

Student Name: Bao Ngoc Truong

Student ID: 34301054

Monash University

FIT5147 Data Exploration and Visualisation

Semester 1, 2024

## Programming Exercise 1: Tableau

### Table of Contents

<b>I. Data checking and cleaning</b>	2
1. How data looks like after loading it in Tableau	2
2. Errors and irregularities	2
Error 1: Abnormal Outliers	2
Error 2: Inconsistencies in numbering categorical variables	2
Error 3: Missing Data after July 2023	2
<b>II. Data exploration</b>	3
1. Question 1	3
a. Year	3
b. Month	3
c. Day of Week	4
d. Hour	4
2. Question 2	5
a. Speed Zone	5
b. Road Geometry	5
c. Degree of urban	6
<b>III. References</b>	7

## I. Data checking and cleaning

### 1. How data looks like after loading it in Tableau

Field Name	Physical Table	Item...
Accident No	Victoria_Accident...	ACCID...
Accident Date	Victoria_Accident...	ACCID...
Accident Time	Victoria_Accident...	ACCID...
Day Of Week	Victoria_Accident...	DAY...
Day Of Week Desc	Victoria_Accident...	DAY...
Accident Type Desc	Victoria_Accident...	ACCID...
Accident Type	Victoria_Accident...	ACCID...
Severity Rank	Victoria_Accident...	SEV...
Light Condition Desc	Victoria_Accident...	LIGHT...
Light Condition	Victoria_Accident...	LIGHT...
Lat Name	Victoria_Accident...	LAT...
Dec User Name	Victoria_Accident...	DEC...

Figure 1: Data after loading into Tableau

### 2. Errors and irregularities

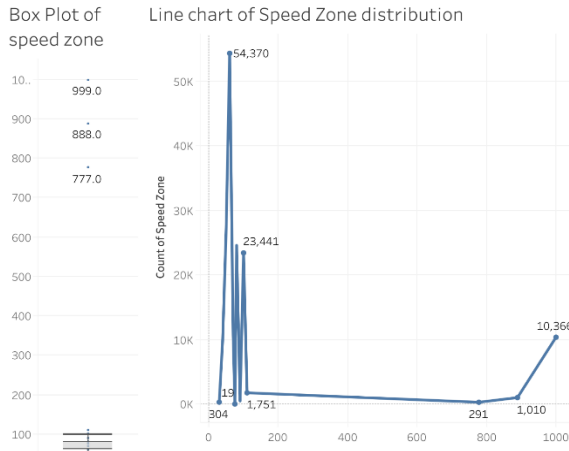
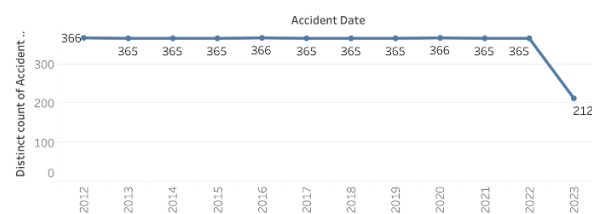


Figure 2: Speed Zone Distribution

#### Error 2: Inconsistencies in numbering categorical variables

The two graphs display variations in incident counts based on the method of recording: one by numerical representation (Day of Week) and the other by descriptive representation (Day of Week description). The discrepancy may stem from data entry errors, where individuals input numbers from 0 to 7 instead of 1 to 7 for Day of Week, resulting in miscounts from Sunday to Saturday. This error leads to inaccurate accident counts and potentially misleading data analysis outcomes. Hence, it is recommended to utilize the descriptive representation of Day of Week when analyzing the likelihood of accidents occurring on different days of the week.

