|  |
| --- |
| Victoria Accident Data Visualisation Project Executive Summary |
| Max Woollons s5349356Reiss Tooze s5046939Kail Terepai s53074742810ICT Software TechnologiesDate: 20/9/23 |

# Abstract

The results of these queries have made clear the dangers of driving and the most common causes for accidents. They have shown that alcohol can dramatically increase the likelihood of a fatality occurring in an accident. The results also have shown the most dangerous days and time of days to drive.

These figures have also made apparent that data figures, such as charts, graphs, and visualisations, offer many benefits for data analysis as they provide a visual representation of complex datasets. They make it easier to understand patterns, trends, and relationships within the data, even for individuals who may not have a strong background in statistics or data analysis.

Visual data analysis like this can also assist with identifying errors and issues with datasets as the charts and graphs produce will make it apparent if there are missing values or data entry errors such as gaps in data for a certain time of day.

# Introduction

The purpose of this executive summary is to analyse the data that has been produced in a graphical format of the following queries:

* For a user-selected period, display the information of all accidents that happened in the period.
* For a user-selected period, produce a chart to show the number of accidents in each hour of the day (on average).
* For a user-selected period, retrieve all accidents caused by an accident type that contains a keyword (user selected), e.g. collision, pedestrian.
* Allow the user to analyse the impact of alcohol in accidents – ie: trends over time, accident types involving alcohol, etc.
* For a user-selected year, display the information of all accidents that occurred on a Victorian public holiday within the selected year.

These queries will cover a 12 month period between the years 2015 to 2020. The goal is that the analysis of this data could identify some key patterns in the cause or effects of road accidents and could potentially help inform future decision making in road safety.

# Analysis 1 - For a user-selected period, display the information of all accidents that happened in the period

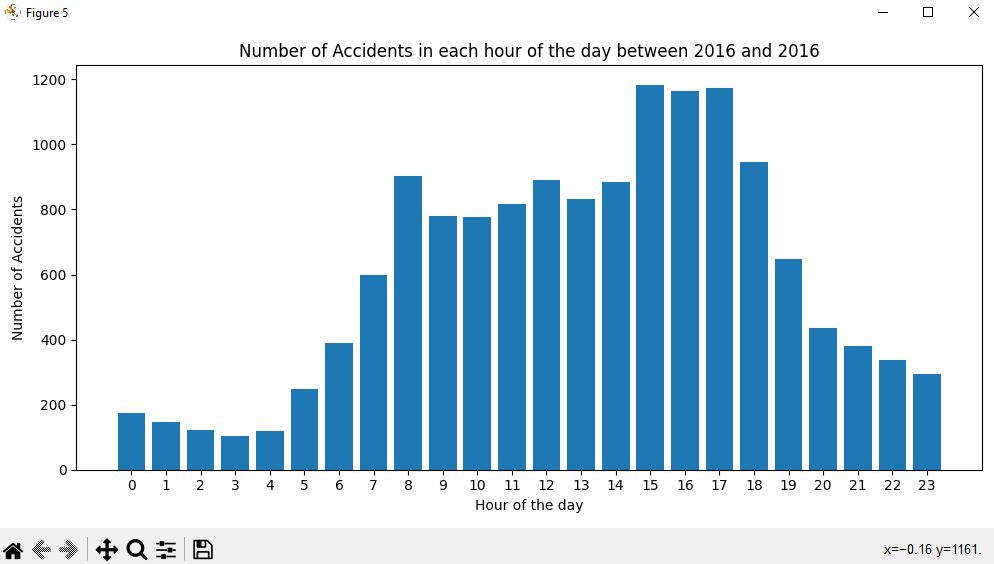
Figure 1: For a user-selected period, display the information of all accidents that happened in the period


The image above is the output for the query: “For a user-selected period, display information on all accidents that happened in the period” with the period selected being the whole year of 2019.

What can be inferred from this data is that the most common cause of accidents was striking an animal which accounted for 41.2% of all accidents in 2019. Knowing that animal collisions pose such a threat to driver’s safety, drivers could reduce this figure by lowering their speed in known high activity times for certain species (most likely kangaroos).

According to the output of the query the types of accidents that are least common are vehicle overturned (no collison), collision with a vehicle and collision with another object. These all shared an equal 5.9% each but cumulatively they accounted for 17.7% which is still not a negligible figure.

# Analysis 2 - For a user-selected period, produce a chart to show the number of accidents in each hour of the day (on average)



The image above is the output for the query: “For a user selected period, produce a chart to show the number of accidents in each hour of the day(on average)” with the selected year for the output being the whole year of 2016.

