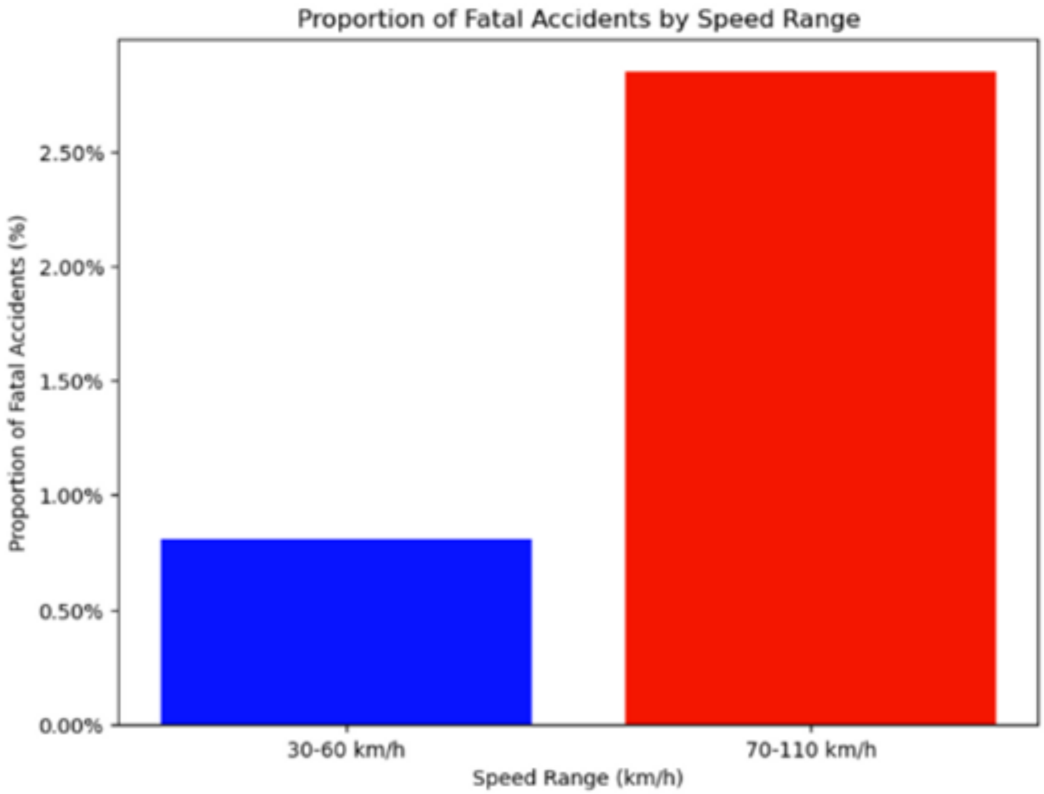


**Figure 4*

Figure 4 changes the original variable of Speed Limit into the Speed Range variable instead, creating a clearer correlation between speed ranges and fatalities. The Car Crash Severity variable was altered to be the Proportion of Fatal Accidents (%) variable instead, as measuring fatalities by a percentage removes the bias of significantly greater car volume on higher speed limit roads.

This model portrays a more evident disparity between low and high speed limit ranges, as the 70-110 km/h speed range experiences an approximately 2% higher fatality rate than the 30-60 km/h speed range. Although we expected this outcome, the particular percentage difference in fatality proportion can allow us to make actionable recommendations to reduce this gap.

Q2: **How can we identify timing and seasonal patterns that can be observed in car crashes between 2017-2021 in NSW?**

To examine underlying trends for both injuries and fatalities in 2017-2021, we determined that a combined line and bar plot would be the most suitable format to display both variables in relation to each other.

