Data Analysis and Visualisation Tool

User Guide

<Group 98>

s5228571 Jaeseok Kim

s5274346 Mausham Kafle

s5105194 Sam Bate

Contents

[**Requirements** 3](#_Toc115979056)

[**Features** 3](#_Toc115979057)

[**Operating Instructions** 4](#_Toc115979058)

[Running the python file 4](#_Toc115979059)

[Using the Features 4](#_Toc115979060)

[Searching by selected period 5](#_Toc115979061)

[Visualising the number of accidents on average 6](#_Toc115979062)

[Searching by selected period and accident type 7](#_Toc115979063)

[Visualising the impact of alcohol in accidents 9](#_Toc115979064)

[Search by selected period and speed zone 10](#_Toc115979065)

[**Appendix – Concise Instructions** 11](#_Toc115979066)

[Searching by selected period 11](#_Toc115979067)

[Visualising the number of accidents on average 11](#_Toc115979068)

[Searching by selected period and accident type 11](#_Toc115979069)

[Visualising the impact of alcohol in accidents 11](#_Toc115979070)

[Search by selected period and speed zone 12](#_Toc115979071)

# **Requirements**

The Crash Statistics Victoria.csv file must be located within same path as the python (.py) file.

The data\_analysis.py file requires a python interpreter with *Pandas* and *MatPlotLib* modules installed to run it.

PyCharm Community Edition 2022.2.2 was used for the purposes of this User Guide.

# **Features**

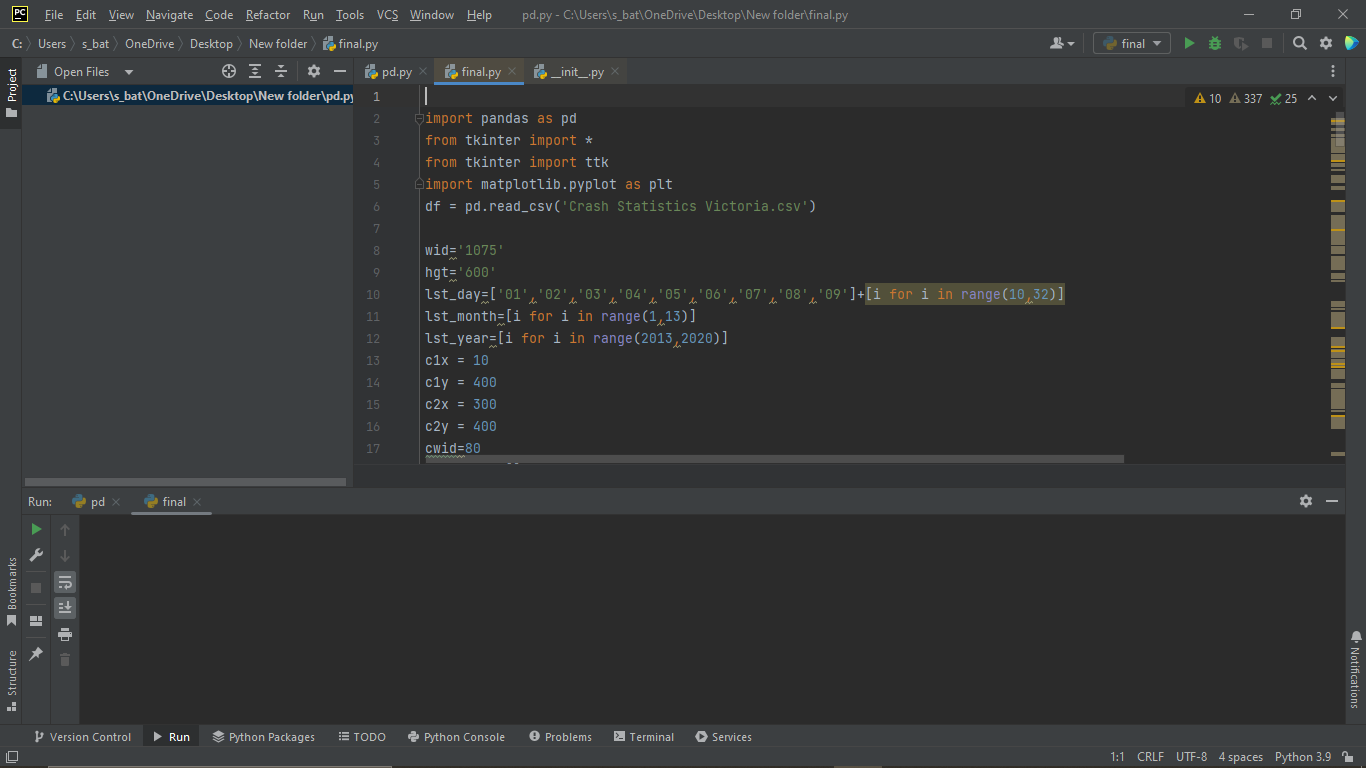
The Data Analysis and Visualization Tool has five search options. They are:

1. **Search by selected period** – a date search with no other filters applied.
2. **Visualise the number of accidents on average** – produces a bar chart to show the average number of accidents for each hour of the day over the date period selected.
3. **Search by selected period and accident type** – is a keyword search.
4. **Visualise the impact of alcohol in** **accidents** – produces two different line graphs to allow users to analyse the impact of alcohol in accidents for a selected date range.
5. **Search by selected period and speed** zone – filters the date search to show all the accidents which occurred within a specific speed zone, during the selected time period.

# **Operating Instructions**

## Running the python file

* Open your Python Interpreter



* Ensure your interpreter has the *Pandas* and *MatPlotLib* modules installed
* Load the file **data\_analysis.py**
* Click on **Run**

A screenshot of a computer

Description automatically generated

The Data Analysis and Visualisation Tool can take up to 60 seconds to load.

Graphical user interface, text

Description automatically generated

## Using the Features

The five search options of the Data Analysis and Visualization Tool are selected by radio buttons.

* **Search by selected period**
* **Visualise the number of accidents on average**
* **Search by selected period and accident type**
* **Visualise the impact of alcohol in** **accidents**
* **Search by selected period and speed zone**

The Tool also has

* **From** and **To** date selectors for performing date search
* free text field for performing a keyword search of Accident Types
* drop-down list of speed zones
* **Start** button

Graphical user interface, text

Description automatically generated

## Searching by selected period

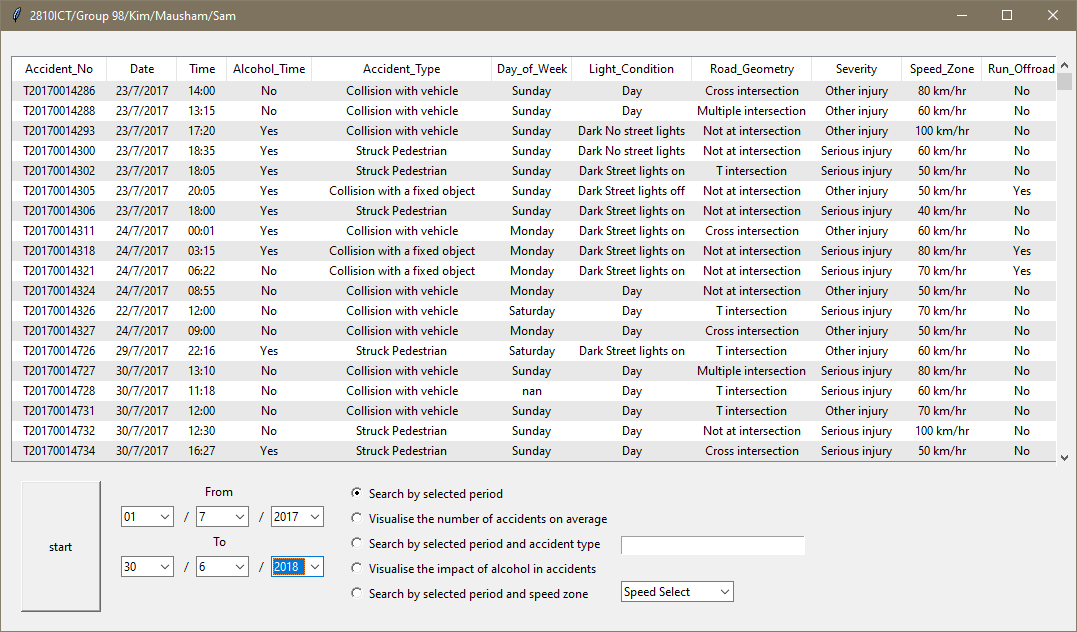
**Search by selected period** is the default search option, so Its radio button is already selected.

