DRIVER SAFETY INTERNSHIP PROJECT

VICTORIA UNIVERSITY IN COLLABORATION WITH DIRECTED TECHNOLOGIES

Report Submitted to **Directed Technologies** by Victoria University Team:

Suraj Awal

Jessica

Jashanjot

15 October 2020

Table of Contents

1. Introduction

Driver safety is the major concern for every fleet operator and organizations who need workers in a role of driver. This concern is to reduce the risk of injury that the workers may face during their shifts while doing their role in the workplace. Such possibility is beneficial for the organization to increase productivity and employees' trust.

The current safe driver monitoring has a positive impact on driver safety. The current system can trace any dangerous event performed by the driver such as harsh braking, harsh speeding, speeding and so on. They also have some limitations including:

- No spatial context to the event (for example: the event harsh braking is more dangerous during rainy days or foggy nights)
- Not implemented any self-learning methodology (for example: considering the previous driving and fines history of the driver, recommend the safe speed range considering current weather and locality)

This project is the initiation to add value to the current driver safety system by attempting to identify patterns that can lead towards eradication of the prevailing limitations faced. This document provides the data analysis and acts as a guide for future references.

2. Business Problems

The main goal during this project is to improve the current system so that it can provide advanced safety for the drivers. Keeping this in mind and based on the Directed's project description, adding the spatial context to the event, and formulating the self-learning mechanism is our primary concern.

Based on the given goal, following are major lacking issues in current system that need to be added:

Environmental Context: The logged events are dangerous but the environmental condition at the time of the event plays the big role on risk factors

Historical Context: Similarly, the history of the driver such as speeding, traffic tickets, accident occurred also credits towards the risk severity of the any dangerous event

Self-Learning and Adaptation: With capability to learn, improve and adapt based on the new data, the system could be more elegant and competitive.

3. Methodology

3.1 Data Sources

Different internal and external datasets have been used to base the analysis of driver safety from the perspective of environmental aspects. The internal datasets used is the event alarm data provided by Directed Technologies while the external datasets used are accident data, weather data, postcode data and so on.

3.1.1 Event Alarm Data

Data Period: 12/07/2020 to 27/07/2020

Source: Vehicle tracking system of Directed Technologies

Data composition:

- Timestamp (timestamp in UTC)
- LocalTime (timestamp in time zone at vehicle location)
- TimeZone (time zone at vehicle location)
- Latitude (Latitude of vehicle location during the alarm trigger event)
- Longitude (Longitude of vehicle location during the alarm trigger event)
- AlarmType (ACC = harsh acceleration, BRK = harsh braking, OSP = speeding)
- VehicleId (Unique identifier of the vehicle)
- MaxSpeed (Maximum speed achieved while speeding)
- Delta (number of kph increase or decrease in one second for harsh driving alarms)

3.1.2 Crash Data of Victoria

Data Period: 2015 to 2019

Source: https://discover.data.vic.gov.au/dataset/crash-stats-data-extract (Victorian Government)

Data composition:

Access full data composition from :

https://data.vicroads.vic.gov.au/Metadata/Attribute_Table_Viewlist_4b251285.asp.html

- Longitude (GDA94 Longitude coordinate - Decimal Degrees)
- Latitude (GDA94 Latitude coordinate - Decimal Degrees)
- ACCIDENT_DATE (Accident date)
- ACCIDENT_TIME (Accident time)
- ALCOHOLTIME (Incidents occurred within Road Crash Information System Definition of Alcohol Times)
- ACCIDENT_NO (Accident Number)
- ACCIDENT_STATUS (Accident status)

- ACCIDENT_TYPE (Accident type)
- DAY_OF_WEEK (Day of week)
- LIGHT_CONDITION (Level of brightness at the time of the accident)
- and more

3.1.3 Weather Station Data

Source: http://www.bom.gov.au/metadata/catalogue/19115/ANZCW0503900703?template=full#distribution-information

Data composition:

Access full data composition from:

https://data.gov.au/data/dataset/82e2ab28-5437-456f-aca2-fd23ce41cd37/resource/35592433-da31-46e4-b0b5-efc15cc0ad5d/download/eta_data_brochure.html#Data-Schema-for-Forecast-and-Observations

- station_number (unique number for weather station)
- parameter (name of weather element)
- valid_start (start of the time period in epoch time that the forecast or observed value corresponds to)
- value (numerical value for the given weather element)
- unit (the unit for the value (e.g. Celsius, mm, %)
- and more

3.1.4 Victoria Postcode Data

Source: https://www.matthewproctor.com/australian_postcodes

Data composition:

- postcode (postcode in numerical format)
- locality (locality of postcode)
- state (Australian state in which the locality is located)
- long (longitude of locality)
- lat (latitude of locality)
- and more

3.1.5 Victoria Suburb Data

Source: https://data.gov.au/data/dataset/af33dd8c-0534-4e18-9245-fc64440f742e

Data Composition:

- Polygon data for coordinate extents of each suburb
- Polygon unique identifier (LC_PLY_ID)
- Suburb ID (LOC_PID)
- Suburb Name (VIC_LOCA_2)

3.2 Data Cleaning and Preparation

3.2.1 Integration of Accident, Weather and Postcode Data

This process cleans accident, weather, and postcode data to retrieve only the relevant fields and to integrate the weather and postcode data in context of accident. Generally, the main aim is to find out the weather condition and locality details for each of the accidents between May 2017 to April 2018.

The cleaning and integration step is performed using Tableau Prep Builder. The flow designed for this process is shown in the given figure.

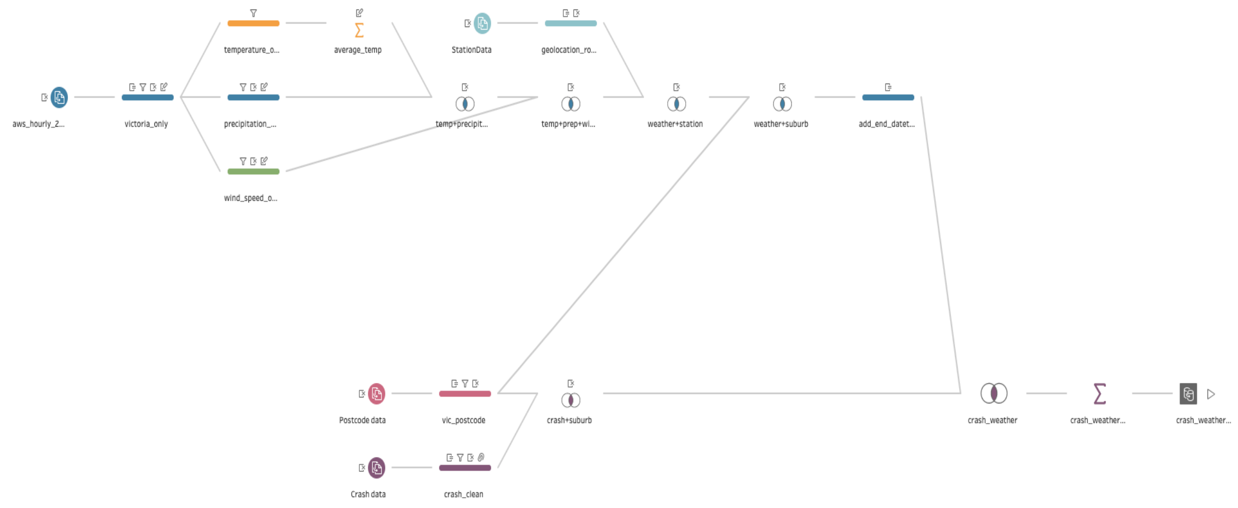


Figure 1 Prep Flow

Step 1: (Cleaning Crash Data)

- Select relevant fields from crash data (ACCIDENT_NO, ACCIDENT_DATE, ACCIDENT_TIME, ALCOHOLTIME, DAY_OF_WEEK, LIGHT_CONDITION, ROAD_GEOMETRY, SEVERITY, SPEED_ZONE, LONGITUDE, LATITUDE, OLD_DRIVER, YOUNG_DRIVER, UNLICENSED)
- Cleaning:
    1. Filter based on Accident Date (2017/05/01 to 2018/04/30)
    2. Create new field accident_datetime_utc by using accident date and accident time
    3. Remove accident date and accident time fields
    4. Create new field accident_datetime_epoch by converting utc datetime into epoch time
    5. Group light condition values (dark no street lights, dark street lights off, dark street lights unknown) into single value as dark
    6. Round latitude and longitude values

7. Exclude any records with ‘null’ value for day of week field or ‘unk.’ for light condition, ‘Camping grounds or off road’, ‘Other speed limit’ and ‘Not known’ for speed zone
8. Create new field driver age based on old driver and young driver fields
9. Remove old driver and young driver fields

Step 2: (Cleaning Postcode Data)

- Select relevant fields from postcode data (suburb, state, lat, lon)
- Cleaning:
    1. Round lat and lon value to create fields as latitude and longitude respectively
    2. Remove lat and lon fields
    3. Filter the value with state as ‘VIC’ only

Step 3: (Cleaning Weather Data)