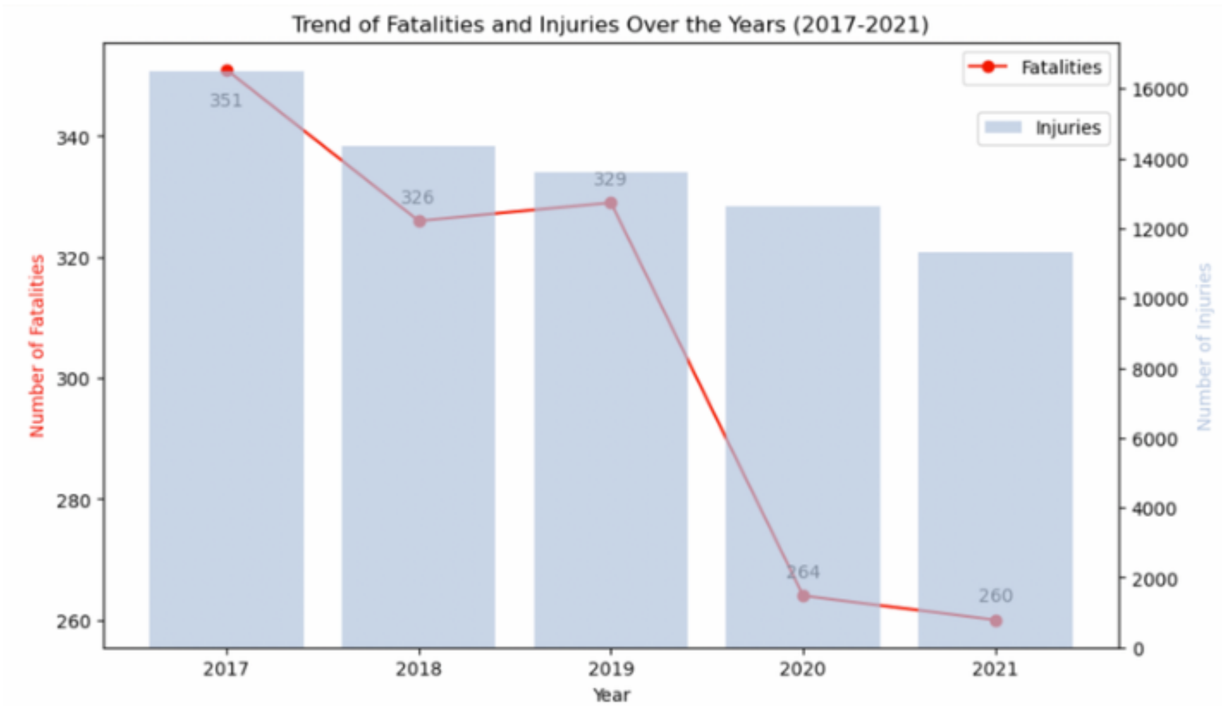


Figure 5

Figure 5 outlines a year-by-year variance with both injuries and fatalities allowing us to identify potentially trends and causes over time.

The primary patterns that can be observed within this graph is a decrease in injuries across all four years and a decrease in fatalities in three of those four years, with a minimal increase of 3 fatalities from 2018-19 being the sole exception.

One particularly noticeable trend is the decrease in fatalities from 2019-2020, with 65, or 19.8%, less fatalities being recorded in this time period. We are hypothesising that this is due to decreased road traffic volume as a result of initial COVID-19 lockdown restrictions, with road traffic falling by approximately 55% in April-May 2020 when compared to directly pre-COVID-19 statistics (Taylor 2021). Additionally, increased NSW Government investment into the NSW Safer Roads Program (Transport for NSW Centre for Road Safety 2020) between 2019-2020 has had an impact on reduced fatalities due to higher quality road safety infrastructure being constructed in that time period.

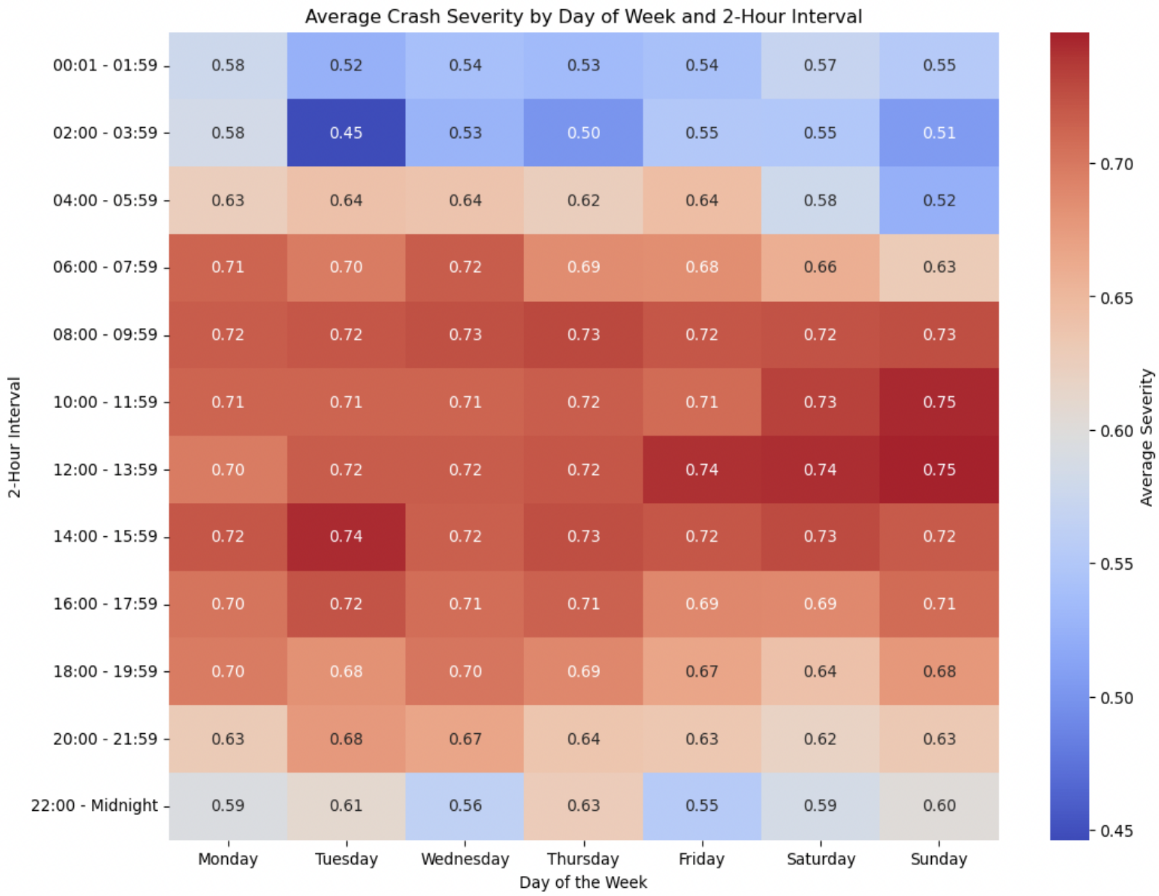


Figure 6

This heatmap works as a visual tool that shows average severity of crashes by day of the week and time interval. The way this was made was that severe crashes (fatal or injury) were encoded to 1 and non-casualty towaways were 0, these were then averaged for time period and day producing the above.

The observation that the most likely time for fatal crashes is between 8 am and 4 pm, particularly on weekends, defies conventional expectations. Prior to making the graph we assumed that late nights would have a higher incidence of severe crashes due to factors like reduced visibility and a higher likelihood of drivers being fatigued or under the influence. The higher daytime crash severity might be attributed to factors such as higher traffic volumes or complacency during commuting hours with drivers being less alert due to the routine nature of their travel.

