2017

Accident Analysis for Victoria

REPORT ON IDENTIFICATION OF INFLUENCING FACTORS OF ACCIDENTS AND ITS MITIGATION

SACHIN PATHIYAN CHERUMANAL

Table of Contents

Introduction

Accidents have become a growing concern of every government. In an effort to reduce accidents, researchers have spent an enormous amount of time and resources trying to analyse accident and its major influencing factors. In this report, we shall use the accident data provided by the Victorian government as the source for performing further analysis that can be used to provide decision support to the stakeholders.

Description of Dataset

Dataset: Crash Stats - Data Extract (Crash Stats, 2014).
Data Source: Victoria Police
Description: The dataset contains accident data collected by the Police, between 2006 and 2017.

The data file has various datasets that give information about the accident, location, road surface condition, light condition atmospheric condition, vehicle-type and so on. Fig 1. gives all the relevant columns and the format of the data involved in the analysis.

Variable	Description
ACCIDENT_NO	Primary Key
ACCIDENTDATE	Date of accident in DD/MM/YYYY format.
LIGHT_CONDITION	'Day', 'Dusk/Dawn', 'Dark Street lights on' etc.
NO_PERSONS_KILLED	Number of persons killed in accident.
ROAD_GEOMETRY	Cross intersection, T intersection etc.
SEVERITY	1-Fatal accident, 2-Serious accident, 3-Other injury accident etc.
ATMOSPH_COND	'Clear', 'Strong winds', 'Smoke' etc.
SURFACE_COND	'Dry', 'Muddy', 'Wet' etc.
LATITUDE	Latitude
LONGITUDE	Longitude
VEHICLE_TYPE	Bicycle, Car, Taxi, Utility, Motor cycle etc
ACCIDENT_TYPE	Collision with vehicle, Collision with a fixed object, Struck Pedestrian etc.

Fig 1: Metadata of Crash Stats - Data Extract

The data was pre-processed, manipulated, explored and analysed using R, where the relationship between the different datasets was established through the primary key, ACCIDENT_NO. The relationship can be seen in Fig 2.

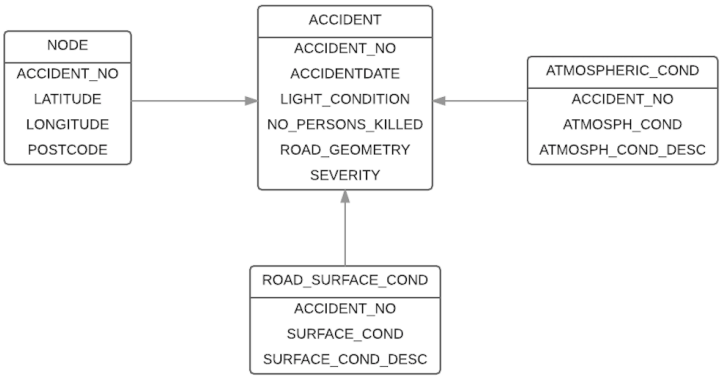

Fig 2: Data files involved in the dataset

Motivation for Analysis

The yearly accident count in Victoria is given in Fig 3. And the Fig 3, shows a linear regression showing the general trend of accident.

Fig 3: Year wise Accidents in Victoria

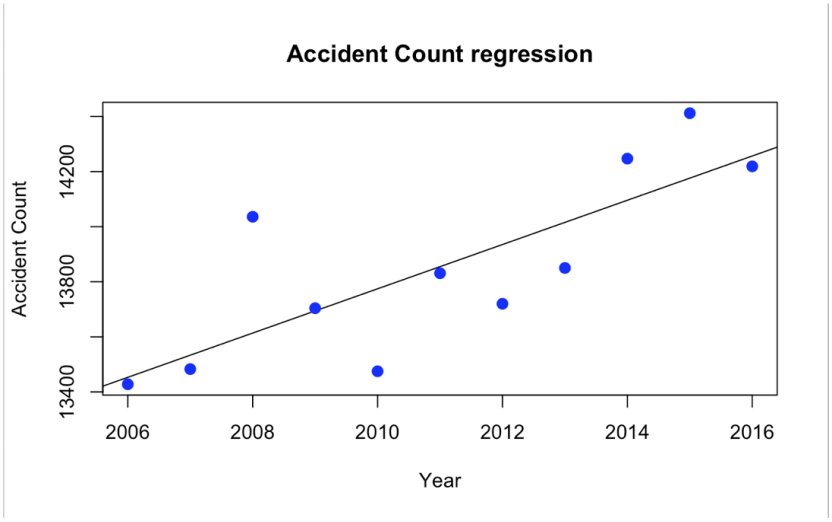

Fig 4: Trend line for Accidents in Victoria

The increasing trend in accidents is the motivation behind the accident analysis. In order to reduce this trend, we will have to identify postcodes with the highest number of accidents (based on various factors) and provide suggestions to mitigate accidents.

