Vehicle Road Crash Analysis in Victoria

Contents

Data checking and cleaning

After loading the dataset into Tableau, it will look like the picture below:

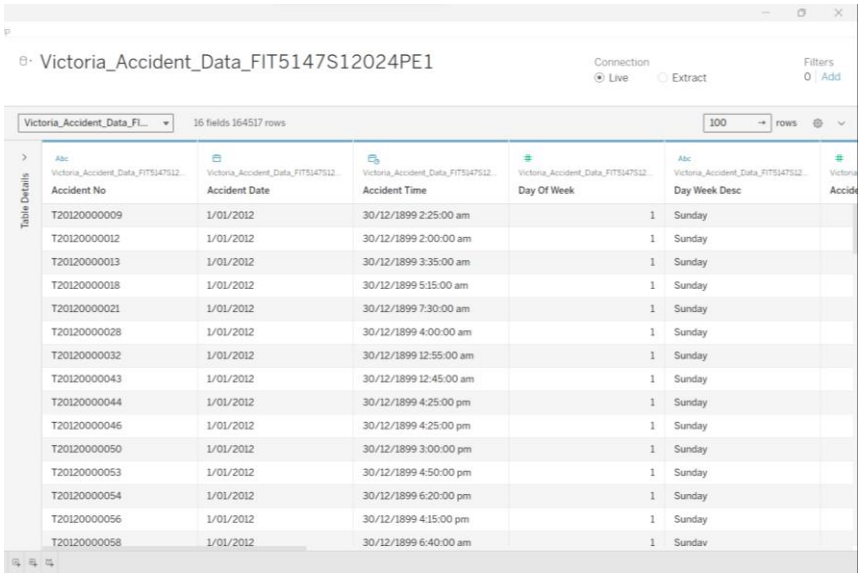


Figure 1 Dataset Overview

First Irregularities

The initial irregularities I encountered pertained to duplicate values for the accident number (ACCIDENT_NO). To solve this issue, I decided to remove the redundant rows using Python. This approach was chosen as duplicate entries offer no additional value to the dataset and may skew analysis results if left unattended.

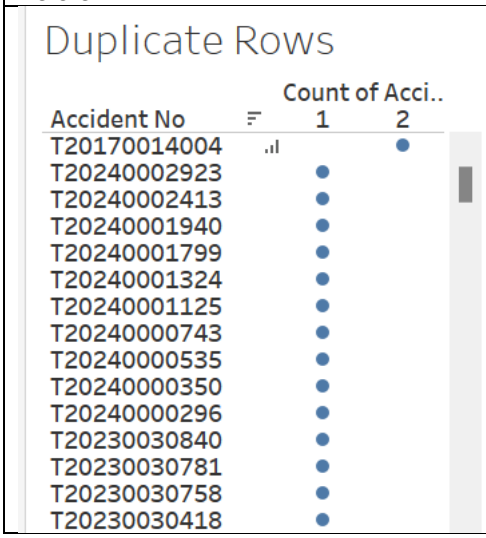
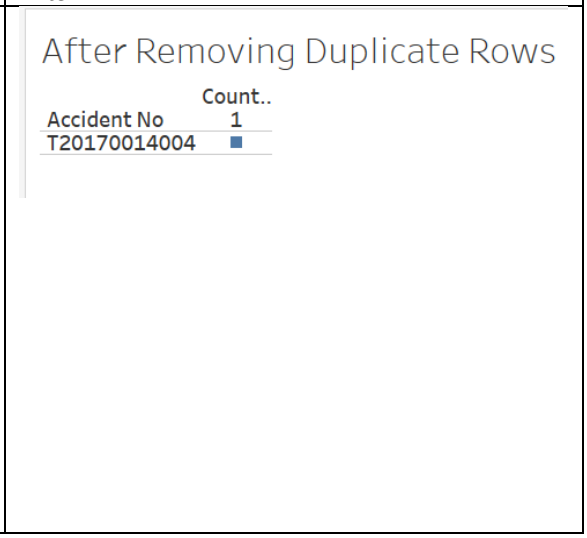
Before	After
	

Figure 2 Duplicate Values (Before and After)

Second Irregularities

I identified an issue related to the SPEED_ZONE column, which had an outlier. Outliers can have a significant impact on statistical analysis and model performance. To address this, I utilized Python to impute the median value for the outlier. By replacing the outlier with the median, we ensure that we have a more representative measure of central tendency and preserve the overall distribution of the data (Horsch, 2021).