Additionally, the impact of the COVID-19 pandemic may be a factor contributing to the results. With many lockdowns and restrictions in NSW in 2020 and 2021 there were changes in traffic patterns. People had fewer reasons to travel at night or activities that lead to fatigue or drinking and were more likely to be staying at home. However during the day they were more likely to engage in exercise bushwalks or picnics which were frequently allowed despite restrictions.

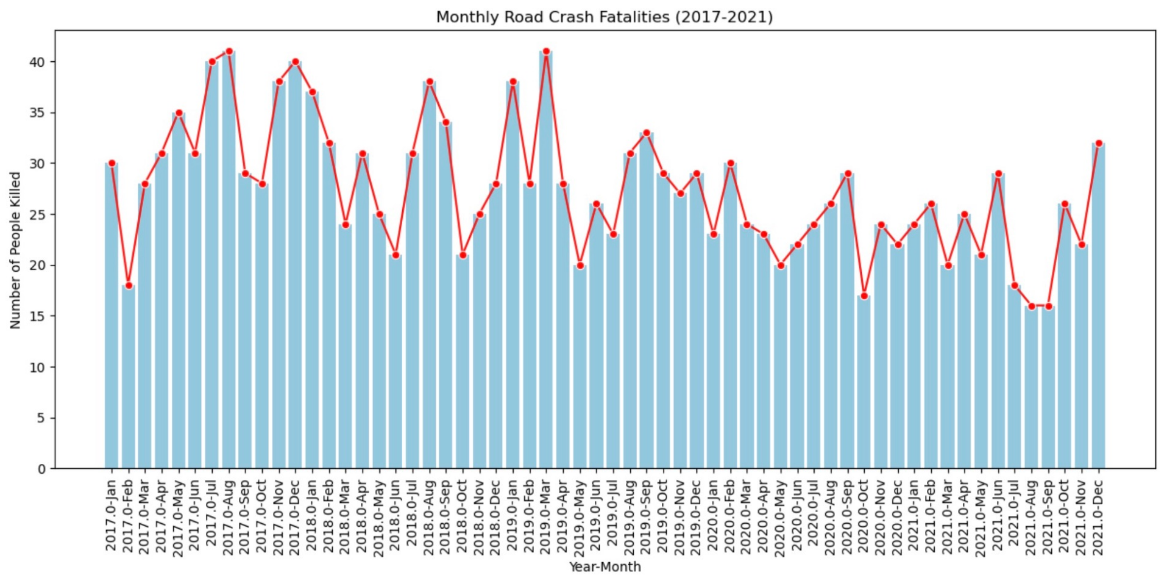


Figure 7

Figure 7 illustrates the fluctuating patterns of road accidents across different months from January 2017 to December 2021. The vertical bars denote the number of people killed in road crashes each month, while the overlaid line graph traces the trend more succinctly, accentuating the rises and falls in the data.

The data suggests that the summer months, typically associated with increased travel and tourism, often led to a higher number of fatalities. This could be due to a greater volume of vehicles on the roads, including a mix of local traffic and out-of-town visitors unfamiliar with the area.

Conversely, the winter months show some elevation in numbers which might be attributed to adverse weather conditions causing more hazardous driving conditions.

There are also spikes at the end of the year around November and December which may be related to the holiday season, where there would be an increase in drivers under the influence of alcohol.

Over the span of the five years, while there are variations in the monthly fatalities, the overall trend does not show a significant decrease or increase, indicating that the factors contributing to these incidents have remained relatively constant. This steadiness calls for a continued or renewed focus on road safety measures, driver education, and perhaps the introduction of more strict regulations to mitigate the risk factors.

Q3. **To what extent does geographical location influence the frequency of car crashes?**

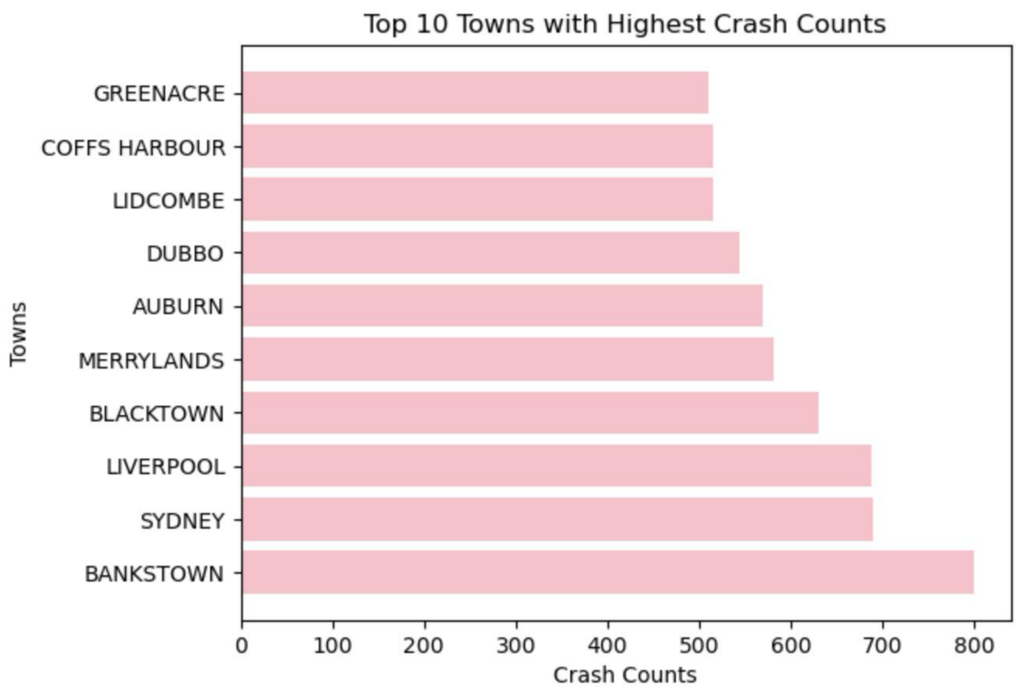


Figure 8

The horizontal bar chart shows the geographic distribution of road traffic incidents. Bankstown leads with the highest number of recorded crashes, however despite Bankstown having the most crashes, most of these incidents are non-fatal. This could indicate better vehicle safety standards, safer road designs allowing for less severe accidents.