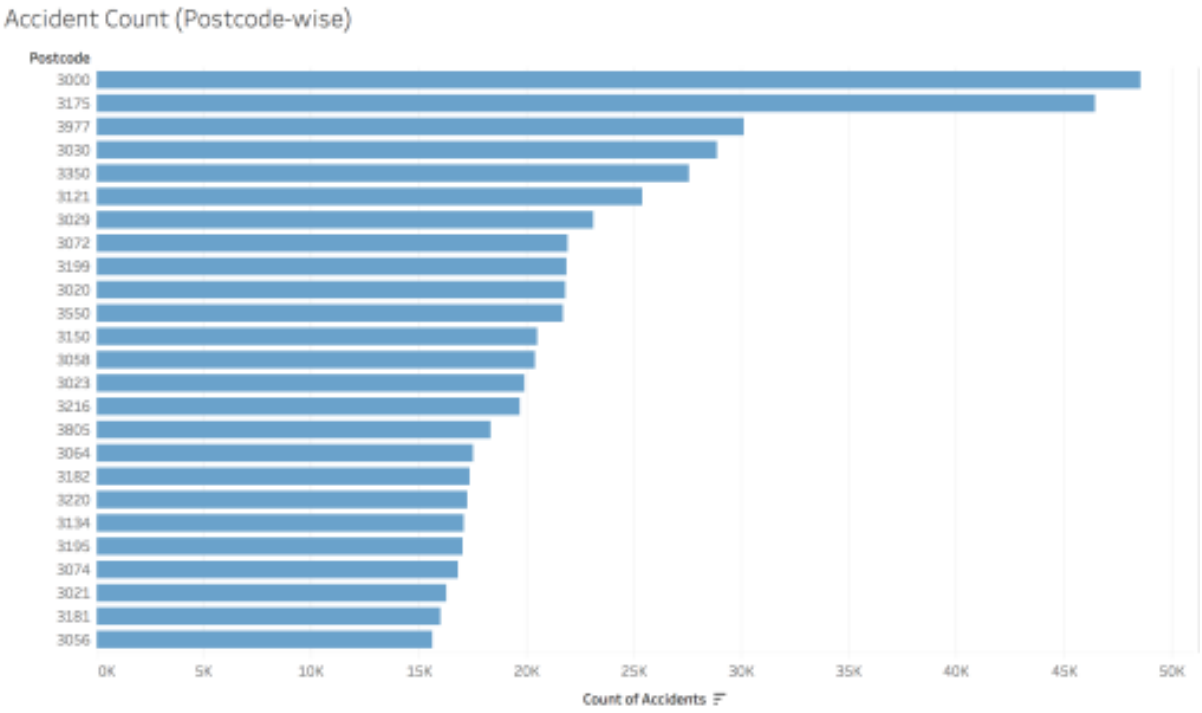

Fig 5: Overall Accident count of Top 25 Postcodes displayed above. Melbourne CBD (3000) the highest. Dandenong (3175) has the second highest accident count.

From Fig 5. it is understood that Melbourne CBD has the highest accidents, however this is because of various factors. We have to further drill down into one of the contributing factors for serious/fatal accidents and identify the postcodes.

This leads us to the first business case that shall be analysed in this report.

Business Case 1: Identify postcodes with accident count based on the highest contributing factors and provide suggestions.

For many, the thought of cycling conjures up images of foreign locations such as France although there are many enticing locations in Australia as well. Victoria is one such state that has networks of cycling lanes. However, Victoria's bike lane budget has lagged behind other states (Carey, 2017). A few regions in Victoria are short of cycling lanes and or at least short of properly planned cycling lanes, due to which accidents occur. We further explore the data and by the end of this report we intend to find suggestions to reduce such accidents in the future.

This leads us to the second business case that shall be analysed in this report.

Business Case 2: Identify postcodes which has highest bicycle accidents, and provide suggestions to mitigate serious bicycle accidents in those postcodes.

In the dataset, the accident is categorized based on its Severity. Fig 6. shows levels of severity. In Fig 6, we consider, "Fatal Injury" and "Serious Injury" for analysing both the business cases.

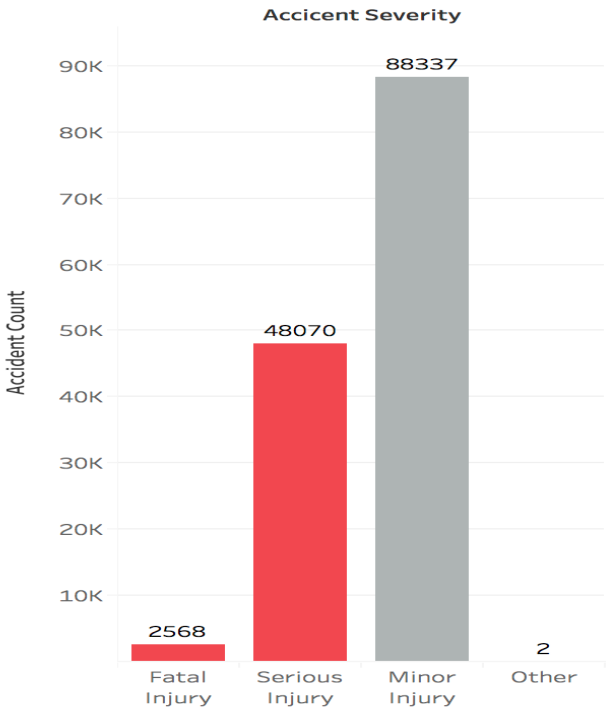

Fig 6: Accident count based on severity level

Methodology – Business Case 1

For identifying the causing factors, we implement Random Forests. In this analysis variables are chosen at random from the dataset and random forests are generated.

Many researches have used Random Forests, for instance, random forest technique has been used in finances to predict K-line (Zou, 2015), this method has also been used for Diabetic retinopathy classification analyses (Casanova, 2014).

In this analysis, Random Forest is used to rank the importance of the variable based on the calculated mean Gini index.