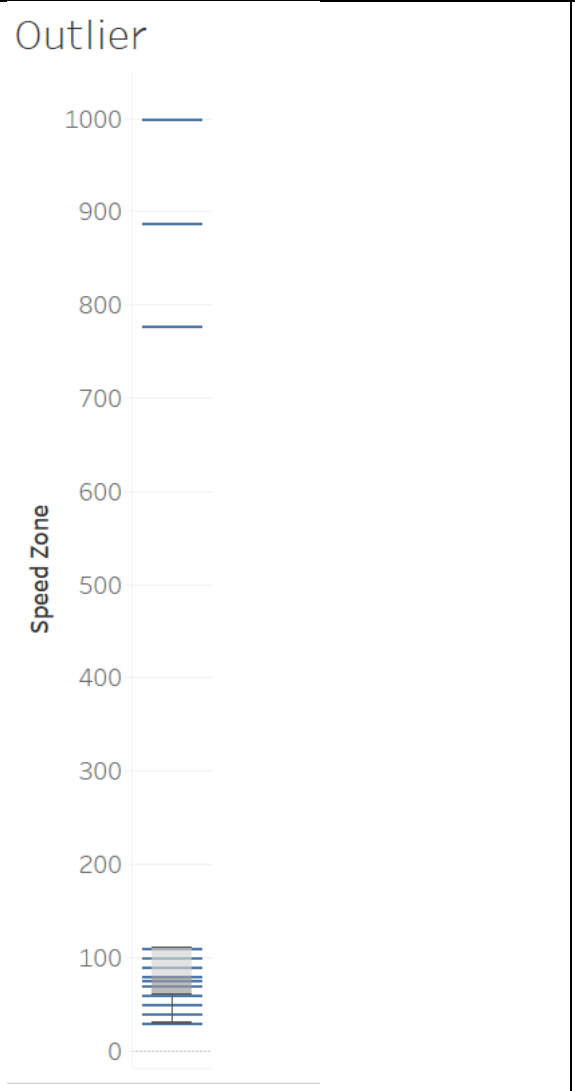
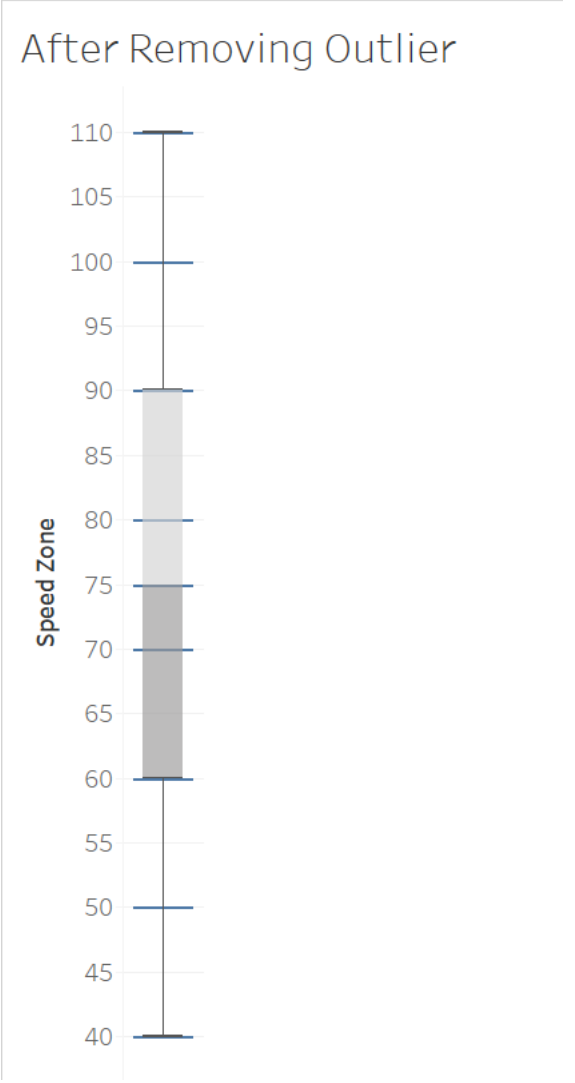
Before	After
	