Out of top 10 towns with the highest number of crashes, 7 of them are in Western Suburb. The western suburbs are known for their rapid expansion and development with high vehicular traffic which consequently raises the probability of road incidents.

This evidence underscores the need for targeted policy interventions and examination of urban planning and road safety strategies in the western suburbs, and Bankstown particularly. These could include enhancing the road network to accommodate growing traffic or implementing smart traffic control systems to manage flow during peak times,

To further inspect the validity of Research Question 3, we graphed the top 10 roads sorted by crash count. Figure 9 provides the opportunity to examine trends relating to the types of roads most frequently crashed on.

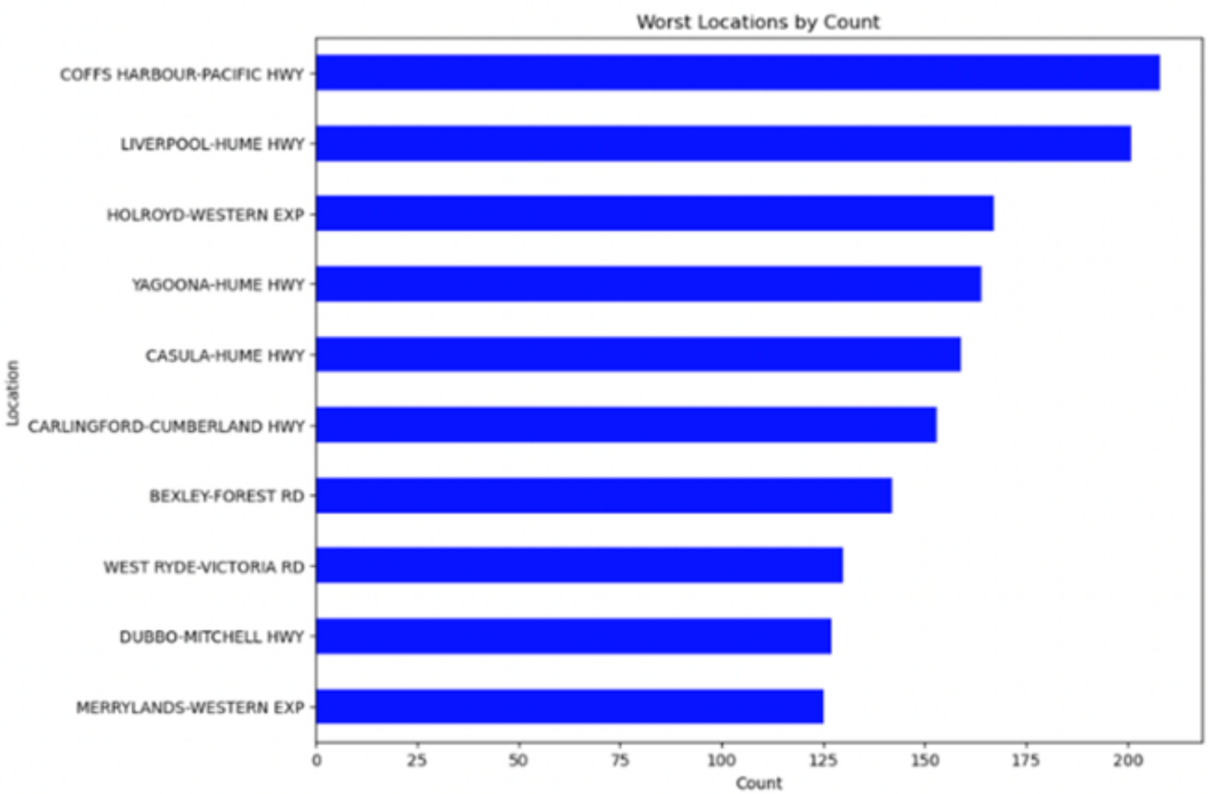


Figure 9

The most noticeable trend is the difference in crashes between the top two locations – "Coffs Harbour-Pacific Hwy" and "Liverpool-Hume Hwy" – and the rest. Although these two locations can be grouped as urban centres with major highways passing through them, there is no clear trend explaining the significantly higher crash count than in other locations in the same group.

Highways and expressways having the highest crash count is expected as these road types consistently contain the most traffic and have the highest speed limits.

However, the presence of two suburban roads, “Bexley-Forest Rd” and “West Ryde-Victoria Rd" is surprising as these types of roads typically experience lower traffic volume. Further research indicates that both of these roads are common alternative routes to avoid Toll Roads in Sydney (Roads Australia n.d), meaning they may receive more traffic than they have the capacity to handle whilst maintaining free-flowing car movement.

Regarding similarities between the above Figure 9 compared to the previous Figure 8, both graphs contain eight out of 10 locations in Sydney, with a further 5-7 locations in the geographical Western Sydney region.

