MONASH UNIVERSITY

# FIT5147 Data Exploration and Visualisation
# Programming Exercise 1: Tableau


**Divya Harishchandra Shetty**

**Student ID: 33496366**

**Course: Master of Information Technology**

**Teaching Associate: Gavin Mooney, Mohit Gupta**

**Class: Applied Session 05**


## Contents

The dataset used for this analysis is derived from the Vehicle Road Crash Data by the Victorian Department of Transport and Planning in January 2024. This report aims to identify and rectify dataset irregularities, utilizing visual analysis to explore different patterns and presenting insights through clear and concise visualizations created using Tableau.

## 1. Data Checking and Cleaning

Firstly, the dataset is loaded into Tableau. Figure 1.1 shows how the data looks. Subsequently, all data cleaning processes were conducted using Excel and Tableau after identifying the irregularities.

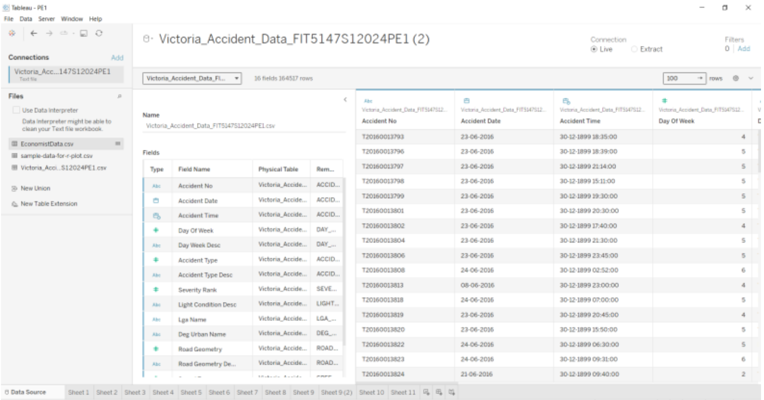


Figure 1.1 - Initial loading of data to Tableau

Upon uploading the dataset into Tableau, it automatically determines the data type of each attribute. The data is visualized to check and find the irregularities in the dataset.

### 1.1 Duplicate Accident Record

While analysing the ACCIDENT_NO, which is a unique id of accident, a duplicate row was identified. This duplication was found by aggregating the data on the ACCIDENT_NO and counting the number for each unique ACCIDENT_NO as shown in Figure 1.2. Accident number T20170014004 has appeared twice in the dataset both having the same details. The duplicate row was deleted to ensure that accident record remained unique and prevent analysis distortion.

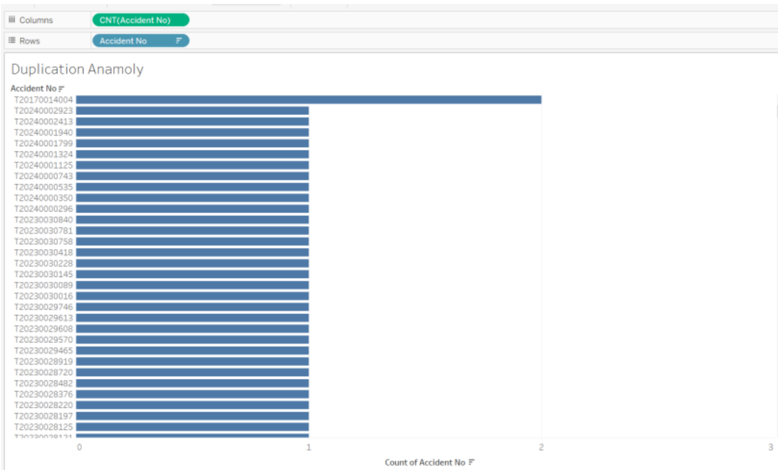


Figure 1.2 – Duplicate Anomaly

### 1.2 Outliers in Speed Zone

On analysing the distribution of speed zone values within the dataset, unusual values of 777, 888, and 999 were noticed which deviates significantly from the standard speed limits recognized in Australia ranging from 30 km/h to 110 km/h as shown in Figure 1.3. Since there were 11,668 entries of outliers, removing them would reduce the size of the dataset which can lead to valuable loss of information and biased analyses. Hence, to maintain the dataset’s integrity for analysis, the outliers were replaced with the median of the speed zone values. Upon filtering out the outliers from the speed zone data and calculating the median in Excel, the resulting value was determined to be 100 km/h. Consequently, the outliers have been substituted with this median value of 100 km/h to ensure consistency and accuracy within the dataset.