* Enter a **From** date and a **To** date by selecting the day, month and year from the drop-down lists.

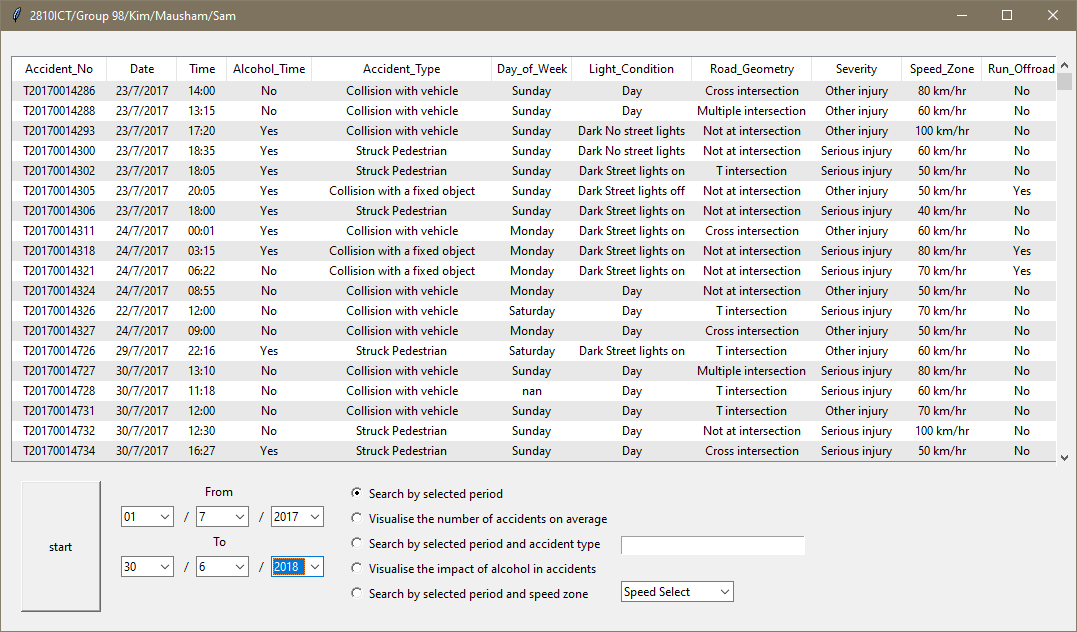
Graphical user interface, text

Description automatically generated

* Click on **Start**



The Data Analysis and Visualisation Tool displays *Accident\_No, Date, Time, Alcohol\_Time, Accident\_Type, Day\_of\_Week, Light\_Condition, Road\_Geometry, Severity, Speed\_Zone* and *Run\_Offroad* for all accidents in the Crash Statistics Victoria database for the selected time period.



**Search by selected period** accepts two inputs, an initial date and an end date and returns accidents where the Accident\_Date falls within that date range. Dates are entered by selecting the Day (1 to 31), Month (1 to 12) and Year (2013 to 2019) from drop-down lists.

The Data Analysis and Visualisation Tool allows users to select dates between 01/01/2013 and 31/12/2019 but the database contains data from 1/07/2013 to 1/02/2019 and can only return data within this range, even if dates outside this range are selected.