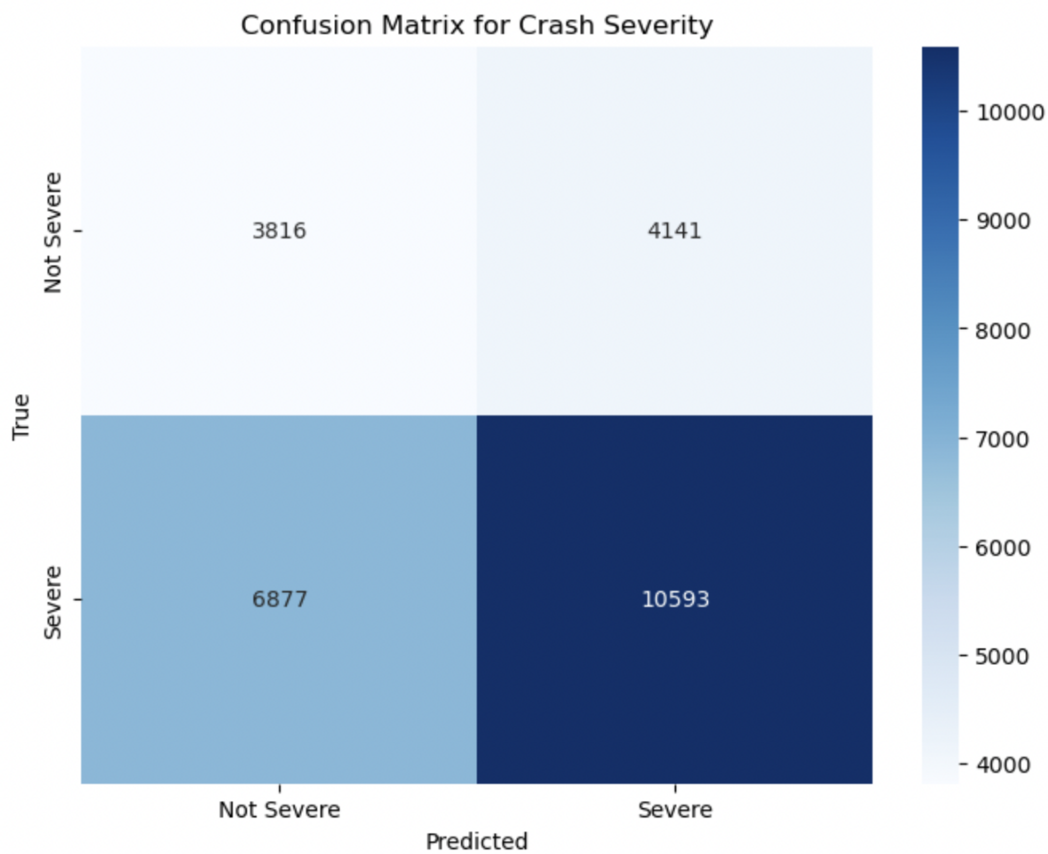


Figure 10

```
Accuracy: 0.5666810870334683
Classification Report:
              precision    recall  f1-score   support

           0       0.36      0.48      0.41      7957
           1       0.72      0.61      0.66     17470

    accuracy                           0.57     25427
   macro avg       0.54      0.54      0.53     25427
weighted avg       0.61      0.57      0.58     25427
```

The logistic regression analysis we conducted aimed to understand how variables like weather, road alignment and road classification impact the severity of crashes. This analysis directly addresses the research questions we had

Regarding the models performance we found insights below;

Accuracy (56.7%): The model's accuracy stands at around 56.7% suggesting it is moderately effective. This level of accuracy indicates that the model can make predictions.

Precision and Recall; The model demonstrates precision (72%) in predicting crashes but lower precision (36%) for non severe ones. Additionally it has a recall rate of 61% for crashes indicating its ability to identify such incidents.

We decided to balance class weights due to an imbalance in the dataset. This adjustment significantly improved the models ability to identify crashes which were underrepresented.

However it is important to note some limitations of the model:

Predictive Power for Non Severe Crashes: The model struggles with classifying severe crashes as indicated by its reduced precision and recall in this category. This could be due to similarities between non severe crashes or a lack of features in the dataset, for effective differentiation.

Feature Limitations:
While the variables we have chosen provide a foundation, for analysis it's important to acknowledge that they might not encompass all the factors that influence crash severity.

To enhance the accuracy and depth of our model we should consider incorporating variables or exploring alternative modelling techniques.

Summing up the logistic regression model offers insights into the factors impacting crash severity and demonstrates a particular proficiency in predicting severe crashes. However it's crucial to recognise its limitations as they present opportunities for improvement and exploration of advanced analytical approaches.

Limitations/Assumptions for Linear Regressions

Categorical values: In Figure 3 and 13 each categorical value was sorted into numerical variables to assess their correlation with other variables. This was performed through the use of a dummy variable, where fatal crashes are represented by 1 and all other crashes are represented by 0. However, this resulted in linear regressions that did not provide accurate insights.