	MeanDecreaseGini
LIGHT_CONDITION	130.9679
ATMOSPH_COND	126.0007
ROAD_GEOMETRY	112.1139

Fig 7: Predictor variables Ranked based on Gini Index

Based on the Fig 7, we understand that Light Condition is the most contributing factor. Hence, we drill down into the data to identify the types of light conditions during accidents (Given in Fig 8.)

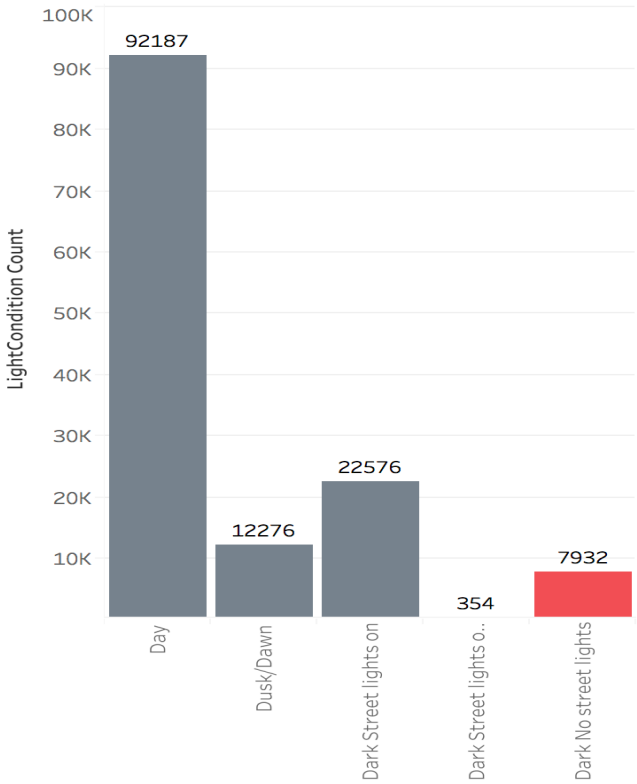

Fig 8: Accident count based on Light Condition

In Fig 8, we understand that the most controllable factor is the street light and hence "Dark No street lights" becomes the most significance to us as we can identify accidents that happen due to insufficient street lighting. We drill down to identify the serious accidents that happened in the dark during absence of street light (Given in Fig 9).

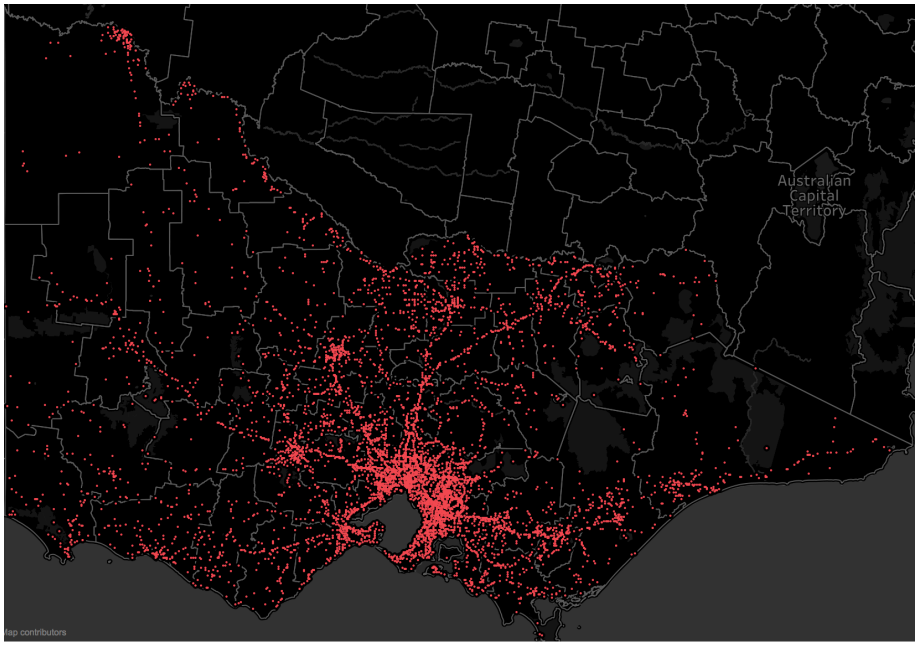

Fig 9: Accident Density in Victoria during dark and in the absence of street light