- Select relevant fields from weather data (area_code, valid_start, station_number, parameter, value)
- Cleaning Part 1:
    1. Filter to select ‘VIC’ only using area code field starts with ‘VIC’
    2. Create new field state with value ‘VIC’
    3. Remove Area code field
    4. Create new field start_datetime_utc by converting valid_start field to utc time
    5. Rename valid start to stat_datetime_epoch
- Cleaning Part 2:
    1. Filter the result of cleaning part 1 to retrieve temperature record only with parameter to be either MaxT or MinT
    2. Group the result: Grouped fields (start_datetime_utc, start_datetime_epoch, state, station_number) and Aggregated fields (value - Average)
    3. Rename value field to temperature
- Cleaning Part 3:
    1. Filter the result of cleaning part 1 to retrieve precipitation record only with parameter equals ‘Precip’
    2. Rename value field to precipitation
    3. Remove parameter field
- Cleaning Part 4:
    1. Filter the result of cleaning part 1 to retrieve wind speed record only with parameter equals ‘MaxWindMag’
    2. Rename value field to wind_speed
    3. Remove parameter field
- Merging Weather Part 1:
    1. Inner Join result of cleaning part 1 and cleaning part 2 on start_datetime_utc, state and station number; Remove irrelevant fields

2. Inner Join result of cleaning part 3 and merging weather part 1 (1) on start_datetime_utc, state and station number; Remove irrelevant fields

Step 4: (Cleaning Weather Station Data)

- Select relevant fields from weather station data (latitude, longitude, station_number)
- Cleaning:
    1. Round latitude and longitude fields

Step 5: (Integrate Weather and Weather Station)

- Inner join output of step 3 and step 4 on station_number
- Remove irrelevant and duplicated fields

Step 6: (Integrate Weather, Weather Station and Suburb)

- Inner join output of step 5 and step 2 on latitude and longitude
- Remove irrelevant and duplicated fields
- Create new field end_datetime_epoch by adding 1 hour to start_datetime_epoch

Step 7: (Integrate Crash and Suburb)

- Inner join output of step 1 and step 2 on latitude and longitude
- Remove irrelevant and duplicated fields

Step 8: (Integrate Crash, Weather and Suburb)

- Inner join output of step 6 and step 7 on suburb, accident_datetime_epoch>=start_datetime_epoch and accident_datetime_epoch<end_datetime_epoch
- Grouping : Grouped Fields (driver_age, accident_datetime_epoch, accident_datetime_utc, ACCIDENT_NO, ALCOHOL_TIME, DAY_OF_WEEK, LIGHT_CONDITION, ROAD_GEOMETRY, SEVERITY, SPEED_ZONE, LONGITUDE, LATITUDE, UNLICENSED) and Aggregated Fields (precipitation - Average, temperature - Average, wind_speed - Average)

Step 9: (Output)

- Generate the output file either .csv or Tableau Data Extract (.hyper) file

3.2.2 Integration of Street in Accident Data

This process is to find out the street name where the accident took place along with the suburb.

The first attempt was to find an intersection in the prepared integrated crash weather data with line geometry provided by the geocoded Victorian street data. But, this attempt was unsuccessful and no intersection could be detected in Tableau Desktop.

The next attempt was to utilize Google Cloud Services Reverse Geocoding API. This api returns the nearest location if geo coordinates are provided as input. The api format is:

https://maps.googleapis.com/maps/api/geocode/json?latlng=<latitude>,<longitude>&key=<api_key>

Important Changes Needed in the Code

geocoding.py

- On line 22, uncomment the line and add your own google api key

reverse_geocode_crash_data.py

- On line 12 and 13, change the path

Running the Code

Install all the requirements

```
pip install -r requirements.txt
```

Run the program

```
python reverse_geocode_crash_data.py
```

3.2.3 Integration of Accident Data and Suburb in Alarm Data

For this process, we will utilize a new tool called QGIS (ArcGIS and other GIS tools will also be suitable, but this process was done on QGIS).

Step 1: Setting up software

- Install QGIS from QGIS Website. Standalone or OSGeo4W, both variants would be suitable for this process.
- Copy various shapefiles, alarm data and accident data to a suitable folder.

Step 2: Integrating the data

- Drag and drop the Suburb Data shapefile into QGIS, select the *WGS 84* coordinate system if asked.

- From the *File toolbar*, select *Layer* and go to *Add Layer -> Add delimited text layer*.
- Select the Alarm Data csv or xlsx file and under geometry definition, select *Point Coordinates* and select *X Field* as Longitude and *Y field* as Latitude. Also set *the geometry CRS* as *WGS 84* if needed.
- From the same *Add layer* option, select the Accident Data csv file, and use the same settings.
- From the layer toolbar (on the left side of window), select the features required and they would be overlayed over each other, thus displaying the suburbs, alarm, and accident data on one map.
- To add symbology, right click on any layer -> *Properties*. Then select the type of symbology from symbology section, and to display labels, select the *Labels* option from same window and choose the Data field that you want to be displayed.

Step 3: Output the data

Data File

- Although this part will not be needed at this stage, to get an integrated dataset, go to *File toolbar -> Vector.*
- In *Vector* first select *Data Management tools -> Create Spatial Index* and select the appropriate layer and click *Run.* This step speeds up the next step.
- Now, from the same menu in *Data Management tools*, select *Join Attributes by Location*. Select the layers that you want to be joined, Accident Data, Alarm Data, Suburb Polygon, or all three. After this step, a new layer would be created.
- To export this newly created layer into a different format, Right Click on the layer and select *Export -> Save Features As*. Select the appropriate file format and click *OK*.

Image File (for Map)

- From the File toolbar, select *Project -> New Print Layout*. Give a name to the print layout.
- A new Print Layout window would appear. To add the current map into the file, select *Add Item* from File toolbar and select *Add Map*.
- Now drag the selection to the area on empty space, and an Image of current map would be created on that. From the *Add Item* various other features can be added, such as a legend, another map (zoomed in or zoomed out version), labels and pictures.
- When done, Select *Layout* from File toolbar and choose the preferred format to export the map.

To get Accident Frequency data from Crash Data

- Add the Crash Data layer to QGIS, go to *Vector-> Analysis Tools -> Count Points in Polygon*.
- After the above step, a new window would appear. Select the Suburb Data as Polygon and the Crash Data as Point based data. This function would then calculate all the points in one polygon or suburb and add a corresponding column to the Suburb Data based on that

calculation. With this data, we can get Heat map of suburbs based on their accident frequency.

3.3 Data Analysis Requirements and Approach

3.3.1 Analysis of Accident Data based on Accident Frequency

This analysis is performed to view insights on how different factors correlate with the occurrence of accidents. This analysis is performed using Tableau Desktop.

Tableau Desktop allows people to make data-driven decisions with confidence, by helping them answer questions more quickly, solve harder problems more easily, and uncover new insights more frequently.

With a couple of clicks, Tableau Desktop connects directly to hundreds of data sources, both on-premises or in the cloud, making it easier to start analysis. Interactive dashboards, drag and drop functionality, and natural language queries help people of all skill levels quickly discover actionable insights, all from an intuitive and visual interface. Ask deeper questions by quickly building powerful calculations, adding trend lines and seeing statistical summaries, or clustering data to see relationships. Better understand your data, make new discoveries, and identify opportunities faster with Tableau Desktop.

3.3.2 Analysis of Accident Data based on Accident Severity

This analysis is performed to view insights on the factors that impact the severity in case of occurrence of accident. This analysis is based on the correlation of accident severity with different factors. The factors that are considered for this analysis are:

- Weather condition
- Rainfall
- Light condition
- Age of driver
- Road geometry
- Speed zone
- Time of day
- Day of week
- Alcohol consumption by driver

This analysis is performed using python programming on Google Colabs. The link to the colabs is:

https://colab.research.google.com/drive/1jtFBiQtZNmcgirxOHGFjJjYuHF-21Ksg?usp=sharing

3.3.3 Analysis of Accident Data w.r.t Directed's Alarm Data

This analysis provides the information about hotspot areas of accidents, with respect to Directed's alarm data. It also shows how certain points on roads have multiple instances of Over speeding, Harsh braking and Harsh acceleration.

This analysis takes data from 3.3.2, based on Accident Frequency and integrates it into Suburbs.

In QGIS software, when all the relevant layers have been added, such as Alarm Data, Accident Data, Suburb Polygon Data, and speed limit/road data, we can proceed to integrate them for analysis.

- Get Accident Frequency data layer from 3.2.3. This data provides information about frequency of accidents per suburb. Now, doing the same process as in 3.2.3 of *Counting Points in Polygon*, select the Accident Frequency data layer as polygon file, and the Alarm data as points file.
- From the setting give weightage to the type of analysis needed, for example, if counting the number of harsh acceleration events in certain suburbs, give weightage to ACC attribute.
- This same process can be followed for all attributes OSP and BRK. The final output will provide three different layers each for OSP, BRK and ACC containing the information regarding frequency of each event per suburb with respect to the accident frequency.