Accident date dip in 2023



Last accident month recorded in 2023

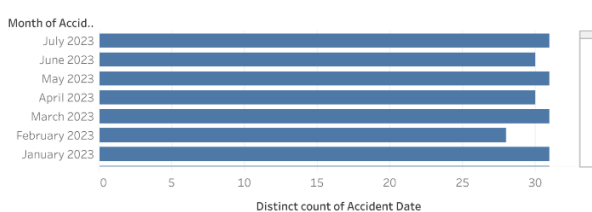
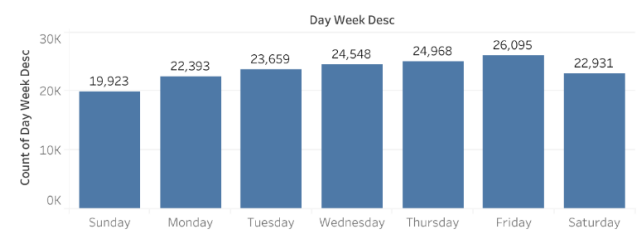


Figure 4: Missing Data in 2023

#### Error 1: Abnormal Outliers

The speed zone data exhibits a wide range, with values ranging from approximately 30 to 999, and a median speed zone of 60. Notably, there are significant outliers, including values of 999 (with 10,366 counts), 777 (with 291 counts), and 888 (with 1,010 counts). The box plot confirms the median speed zone at 60, with quartiles at 40 and 100. These substantial outliers suggest a positively skewed distribution, which could potentially impact data analysis results. Hence, it's advisable to exclude these outliers when analyzing the dataset.

Day of week description distribution



Day of week distribution

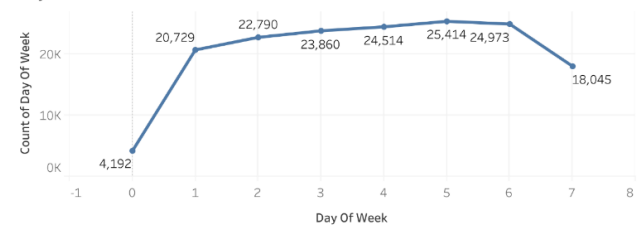


Figure 3: Day of week distribution

#### Error 3: Missing Data after July 2023

The dashboard indicates that the data is only available up to July 2023. The absence of data beyond this period may create misconceptions, such as the belief that there were fewer accidents in 2023 and a misinterpretation of decreased accidents from September to December. To ensure accurate analysis and visualization, it is advisable to filter out the data from 2023.

## II. Data exploration

### 1. Question 1

#### a. Year

Accidents by Type and Year

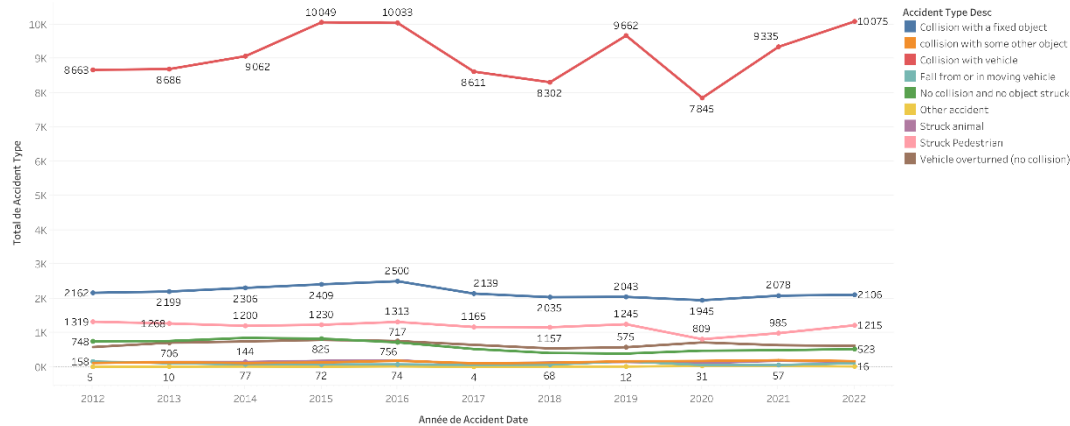


Figure 5: Accident by Type and Year

lockdown in 2020, which significantly reduced traffic volume, consequently lowering accident rates.

In April 2020, there was a sharp decrease in the number of accidents following the Australian government's lockdown policy. However, in May, as restrictions eased, there was a noticeable increase in accidents. Victoria's second COVID-19 lockdown in July led to a decrease in accidents, which continued until September when restrictions were relaxed, resulting in another surge in accidents. The second lockdown officially ended in October 2020, coinciding with a significant rise in accident numbers (Knowlton, 2023).