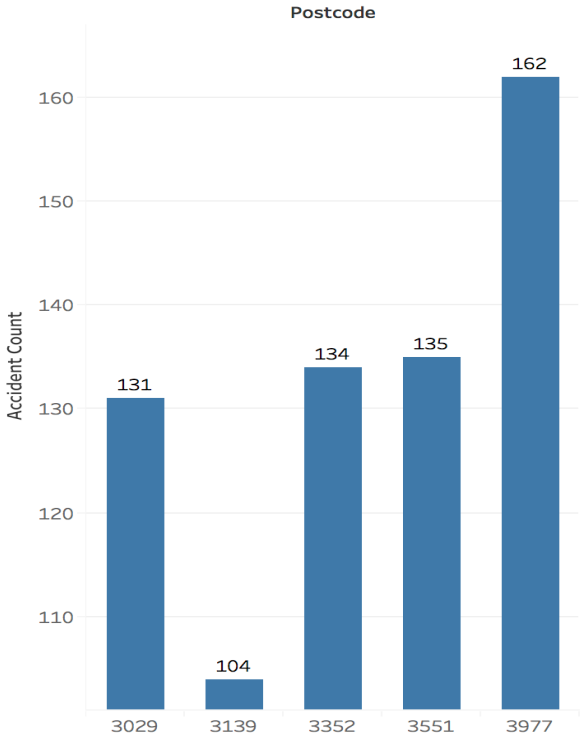

Fig 10: Top 5 Postcode with accident count in the absence of street light

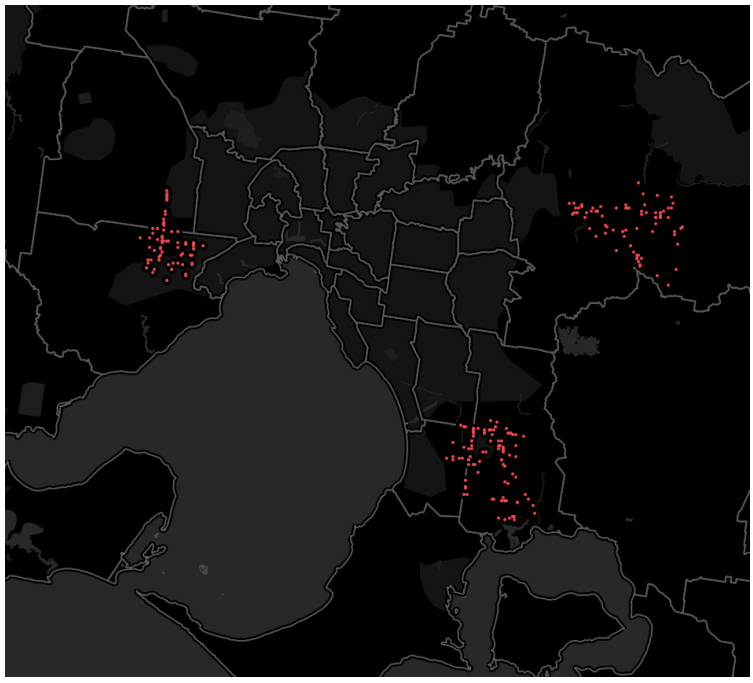
Fig 11: Top 5 Postcode Accident locations in the absence of street light

Although the postcode 3000 (Melbourne CBD) has the highest number of accidents (as shown in Fig 5), it is clear from the above analysis that, Melbourne CBD accidents are not due to street light insufficiency. Instead, the postcodes 3977, 3551, 3352, 3139 and 3029 that are away from the city, are the ones whose accidents are majorly due to insufficient street lighting.

Methodology – Business Case 2:

With the serious bicycle accidents from the previous years we forecast the accident count for a year ahead (given in Fig 12).

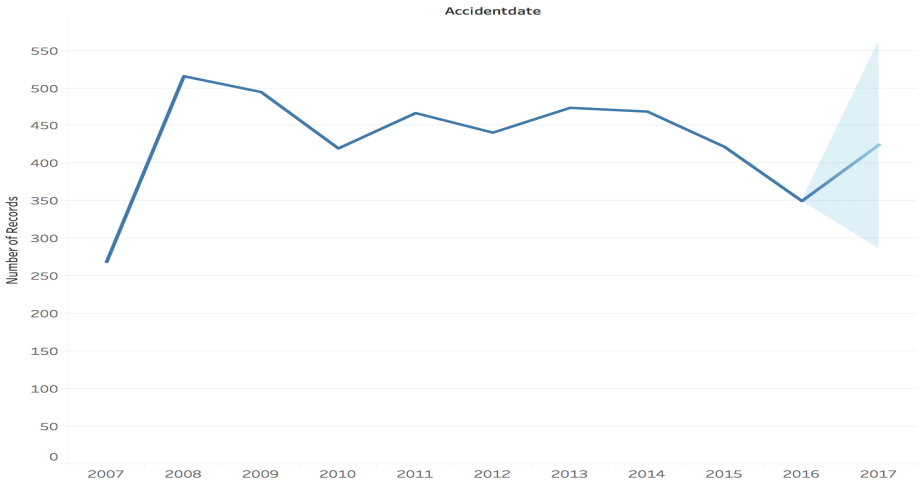

Fig 12: Forecast of serious bicycle accidents. The blue shaded region in the graph represents the forecast for 2017.

In order to mitigate such serious accidents, we need to identify the locations where it happens and find a pattern in the accidents.

The dataset includes a lot of vehicle types that are involved in accidents (refer Fig 13), however to analyse our business case, all the accidents that involved "Bicycle" as the vehicle type was chosen as a subset.

