**CP5046** ASSESSMENT TASK 3: **Project Audit**

This assessment task has been prepared by Dr. Dmitry Konovalov for James Cook University. Updated 26 May 2020.

**© Copyright 2019**

This publication is copyright. Apart from any fair dealing for the purpose of private study, research, criticism, or review as permitted under the Copyright Act, no part may be reproduced by any process or placed in computer memory without written permission.

**ASSESSMENT TASK 3: PROJECT AUDIT**

|  |  |
| --- | --- |
| **Aligned subject learning outcomes** | All |
| **Group or individual** | Individual, ***Students from the same team may be awarded different marks for this assessment item*** |
| **Weighting** | 50% |
| **Due date** | Week-13 Friday 9am 31-May-2019 |

**ASSESSMENT TASK 3: DESCRIPTION**

This task is the Audit of: (i) your group ICT solution delivered to the client, and (ii) ***individual contributions*** to the overall team effort and results. In terms of the Agile Software development, this is the ***iteration-2***. It delivers the ***alpha release*** to the client, and makes planning for ***iteration-3*** (***beta release***), which is due middle of CP5047.

**ASSESSMENT TASK 3: CRITERIA SHEET**

***NOTE! Students from the same team may be awarded different marks for this assessment item. To arrive at the individual student marks, your lecturer may consider all or some of the following contributing factors: your team overall marks; team feedback forms; personal observations of the student project contribution; student workshop attendance and subject participation; student competency during workshops, this audit and the project presentation.***

The following is the list of items which are required to be completed in this assignment. Maximum possible marks are given in brackets at the beginning of each item.

**[\_\_\_\_/Prerequisite for marking]** Assignment is completed using electronic copy of ***this*** document and submitted to LearnJCU electronically.One submission per team.

**[\_\_\_\_\_/20 marks] *Individual* Team Feedback form is submitted to LearnJCU electronically.** One submission per student. Assignment is done in a group with 2-4 students.

Write group members here:

|  |  |  |
| --- | --- | --- |
| **Student Name** | **Project role** | **Project contribution** |
| Vikash Singh | Project Manager | Responsible for setting up GitHub repository, scheduling recurring meetings and delegating tasks. |
| Willis Lin | Document officer | Responsible for data gathering, version control of the documents and testing of the solution. |
| Kevin  Vathalloor James | Data Analyst | Responsible for dataset pre-processing using python and unit testing. |
| Harish Shetty | Development & Support Engineer | Responsible for writing Python code in Kaggle Notebook, finding patterns for road crashes and unit testing. |

Note: Each student ***must be present*** at the audit and must be ready to answer any questions regarding the student’s individual contribution.

**[\_\_\_\_\_/40 marks] Report and demonstrate the *ACTUALLY* delivered alpha-release (see your user stories in iteration-1).** Any deviations from the alpha-release-iteration-1 must be documented and briefly explained. Screenshots (or illustrations) of running alpha-release with comments or explanations.Write here: Minimum **TWO** pages, maximum **TEN** pages.

The actual Alpha-Release was delivered and demonstrated to our client on the 29th of April with the following user stories and prototype on Kaggle Notebook.

<https://www.kaggle.com/kevinjames1993/australian-fatal-crashes-report-from-1989-to-2019>

1. **Data Cleaning**

**Effort Estimate:** 9 days

Data can have a lot of missing values or irrelevant parts. To handle these parts, data cleaning needs to be done for missing and noisy data.

1. **Fatality rate in Australia from 2015 to 2020**

**Effort Estimate:** 4 days

Representing fatality rate in each state from 2015 to 2020.