4. Findings

4.1 Accident Data

4.1.1 Overall Summary

For the Traffic Accident events data, we try to find relationship between traffic accidents with weather (in term of Temperature, Precipitation, Windspeed.), road condition (in term of Light condition, Speed zone condition, Geometry of Road.), time (in term of month , day, hours), and location (suburb and street).

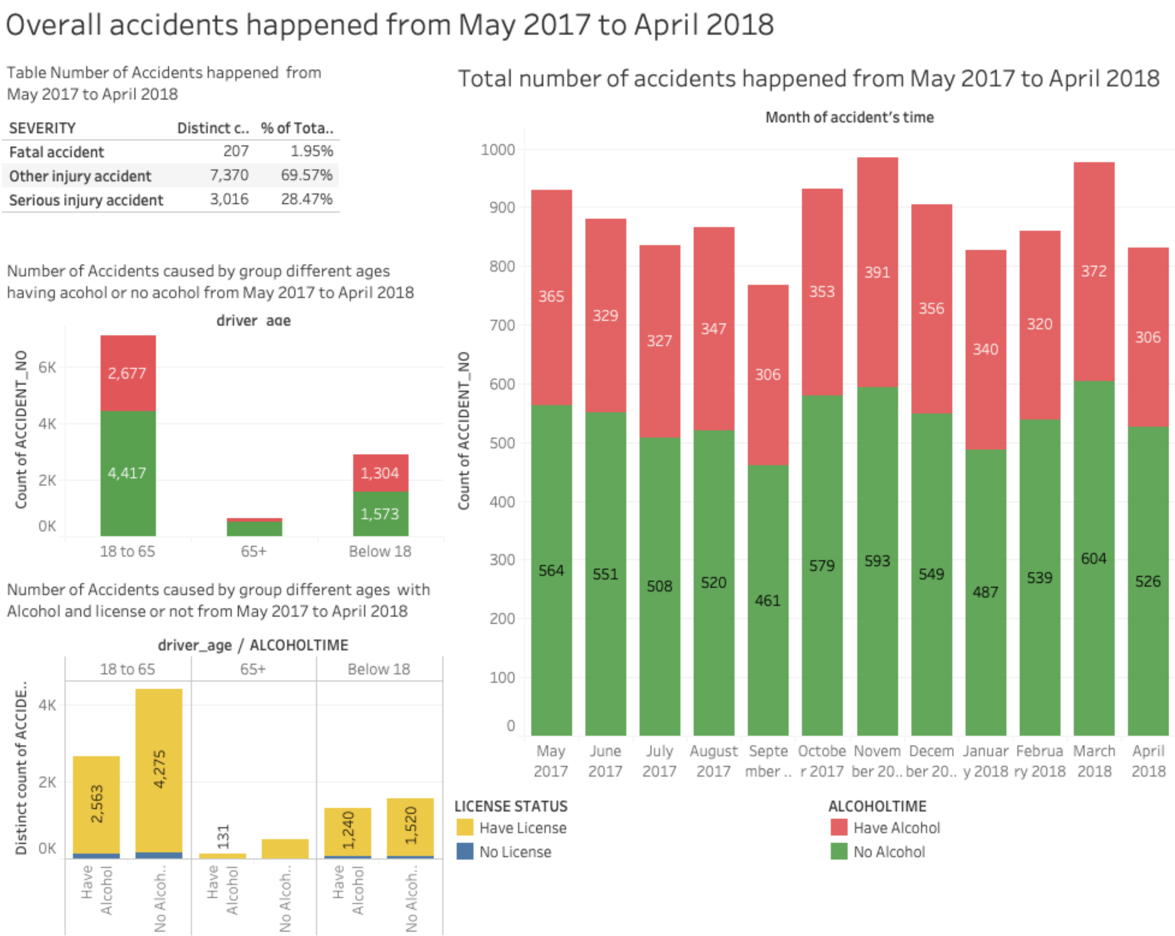


Figure 2 Accident Analysis

Overall, from May 2017 to April 2018, Victoria had a total of 207 fatal accidents, accounting 1.95%, 3016 serious injury accidents with 28.47% and 7370 other injury accidents. It can be observed that around 500 to 600 accidents happen each month and nearly 40% of these accidents are caused by drivers with alcohol consumption. Moreover, most of them belonged to the 18 to 65 age group and possess a license.

4.1.2 Weather Conditions

Figure 3 Weather Conditions Analysis

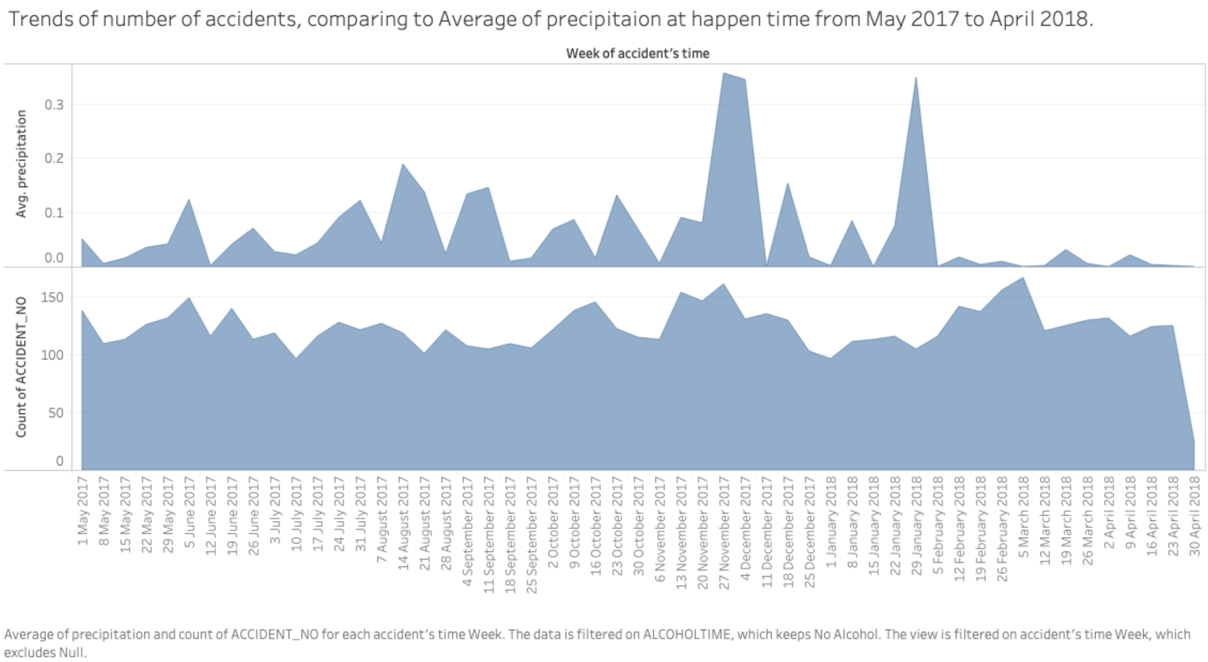


Figure 4 Accident analysis w.r.t Precipitation

Regarding weather conditions, no relationship was found between accidents with the temperature and wind speed at the time when accidents occur. However, for the precipitations, it has the little same shape on chart with several accidents, so it seems that if the day is rainy and higher precipitations, it will have more accidents on this day. However, it's not strong links, and if we want to support this result, we need to analyse more data, at least 5 years data. In this project, we just use data over a single year.