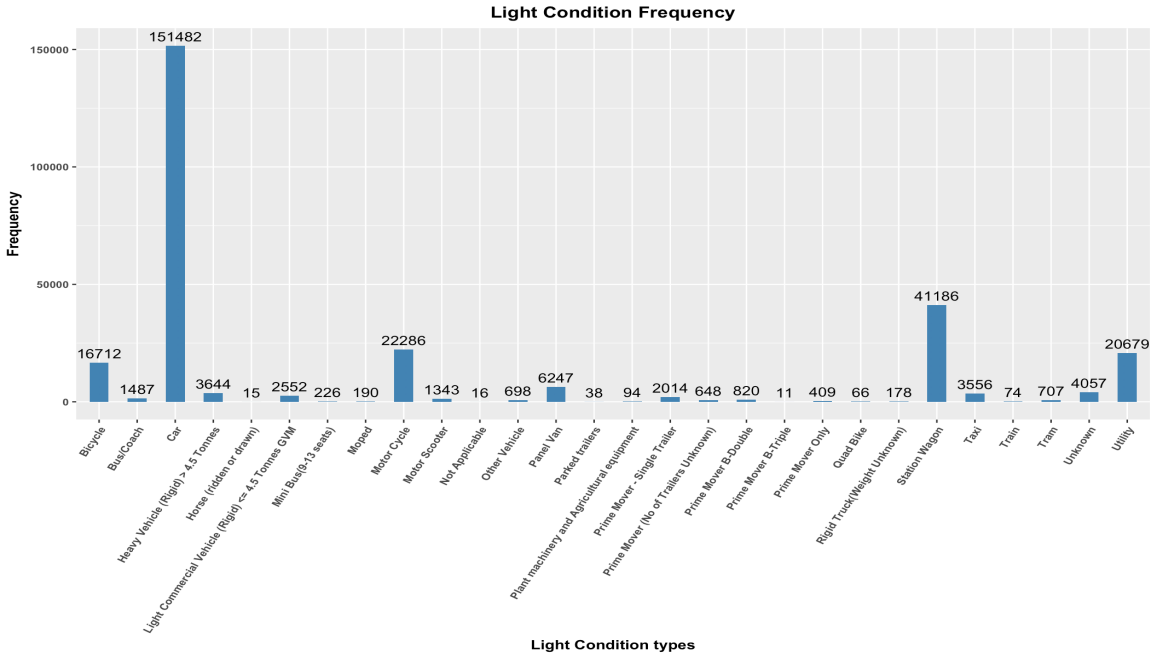

Fig 13: Accident counts based on Vehicle Type, **Bicycle** is chosen as our point of interest.

We are trying to identify accidents that happened between a cycle and another vehicle, hence, we consider only the bicycle accidents caused due to **Collision with vehicle** and ignore the rest (Refer Fig 14.)

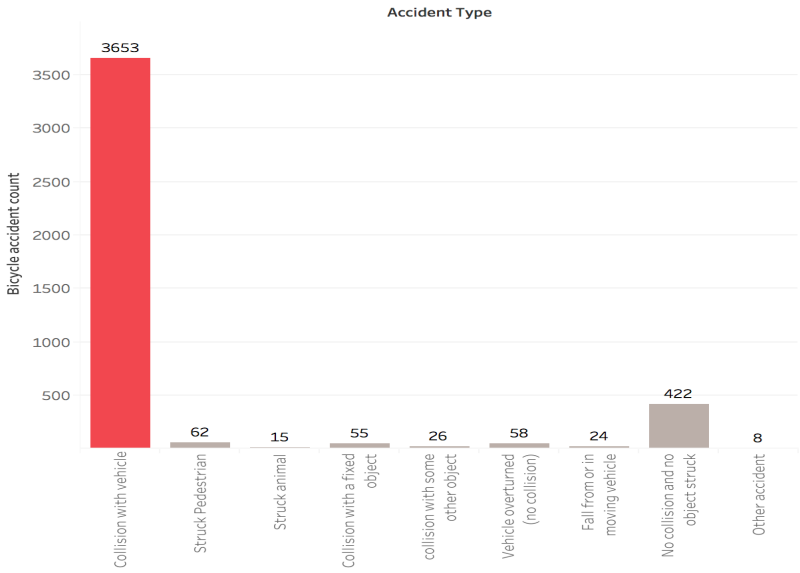

Fig 14: Serious bicycle accident count categorized based on accident type. **Collision with vehicle** chosen as our point of interest.

We identify the postcodes where the highest number of serious bicycle accidents take place and then identify a pattern in the bicycle accidents that happened in those postcodes.

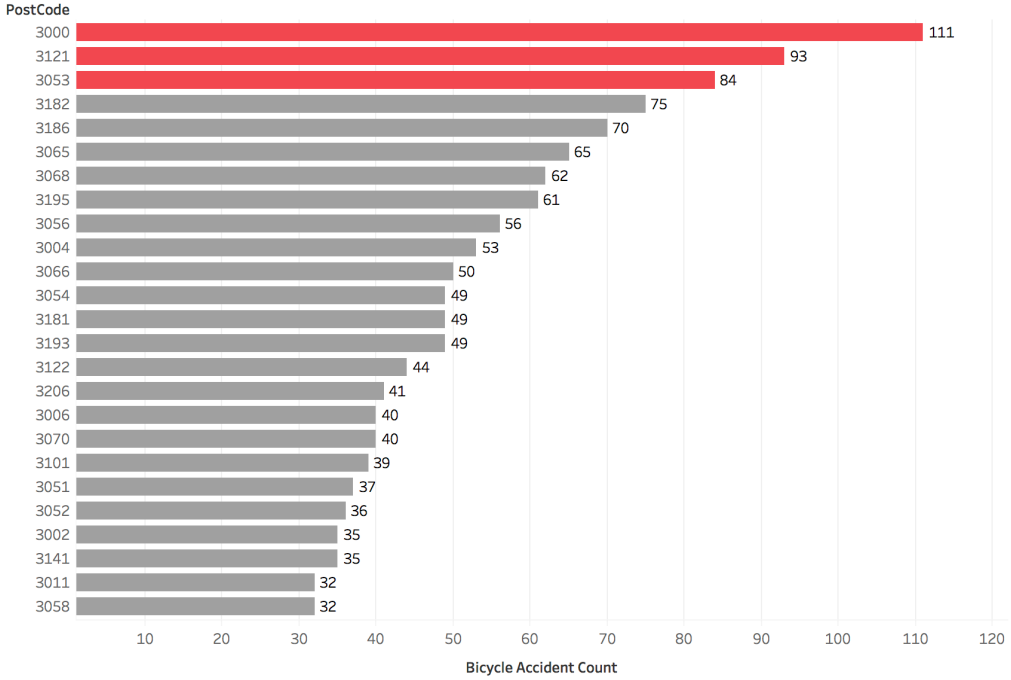

Fig 15: Postcodes with highest bicycle accident counts due to collision with vehicle. Top 3 postcodes are our points of interest in this analysis.