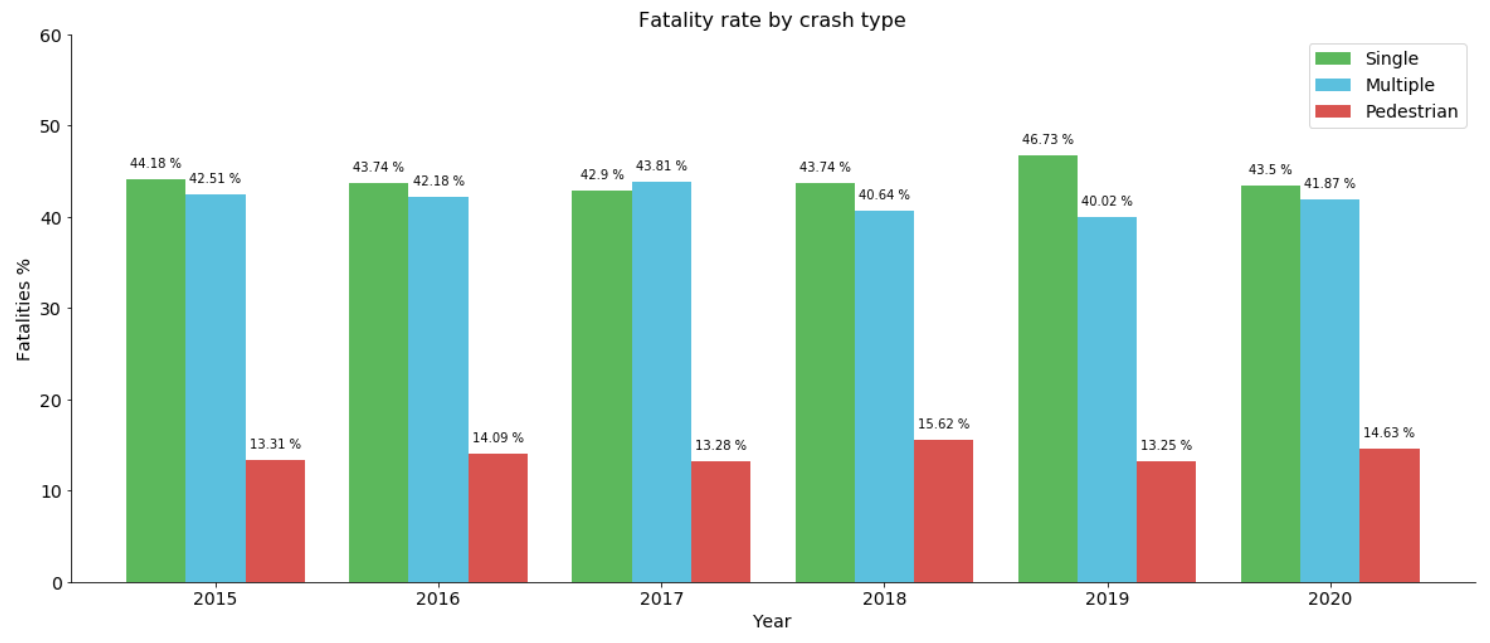
**A picture containing text, map

Description automatically generated**

1. **Fatality rate by crash type from 2015 to 2020**

**Effort Estimate:** 4 days

To find the percentage of each crash type for pedestrians, single, multiple annually in each state from 2015 to 2020.



1. **Pedestrian Fatality Rate from 2015 to 2020**

**Effort Estimate:** 3 day

To reduce the minimal fatalities, the count of fatality by pedestrians is represented compared with the rate over the year from 2015 to 2020.

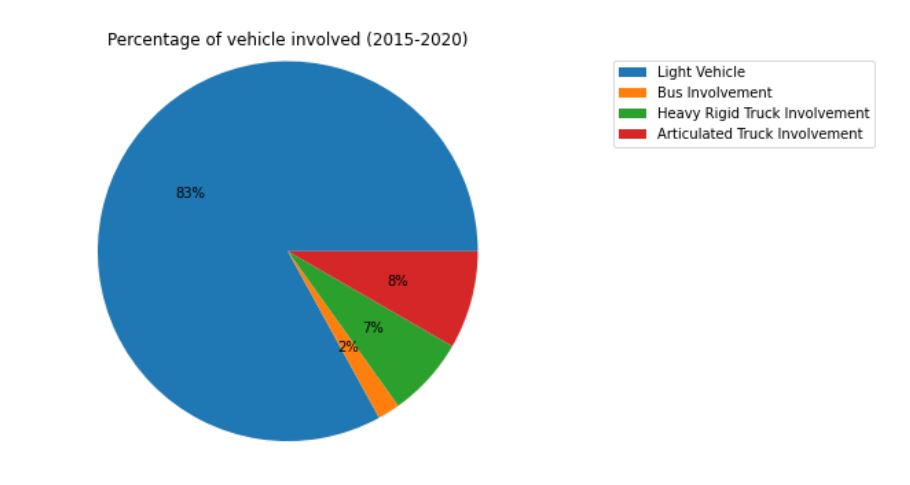
**A picture containing umbrella

Description automatically generated**

1. **Types of Vehicle Involvement from 2015 to 2020**

**Effort Estimate:** 4 days

The involvement of various types of vehicles with the road fatalities to find the majority rate is represented in percentage from 2015 to 2020.

****

1. **Road User Fatality Rate from 2015 to 2020**

**Effort Estimate:** 4 days

To uncover the percentage of fatality by road users, including passenger, driver, cyclist, motorbike rider from 2015 to 2020.

**A close up of a logo

Description automatically generated**

1. **Fatality by Speed Limit from 2015 to 2020**

**Effort Estimate:** 3 days

Fatal crashes rate on each speed limit zones from the year of 2015 to 2020.

**A screenshot of a cell phone

Description automatically generated**

1. **Fatal Crash by Day/Night from 2015 to 2020**

**Effort Estimate:** 5 days

To understand the percentage of fatal crashes rate happened during the day/night in each year from 2015 to 2020.

**A screenshot of a cell phone

Description automatically generated**

1. **Fatality Rate by Day of the Week from 2015 to 2020**

**Effort Estimate:** 3 days

To understand which day of the week has higher percentage on road crashes from 2015 to 2020.