Figure 3 Outlier (Before and After)

Third Irregularities

The final issue I encountered was inconsistency between the Day of Week (DAY_OF_WEEK) and Day Week Description (DAY_WEEK_DESC) columns. To tackle this, I used Python to standardise the Day of Week column by assigning Monday as 1, Tuesday as 2, and so forth. This alignment ensures uniformity in the dataset.

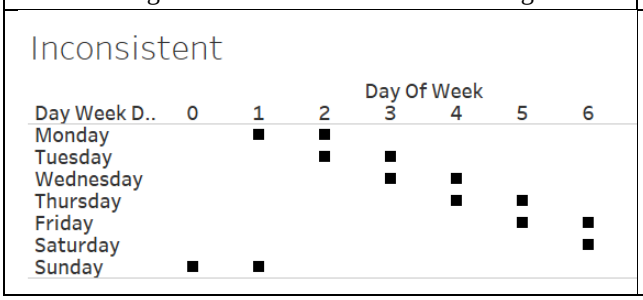
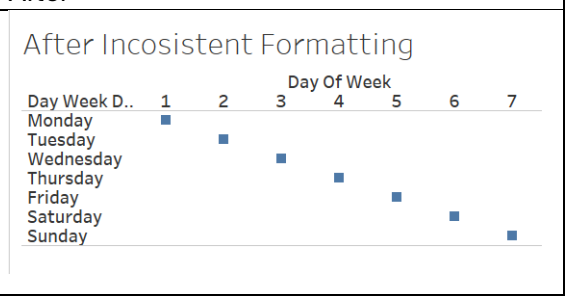
Error & Irregularities 3: Inconsistent Formatting Before	After
	

Figure 4 Inconsistent Formatting (Before and After)

The image's left side depicts the dataset before preprocessing, where duplicate values, outliers, and inconsistent formatting are evident. The right side shows the cleaned dataset after addressing these issues, resulting in a more organised and reliable dataset for analysis.

Data exploration

Question no 1

Compare and contrast when and what type of accidents occur over time (2012-2024). Consider different time periods: hour of day, day of week, month of year, year in the dataset. What do you notice about the pattern over time? What do you know about that time period in order to provide some explanation of increases, decreases, and similar numbers of accidents?

