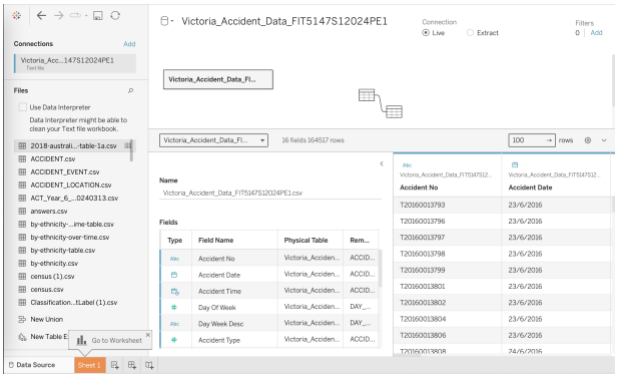
# Victoria Accident Analysis and Visualisation

## Table of Contents

## Data Loading

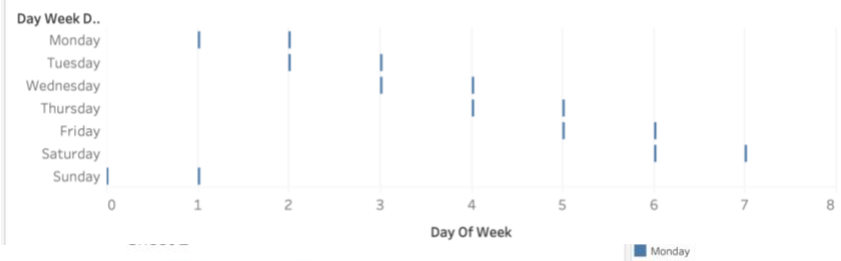
First, we import data to Tableau using the text file option. The data will look like this:

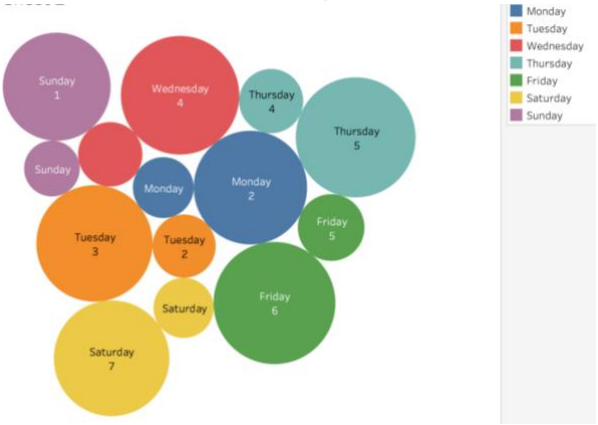



## Data Checking and Cleaning

### 1. Inconsistency in Day of the Week Data

The plot created to investigate the consistency between the 'Day of Week' and 'Day Week Desc' columns revealed discrepancies. Each day's name was associated with two different numerical day indexes. To determine the correct index for each day, I utilized a bubble chart. This visualization indicated that the higher index is likely accurate for each day, as it appeared 3-4 times more frequently than the lower one.

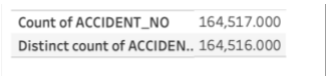



**Correction:** Based on the assumption that the majority occurrence would more likely represent the correct data I mapped the correct indices to their respective day names using a dictionary and replace the incorrect ones. To be sure that this is the correct approach I manually checked some dates and saw which day of the week was that day.

### 2. Duplicate Error

The issue was spotted when comparing two different counts in Tableau: the total number of "ACCIDENT_NO" records, which was 164,517, and the number of unique "ACCIDENT_NO" records, which came out to be 164,516. This means there is one record in the dataset that has been counted twice, which should not happen since each "ACCIDENT_NO" is meant to be unique for each accident.



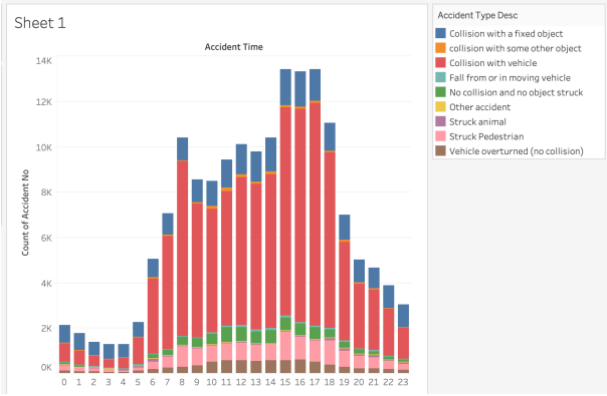
**Correction:** I used a Python script to first identify the duplicate record. Then, after finding it, I compared the other values in these two records and, seeing that they were identical, I removed one of the duplicated rows.

## Data Exploration

**Analysis of Accident Types Over Hours of the Day**

**Findings:** The chart shows two peaks in accidents: a morning peak (7:00-9:00) and a late afternoon peak (15:00-18:00), coinciding with work commutes. The increased traffic likely leads to more accidents, with 'Collision with vehicle' being most common during these times. 'Collision with a fixed object' accidents are notably frequent in early hours (00:00-04:00), when visibility is low and traffic is sparse.