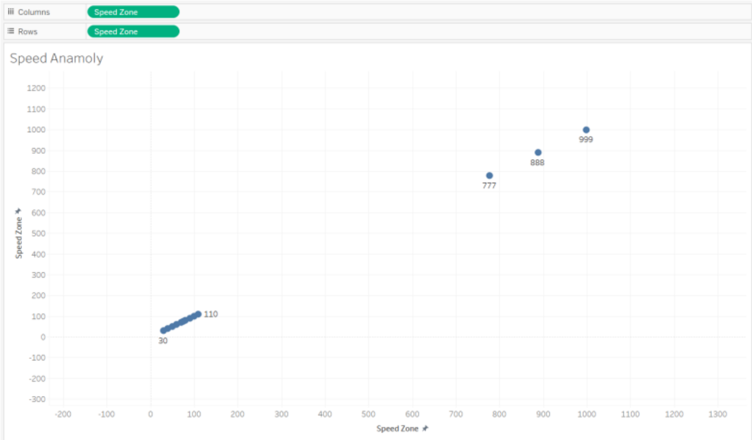


Figure 1.3 – Representation of outliers

## 1.3 Logical anomalies

Upon reviewing the data, an inconsistency between the numerical DAY_OF_WEEK and corresponding textual description (DAY_OF_WEEK_DESC) was identified. This anomaly was detected by simply plotting DAY_OF_WEEK against DAY_OF_WEEK_DESC as shown in Figure 1.4. It can be clearly seen in the below figure that days have been mapped twice from the range 0-7. This discrepancy undermines the dataset's integrity, potentially skewing any time-sensitive analyses leading to inaccurate conclusions about daily accident patterns. To rectify this issue, the days of the week description was mapped according to its numerical value i.e. 1 – Sunday, 2 – Monday, 3 - Tuesday, 4 – Wednesday, 5 – Thursday, 6 – Friday and 7 – Saturday. In this case 1-7 range is considered because the number of Sunday mapped to 1 is more than that of mapped to 0.


Figure 1.4 – Logical Anomaly

## 1.4 Cleaned Data

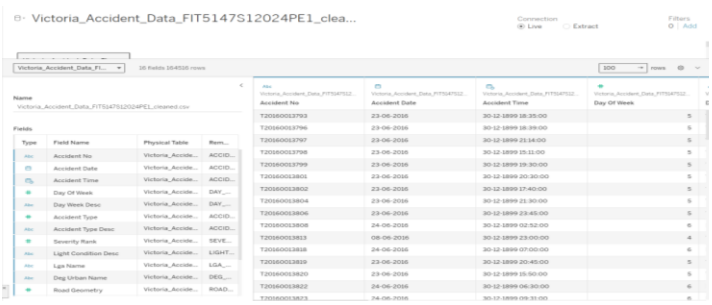

Figure 1.5 – Cleaned Data

## 2. Data exploration

**Q1: 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.**

In this analysis, the temporal patterns of traffic accidents from 2012 to 2023 is examined to understand the types of accidents that occur frequently. By examining these incidents across various timeframe – hourly, weekly, monthly, and annually, significant trends and factors contributing to fluctuations in accident occurrences can be identified. Understanding these patterns is crucial for developing targeted strategies to enhance road safety and mitigate risks.

A stacked bar graph is used to visualise the frequency of the accidents by hour of the day, along with data on accident severity as shown in Figure 2.1. This not only tells us about accident occurrences but also their severity, offering an understanding of road safety challenges at different times of the day.

From this analysis, we can see that accident rates often peak during morning and evening rush hours due to increased traffic volume. Also, late evenings to night hours see a concentration of more sever accidents. From Figure 2.2 we can analyse that spike in accidents during rush hour is typically due to collision with vehicle and potentially hurried or distracted driving as commuters travel to and from work.