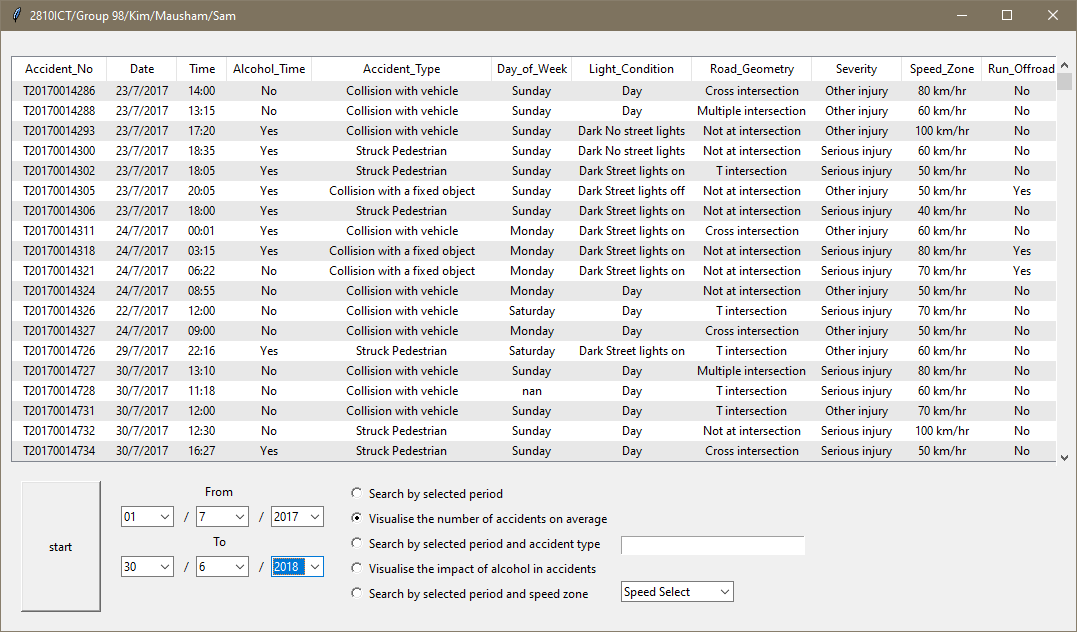
If both the selected initial and end dates fall outside the database range, no data will be returned. Likewise, a range with an end date earlier than the start date will also return no data.

It is possible to select dates which don’t exist (eg: 31/02/2014). In this instance, it will return data from the next, real date, after the non-existent date, to the last, real date before the non-existent date, which contain data.

If the selected date range falls within the database date range, and the selected dates do exist and the tool still returns no results, it means no accidents occurred in the selected time period.

## Visualising the number of accidents on average

* Select the **Visualise the number of accidents on average** radio button



* Enter a **From** date and a **To** date by selecting the day, month and year from the drop-down lists.

A screenshot of a computer

Description automatically generated with medium confidence

* Click on **Start**

Chart, histogram

Description automatically generated

**Visualise the number of accidents on average** plots a graph, showing the average number of accidents over each hour of the day, over the selected time period.

The graph has controls which (from right to left) save the graph, adjust subplot parameters, zoom in and out on the graph and pans left, right, up and down.

Users can scroll forwards and backwards between multiple views of the graph. The Home button resets the current view to the default.Chart, histogram

Description automatically generated

**Visualise the number of accidents on average** sorts the selected data by Accident\_Time and then graphs the results.

The Data Analysis and Visualization Tool can only generate charts if there is data to graph. If a date search returns no data, no chart will be generated.

## Searching by selected period and accident type

* Select the **Search by selected period and accident type** radio button

Graphical user interface, text

Description automatically generated

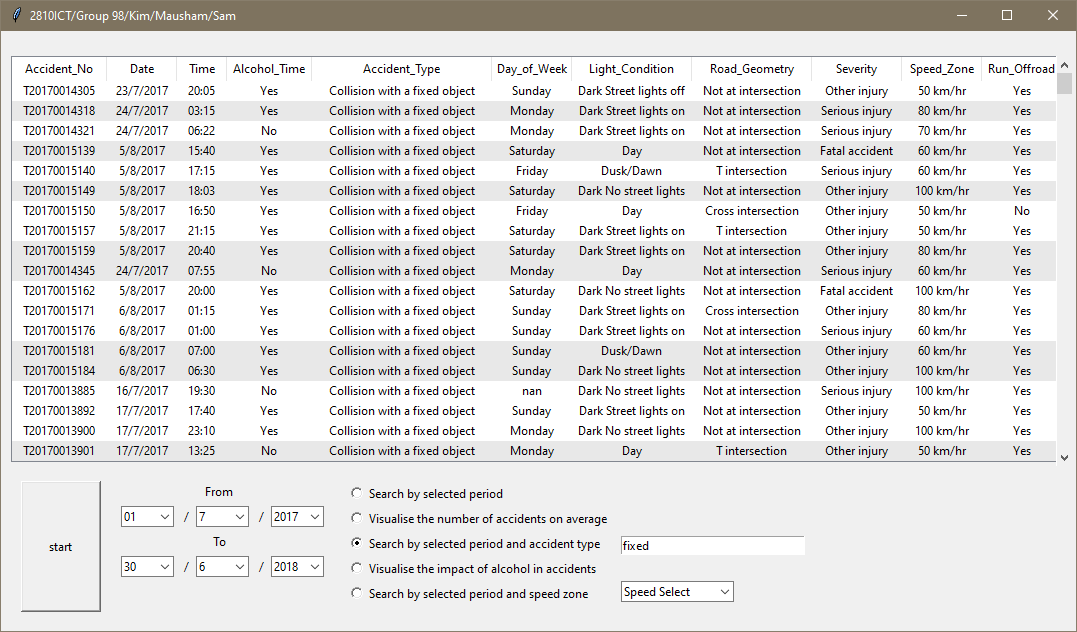
* Enter a **From** date and a **To** date by selecting the day, month and year from the drop-down lists.