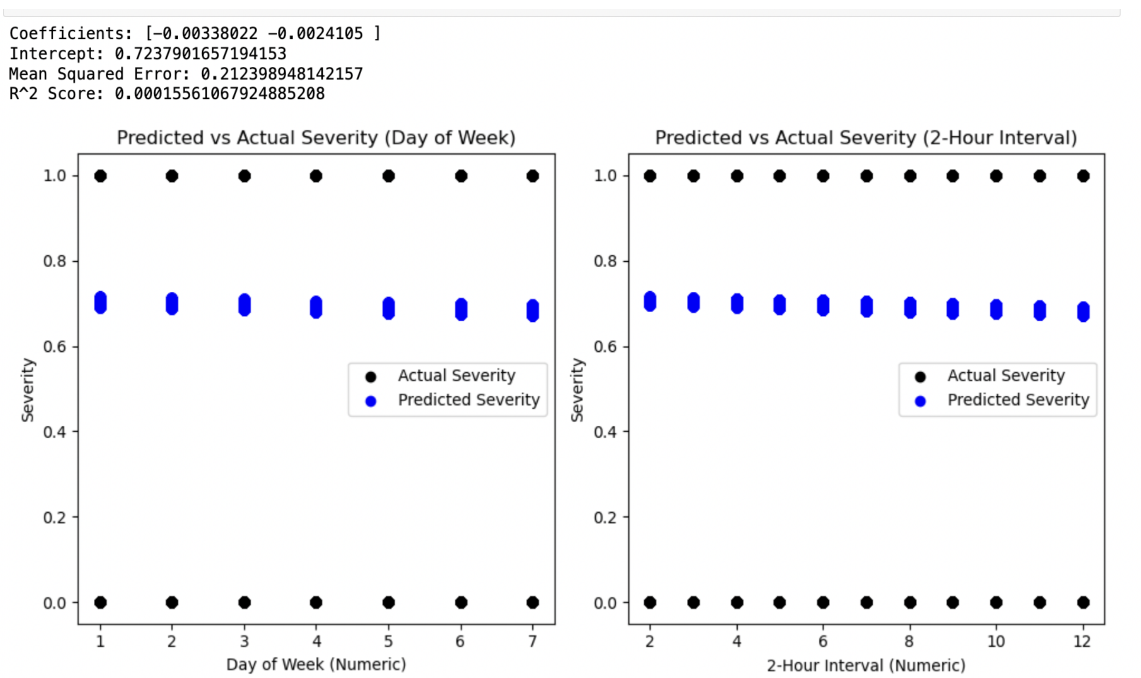


Figure 13 - Linear Regressions for Day of the Week and 2 Hour Interval vs Severe Crashe

<u>Decision Tree Classification</u>

Accuracy: 0.3244518729721758

Classification Report:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	471
1	0.00	0.00	0.00	5686
2	0.29	0.40	0.34	8536
3	0.35	0.61	0.45	9505
4	0.32	0.10	0.16	6315
accuracy			0.32	30513
macro avg	0.19	0.22	0.19	30513
weighted avg	0.26	0.32	0.27	30513

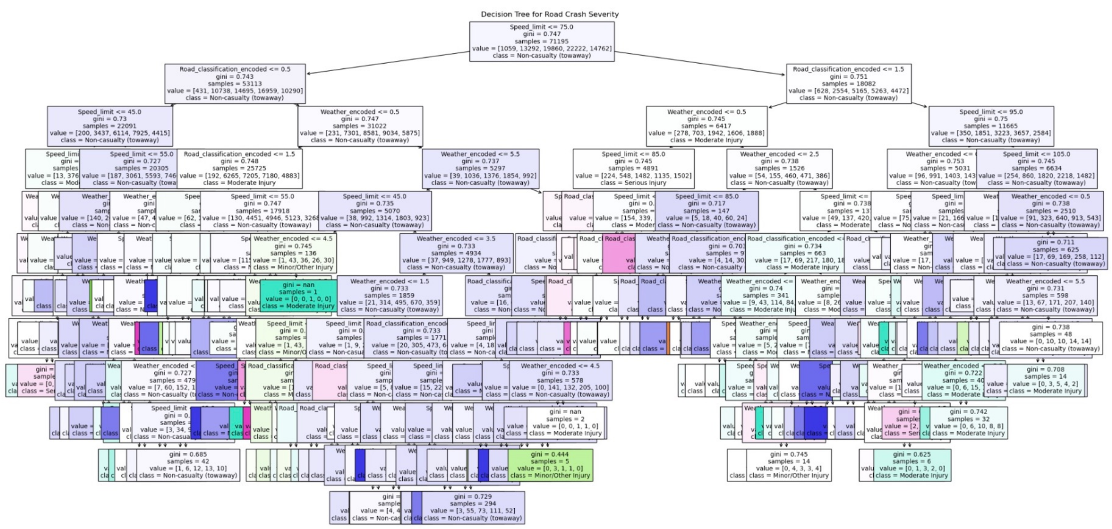


***Figure 11**

The decision tree for road crash (Figure 11) severity in NSW reveals crucial insights into the factors contributing to accidents. The Gini impurity scores gauge the homogeneity of the data following a split, indicating a strong correlation between certain features and crash occurrences. Specifically, the Gini score of 0.747 at nodes splitting on speed limit and weather conditions suggests that these variables are significant predictors of crash outcomes. This high Gini value implies a substantial level of disorder, meaning that variations in speed and weather are associated with different levels of crash severity. Simultaneously, road classification emerges as another influential factor with a Gini score of 0.743. While slightly lower, this figure still underscores a significant relationship with crash severity, pointing to the influence of road type in the likelihood and impact of traffic incidents. Together, these metrics underscore the multifaceted nature of road safety and the importance of speed management and environmental considerations in mitigating crash severity in NSW.