Accidents by Type and Month (2020)

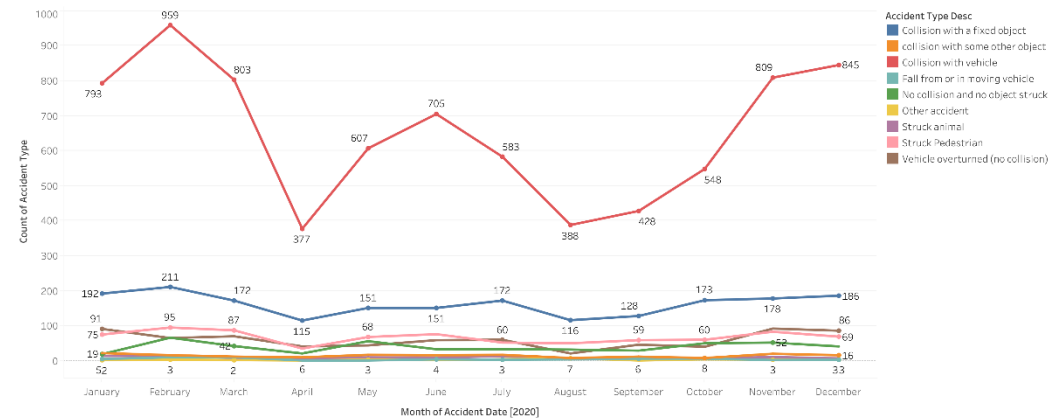


Figure 6: Accident by Type and Month in 2020

#### b. Month

Accidents by Type and Month

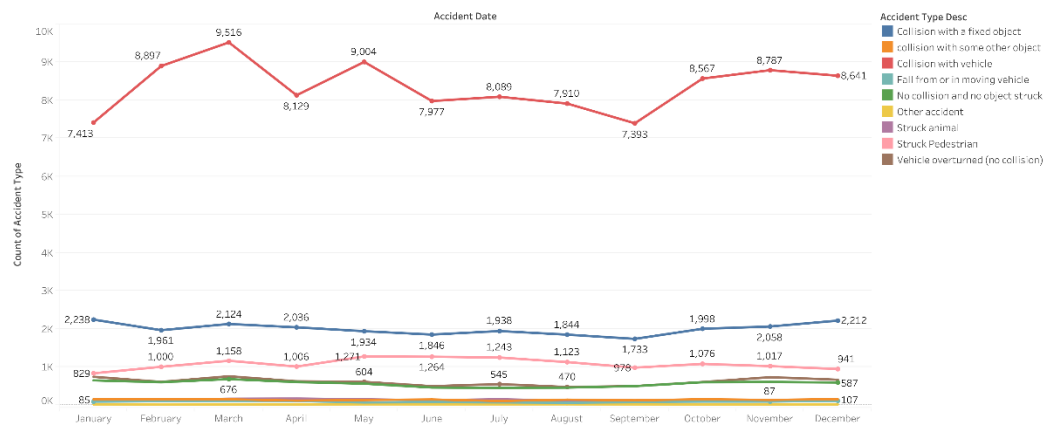


Figure 7: Accident by Type and Month

The chart of accidents by type and year (2012-2022) reveals that the most prevalent accident type is collision with a vehicle, peaking at 10,075 incidents in 2022 and reaching its lowest point in 2020 with 7,845 cases. While collisions with vehicles show fluctuations across years, other types of accidents remain relatively consistent. The variation in collisions with vehicles can be attributed to the COVID-19

Similar to the chart of accidents by type and year, the most frequent accident type recorded from 2012 to 2022 is collision with a vehicle. March has the highest number of accidents among all months, totaling 9,516 cases.