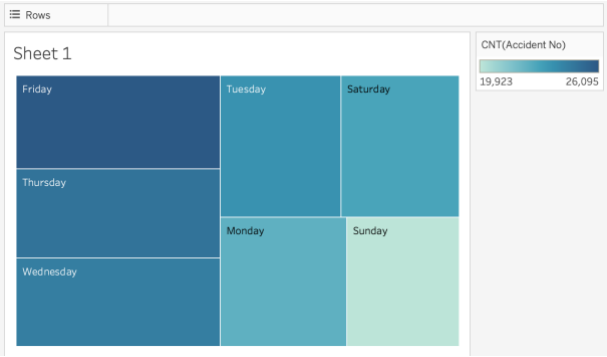
**Explanation for Visualization Choice:** The stacked bar chart was selected for its ability to showcase multiple accident types across different times, with color coding for quick type identification, highlighting specific time-based patterns.




**Analysis of Accident Frequency by Day of the Week**

**Findings:** The visualization indicates a clear pattern regarding the frequency of accidents on different days of the week. Weekdays generally see a higher number of accidents, with Friday being the peak day at 26,095 accidents. The numbers gradually increase from Monday, suggesting a possible buildup of fatigue or other compounding factors over the week. In contrast, the weekend, especially Sunday, has the lowest count of accidents, which may reflect reduced commuter traffic and overall vehicle movement.

**Explanation for Visualization Choice:** A treemap helps us easily compare how common accidents are on different days of the week. Bigger and darker rectangles show us days with more accidents. This makes it simple to see that Fridays have the most accidents and Sundays have the fewest just by looking at the size and color of the rectangles.

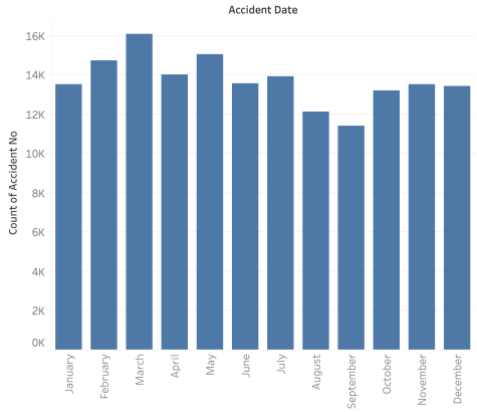



**Month of Year Analysis**

**Findings:** The bar chart indicates a relatively consistent distribution of accidents throughout the year, with slight variations. Notably, March shows the highest number of accidents, while September presents with the lowest. The reason that March is the month with a high accident count, might coincide with changing seasons, potentially leading to more challenging driving conditions.

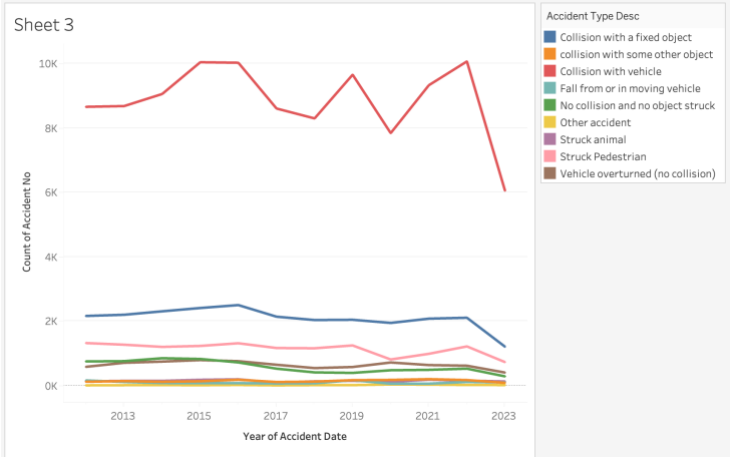
**Explanation for Visualization Choice:** The bar chart is effective for displaying the frequency of accidents in each month as it allows for straightforward month-by-month comparison. It is one of the most accessible

visualization types, making it easier for a wide audience to understand the distribution of accidents throughout the year.

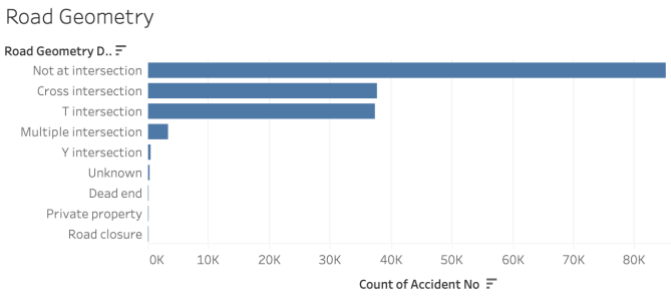



## Years Analysis

**Findings:** Observing the graph, there is an apparent peak in accidents involving a vehicle (red line) around the years 2015-2016, followed by a notable decline then it increased and there was another peak in 2022. Struck pedestrian accidents (pink line) remain relatively stable across the years, with a slight downward trend. Accidents involving collision with a fixed object (blue line) and vehicle overturns (tan line) show a steady pattern without significant fluctuations. The reason that in 2023 all type of accident is decreased is that the collected data is until July 2023 and it is not for the 2023 completely.
**Explanation for Visualization Choice:** A line graph is suitable for displaying data trends over time. It allows us to track changes and compare the frequency of different accident types across multiple years. This visualization makes it easier to identify upward or downward trends in accident rates and potential anomalies in the data.