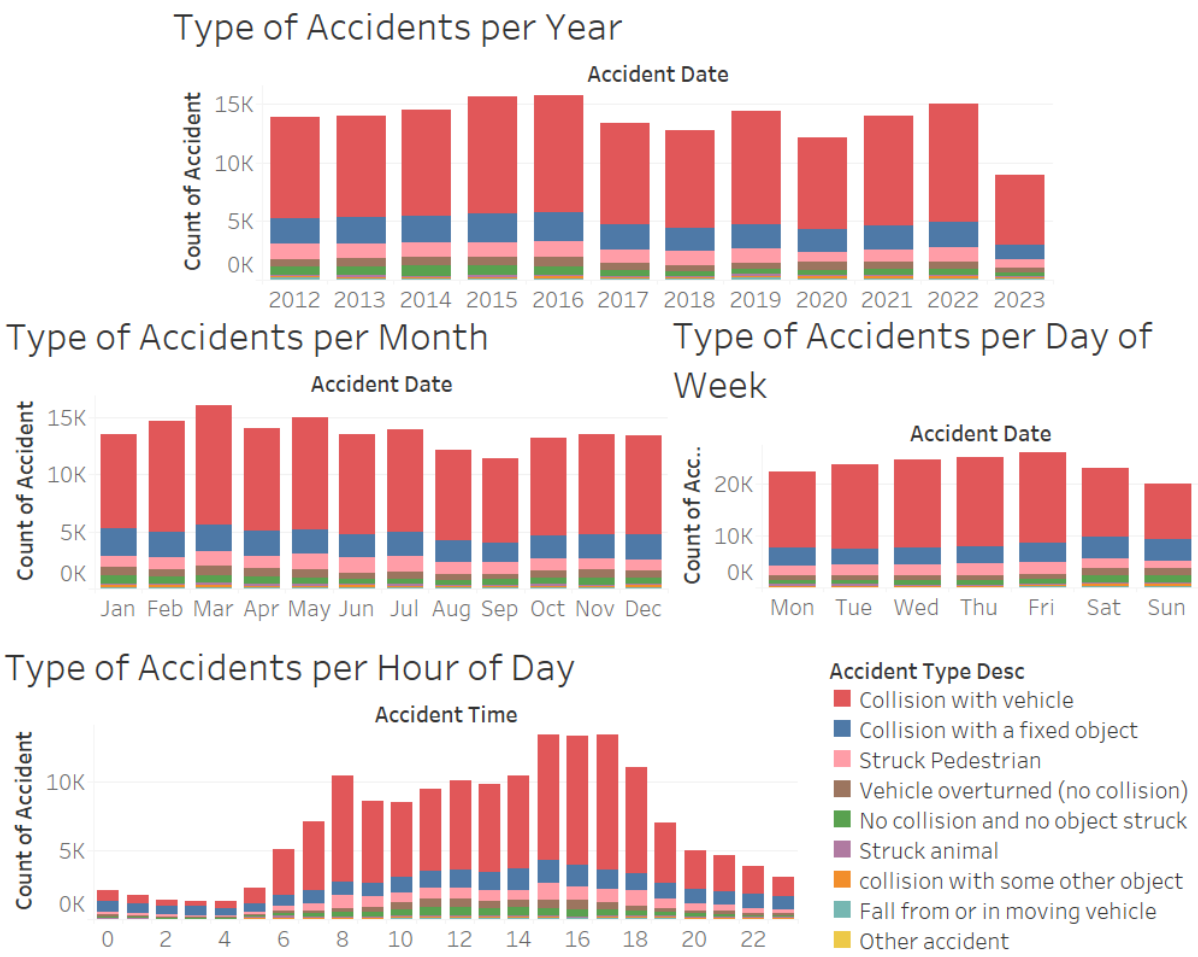


Figure 5 Type of Accidents Over Time (2022-2023)

By analysing data from 2012 to 2024, there were consistent patterns across different timeframes: year, month, day, and hour of the day. Notably, collisions with vehicles and fixed objects are the most frequent accident types throughout this period.

Regarding the yearly distribution, from 2012 to 2014, the number of accidents remained relatively stable. However, a noticeable increase occurred between 2015 and 2016, reaching its peak. Subsequently, there was a decline in the number of accidents, followed by another increase in 2019 and a subsequent rise again in 2022. Regarding the monthly distribution of accidents, March and May stand out as the months with the highest occurrences of accidents, while other months remained stable. These months may coincide with seasonal changes, increased travel activities, or other environmental factors contributing to elevated accident rates.

Regarding the day of the week, it reveals a tendency for accidents to be more prevalent during weekdays compared to weekends. This observation aligns with expectations, as weekdays typically experience higher

levels of vehicular traffic due to work commutes and other weekday activities. Furthermore, examining accidents by the hour of the day highlights two prominent peaks: one in the morning around 8:00 am and another in the late afternoon, between 3:00 pm and 6:00 pm. These peaks likely correspond to rush hour periods.

Question no 2

Compare and contrast where accidents occur geographically. Consider in particular the geometry, speed zone of the road and the urban/rural aspect of the location. How does the spatial mapping of the data and additional information about the road help support, challenge, or change your conclusions to the first question?