4.1.3 Relationship of Accident Occurrence with Road Conditions

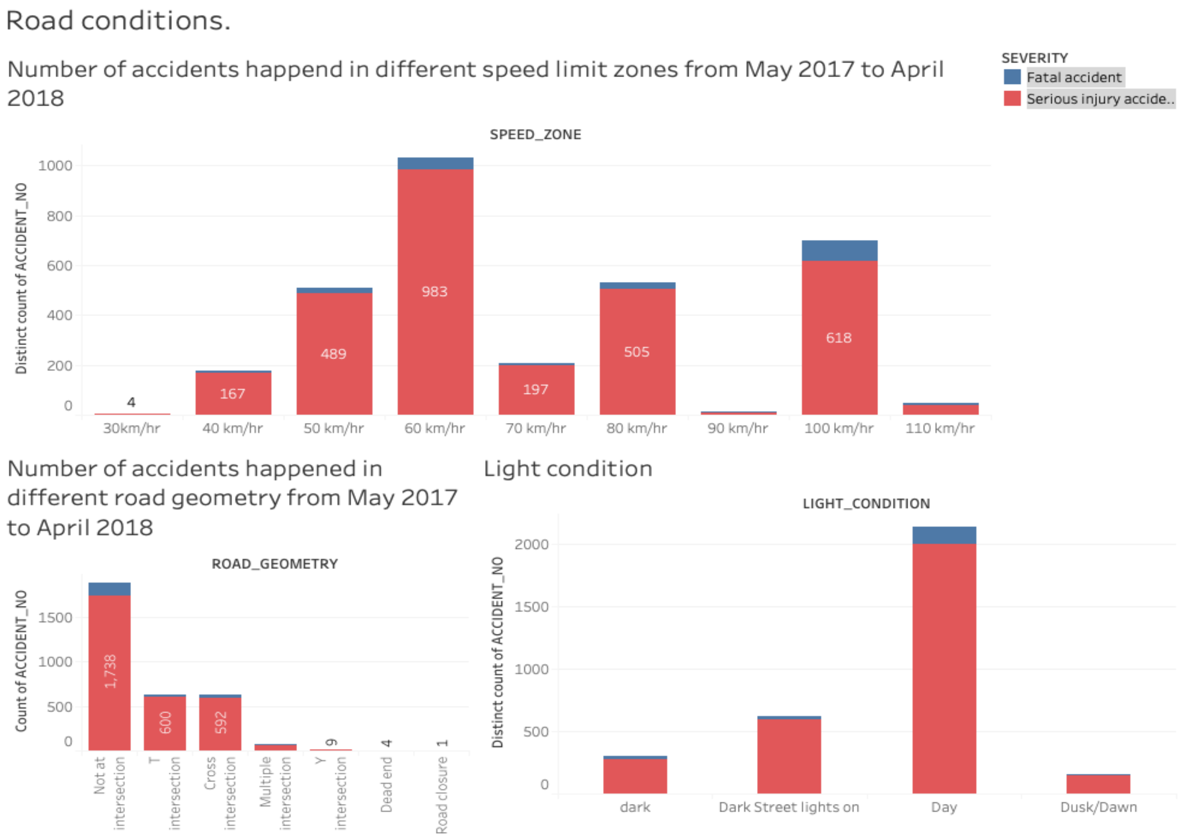


Figure 5 Accident analysis w.r.t Road Conditions

Most accidents tend to happen in 60km/h and 100 km/h limited speed zones and most fatal accidents happen in 100 speed limit zones. Compared to different road geometry, 1,738 accidents happened on road without intersection; nearly 600 accidents happened at T and Cross intersection. It is clear that most accidents happen in daylight because of the high volume of traffic. However, another interesting insight is that the number of accidents happening during the condition 'dark with streetlights on' is higher than the ones with dark and dusk/dawn.

4.1.4 Traffic Volume - Accidents in different time

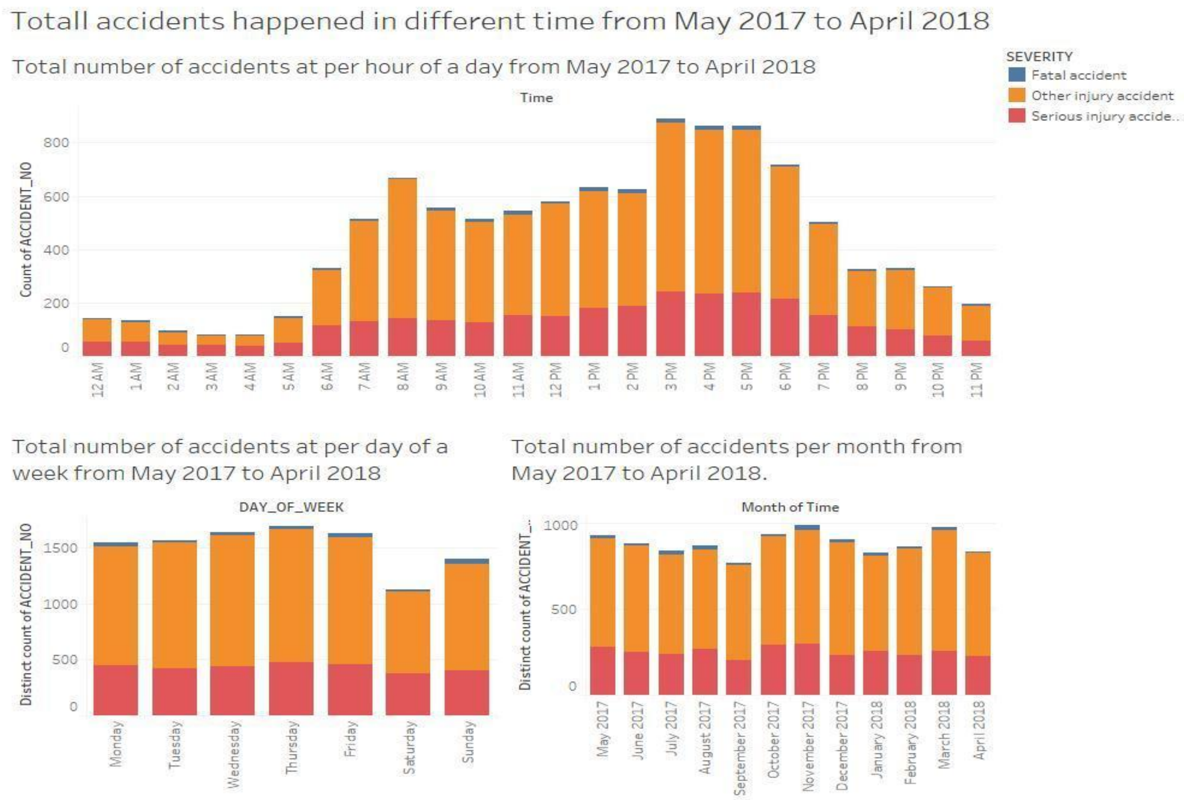
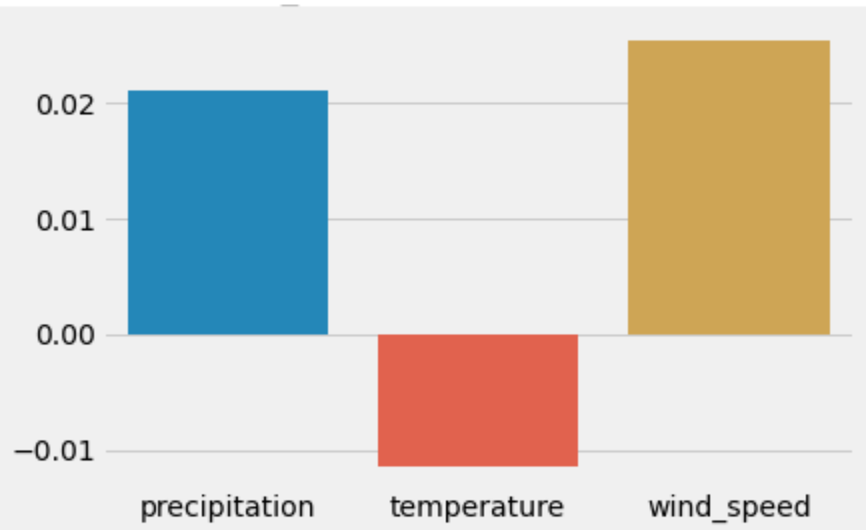


Figure 6 Accident Analysis w.r.t Time of Day

It is very clear that the total number of accidents in peak time is highest, from 3pm to 6pm and 8am. The numbers of accidents on Saturday and Sunday are relatively lower than other days. The reason is that if the traffic volume is high, the number of accidents will be high.

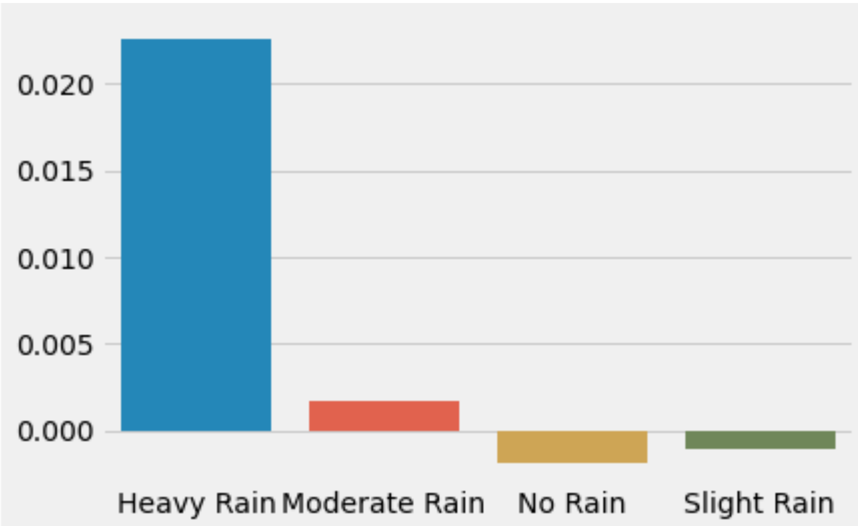
4.1.5 Accident Severity

Weather Condition vs Accident Severity



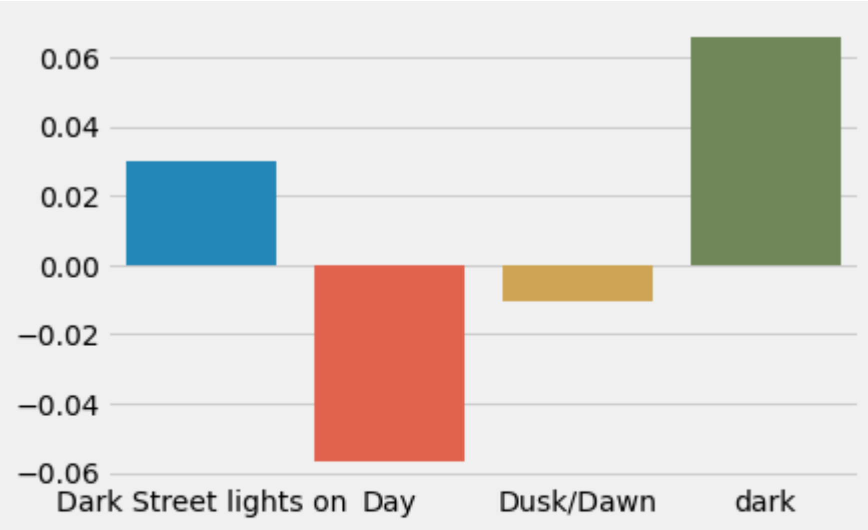
This shows that the accidents are more severe during high precipitation or high wind speed.