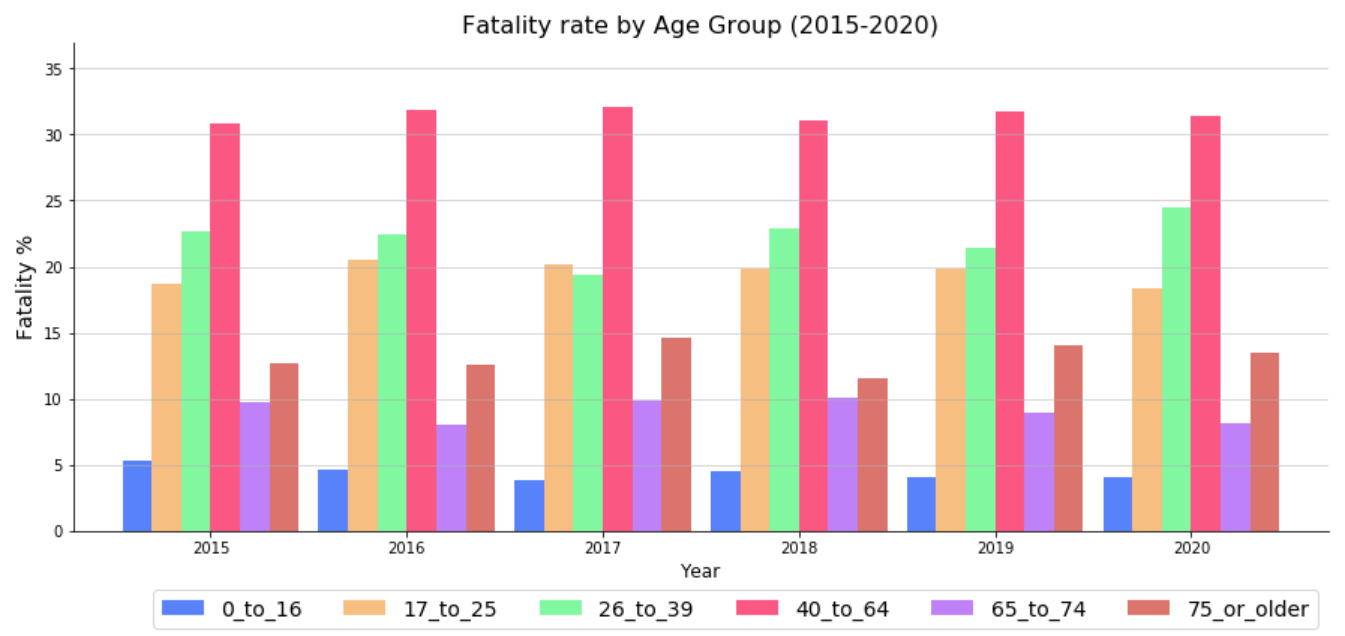
A screenshot of a cell phone

Description automatically generated

1. **Fatality Rate by Age Group from 2015 to 2020**

**Effort Estimate:** 5 days

To provide accuracy in the fatality rate by age group that attributed from 2015 to 2020.



1. **Number of fatal crashes by gender from 2015 to 2020**

**Effort Estimate:** 3 days

To find the fatality rate pattern the accidents by gender from 2015 to 2020.

**A screenshot of a cell phone

Description automatically generated**

1. **Creating the report**

**Effort Estimate:** 5 days

Creating a report through analysing the dataset, obtaining meaningful patterns, illustrating graphs and charts to assist the client’s understanding.

1. **Creating a user manual**

**Effort Estimate:** 5 days

Creating a user manual for the client so that the client knows how to navigate Kaggle Notebook and to interpret the results and findings.

1. **Final Submission**

**Effort Estimate:** 2 days

Upon successful validation of the report, the project documentation is submitted through LearnJCU / [Kaggle](https://www.kaggle.com/kevinjames1993/australian-fatal-crashes-report-from-1989-to-2019).

**[\_\_\_\_\_/20 marks] Client signed acceptance of the alpha-release, and the proposed beta- and final-releases**. Any changes from iteration-1 are approved by the client. Write here: Minimum **TWO** pages, maximum **TEN** pages.

* [\_\_\_\_/5 marks] Provide the burn-down and velocity charts/values for iteration-2, and how they are used to plan beta- and final-releases.

The following is the burn-down charts for iteration 2 and how they are used to plan beta and final release of our project.

Our initial term was for 12 weeks and we divided our project into 6 sprints.

Each sprint consists of two weeks. Thus, the total number of sprints is 6 with a total of 12 weeks.

In sprint 1, our target was 80 and we were able to finish it by 72 hours.

A close up of a map

Description automatically generated

In sprint 2, our target was 80 and we were able to finish it by 74 hours. By Finishing sprint 2, we were able to achieve the first milestone which is "Load data to Kaggle and clean up noise data ".

A picture containing table, white

Description automatically generated

In sprint 3, our target was 88 and we were able to finish it by 84 hours. By finishing sprint 3, we were able to achieve the next milestone "Design initial feature on Kaggle " which is the alpha release.

A close up of a map

Description automatically generated

In sprint 4, our target was 96 and we were able to finish it by 108 hours. This was more challenging and needed time for modification as per the feedback from the client.

A screen shot of a map

Description automatically generated

In sprint 5, our target was 64 and we were able to finish it by 62 hours. This is the beta release and completion of milestone 3 "Develop several charts and graphs to identify focuses on interesting patterns ".

A screenshot of a computer

Description automatically generated

In sprint 6, our target was 96 and we were able to finish it by 108 hours. this is the final release and the completion of milestone 4 "Finalise reports and design a user manual ".

A close up of a map

Description automatically generated

## JCU_Logo_RGB.jpgClient Consultant Engagement

## For ICT -1

## Between: Dmitry Konovalov\_

## (including its successor ‘the client’)

## And: Kevin Vathalloor James, Sai Harish Dachepalli, Vikash Singh, Wei-Chih Lin (Willis)

## (including its successor ‘the consultant’)

|  |  |
| --- | --- |
| Project:ICT - 1 | Location:JCU Townsville |
| Client’s Representative:Name: Mr. Dimitry KonovalovContact Details: [Dmitry.konovalov@jcu.edu.au](mailto:Dmitry.konovalov@jcu.edu.au) | Consultant Representative for the group:Name: Mr. Vikash SinghContact Details: [Vikash.singh@my.jcu.edu.au](mailto:Vikash.singh@my.jcu.edu.au) |
| Scope and nature of the Service:Analysing the dataset provided by the Bureau of infrastructure, transport and regional economics (BITRE) - a government body of Australia, using data mining techniques to generate common causes for road crashes in several categories such as speed limits, gender, age and holiday periods, etc. The complete report will be provided to the client for improvement of Australia Road Safety and Regulations. | |
| Information and Service to be provided by Client:Specifying the requirements and providing details for the source and data required for analysis and publication online. | |
| The Client engages the consultant to provide the service described above and the consultant agrees to perform the services (including as may be set out in any relevant statement of work provided to the consultant by the client). Both parties agree to be bound by the provision of this client consultant engagement. Once signed, this agreement will replace all or any oral agreement previously reached between the parties. | |
| Client Authorized Signature:Print Name: Dimitry KonovalovDate: | Consultant Authorized Signature:Print Name: Vikash SinghDate: |