Two of the cycling lanes available in Victoria are Dedicated cycling lanes and Bicycle-friendly roads. Google APIs provide the cycling lane information. An instance of a bicycle lane is given in the Fig 16.

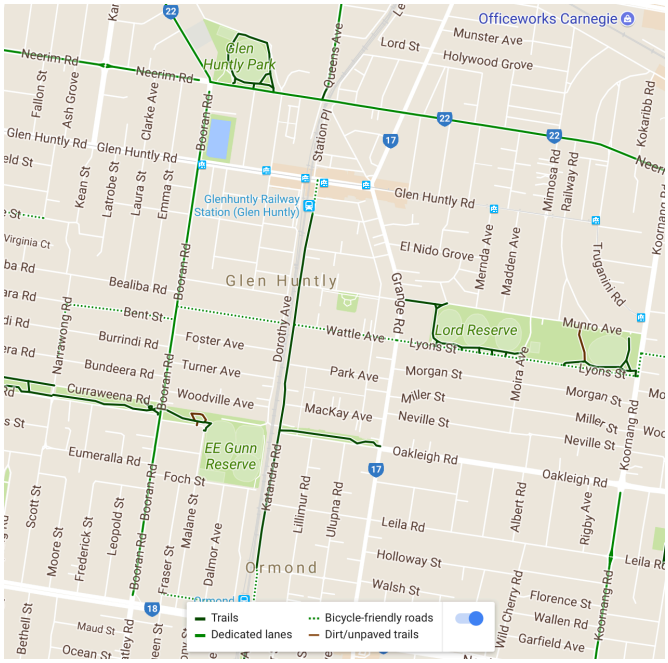

Fig 16: Bicycle lane information at Glen Huntly provided by Google.

We further drill down into each of the top 3 postcodes highlighted in Fig 15 and compare the accident locations with the bicycle lane information provided by the Google API.

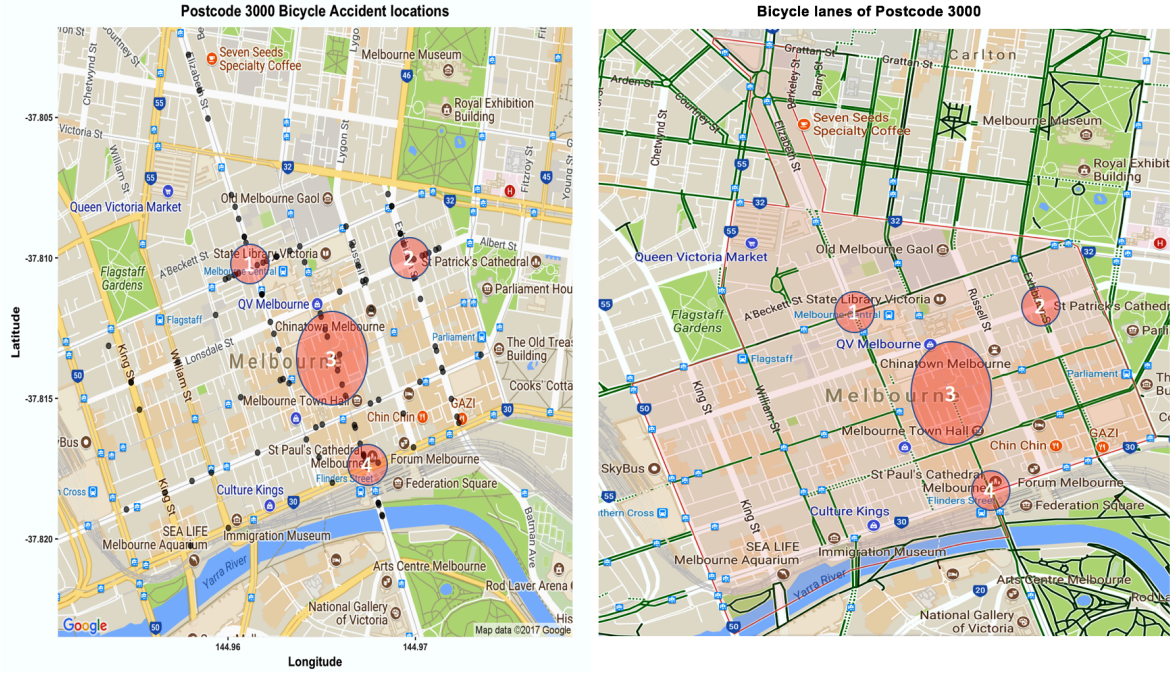

Fig 17: Comparison between bicycle lanes and bicycle accident locations for Postcode 3000. The red circles in the map are zones of interest.

Zone	Description
1	Intersection between bicycle friendly roads, non-cycling lanes and dedicated cycling lanes
2	Intersection between dedicated bicycle lanes and non-cycling lanes.
3	Non-cycling lanes
4	Intersection between dedicated bicycle lanes and non-cycling lanes.