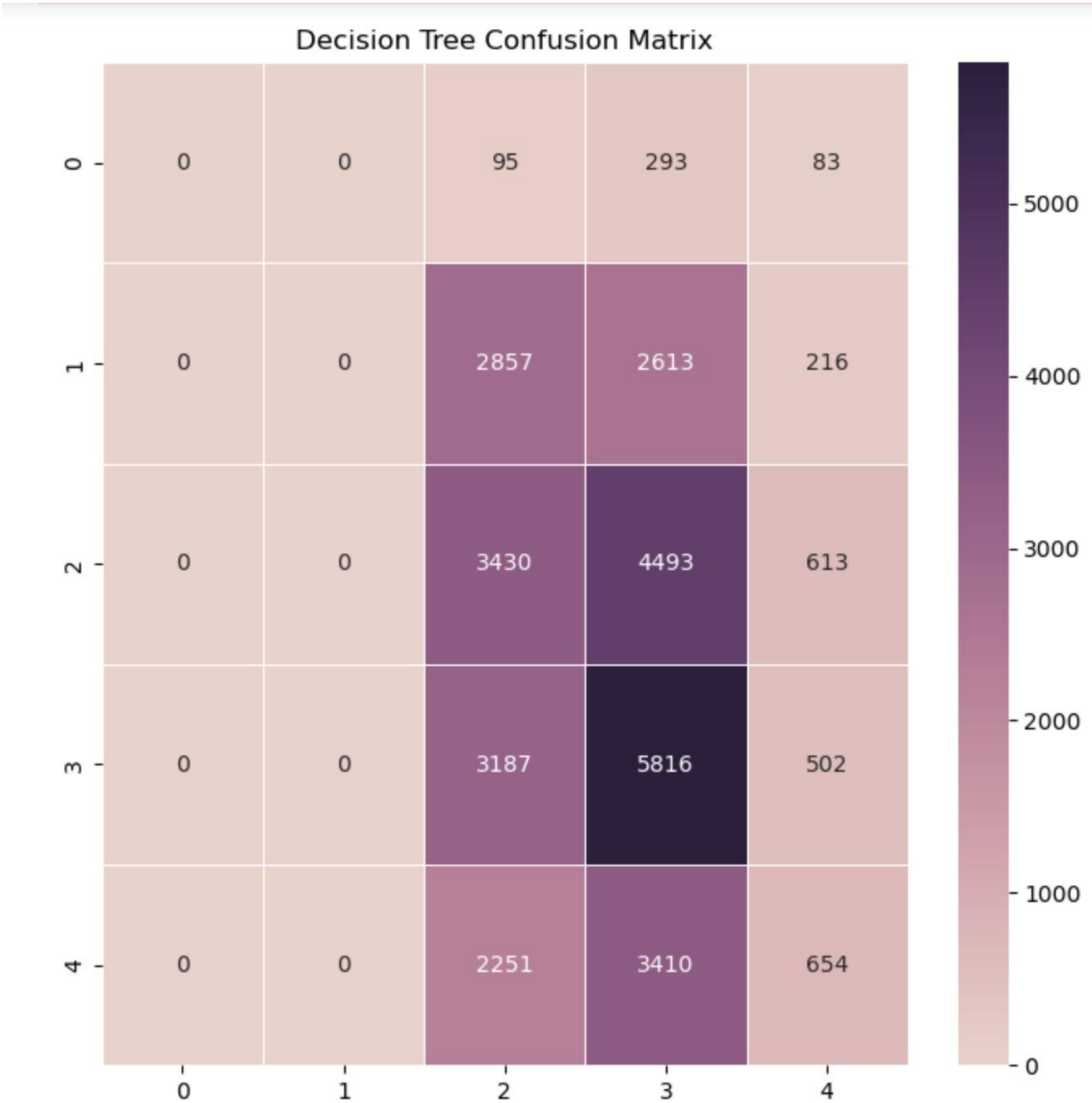


***Figure 12**

Likewise, Figure 12 (Confusion Matrix for Decision Tree) is used to further validate this assumption. Figure 12 shows that all factors are at a significant rate, ranging from 2251 to 5816 which is highly significant. Therefore, we can take into account speed limits, weather conditions, and road classification in order to design a good measure to prevent crashes that are likely to happen in the future.

```
Accuracy of the Decision Tree classifier: 0.6729590666273392
```

Recommendations and Conclusion

To optimise their campaign budget we recommend that Transport For NSW should focus on the below strategies:

- **Daytime Driving Education:** Contrary to common beliefs, the data indicates a higher fatality rate during daytime. Campaigns should therefore educate the public about daytime driving risks, emphasising vigilance and the dangers of high traffic volumes.
- **Speed Management:** With their being a strong correlation between higher speeds and greater crash fatalities, we recommend the campaign focuses on public awareness campaigns of the risks associated with high speed driving
- **Safety Measures in High-Incidence Areas:** In regions where accidents frequently occur such as Bankstown and Liverpool or the Pacific Highway near Coffs Harbour the adoption of advanced traffic safety strategies is paramount. This includes increasing police presence, deploying speed and mobile phone detection cameras, and enhancing 'drive safely' signage.

By focusing on the above strategies, Transport For NSW can effectively use their resources to decrease road incidents and advance the goals they set out in the 2026 action plan. The analysis of external factors such as weather conditions and speed limits and their influence on crash severity shows a need for speed management and targeted safety measures. Moreover, the investigation into timing and seasonal patterns of crashes underlines the importance of daytime driving awareness campaigns. Additionally, the geographical analysis of crash occurrences emphasises the necessity of focusing on high-risk areas as pointed out earlier. Predictive models used in the study such as logistic regression and decision tree classification reinforce these strategies by providing factors that contribute to crashes. These data-driven insights and analytical approaches will enable Transport for NSW to adopt a proactive and informed strategy in reducing road accidents in order to achieve their Towards Zero Goals.