**[\_\_\_\_\_/20 marks] Demonstrate the Project development and release ICT infrastructure.** This must include development environment, programming languages, source code repositories (Configuration Management), project collaboration tools, and development tools. Write here: minimum **TWO** pages, maximum **TEN** pages.

* [\_\_\_/10 marks] **Configuration Management**/version control, e.g. git, github, heroku, bitbucket; Project tools. Programming languages/IDEs. Building tools/procedures. How to set-up your development/release environment for a new team member.
* [\_\_\_/10 marks] **Prototypes are demonstrated to justify the proposed beta-release**.

As part of the ICT project 1, the version control that we use in this data scenario is Kaggle. Kaggle is an online community that contains enormous datasets that can be used for data manipulation and data mining. The Kaggle can also be used to publish a dataset, explore and build models that can be helpful for accessing it to a web-based environment. Another important part of Kaggle is the Kaggle notebook which is a programming environment for data processing and report generating. It is a free platform known as Kaggle kernel that is a Jupyter notebooks in a web browser.

Developing a report from a bunch of excel files manually is a tedious process. To make reporting easier and simple, one of the best solutions is to get it done in a program that generates reports according to the parameter provided. The program that we use in this case is python 3.8. It is the newest major release of the Python programming language, and it contains many new features and optimizations. Python is one of the most popular programming languages that is used for machine learning and data exploration.

Our team hold a weekly update with our client to demonstrate the proposed prototypes, fulfil changes to justify the proposed beta-release according to the client’s requests. The following is the final agreed version for the results and findings. The prototype for our final release is available on Kaggle Notebook via this link [Australia Fatal Crashes Report from 1989 to 2019](https://www.kaggle.com/kevinjames1993/australian-fatal-crashes-report-from-1989-to-2019).

### Fatality per year per 100,000 people

As we can see in Figure 1, the number of fatalities all over Australia is represented for the population. In this instance, the fatality per 100,000 people is plotted and the rolling average of 3 years is plotted adjacent to the red line. It is so obvious that the fatality over the years has decreased to a level from more than 16% in 1989 to less than 6% by the end of 2019. It might be due to the strict laws and traffic regulations altered or changed over the years.

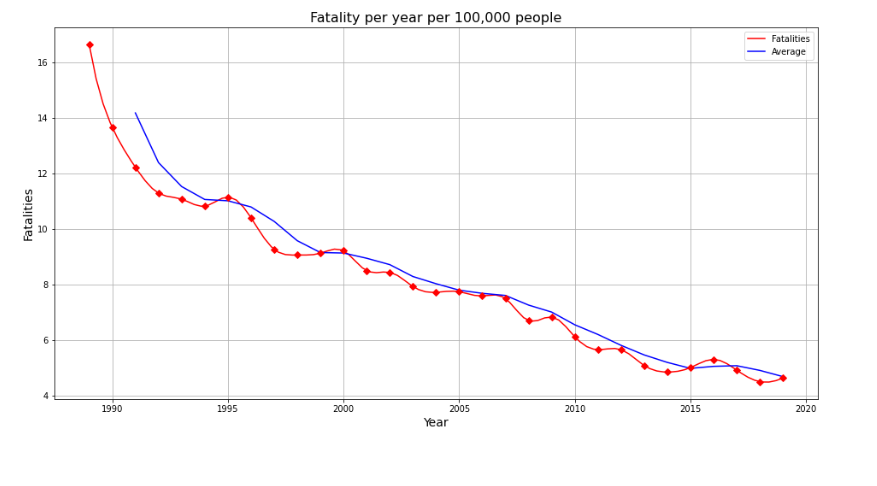


Figure 1. Fatality per year per 100,00 people

### Fatality by State per year per 100,000 people

All state-wise fatality per 100,000 people is represented in Figure 2. It can be seen from the graph that the fatality is 0 in ACT from mid-90’s until mid-2000’s. The data was not available to represent the fatalities in that period. However, it is surprisingly evident that Northern Territory (NT) has around 15 fatalities by 2019. All other states are demonstrating a downward trend to the fatalities.