What immediately stands out from this data is that the most hazardous travel time is the “rush hour” period of 1500 to 1700 accounting for nearly 1200 accidents per hour on average. This high accident figure would be caused by the cumulative effect of people finishing work and parents collecting their children from school resulting in a high influx of drivers on the road at the same time.

Another notable hazardous hour for driving is 0800 which, similarly to the afternoon, would be a result of many drivers on the road going to work and dropping their kids off to school.

# Analysis 3 - For a user-selected period, retrieve all accidents caused by an accident type that contains a keyword (user selected)

# 

The images above show the outputs for the query “For a user-selected period, retrieve all accidents caused by an accident type that contains a keyword (user selected)” with the chosen keywords being: struck pedestrian, collision with a fixed object, collision with vehicle and no collision and no objects struck.

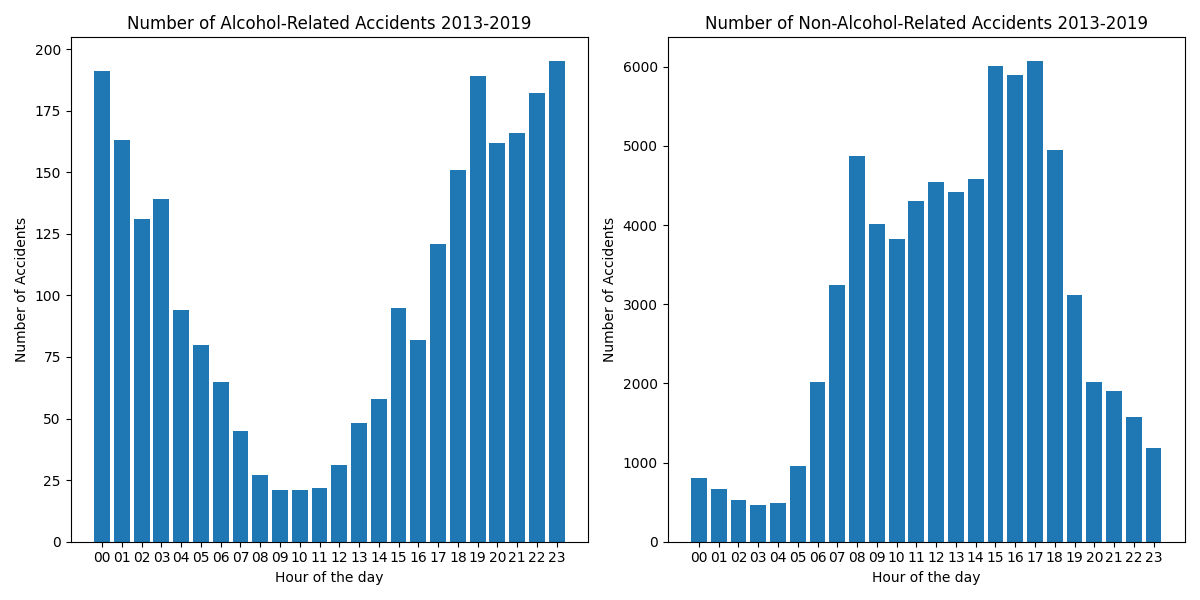
For the keyword “struck pedestrian” the figure shows us that the monthly average accidents involving struck pedestrians sits at around 110 accidents with the high outlier being July at over 140 and the low outlier being December at around 70 accidents.