Accidents by Type and March Month

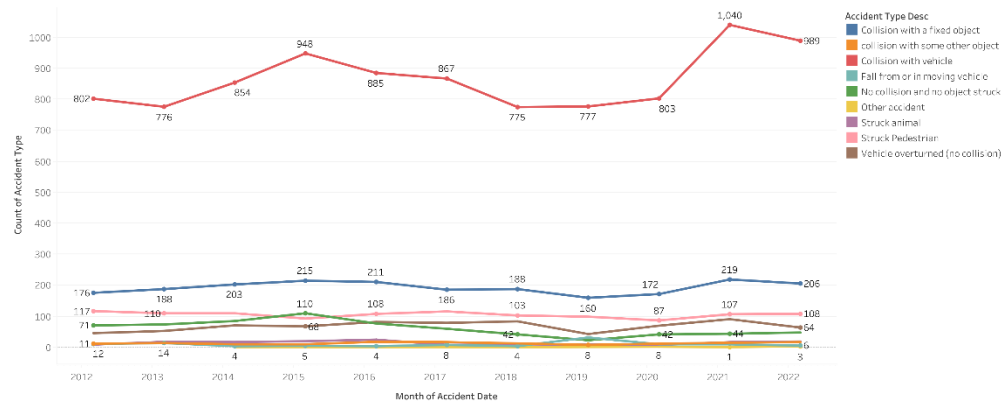


Figure 8: Accidents that happen in March by Type and Year

The graph depicts accidents occurring in March over the years, with the highest number recorded in March 2021. One potential factor contributing to this spike is the administration of the first doses of COVID-19 vaccination to people in Victoria on February 22, 2021 (Monash Health, 2022). As a result, individuals may have felt less concerned about their safety when venturing outdoors, potentially leading to a significant increase in accidents.

### C. Day of Week

Accidents by Type and Day of Week

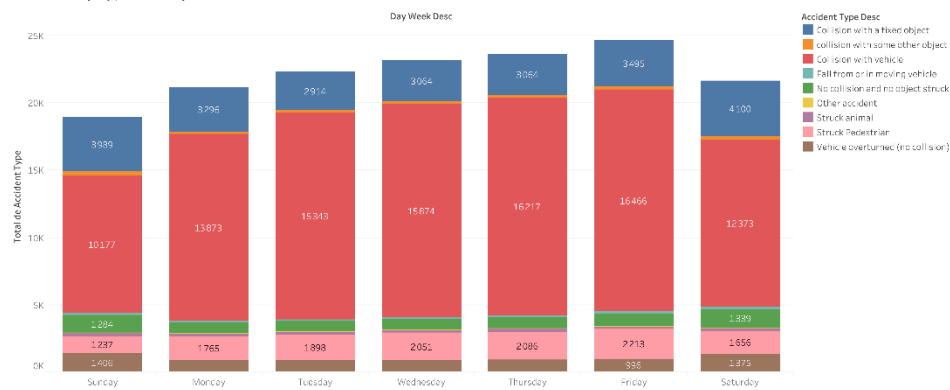
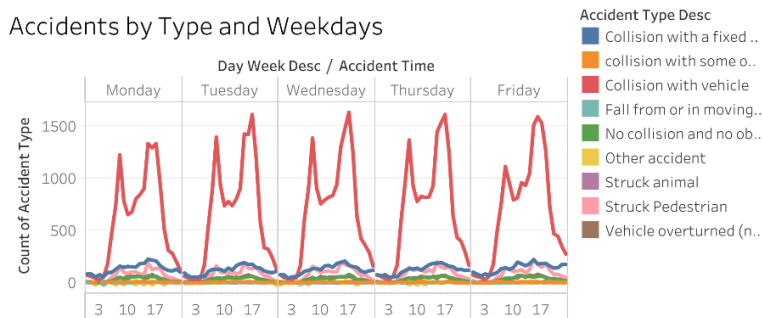


Figure 9: Accident by Type and Day of Week

In terms of the Day of the Week, Friday exhibits the highest rate of accidents. Sundays and Saturdays generally have the lowest accident rates, but their incidence of accidents involving overturned vehicles, no collision, no object struck, and collisions with fixed objects is higher compared to weekdays. Conversely, weekdays experience a higher frequency of collisions between vehicles compared to weekends. This might be due to different activities people do during weekdays and weekends.