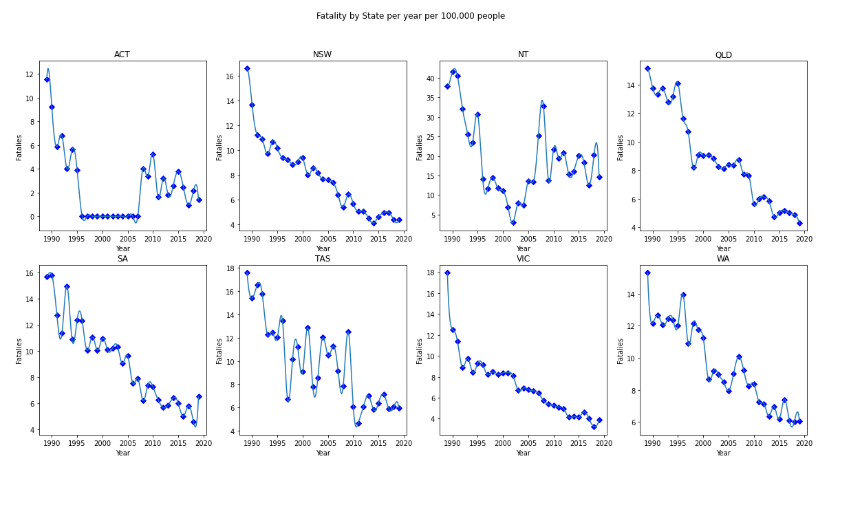


Figure 2. Fatality by state per year per 100,000 people

### Fatality Rate by Crash Type

From the year 1989 to 2019, there are three categories in Figure 3 representing the fatality rate over the years. They are Multiple, Pedestrian and Single in each year. From multiple perspectives, it fluctuates between 40.02% and 43.81% with a margin of less than 4% over the years. The lowest rate was recorded in 2019 and highest in 2017. On the Pedestrian category, the percentage is twice lower than the Multiple, it looks quite steady between 13.25% and 15.62% which has less than 3% difference over the years. The lowest was recorded in 2017 and 2019 while the highest in 2018. The single fatality rate, on the other hand, has the highest percentage recorded amongst the three categories. The graph shows that between 42.9% and 46.73% which is almost three times greater than the Pedestrian. The fluctuation rate is just below 4% throughout the years, the lowest was recorded in 2017 and highest in 2019.

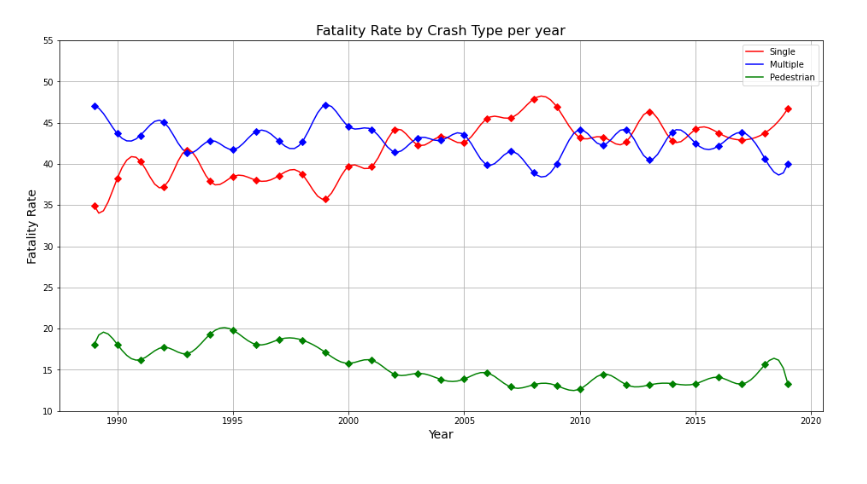


Figure 3. Australia Fatality Rate by Crash Type 1989 to 2019

### Fatality Rate by Road User

Knowing types of road users enable the government to make the road safer according to the user type. Figure 4 representing the fatality rate by crash type. There are six types of road users in our dataset, each type has its significance to the fatality rate. Begin from the highest to the lowest, Driver, represents 47%, the biggest contributor to fatality rate. The second is the Passenger category, which has 18%. The third is the Motorcycle Rider which is only 1% lower than the Passenger. The Pedestrian is the fourth with 14% in total. The fifth is Pedal Cyclist which has only 3% among all other categories. Lastly, there is 1% unspecified according to the dataset. From the graphs, it is evidently clear that the pedestrian fatality rate is more than 10% all around the course of time. This area must be taken into consideration while implementing the counter measures.