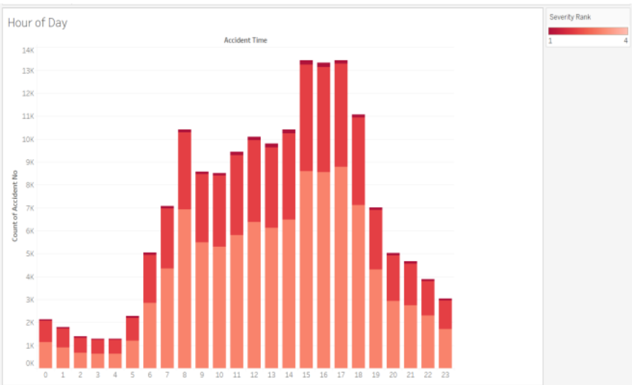


Figure 2.1 – Number of accidents through hour of Day

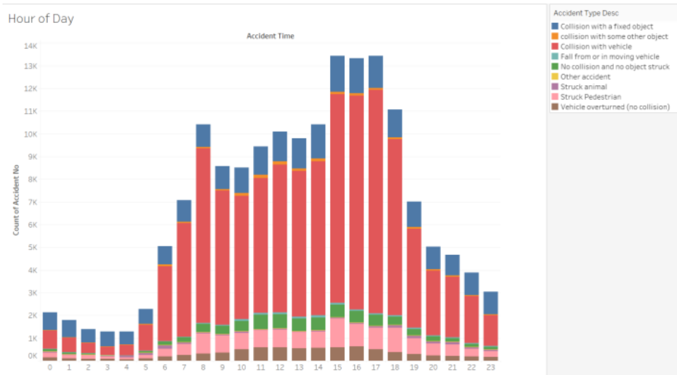


Figure 2.2 – Accident type through hour of Day

To analyse accidents occurring on different days of the week, a stacked bar graph is used as shown in Figure 2.3. This visualization clearly reveals that accidents are more prevalent on weekdays, with a distinct spike observed on Fridays. This uptick likely corresponds with the onset of weekend activities, as individuals embark on travels or engage in social events. Conversely, while the weekend shows a reduction in the total number of accidents, those that do occur tend to be of higher severity. This pattern suggests that weekend accidents may be more likely to involve factors such as alcohol consumption, associated with leisure activities.

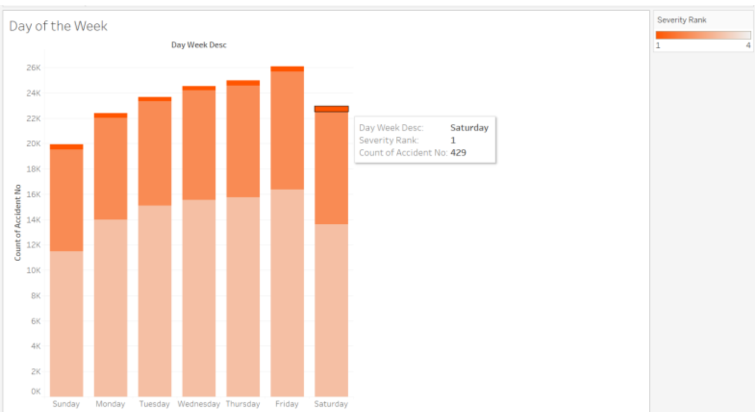


Figure 2.3 – Number of Accidents through Days of the Week

Utilizing a stacked bar graph, as depicted in Figure 2.4, reveals notable seasonal variations in road accidents throughout the year. March, situated in Victoria's autumn, emerges as the month with the highest accident frequency, likely influenced by fluctuating daylight and unpredictable weather, including rainfall. Conversely, September a time of early spring, shows fewer incidents, attributed to the gradual shift from winter, potentially leading to decreased traffic and accidents due to the reduced outdoor activities. The consistent accident figures in November and December align with the onset of summer and festive activities, possibly escalating due to events that involve alcohol consumption. Moreover, the period between May and June, marking early to mid-winter, experiences a peak in accidents, primarily due to reduced visibility from insufficient street lighting and the inherent darkness of the season.