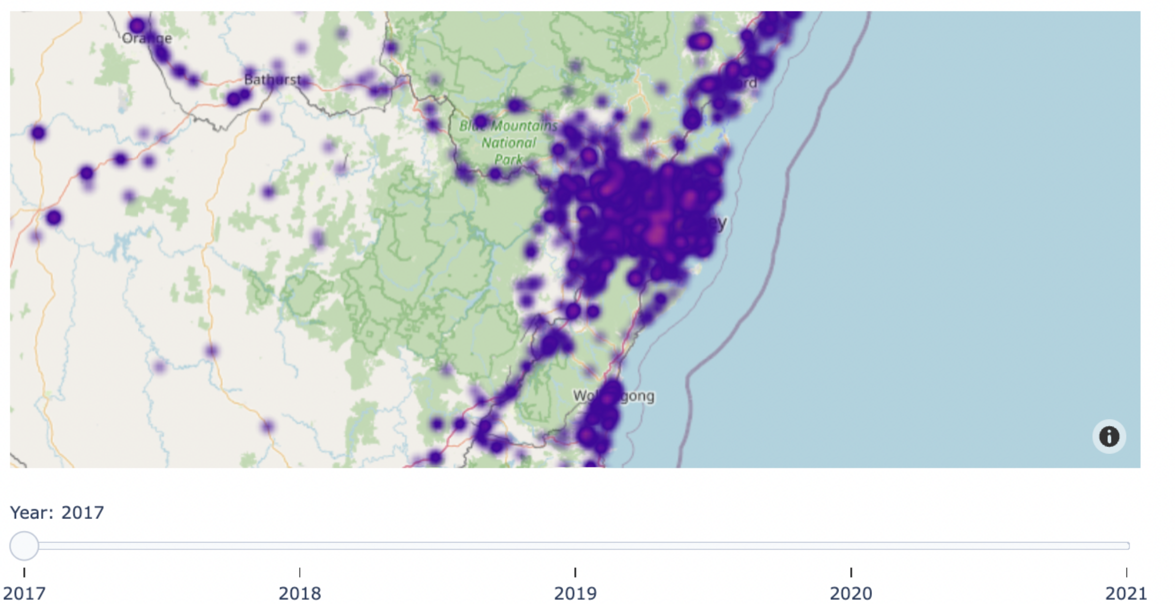
References:

2026 Road Safety Action Plan | Towards Zero 2021, Nsw.gov.au, viewed 15 November 2023, <https://towardszero.nsw.gov.au/roadsafetyplan>.
Transport for NSW 2026 Road Safety Action Plan Toward zero trauma on NSW roads 2021.
Taylor, A 2021, ‘Road traffic almost back to pre-COVID levels as commuters shun public transport’, Sydney Morning Herald, 1 February, accessed 15 November 2023, <https://www.smh.com.au/national/road-traffic-returns-to-pre-covid-levels-as-commuters-shun-public-transport-20210129-p>
Transport for NSW Centre for Road Safety, NSW Road Safety Progress Report 2020, accessed 15 November 2023, <https://www.transport.nsw.gov.au/system/files/media/documents/2023/Road%20Safety%20Progress%20Report%202020.pdf>
Roads Australia n.d, Sydney’s Guide to Toll Roads, Weebly, accessed 15 November 2023, <https://roadsaustralia.weebly.com/sydneys-guide-to-toll-roads.html>
NSW ROAD SAFETY STRATEGY 2012, accessed 15 November 2023
NRMA finds that the public backs more police as NSW road toll skyrockets 2023, Mynrma.com.au, viewed 15 November 2023, <https://www.mynrma.com.au/media/press-releases/2023/nrma-finds-that-the-public-backs-more-police-as-nsw-road-toll-skyrockets#:~:text=Wednesday%2013%20 September%202023%3A%20There,the%20same%20time%20 last%20 year.>.
6. Speed, distraction, and intoxication 2018, Aph.gov.au, viewed 15 November 2023, <https://www.aph.gov.au/Parliamentary_Business/Committees/Joint/Former_Committees/Road_Safety_2021/RoadSafety2021Joint/Report/section?id=committees%2Freportjnt%2F024922%2F79479>.

Appendix:

Road Crash Heatmap

Road Crash Density Heatmap in NSW



The above heatmap displays crash data as a magnitude of sum number of people killed and injured for Australia between 2017 and 2021 backing up the geographical graphs from earlier.

Key Information from 2026 Road Safety Action Plan

  - Goals: half deaths and reduce serious injuries by 30% of 2018-2022 levels by 2030
    - And optimistic goal of 0 fatalities and serious injuries on roads by 2050
- Internationally recognised Safe System (e.g)
  - To prevent and reduce the severity of crashes - the energy in a crash needs to be managed to save lives
    - designing of roads speeds can increase the amount of time available to a driver to react to a hazard or correct a mistake - reducing the likelihood of a crashed safer travel
    - Safer speed also means less energy absorbed by human body in a crash which reduces severity of crash
- Current government plans:
  - Accelerate safety features in vehicles

    - Using mobile detection cameras to enforce seat belt laws *(Transport for NSW 2026 Road Safety Action Plan pg 2)*
- Developing a new online NSW Road Safety Education Centre -
    - to improve access to information and resource for all road users (*(Transport for NSW 2026 Road Safety Action Plan pg 3)*
- Creating safer country roads and urban places -
    - Towards Zero Safer Roads Program by 2030 - systematically build a safer road network through safety infrastructure and speed management (pg6)
- Making safer choices on our roads:
    - Establish a Drug and Alcohol Road Safety Advisory GRoup to consider deployment of alcohol and drug enforcement and testing processes and recommend ways to achieve greater efficiency
    - Expand use if approved mobile phone detection cameras to enforce seat belt laws
- Embracing technology and innovation:
    - Developments in vehicle safety, smarter vehicles, smart traffic systems, use of applications to assist drivers
- Someone is killed or hospitalised every 46 minutes because of a crash on NSW roads (pg10)
    - Last year 270 people were killed and 10,412 seriously injured while using roads in NSW - costing the community over $9 billion each year
- Challenges for Transport NSW:
    - ⅔ of NSW road fatalities happen on country roads where the fatality rate is approx. 4 times higher than on urban roads
        - However in contrast - almost ⅔ of serious injuries happen on metropolitan roads
- Behavioural factors in NSW fatalities:
    - 41% caused by speeding
    - 24% illicit drugs present
    - 18% drink driving
    - 17% tired drivers
    - 14% seatbelt non-usage