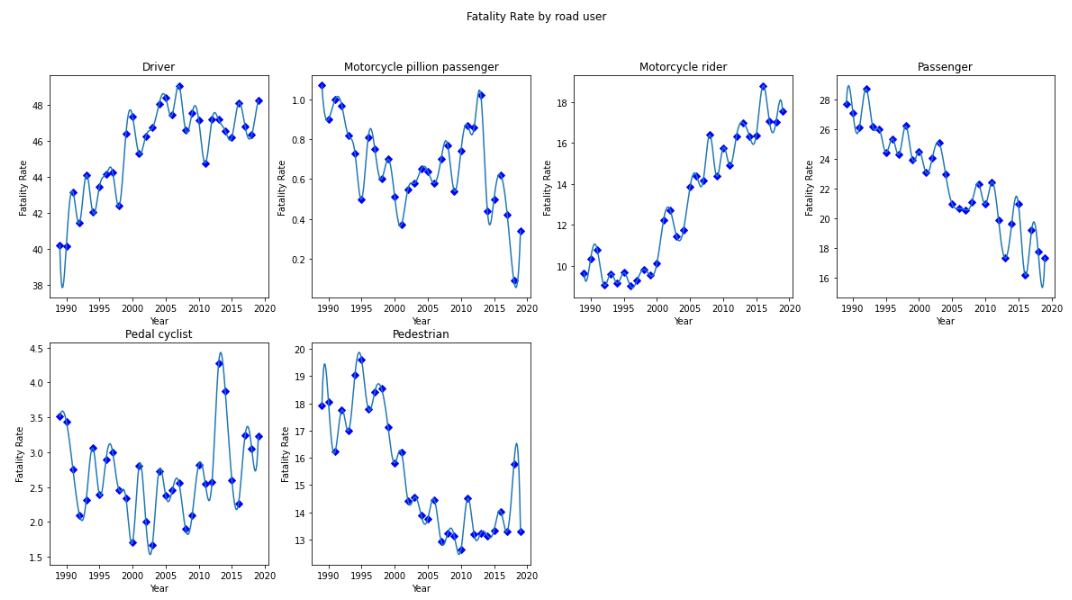


Figure 4. Road User Fatality Rate from 1989 to 2019

### Fatality Rate by Speed Limit

The speed limit is a substantial factor contributing to road crashes. Figure 5 shows that fatal crashes are more likely to happen above 75 km/h speed range than driving below 75 km/h. The fatality rate is more than 50% from Above 75-speed limit and this speed limit area might be a motorway that people attempt to drive faster. On the other hand, the local roads that have the speed limit of less than 75 were around 48%, in the beginning, and it is decreasing now and was 35% in 2019. This downtrend is due to the advanced speed cameras installed in different areas but limited to motorways.

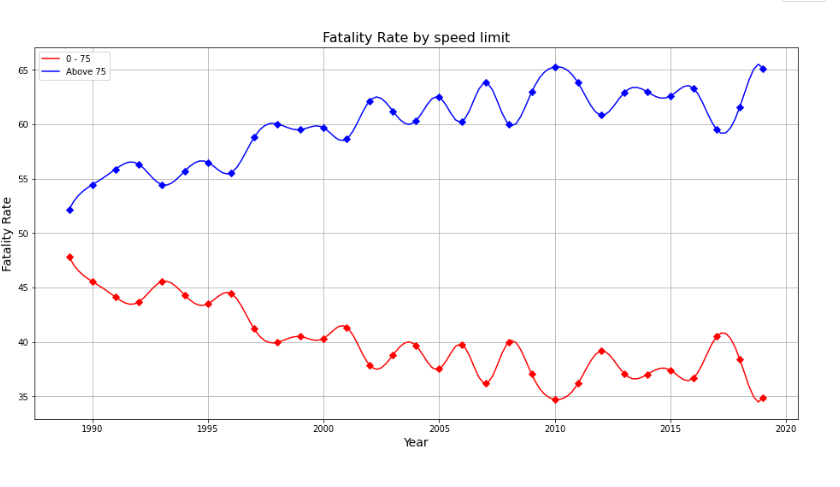


Figure 5. Fatality Rate by speed limit from 1989 to 2019

### Fatality Rate by Time Range

It is important to know the time frame in which the accidents are mostly happening. In order to achieve that, we have divided 24 hours into 4-time frames. From the graph of Figure 6, it is evident that the fatality rate is high during 12 PM-6 PM contributing more than 365 in 2019. In second, the fatality rate is high during 6 AM-12 PM which was initially less than 205 in 1989. Both time frames above mentioned are considered as the peak business hours in Australia. To further analyse and verify the above results, the rate of accidents every hour is plotted in a histogram form (00:00 to 23:00) every day from 1989 to 2019.

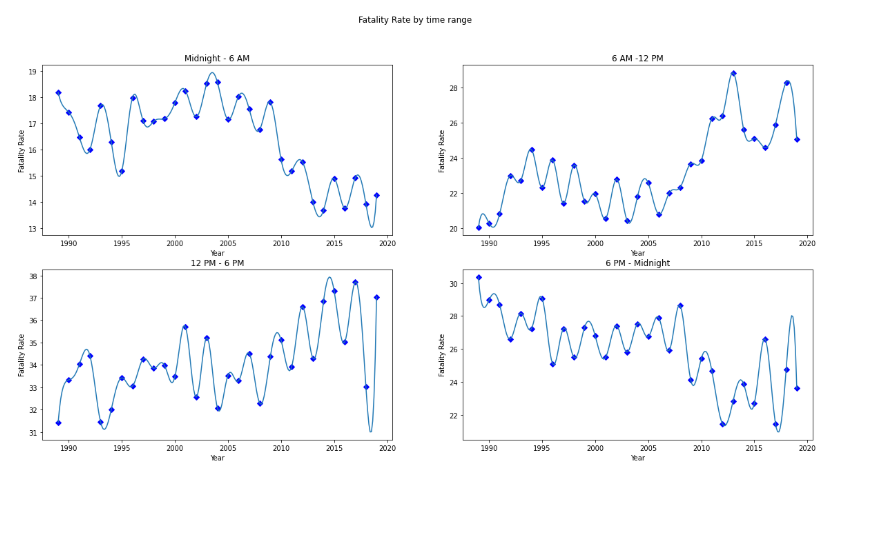


Figure 6. Fatality Rate by Time Range

In Figure 7, the fatalities of every hour are represented to verify the results. It is thus evident from the histogram that the fatality rate is increasing from 6 AM and reaches its maximum at 15:00 (3:00 PM) and starts to reduce from that point. However, the minimum fatality is approximately 1,000 at 04:00 AM and maximum is around 3,500 at 15:00 (3:00 PM).