Rainfall vs Accident Severity



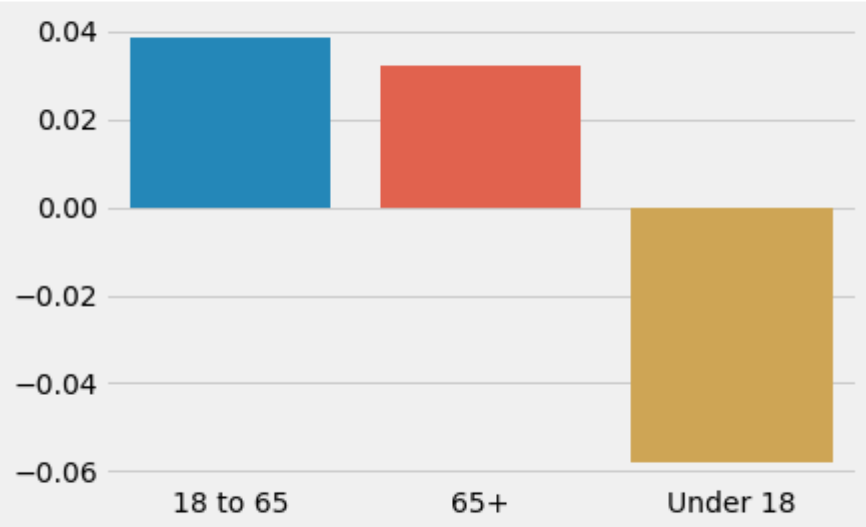
This shows that the accidents are more severe during heavy rainfall. During moderate rain, slight rain or no rain, the accident severity is very low.

Light Condition vs Accident Severity



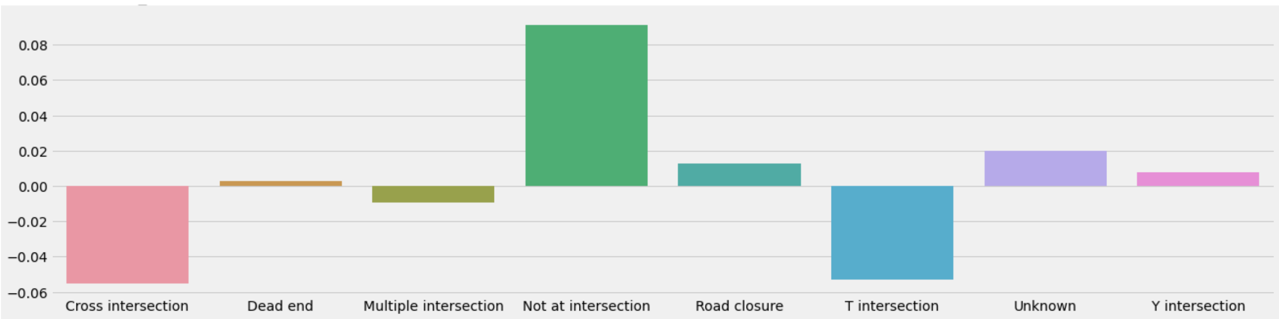
During the dark condition, the accident severity is high. In addition, in daytime the correlation is negative, that means the accident severity is very low.

Driver Age vs Accident Severity



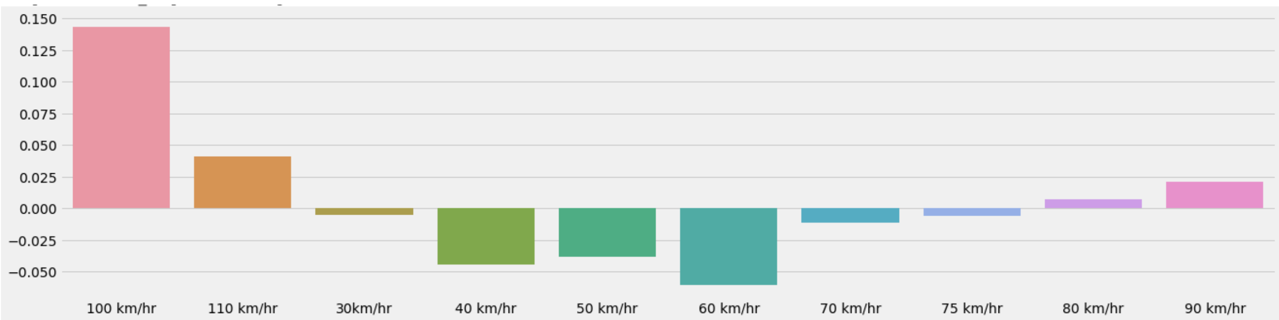
The accident severity is high for the accidents involving drivers over the age of 18. This may be since under 18 drivers are generally supervised.

Road Geometry vs Accident Severity



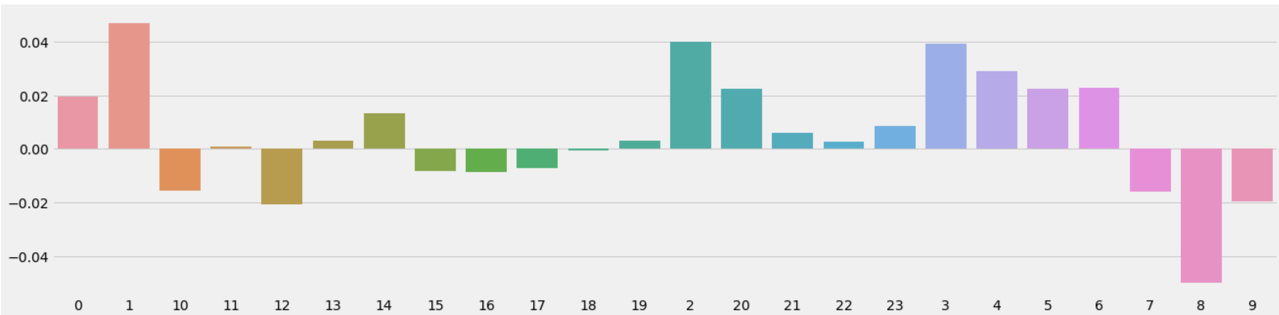
The accidents that occurred at roads without intersection are directly related to the accident severity. This is because of the high speed at non-intersections.

Speed Zone vs Accident Severity



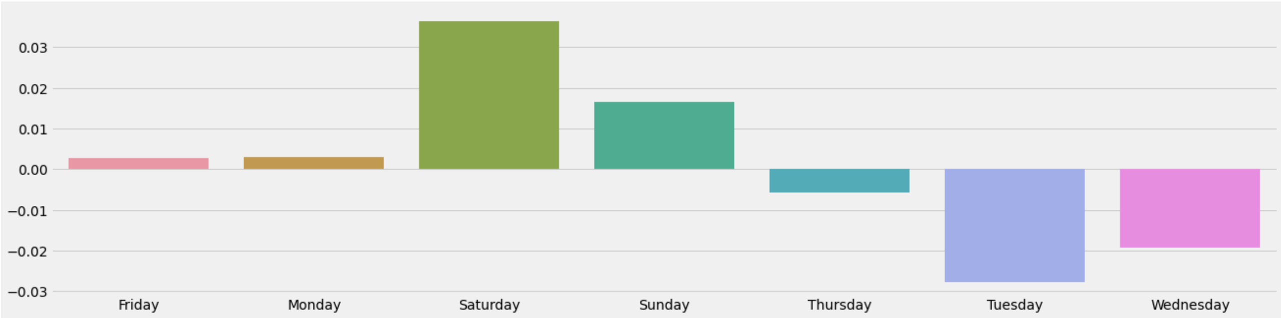
This shows that accidents that occur at high speed areas above 80 km/hr are more severe.