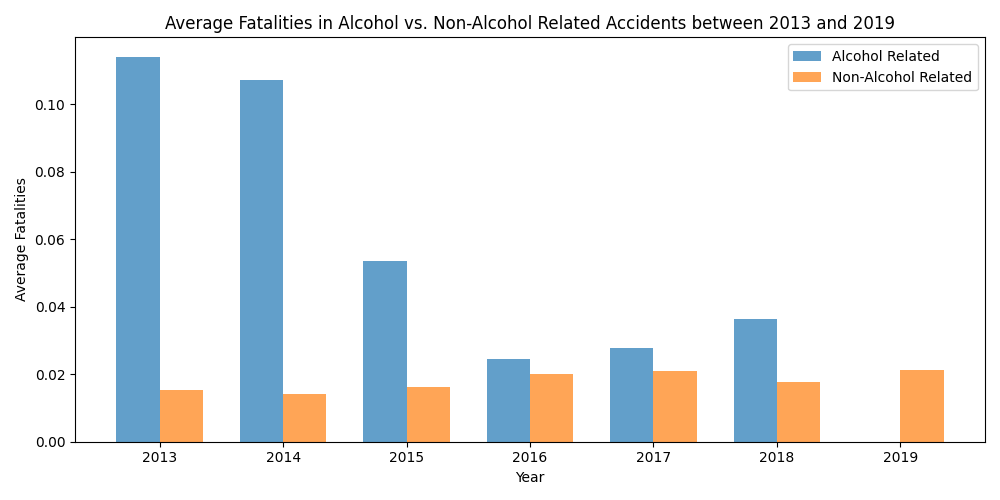
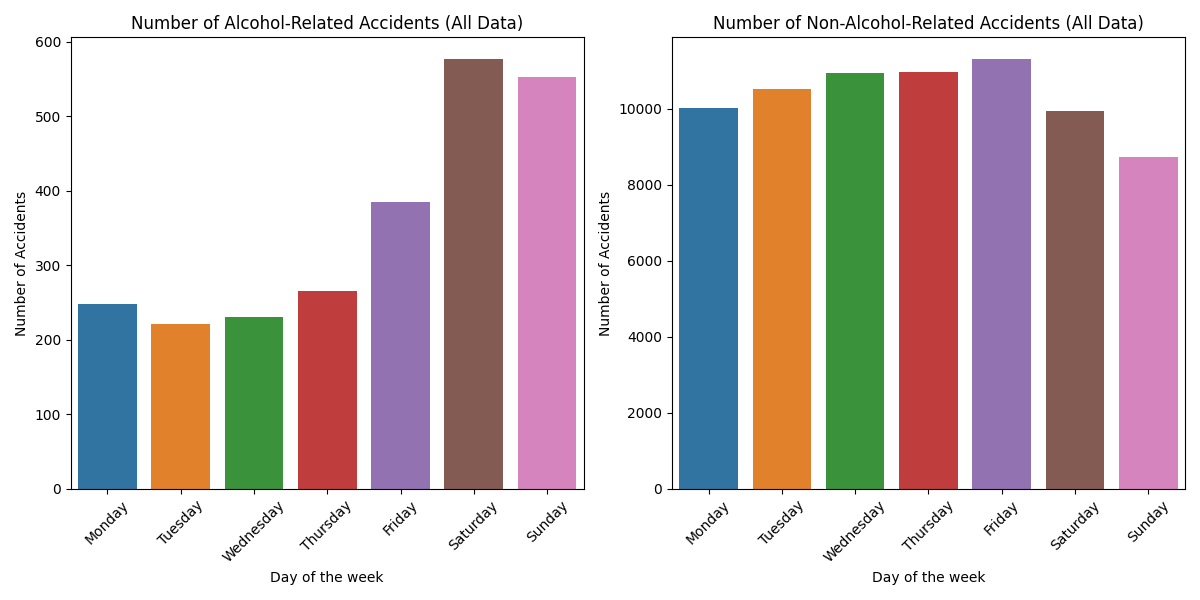
For the keyword “collision with a fixed object” the figure shows us that the monthly average accidents involving collisions with a fixed object sits at around 170 accidents with the high outliers being December and January at over 210. The higher number of accidents of this type during these months specifically could be due to the summer holidays and increased vehicle density due to these holidays.

For the keyword “collision with vehicle” the figure shows us that the monthly average accidents involving collision with vehicles sits at around 750 accidents with the highest February to May at over 800 accidents respectively and the lowest being January at just over 600. The monthly accidents for this accident type does not vary greatly to infer any kind of reasoning for the differences.

For the keyword “no collision and no objects struck” the figure shows us that the monthly average accidents involving no collisions and no objects struck sits at around 60 accidents with the high outlier being January at just below 80 and the low outlier being July at just over 30 accidents.

# Analysis 4 - Allow the user to analyse the impact of alcohol in accidents – ie: trends over time, accident types involving alcohol, etc





The images above show the outputs for queries that visualise accident data for accidents involving alcohol using all time periods of the data (2013 - 2019).