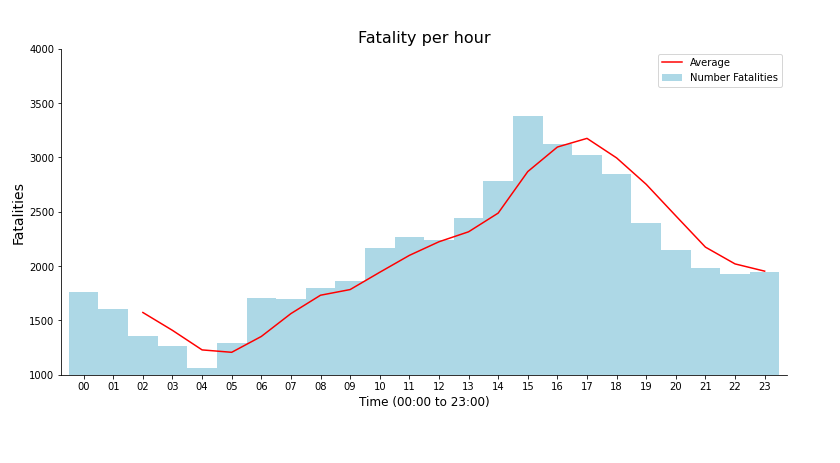
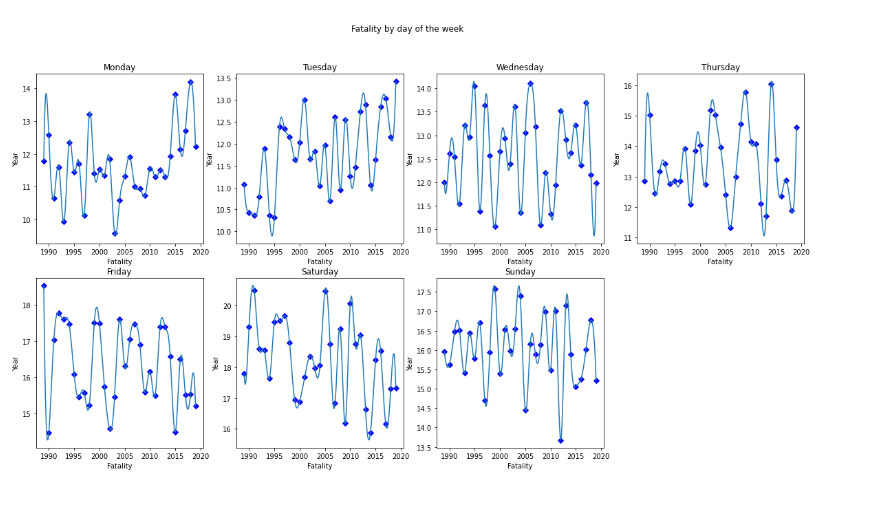


Figure 7. Fatality per hour

### Fatality Rate by Day of the Week

When it comes to understanding which day of the week fatal crashes is most likely to happen, the above bar chart illustrates evidently. Figure 8 representing fatality by day of the week. It reveals that from Friday to Sunday where the highest fatality rate was recorded from 15.44%, 17.53% and 15.57% respectively. The rest of the days during the week were below 13.1% in general, within them, however, Saturday has the highest fatality rate of 17.53%.

Figure 8. Fatality Rate by Day of the Week

### Fatality Rate by Age Group

Driver’s age always plays a vital role when it comes to road crashes. Figure 9 shows the different age groups fatality rate over the years from 1989 to 2019. It can be easily observed that the age group between 51 and older has the highest fatality rate of over 37% by 2019 with an initial value of less than 25% in 1989. It also shows that the age group between 0 and 18 reduced from more than 18%in 1989 to 8% in2019. One interesting result should be noted that in the age group between 31 to 50 has increased dramatically and even peaked to more than 30% in 2016 and is now showing a downward trend. The age group between 19 and 30 was in its maximum of more than 34% and reduced with a sudden spike to more than 305 in 2005 and showed a decrease in the graph until now.

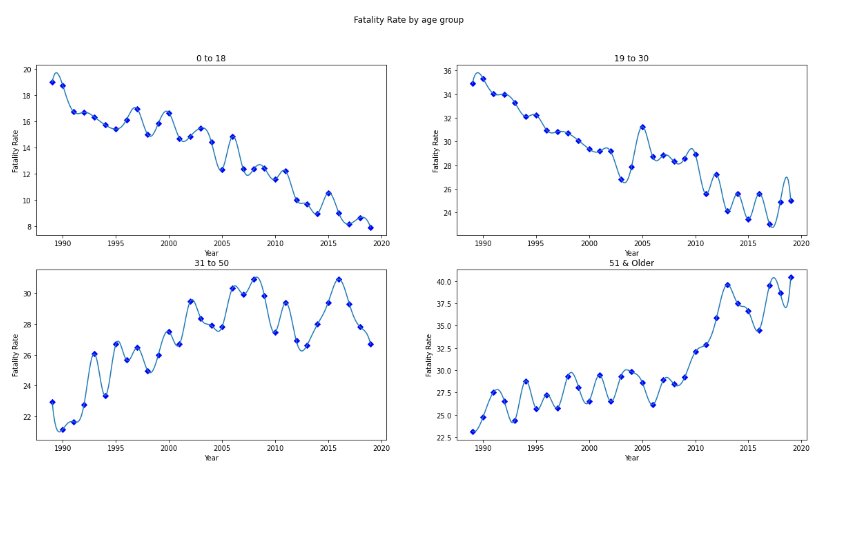


Figure 9. Fatality Rate by Age Group

### Fatality by Gender

In this category, the fatality rate between male and female during road crashes can be clearly seen in Figure 10 and 11. In general, female has a lower fatality rate than male. However, the gap between male and female has a significant difference. The fatality rate of male is almost three times larger than female in each year nationwide. The lowest fatality rate in the female category was recorded in 2019 with 23.43% while the highest in 2017 with 26.54% respectively. On the other hand, in the male category, the lowest fatality rate was recorded at 73.46% in 2017 while the highest in 2019 with 76.57% respectively. The margin over the years for both Male and Female categories is just over 3% which is considered steady. A rolling average on each graph is also represented to get a familiar view in the course of 3 years.