### d. Hour

Accidents by Type and Weekdays



Accidents by Type and Weekends

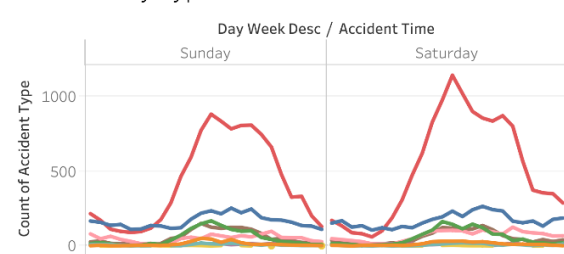


Figure 10: Accidents by Type on weekdays and weekends

From the dashboard, it's evident that accidents, particularly collisions between vehicles, spike at different times on weekdays and weekends due to varying social activities. On weekdays, accidents peak at 8 AM and 3-5 PM, corresponding to rush hours when people commute to and from work or school. Conversely, accidents occur most frequently at 12 PM and 4-5 PM on weekends, coinciding with brunch and dinner outings. Additionally, the graphs indicate higher activity levels on Saturdays compared to Sundays, with over 1000 accidents recorded around 12 PM on Saturdays compared to approximately 800 cases on Sundays.

2. Question 2

a. Speed Zone

The graph of accidents by speed zone and degree of urbanization reveals that in Melbourne's suburban areas, the most frequent accidents occur at 60 km/h, totaling 40,701 cases. Additionally, the second and third most common speed zones for accidents in urban areas are 50 km/h and 80 km/h, respectively. The 50 km/h speed zone typically denotes built-up areas or areas without speed limit signs, while the 80 km/h speed zone likely represents highways near residential areas (Vic Roads, n.d.).

Conversely, in rural areas of Victoria, the majority of accidents occur in the 100 km/h speed zone, with the second most common speed zone being 80 km/h. This trend is attributed to the proliferation of highway roads in regional Victoria.