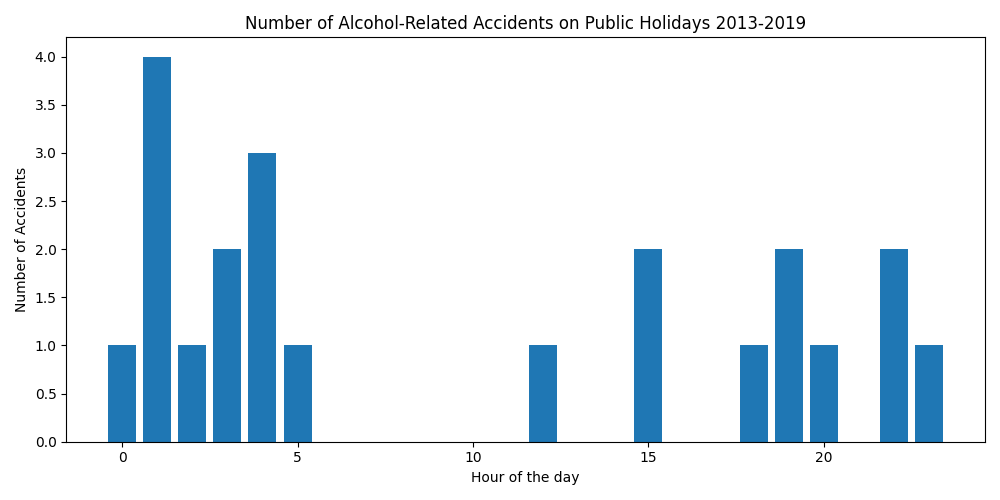
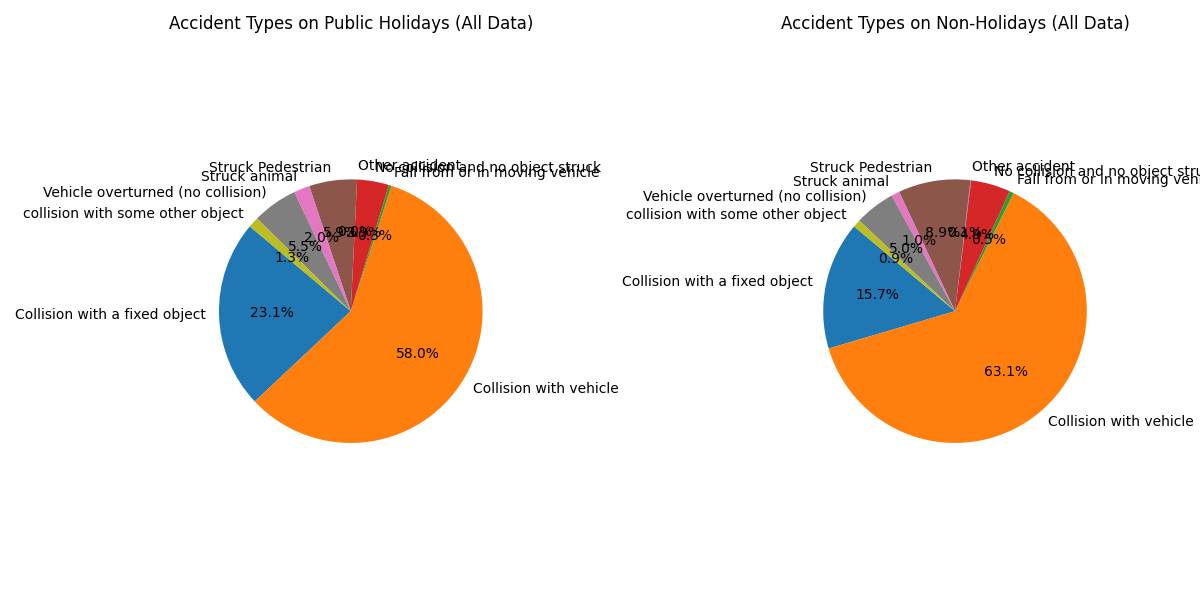
The first figure shows the amount of accidents involving alcohol for each hour of the day and compares it to accidents that do not involve alcohol. The data in this figure indicates that most accidents involving alcohol occur between the hours of 1700 and 0300 with the least amount of accidents occurring between 0800 and 1200. This aligns with the time of day you would expect most people to drink alcohol.

The second figure compares alcohol vs non-alcohol related accidents for each day of the week. Unsurprisingly accidents involving alcohol increase on a Friday but mostly occur on a Saturday and Sunday whereas accidents not involving alcohol do not vary greatly day to day.

The final figure compares alcohol vs non-alcohol related accidents where a fatality occurs for each year of the available data (2013 to 2019). It shows a dramatic reduction in fatalities from a highest average of 10% in 2013 to a lowest of 2% in 2016 then increasing again to 4% in 2018. This reduction could be the result of a combination of factors including newer cars using more safety equipment, higher driver education or increased policing.

This final figure also makes it apparent that accidents involving alcohol are much more likely to have a fatality than an accident not involving alcohol.

# Analysis 5 - For a user-selected year, display the information of all accidents that occurred on a Victorian public holiday within the selected year



The final analysis is the query “For a user-selected year, display the information of all accidents that occurred on a Victorian public holiday within the selected year.”

The first figure compares the occurrence of all accident types on public holidays vs non-public holidays. This comparison shows that on public holidays there are collisions with other vehicles but more collisions with fixed objects. If one assumes that alcohol consumption increases on public holidays, then that may be the cause of this variance.

The second figure shows the average number of accidents for each hour of the day on public holidays. We can see the most dangerous hour of the day is 0100. The gaps in this figure are a result of the limited data in the dataset regarding public holidays.