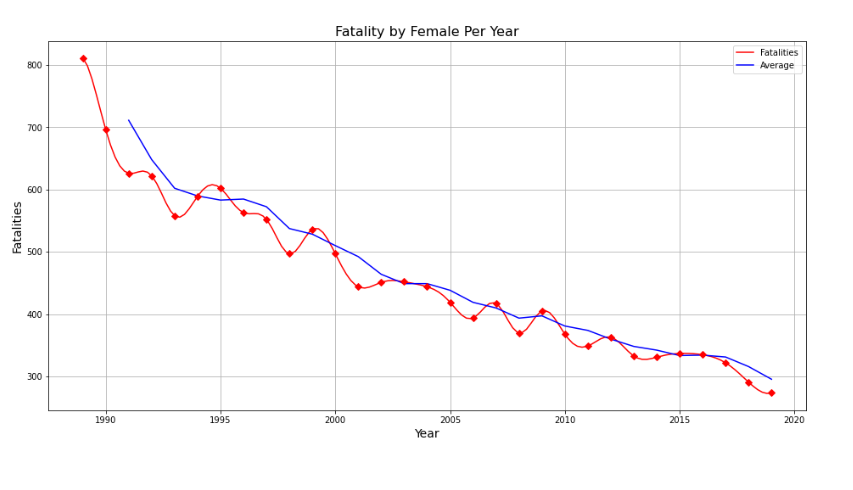


Figure 10. Fatality of Female per year

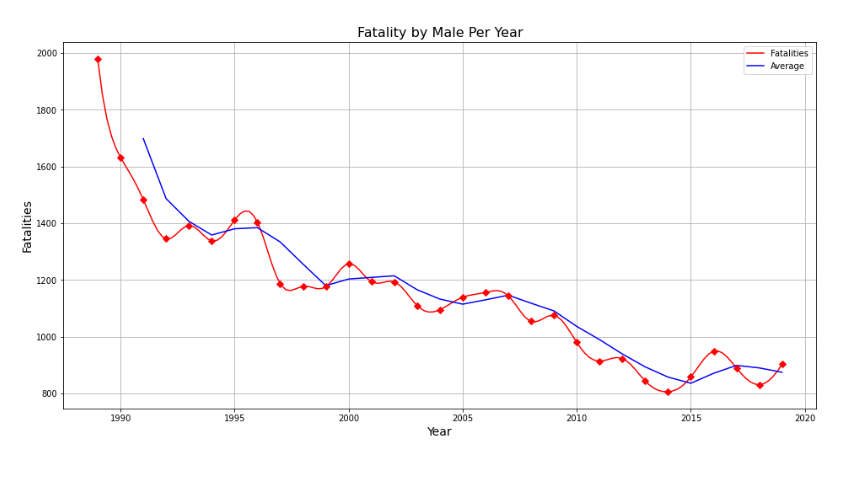


Figure 11. Fatality of Male per year

## Recommendation of Future Work

This report has identified several areas for the improvement of related work in the future. Firstly, within the datasets collected by ARDD, there are a considerable amount of missing values in several attributes which affects the integrity and accuracy of the result. Therefore, it is recommended that the recording of the dataset should be more thorough to improve the analysis of road crashes. Secondly, a new area of research which is called Machine Learning has been gradually introduced to this type of work. The clustering and association rules are two important machine learning techniques in finding patterns and categorising the similarities and dissimilarities which would be highly beneficial for the prevention of road accidents and the improvement of road safety. Thirdly, based on the results and findings, the speed limit was one of the significant factors for fatal crashes. It is recommended that more speed cameras should be installed in areas where accidents have been increased over the years. The recommendation also suggests that traffic rules should be regularly reviewed according to the improvement of public infrastructure and means of transportation.

## Conclusion

The land of Australia is vast, but the population is low. The demographic is therefore highly densified in one area and scattered in another. This report provides us with many insights in fatalities related to road crashes. It tells us that New South Wales has the highest fatality rate, pedestrians have the lowest fatality rate in terms of road crashes. In relation to vehicle involvement, Light vehicles contributed most road crashes which is four times larger than the rest of other vehicles combined.

One interesting finding associated with the speed limit was that the fatality rate between speed limit above 75km/h and below 75km/h were almost identical. There was only a 2% difference between the two and the fatality rate was gradually dropping on the speed limit below 75km/h due to modern speed cameras. In terms of when the road crashes are more likely to happen, the result indicates that during daytime it is 1.5 times higher than night. In addition, the fatality rate is greater during weekends. For example, from Friday to Sunday with Saturday having the maximum fatality.

Age group and gender are important indicators when it comes to an insurance premium. According to the analysis, the most vulnerable age group is between 51 and older because they contributed over 30% of fatal crashes over the years. Female drivers are almost three times likely to be involved in fatal crashes compared to their counterparts.

## References

1. Bureau of Infrastructure, Transport and Regional Economics. (2020). Australian Road Death Database (ARDD). Retrieved from <https://data.gov.au/dataset/ds-dga-5b530fb8-526e-4fbf-b0f6-aa24e84e4277>
2. Budget Direct. (2019). Car Accident Statistics 2019. Retrieved from <https://www.budgetdirect.com.au/car-insurance/research/car-accident-statistics.html>
3. Harish S., Kevin J., Vikash S., Willis L. (2020). Australian Fatal Crashes Report from 1989 to 2019. Retrieved from <https://www.kaggle.com/kevinjames1993/australian-fatal-crashes-report-from-2015-to-2020>