Fig 18: Analysis/Comparison for Postcode 3000 (Refer Fig 17)

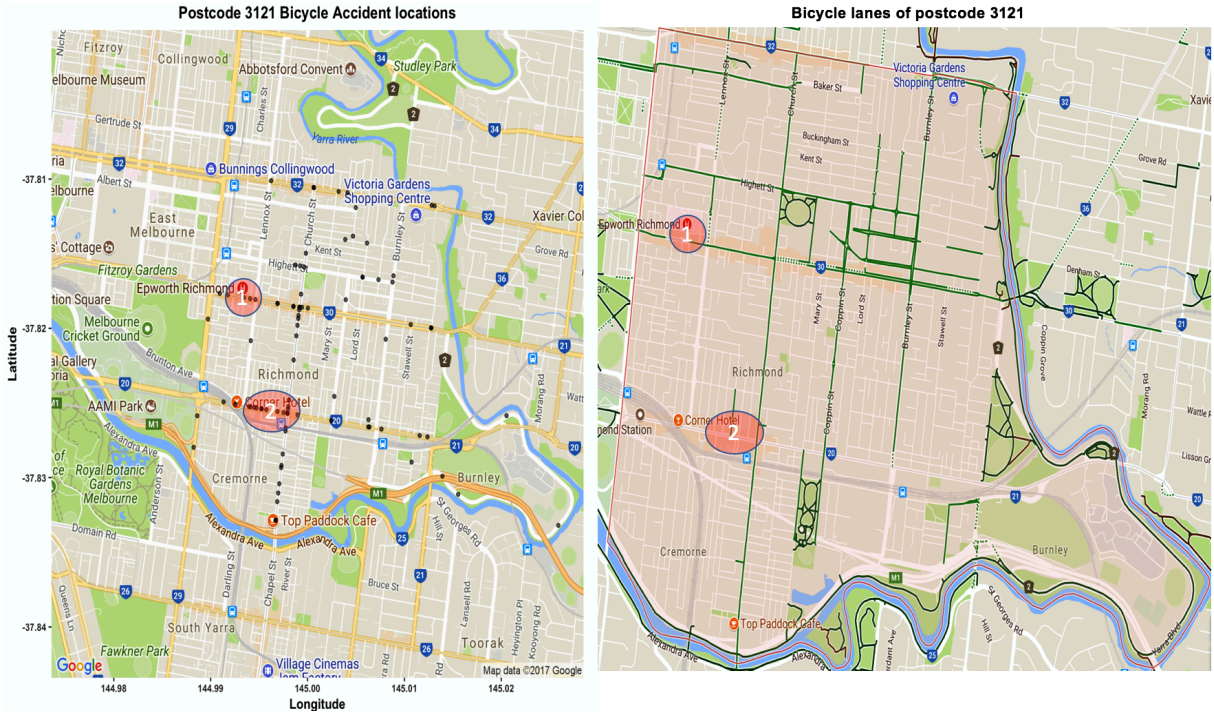

Fig 19: Comparison between bicycle lanes and bicycle accident locations for Postcode 3121. The red circles in the map are zones of interest.

Zone	Description
1	Intersection between bicycle friendly roads and dedicated cycling lanes
2	Intersection between dedicated bicycle lanes and non-cycling lanes.

Fig 20: Analysis/Comparison for Postcode 3121 (Refer Fig 19)

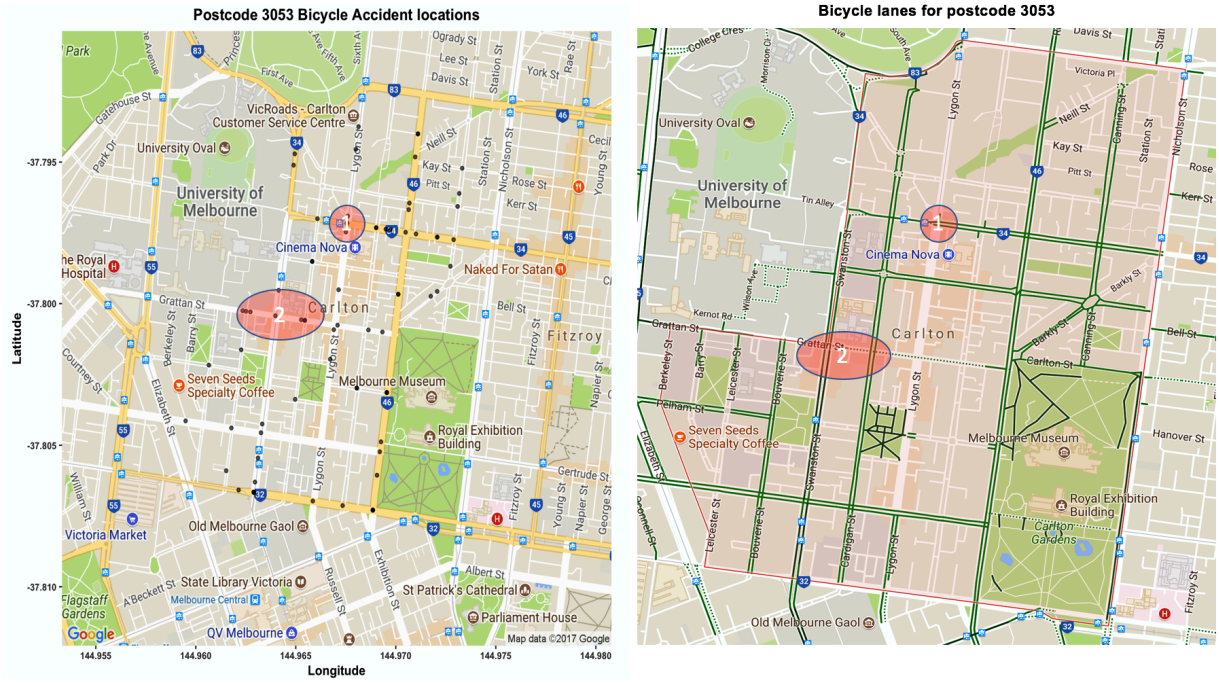

Fig 21: Comparison between bicycle lanes and bicycle accident locations for Postcode 3053. The red circles in the map are zones of interest.