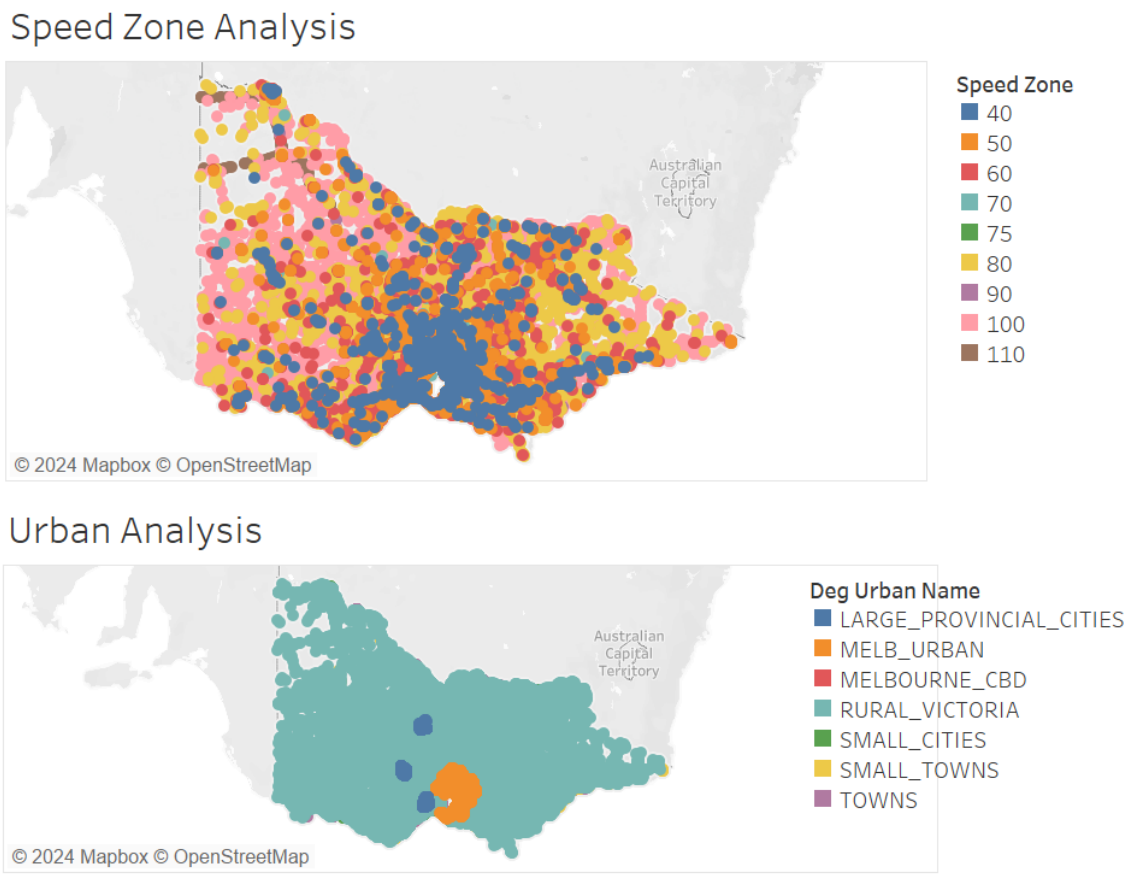


Figure 6 Spatial Analysis of Road Accidents (Speed Zone and Urban)

Based on the analysis of speed zones, it has been found that a considerable number of accidents are taking place in areas where the speed limit is around 100 and 110 km/h. This indicates that speed significantly contributes to road accidents (McPherson, 2024).

The map clearly illustrates that most vehicle crashes occur in rural Victoria. Consequently, we need to explore additional factors such as driver behaviour and road safety enforcement in rural regions to gain a comprehensive understanding of this phenomenon.