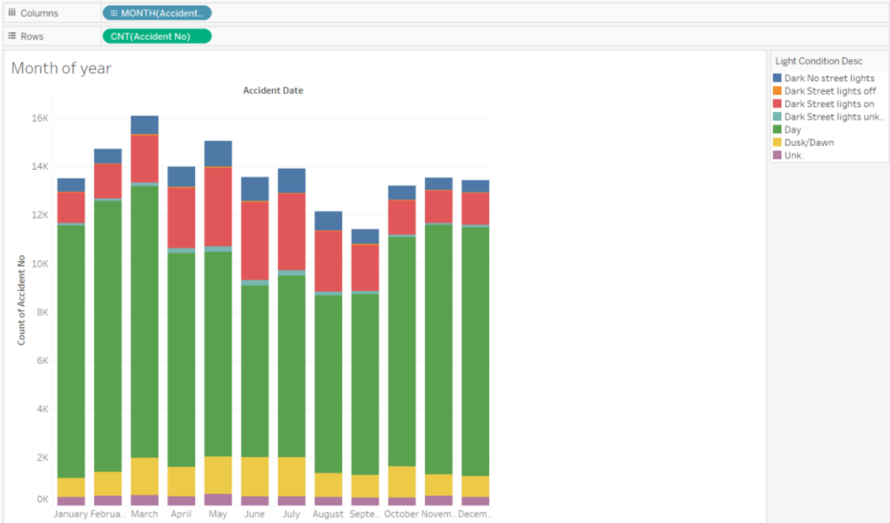


Figure 2.4 – Number of Accidents through Month

Analysing the accident trends over the years (2012-2023) using a stacked bar graph as shown in Figure 2.5, we can see that highest number of accidents occurred in the year 2015 and 2016 mostly due to collision with vehicles. This peak is followed by a notable decrease in incidents starting in 2017. This indicates improvements in vehicle safety features and effective road safety campaigns [1] . 2020 experienced the lowest accident rates, primarily due to the global pandemic's restrictions on mobility. It is important to note that the accident data for 2023 is compiled only up until July, indicating a partial year's analysis.

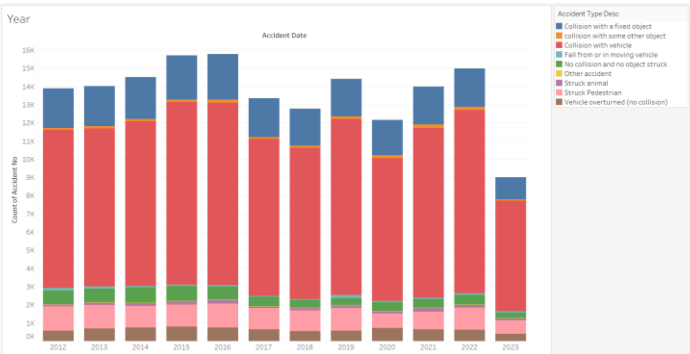


Figure 2.5 – Number of accidents over the year (2012-2023)

**Q2: 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?**

The spatial analysis of road geometry causing accidents as shown in Figure 2.6 shows that highest number of accidents occur at ‘not at intersection.’ This suggests that straight roads, often associated with uninterrupted driving, are more prone to accidents, challenging the common perception that intersections pose higher risk.

The visual representation as shown in Figure 2.7 shows where accidents occur most frequently within various speed zones. Rural regions marked in red appears to have a significant number of accidents, potentially due to factors such as higher speed limits which is common in less populated areas. Conversely, the urban areas of Melbourne, highlighted in lighter shades of red, also show a high occurrence of accidents, which could be sue to higher traffic density and different road geometry (Cross intersection).

In rural zones, accidents may be more dispersed, but can have a higher severity as shown in Figure 2.8 and Figure 2.9 due to factors like higher travel speeds on country roads and longer distance to medical care. The higher density of accidents in the urban areas can be attributed to various factors like higher traffic volumes and more complex road networks.