Accidents by Speed zone and Degree of Urban

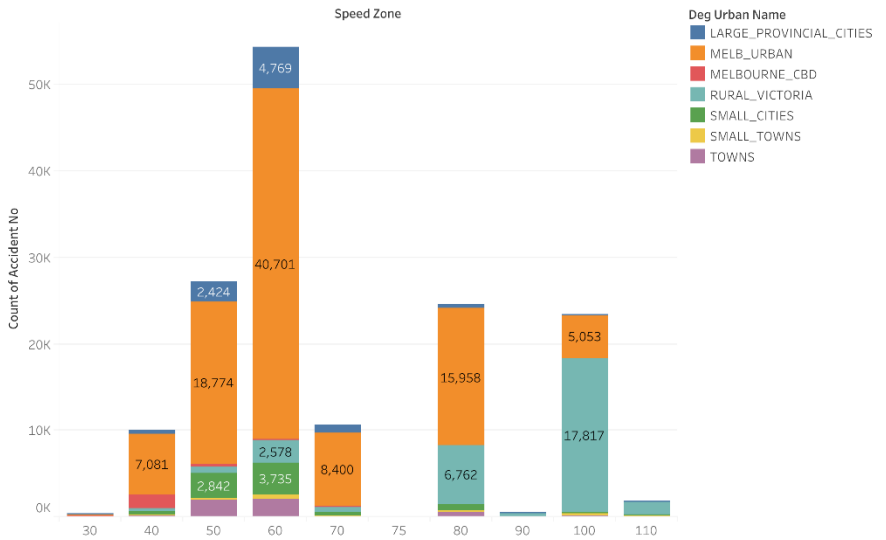


Figure 11: Accidents by Speed Zone and Degree of Urban

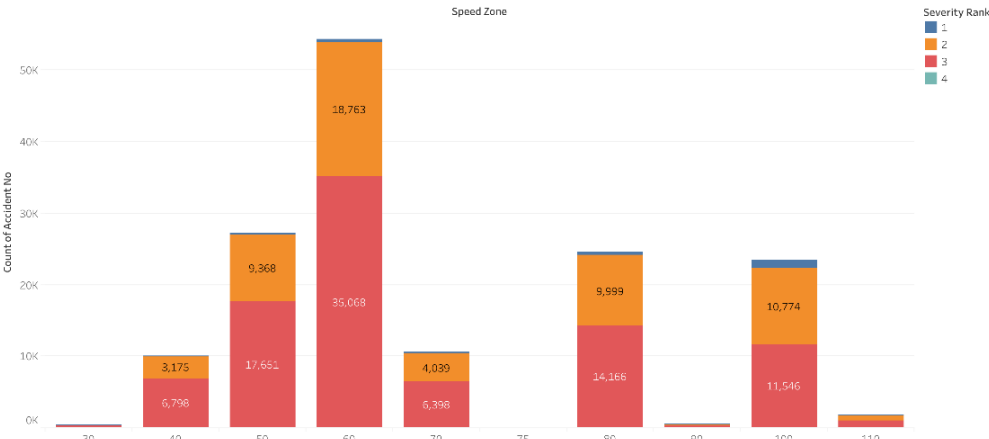


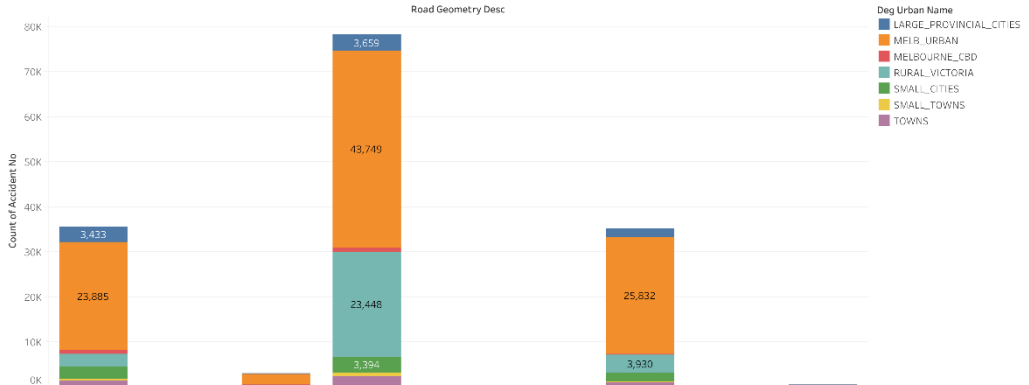
Figure 12: Accidents by Speed Zone and Severity Rank

Accidents in lower speed zones, such as 60 km/h and below, generally result in less severity compared to those in higher-speed zones. Notably, fatal accidents are more prevalent in 100 km/h speed zones, particularly in rural Victoria. This pattern suggests that high-speed areas like highways and freeways are more prone to severe accidents, whereas urban streets with lower speed limits may have more minor incidents but a higher likelihood of pedestrian involvement.

b. Road Geometry

According to the bar chart analyzing accidents by Road Geometry and degree of urbanization, the primary locations for accidents are not at intersections. Roads without intersections, such as highways and freeways designed for continuous traffic flow, contribute significantly to accident rates. Despite both urban and rural areas experiencing high accident numbers, Melbourne's urban areas record nearly double the accidents

Accidents by Road Geometry and Degree of Urban



compared to rural Victoria, largely due to higher traffic volume. Cross intersections and T intersections rank as the second most common road geometries for accidents, with Melbourne's urban areas responsible for over two-thirds of accidents in these locations. These types of intersections are more prevalent in local roads, urban streets, and residential areas.