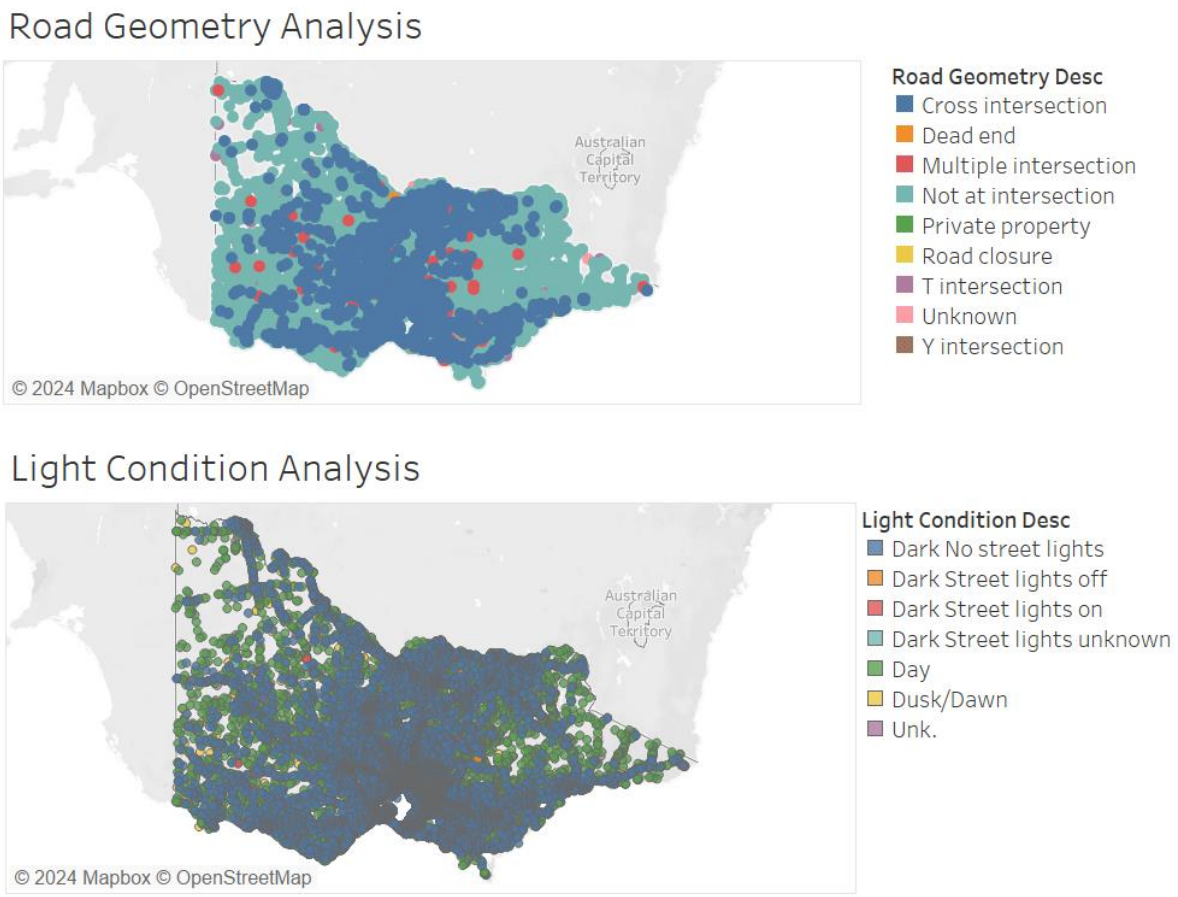


Figure 7 Spatial Analysis of Road Accidents (Road Geometry and Light Condition)

The Choropleth map also reveals that accidents predominantly occur outside intersections, particularly in areas with limited lighting, such as dark roads without streetlights. It proves that the most common case is because of minor mistakes such as being distracted or fatigued (QBE, 2017).

In summary, integrating spatial mapping with additional information about road infrastructure and environmental factors enhances understanding of where and why accidents occur geographically.

References

Horsch, A. (2021, February 15). *Detecting and Treating Outliers In Python — Part 3*. Medium. https://towardsdatascience.com/detecting-and-treating-outliers-in-python-part-3-dcb54abaf7b0

McPherson, E. (2024, March 12). *The common factor first responders say they see in almost all fatal road crashes* [Review of *The common factor first responders say they see in almost all fatal road crashes*]. 9 News; 9 News. https://www.9news.com.au/national/road-toll-australia-the-one-common-factor-first-responders-say-they-see-at-fatal-road-crashes/4b1c7ab2-1583-45ab-b517-97cfe43f8908

QBE. (2017). *The most common causes of car accidents in Australia | QBE AU*. Australia. https://www.qbe.com/au/news/the-most-common-causes-of-car-accidents