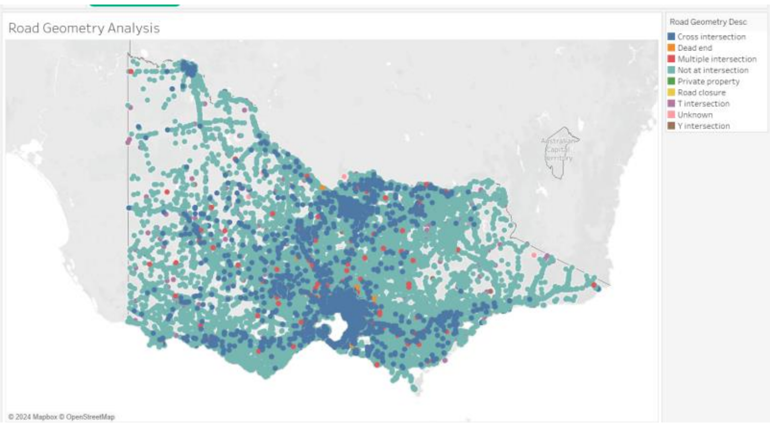


Figure 2.6 – Road Geometry Analysis

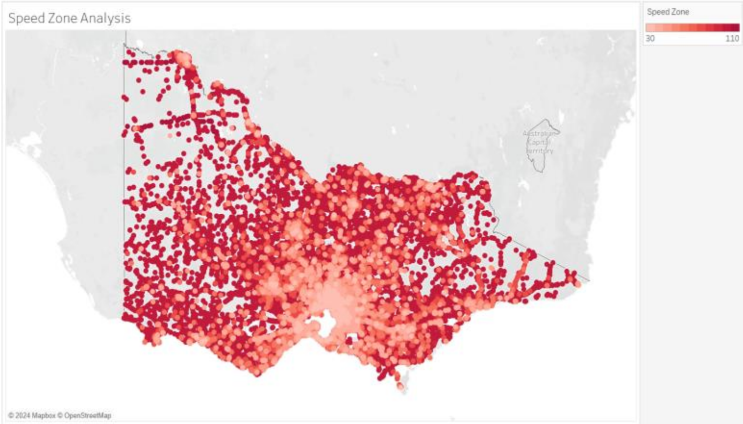


Figure 2.7 – Speed Zone Analysis

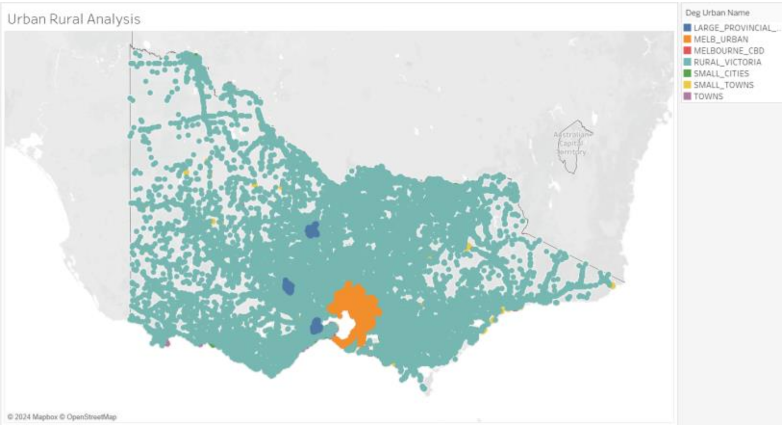


Figure 2.8 - Urban Rural Analysis

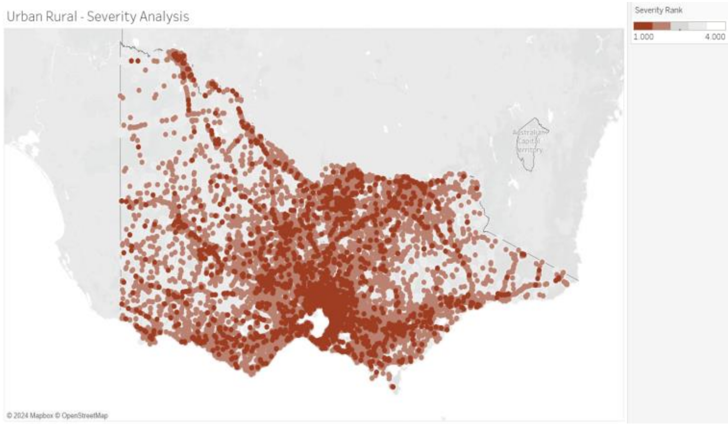


Figure 2.9 – Severity Analysis

Spatial mapping enhances our understanding of accidents by providing context to temporal patterns observed from 2012-2023. While temporal data indicated more accidents during rush hours, spatial analysis reveals that straight roads in rural areas, not just busy urban intersections, are significant sites of accidents. This challenges the notion that time-specific congestion is the primary risk factor, suggesting that road type and speed also play pivotal roles. Additionally, the distribution of accidents within various speed zones shows that rural areas, often with higher speed limits, have a comparable incidence of accidents to urban centres, indicating that high speeds contribute to risk across both settings. Hence, the combination of temporal and spatial data can influence the development of road safety measures.

## 3. References

[1] Road Safety Road Rules 2017, Authorised Version No. 001, S.R. No. 41/2017, Authorised Version as at 1 July 2017. Available: https://www.legislation.vic.gov.au/in-force/statutory-rules/road-safety-road-rules-2017/001