Time of Day vs Accident Severity



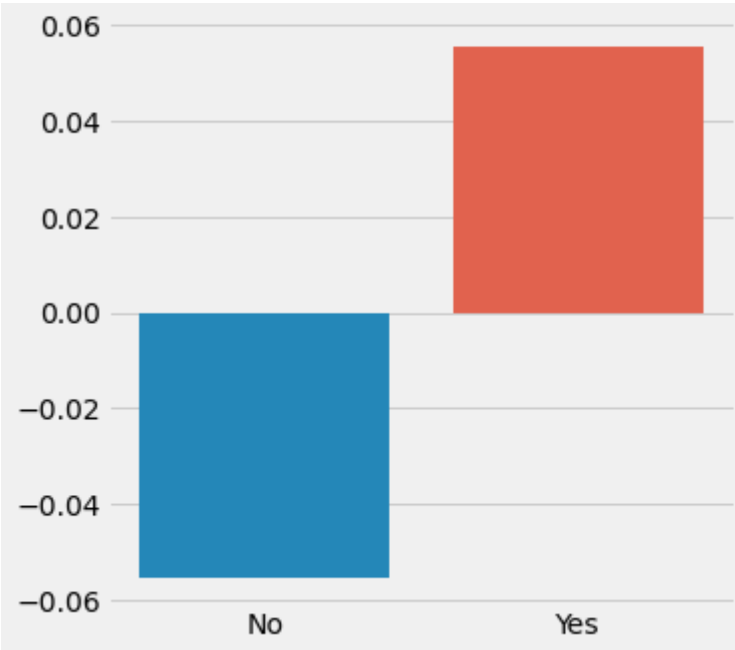
Severe accidents occur during non-peak hours generally between midnight to 6 in the morning.

Day of Week vs Accident Severity



This shows that the accidents are more severe during weekends.

Alcohol Consumption vs Accident Severity



This shows that the severe accidents take place when the driver has consumed alcohol.

4.1.6 Location: Street

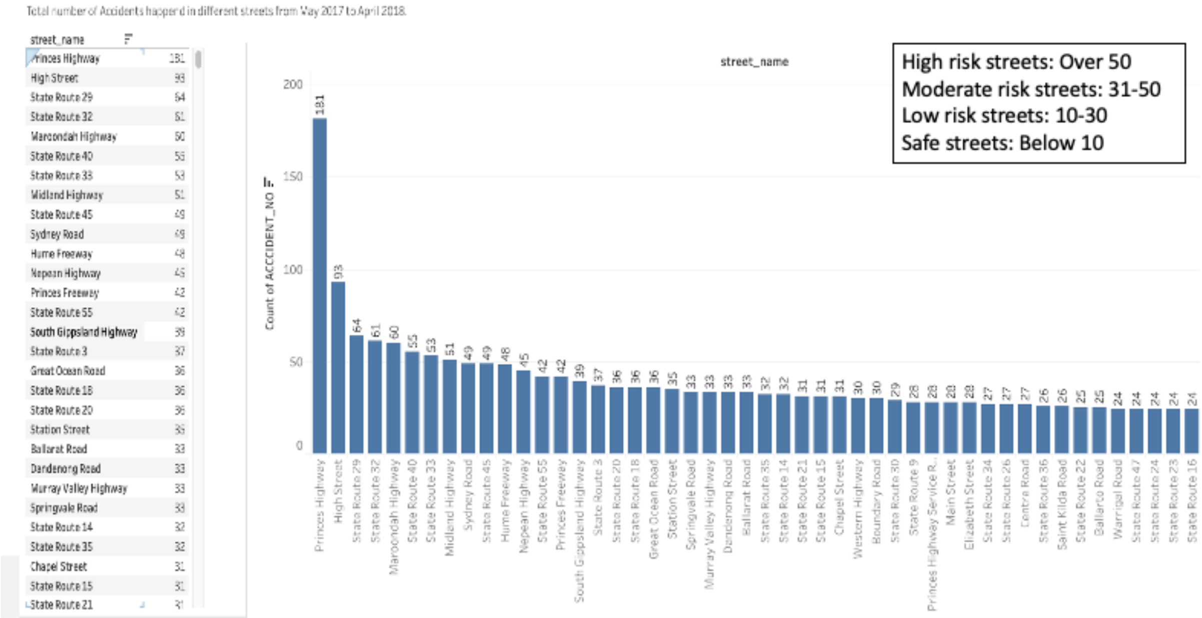


Figure 7 Roads with Highest Accident occurrence

Based on the location of traffic accidents, we can define the topmost dangerous street in Victoria. After that the streets are divided into 4 groups: high risk street (if this street has more than 50 accidents in this particular period reflected by data time frame), moderate risk street (for total number around 31-50), low risk street (10-30 accidents) and safe street (below 10 accidents).

Moreover, have a look at the topmost dangerous streets, we can see most of them are on highways. However, because we do not have data about the type of road, we do not know exactly what type of road, for example freeway, highway and so on, have more accidents than others.

4.1.7 Location: Suburb

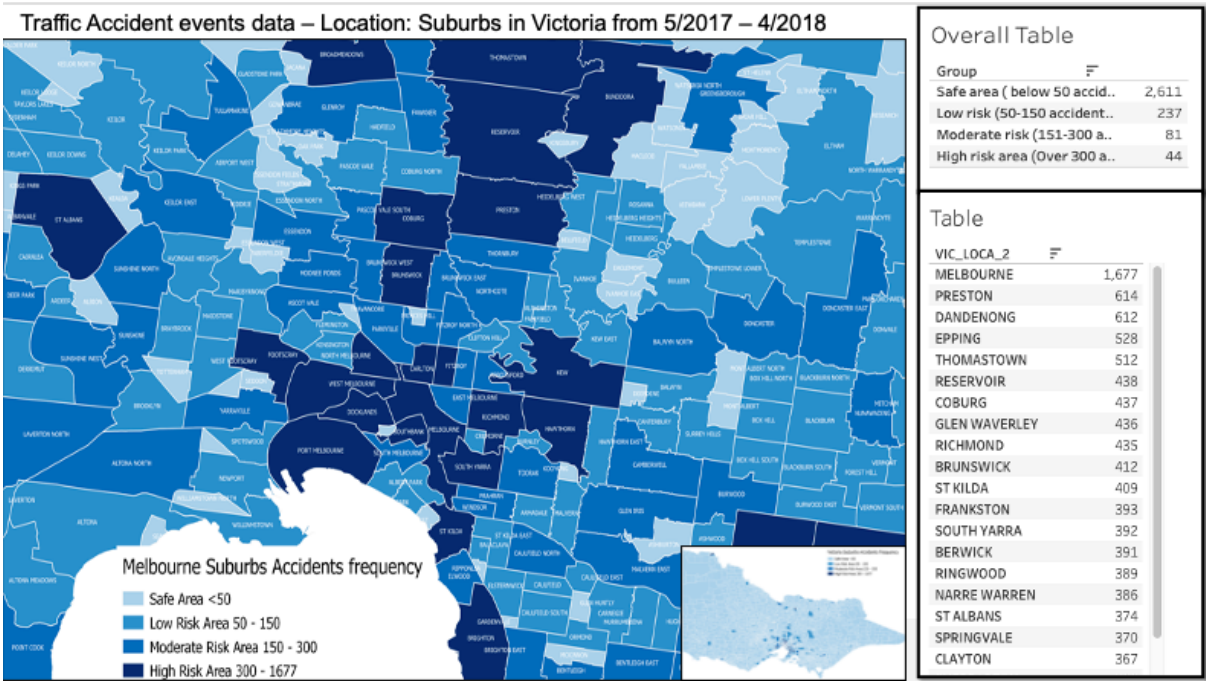


Figure 8 Melbourne Suburb Accident Frequency Map

The same way with location with street, we also define the total accidents that happen in all suburbs in Victoria and divide them into 4 groups: Safe Area < 50, Low Risk Area 50- 150, Moderate risk Area 150-300 and high risk Area with over 300. For the hotspot areas, there are 44 suburbs in the high-risk area with more than 300 accidents. And we see all of them are suburbs having high populations in Victoria, so again, it supports the relationship between traffic volume and accidents above.