c. Degree of urban

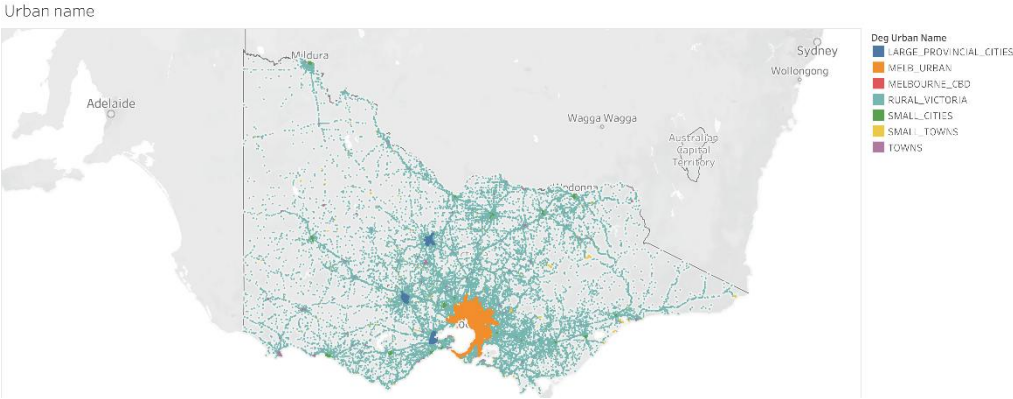


Figure 15: Map of accidents by degree of urban

The map of accidents by degree of urbanization reveals that the majority of accidents occur in and around Melbourne's urban areas and nearby rural Victoria. In rural Victoria, accidents are concentrated on highways connecting Melbourne urban areas to provincial and small cities. These findings corroborate the conclusion in the first question. Specifically, most accidents happen in urban areas

due to higher traffic density compared to rural areas of Victoria. Additionally, on weekdays, heightened traffic flow towards the Melbourne CBD and urban areas for work contributes to the increased accident rate in Melbourne's urban areas.

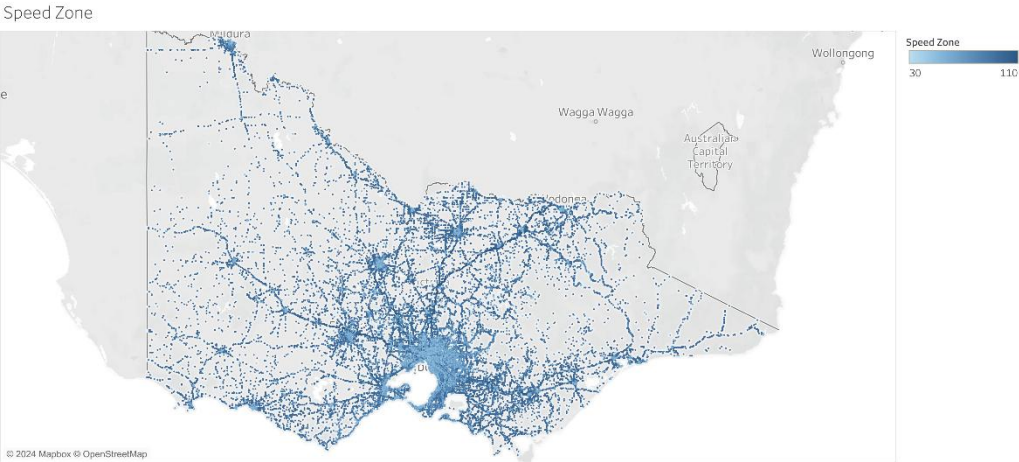


Figure 16: Map of accidents by speed zone

The speed zone map highlights a notable contrast in accident distribution between Melbourne's urban areas and rural Victoria. While accidents in Melbourne urban areas are primarily concentrated in low-speed zones due to the prevalence of workplaces, schools, and residential areas prioritizing pedestrian safety, rural Victoria sees many accidents occurring in high-speed zones. This difference can be attributed to the prevalence of highways, arterial roads, and

collector roads in rural areas, which serve as crucial transportation routes for residents, commuters, and businesses. These roads, characterized by higher speed limits, contribute to the increased likelihood of accidents in rural Victoria.

### III. References

Knowlton, C. (2023, February 20). *A timeline of Covid-19 in Australia in wake of grim milestone*. Time Out Melbourne.

<https://www.timeout.com/melbourne/things-to-do/a-timeline-of-covid-19-in-australia-two-years-on>

Monash Health (2022, February 22). Celebrating 12 months since Victoria's first COVID-19 vaccination at Monash Health. *Monash Health*.

<https://monashhealth.org/latest-news/2022/02/22/celebrating-12-months-since-victorias-first-covid-19-vaccination-at-monash-health/>

Vic Roads. (n.d.). *Speed limit*. <https://www.vicroads.vic.gov.au/safety-and-road-rules/road-rules/a-to-z-of-road-rules/speed-limits>

[limits](https://www.vicroads.vic.gov.au/safety-and-road-rules/road-rules/a-to-z-of-road-rules/speed-limits)