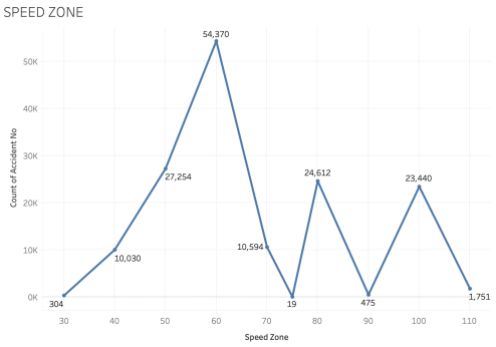

## Road Geometry Analysis




From the bar chart, it appears that accidents most frequently occur in areas 'Not at intersection,' suggesting that stretches of road without intersections may have factors contributing to accidents, such as higher speeds or fewer controls like stop signs or traffic lights. 'Cross intersections' and 'T intersections' follow in frequency,

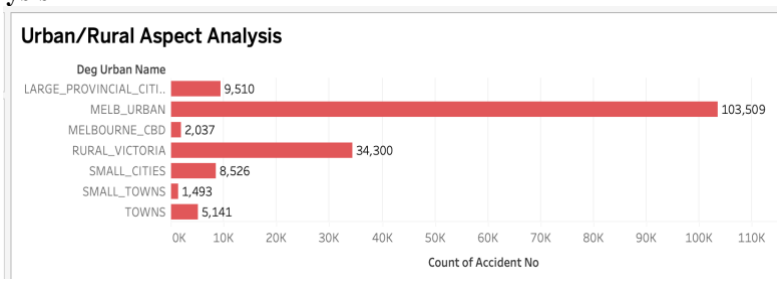
which could be due to the complexity of navigating these points, where multiple traffic directions intersect, increasing the potential for collisions.

## Speed Zone Analysis



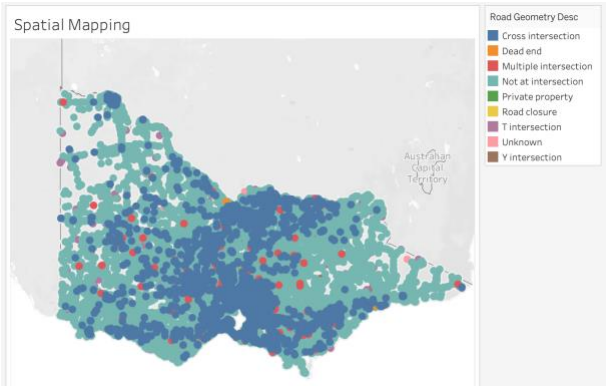

The line chart for Speed Zone indicates a significant peak in accidents within the 60 km/h zone, followed by notable frequencies in the 50 km/h, 80km/h and 100 km/h zones. But we can't really say which speed zone is safer without knowing how many roads have each speed limit. For example, the chart shows more accidents in the 60 km/h zones, but this might be because there are just more roads with that speed limit.

## Urban/Rural Analysis

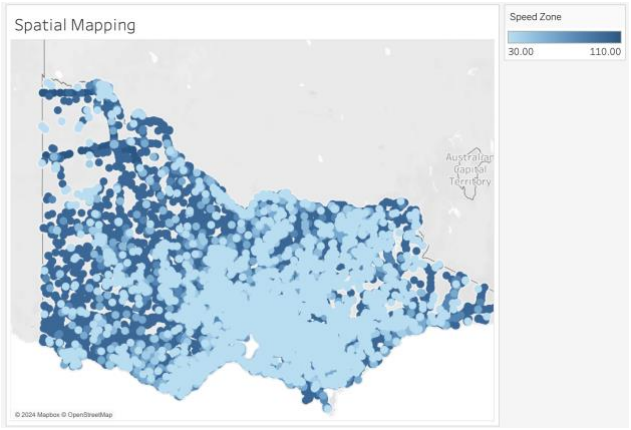



The bar chart shows that most accidents occur in the 'Melb_Urban' area, which is much higher than in any other area. Rural areas have fewer accidents. This might be because there are more cars and people in the city, making accidents more likely. Small towns have the fewest accidents. Even though the central business district (CBD) of Melbourne has fewer accidents, its small area and high density make this number quite significant.

## Spatial Mapping Analysis




The map supports the initial conclusions by showing the spread and frequency of accidents across various road geometries.




At first look, the bellow map might seem to contradict our earlier statement by appearing to show more accidents in Rural Victoria than in Melbourne's urban areas. However, this interpretation would be misleading. The larger geographic area of Rural Victoria might give the impression of more accidents, but a closer look reveals a higher density of accidents in the more compact Melbourne urban area.