4.2 Event Alarm Data

4.2.1 Time

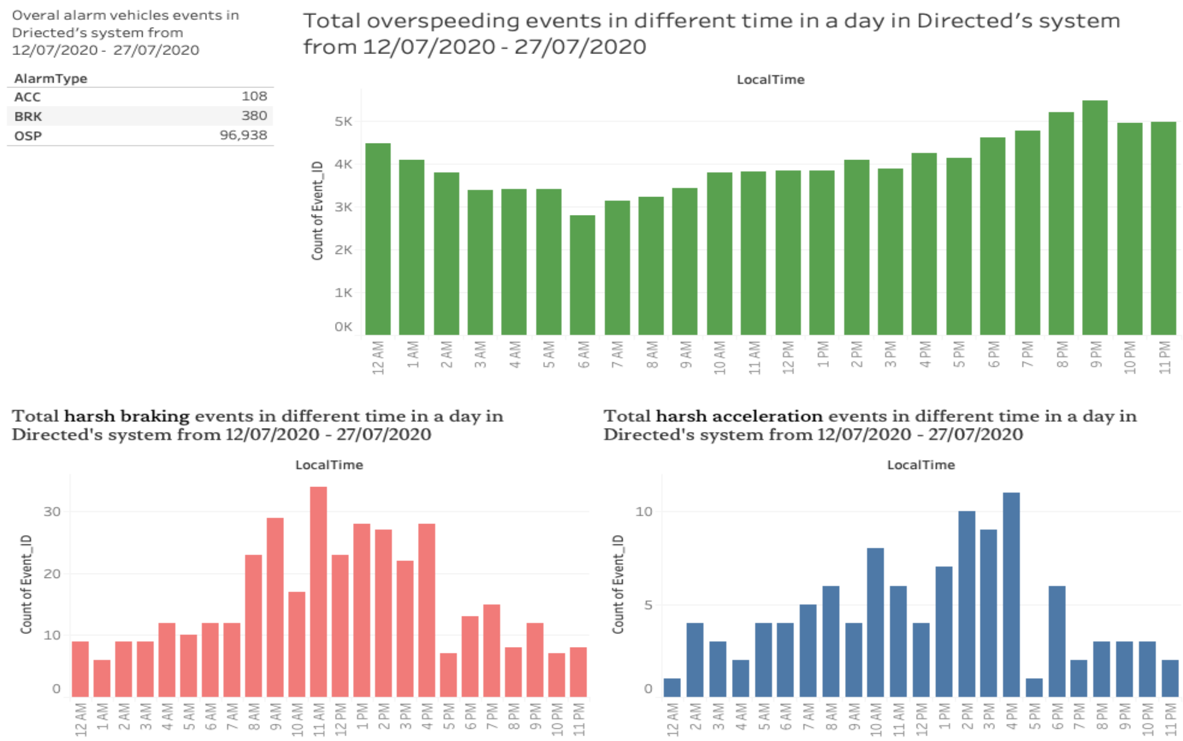


Figure 9 Alarm Data analysis w.r.t Time of Day

Look at this table, sample data have 108 ACC (harsh acceleration events), 380 BRK (harsh braking events), and near 100,000 over speeding events. So, in terms of this, we just focus on speeding events, because other events do not have enough data to analyse.

For over speeding, there are many reasons why drivers tend to overspeed, they may enjoy driving fast, may be in a hurry, depending on their gender, their age. The green chart shows that people tend to overspeed at night from 7pm to midnight.

4.2.2 Location: Street

Table

street_name	
Hume Freeway	3,712
Western Freeway	1,067
Princes Highway	1,016
Western Highway	569
Calder Freeway	351
Princes Freeway	214
Hume Highway	208
Western Ring Road	202
Broadford-Wandong Road	179
Marchbanks Road	177
Geelong Ring Road	145
Old Western Highway	120
Whiteside Street	118
Pentland Hills Road	112
Cobden-Stonyford Road	111
Lofven Street	98
Bruthen-Nowa Nowa Road	96
Zenith Circuit	91
Midland Highway	89
Northern Highway	88
Western Ring Path	85
Haugh Street	81
Craigieburn Bypass	80
Fullarton Road	71
Jefferies Road	68
Great Alpine Road	67
Bourkes Road	64
Holts Lane	63
Wicklow Drive	61
Dartmoor-Hamilton Road	55
Maroondah Highway	53
Brooklyn Road	51
Orbis Drive	50
Calder Highway	49
Deverall Road	49
Dicksons Road	49

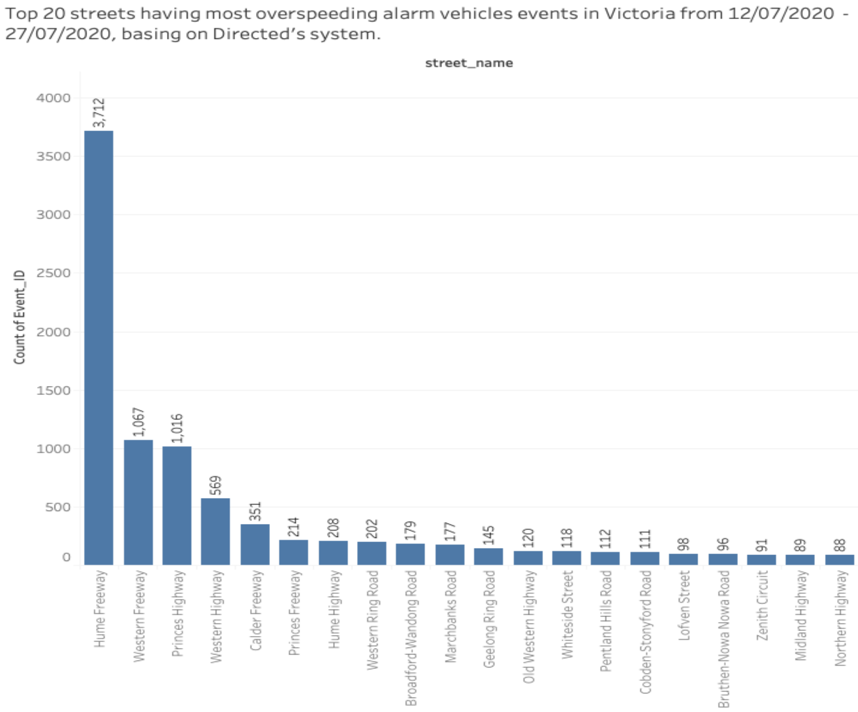


Figure 10 No. Alarm Data events w.r.t Roads

Based on the location of Vehicle alarm events data, we can define the top street in Victoria that people tend to overspeed.

Have a look at top streets having the most over speeding event, we can see most of them also are on highways and freeways. However, the same with accidents data, because we don't have data about the type of road, so we don't know exactly what type of road , for example freeway, highway and so on, have more drivers that tend to overspeed.

4.2.3 Location: Suburb

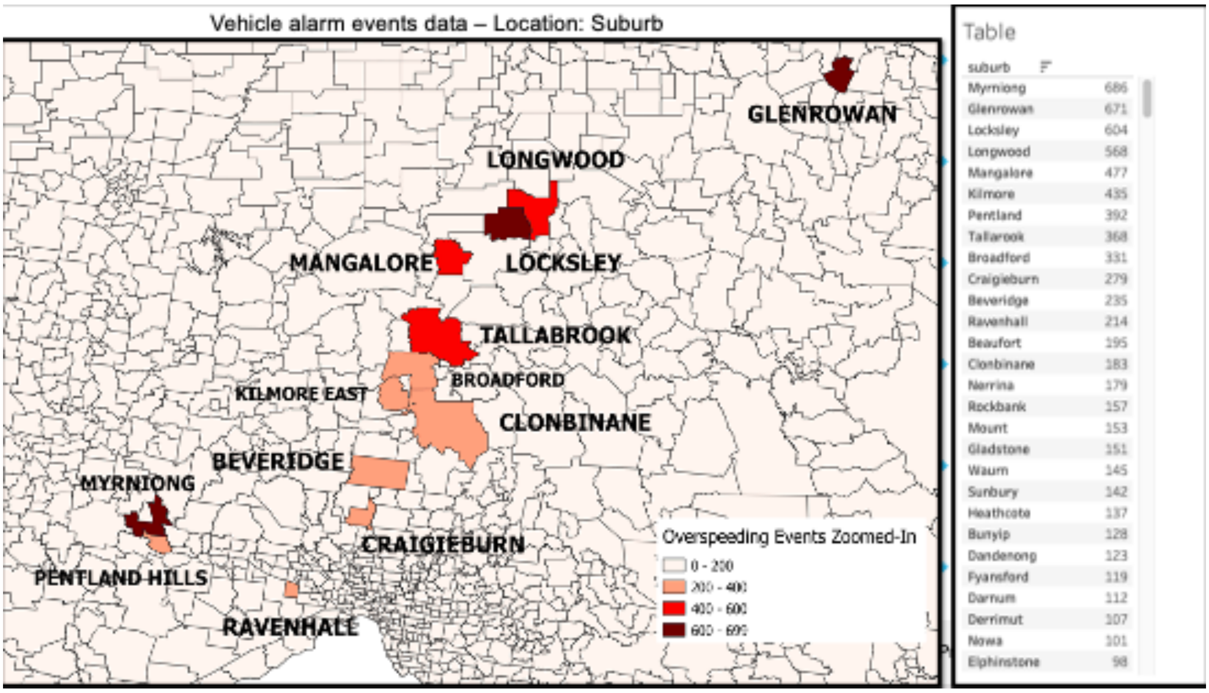


Figure 11 Highest over speeding events w.r.t Suburbs

Over speeding has many risks to cause an accident. We also define the total over speed events happening in all suburbs in Victoria and divide them in this term of research to 4 groups: Suburb have below 200 events , second group is from 200-400 events, third group from 400-600, last group in over 600 events like the maps. Then we can know which suburbs are more dangerous based on the total number of over speeding events.

5. Proposed Architecture

Based on the analysis results, we propose the architecture as shown in figure below to gain our objectives to improve the current system.