Zone	Description
1	Intersection between non-cycling lane and dedicated cycling lanes
2	Intersection between dedicated bicycle lanes and bicycle-friendly lane.

Fig 22: Analysis/Comparison for Postcode 3053(Refer Fig 21)

Out of curiosity we decided to pick up one more postcode (3195) from Fig 15 at random and perform an analysis. The comparison is given in Fig 23.

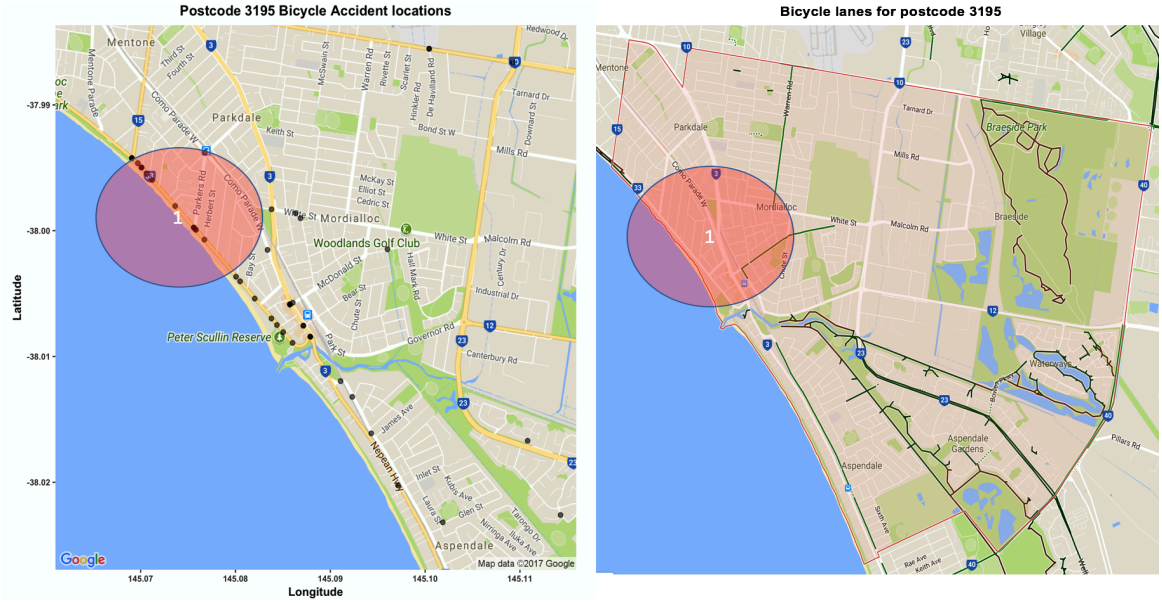

Fig 23: Comparison between bicycle lanes and bicycle accident locations for Postcode 3195. The red circles in the map are zones of interest.

Zone	Description
1	Non-cycling lanes

Fig 24: Analysis/Comparison for Postcode 3195

From all the analysis done above on each of the postcodes (3000, 3121, 3053 and 3195) we identify the zones that require attention in terms of either new dedicated cycling lanes or improvement in the structuring of cycling lanes.

Conclusion

The overall idea of this report is to reduce the accidents occurring in Victoria, Australia. We considered two of the many aspects and drilled down into each to identify the influencing factors of these accidents.

In **Business Case 1**, we identified that poor street lighting is one of the factors causing serious accident in certain regions of Victoria. The postcodes 3977, 3551, 3352, 3139 and 3029 are the ones, city council should concentrate on to mitigate accidents due to poor street lighting. Either of the following steps can be taken:

- Improve street lighting in these postcodes.
- Digital warning signs at the roadside to alert the drivers about low visibility due to poor street lighting.

In **Business Case 2**, we identified that most of the serious bicycle accidents happen either at the intersection of a dedicated cycling lane and a non-dedicated cycling lane or at locations where there are no cycling lanes. These accidents can be mitigated by the following steps:

- Construct new dedicated or bicycle-friendly lanes
- Proper sign boards at the intersection of a dedicated cycling lane and a non-dedicated cycling lane.
- Conducting awareness camps on cycling in Victoria.

References

Crash Stats - Data Extract - Victorian Government Data Directory. (2014, August 1). Retrieved September 13, 2017, from https://www.data.vic.gov.au/data/dataset/crash-stats-data-extract

Carey, A. (2017, May 18). Victoria's bike budget lags behind most other states, report finds. Retrieved November 01, 2017, from http://www.theage.com.au/victoria/victorias-bike-budget-lags-behind-most-other-states-report-finds-20170515-gw56rz.html

Casanova, R., Saldana, S., Chew, E. Y., Danis, R. P., Greven, C. M., & Ambrosius, W. T. (2014). Application of random forests methods to diabetic retinopathy classification analyses. PLoS One, 9(6), e98587.

Zou, Z. B., Peng, H., & Luo, L. K. (2015). The Application of Random Forest in Finance. In Applied Mechanics and Materials (Vol. 740, pp. 947-951). Trans Tech Publications.