Graphical user interface, text

Description automatically generated

* Type a keyword in the free text field to search for Accident Type

The Keyword search is case sensitive



* Click on **Start**

A screenshot of a computer

Description automatically generated with medium confidence

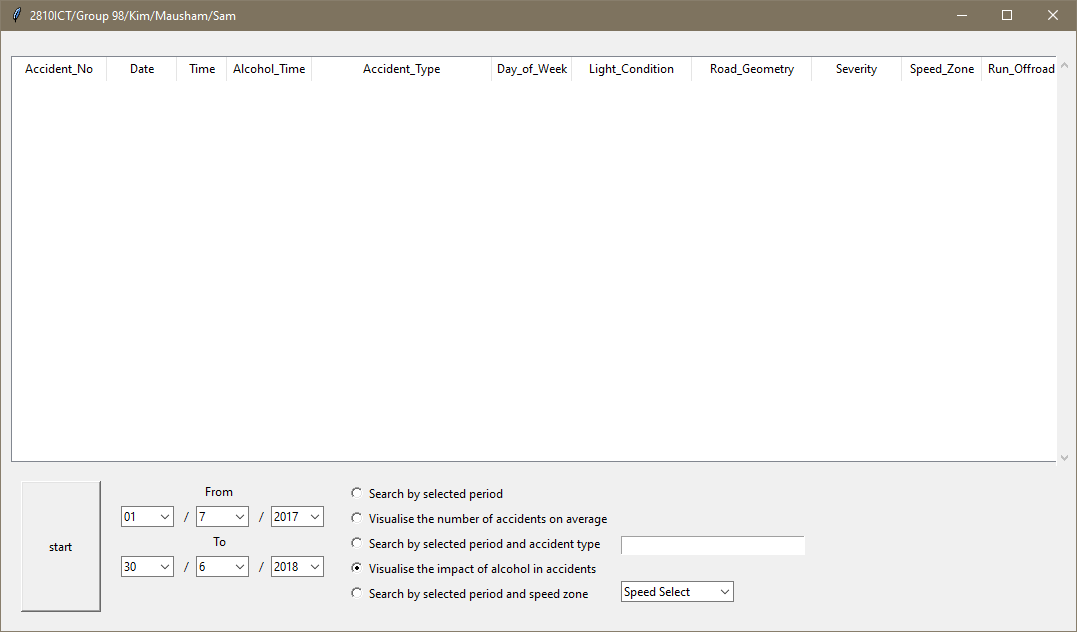
The keyword search returns *Accident\_No, Date, Time, Alcohol\_Time, Accident\_Type, Day\_of\_Week, Light\_Condition, Road\_Geometry, Severity, Speed\_Zone* and *Run\_Offroad* for all accidents in the Crash Statistics Victoria database for the selected time period, where *Accident\_Type* contains the Keyword.

**Search by selected period and accident type** filters the date search to retrieve all accidents caused by an accident type that contains a user selected keyword (e.g. collision, pedestrian). Option 3 accepts input from the user which is saved as a tuple. It then uses the tuple to search the Accident\_Type field for the date range selected.

If the date search returns no data, the keyword search will also return no data as there is nothing to perform the keyword search on. The keyword search will also return no data if the keyword does not appear in the Accident\_Type field.

## Visualising the impact of alcohol in accidents

* Select the **Visualise the impact of alcohol in accidents** radio button



* Enter a **From** date and a **To** date by selecting the day, month and year from the drop-down lists.