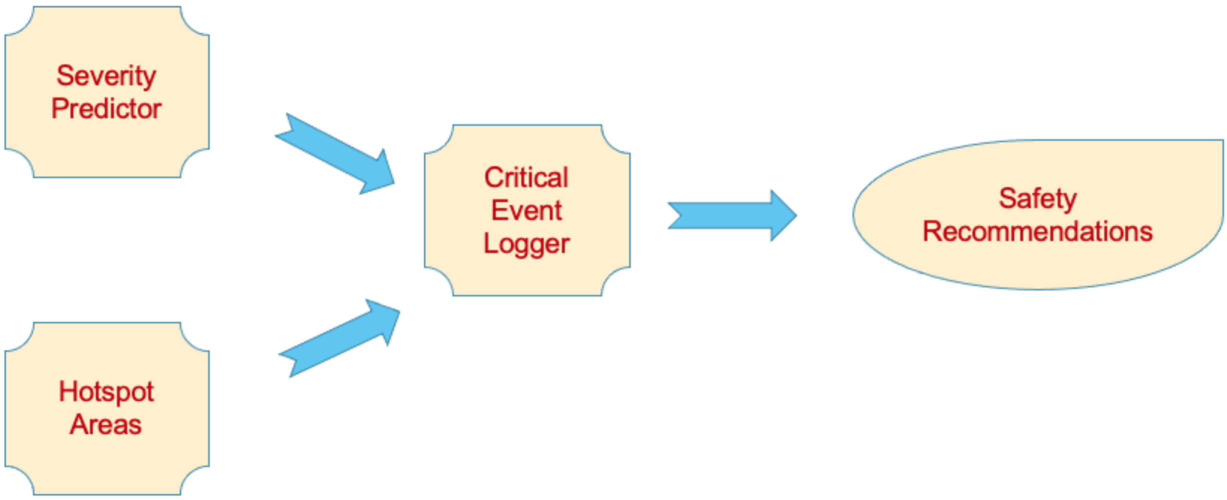


Figure 12 Proposed Architecture

The proposed architecture consists of 4 modules:

- Critical Event Logger (Current System)
- Severity Score Predictor (Based on the environmental conditions and situational factors, predicts the severity in case of occurrence of accident.)
- Hotspot Area Categorizer (Based on the historical accident frequency data, this module provides the area category in which the driver is currently driving.)
- Safety Instruction Recommender (Based on the severity score and area category, recommend course of safety instructions including speed, hazardous actions, etc to the driver)

Example Scenario:

Consider a driver driving on Princess Highway in Dandenong suburb when raining heavily. On the way, the driver triggers harsh acceleration followed by harsh braking within a few minutes.

Based on the severity analysis, the current environmental situation generates a high severity score. Similarly, the current area belongs to a high-risk zone. Considering these factors, the critical events performed by the driver may lead to serious accidents. So, the system should provide instruction to the driver to drive slowly and try not to brake or accelerate harshly that may cause the accident to occur and if it occurs there is a high possibility of fatality.

5.1 Severity Score Predictor

Severity score predictor is to be implemented using machine learning. A ML model is to be trained using the available data. The input factors would be light condition, rainfall, windspeed, day of week, time of day, speed zone, road geometry and driver age. The model should be trained to generate a single output i.e. severity score that ranges from 0 to 1 (0 being less severe and 1 being fatal).

Based on the requirements and data, the candidate algorithm to be used for training a model are:

-Multiple Linear Regression

From our analysis, we obtain 8 features and with more data, this number of features can issue the processing capability constraints. In such a case, the above algorithm can be combined with dimensionality reduction algorithms like PCA to only select the principal features automatically.

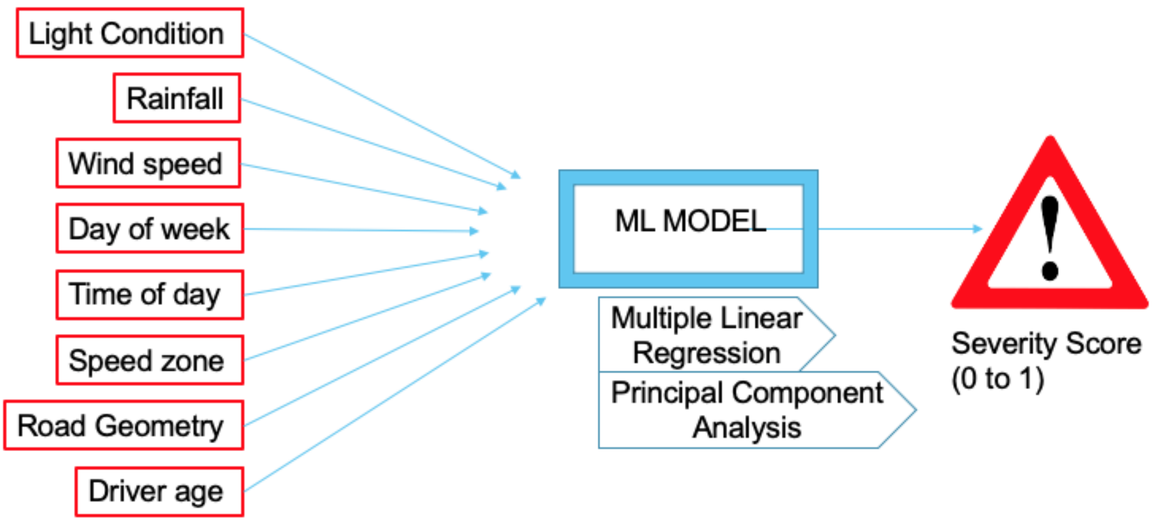


Figure 13 Severity score predictor

5.2 Hotspot Area Categorizer

Hotspot area categorization is to categorize the location based on accident frequency and critical events frequency in the past. Based on these factors, the module categorizes any location into 4 risk areas:

- Area 1: Low Risk
- Area 2: Moderate Risk
- Area 3: High Risk
- Area 4: Very High Risk

This module should implement the self-adaption capability with category update when new data are available.

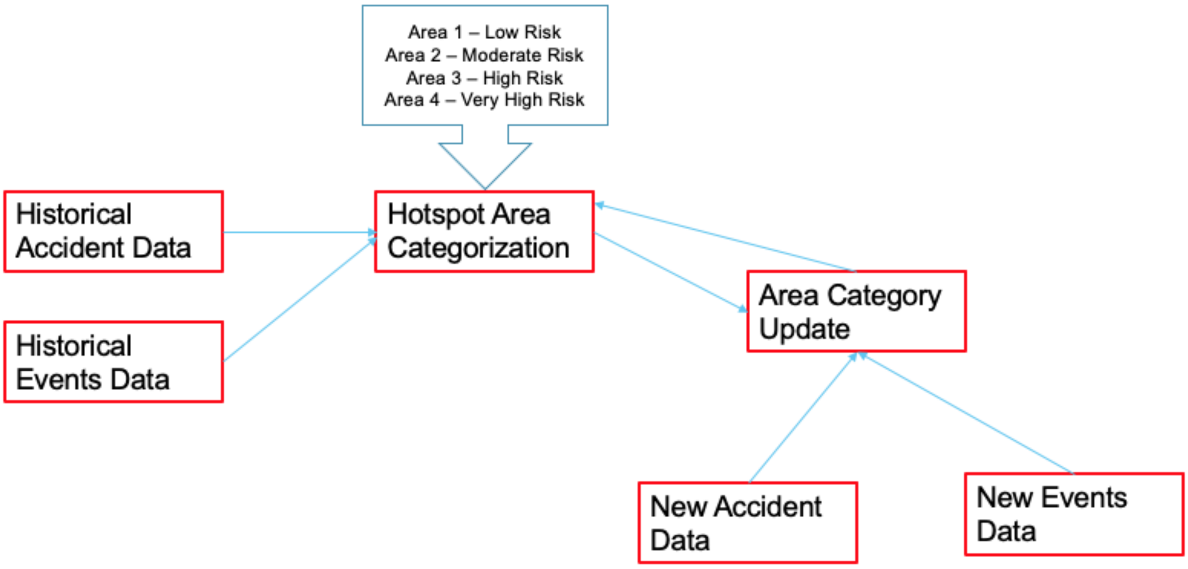


Figure 14 Hotspot Area Categorizer

6. Limitations and Conclusions

During these 12 weeks internship project, the team faced many challenges that hinders the team's ability to perform at their best level. Because of the limitations imposed during the work, the scope of this project is limited to the improved system architecture proposal only; restricting the team to work on building the improved system and test it.

The major limitations are as follows:

1. **Resource Constraint:**

   The project members use their own personal computers during the entire project. Due to limited processing capabilities in the resource, it takes a lot of time for data cleaning and preparation steps when using Tableau Prep Builder. Similarly, a significant amount of time must be spent during the analysis process as well when using Tableau Desktop.

2. **Data Constraint:**

   The analysis performed and the findings obtained during this project is based on the accident data of a year (from May 2017 to April 2018) and the event alarm data of approximately a month (July 2020). The amount of data used is relatively less to base our findings to be completely valid.

3. **Data Time Frame Mismatch:**

   The time frame of the data acquisition of accident data and event alarm data is different. Because of this, the team is unable to perform any analysis based on the correlation between these two data sources.

4. **Time Constraint:**

   The overall internship duration being only 12 weeks, out of which only 8 weeks has been used towards the actual work. This duration is relatively low to achieve complete objectives as per the initial plan. This forces the team to rest the work and limits the scope to analysis and suggestions only.