Graphical user interface, text

Description automatically generated

* Click on **Start**

Chart, line chart

Description automatically generated

**Visualise the impact of alcohol in accidents** produces two different line graphs to allow users to analyse the impact of alcohol in accidents for a selected date range. One graph shows the daily, hour-by-hour trends of accidents where alcohol was a factor compared against accident where alcohol wasn’t a factor.

The graph has controls which (from right to left) save the graph, adjust subplot parameters, zoom in and out on the graph and pans left, right, up and down.

Users can scroll forwards and backwards between multiple views of the graph. The Home button resets the current view to the default.Chart, histogram

Description automatically generated

**Visualise the impact of alcohol in accidents** sorts the selected data by Accident\_time and graphs the results. The other graph shows the weekly, Monday to Sunday trends of accidents where alcohol was a factor compared against accident where alcohol wasn’t a factor. This graph is generated by sorting the selected data by Day\_of\_the\_Week.

This works in much the same way as Visualise the number of accidents on average and has the same requirements and restrictions.

## Search by selected period and speed zone

* Select the **Search by selected period and speed zone** radio button

Graphical user interface, text

Description automatically generated

* Enter a **From** date and a **To** date by selecting the day, month and year from the drop-down lists.

Graphical user interface, text

Description automatically generated

* Select Speed Zone from the drop-down list

Graphical user interface, text

Description automatically generated

* Click on **Start**

Graphical user interface, text

Description automatically generated

**Search by selected period and speed zone** returns *Accident\_No, Date, Time, Alcohol\_Time, Accident\_Type, Day\_of\_Week, Light\_Condition, Road\_Geometry, Severity, Speed\_Zone* and *Run\_Offroad* for all accidents in the Crash Statistics Victoria database for the selected time period, where *Speed\_Zone* matches the selected speed zone.

**Search by selected period and speed zone** This option accepts user input, in the form of a selection from a drop-down list of all the different speed zones enforced throughout Victoria. It saves the selection as a tuple and returns all the accidents within the selected data where Speed\_Zone matched the tuple.

If the date search returns no data, the Speed\_Zone search will also return no data as there is nothing to perform the Speed\_Zone search on. The Speed\_Zone search will also return no data if there are no accidents with the selected Speed\_Zone during the selected time period.

# **Appendix – Concise Instructions**

## Searching by selected period

**Search by selected period** is the default search option, so Its radio button is already selected.

* Enter a **From** date and a **To** date by selecting the day, month and year from the drop-down lists.
* Click on **Start**

**Search by selected period** displays all accidents in the Crash Statistics Victoria database for the selected time period.

## Visualising the number of accidents on average

* Select the **Visualise the number of accidents on average** radio button
* Enter a **From** date and a **To** date by selecting the day, month and year from the drop-down lists.
* Click on **Start**

**Visualise the number of accidents on average** plots a graph, showing the average number of accidents over each hour of the day, over the selected time period.

## Searching by selected period and accident type

* Select the **Search by selected period and accident type** radio button
* Enter a **From** date and a **To** date by selecting the day, month and year from the drop-down lists.
* Type a keyword in the free text field to search for Accident Type

The Keyword search is case sensitive

* Click on **Start**

**Search by selected period and accident type** returns *Accident\_No, Date, Time, Alcohol\_Time, Accident\_Type, Day\_of\_Week, Light\_Condition, Road\_Geometry, Severity, Speed\_Zone* and *Run\_Offroad* for all accidents in the Crash Statistics Victoria database for the selected time period, where *Accident\_Type* contains the Keyword.

## Visualising the impact of alcohol in accidents

* Select the **Visualise the impact of alcohol in accidents** radio button
* Enter a **From** date and a **To** date by selecting the day, month and year from the drop-down lists.
* Click on **Start**

**Visualise the impact of alcohol in accidents** produces two different line graphs to allow users to analyse the impact of alcohol in accidents for a selected date range. One graph shows the daily, hour-by-hour trends of accidents where alcohol was a factor compared against accident where alcohol wasn’t a factor.

## Search by selected period and speed zone

* Select the **Search by selected period and speed zone** radio button
* Enter a **From** date and a **To** date by selecting the day, month and year from the drop-down lists.
* Select Speed Zone from the drop-down list
* Click on **Start**

**Search by selected period and speed zone** returns all accidents in the Crash Statistics Victoria database for the selected time period, where *Speed\_Zone* matches the selected speed zone.