FIT5147 Data Exploration and Visualisation

Semester 1, 2024

Programming Exercise 1: Tableau (5%)

Please carefully review all the requirements below to ensure you have a good understanding of what is required for your assessment.

1. **Instructions & Brief**

2. **Assessment Resources**

3. **Assessment Criteria**

4. **How to Submit**

5. **Word Count & Penalties**

---

1. Instructions & Brief

In this assignment you are required to read in some data and explore and visualise it using Tableau Public/Desktop, then submit a brief report showing your findings and the visualisations you used. **It is an individual assignment and worth 5% of your total mark for FIT5147.**

Relevant learning outcomes for FIT5147:

1. Perform exploratory data analysis using a range of visualisation tools;

6. Implement interactive data visualisations using R and other tools

Data Source:

The data set used in the assignment is based on the Vehicle Road Crash Data released by the Victorian Department of Transport and Planning in January 2024. However, we highly recommend you do not explore the original data or related studies as part of this assessment task as the files used for the S1 2024 programming exercise assignments have been heavily edited especially for this unit. This includes omitting some fields and rows, adding full or partial data, and editing the values.

Details of task:

For this PE1 assignment, the resulting data describes when and where accidents occurred in Victoria, relating to the location, date and time, accident type, severity, light conditions, road geometry and speed zone.

Column	Description
ACCIDENT_NO	Unique Id of accident
ACCIDENT_DATE	Date of Accident
ACCIDENT_TIME	Time of Accident in local time
DAY_OF_WEEK	Day of Week (numerical)
DAY_WEEK_DESC	Day of Week Name
ACCIDENT_TYPE	code for type of accident
ACCIDENT_TYPE_DESC	Description for type of accident
SEVERITY	Estimation of the severity or seriousness of the accident (1 Fatal accident 2 Serious injury accident 3 Other injury accident 4 Non injury accident)
LIGHT_CONDITION_DESC	Description of lighting conditions
LGA_NAME	Name of Local Government Area of location of accident e.g. DANDENONG
DEG_URBAN_NAME	Category for degree of urban
ROAD_GEOMETRY	Code for layout of the road where the accident occurred
ROAD_GEOMETRY_DESC	Description for the layout of the road where accident occurred
SPEED_ZONE	Speed zone of location where accident occurred
LATITUDE	Geographical Location (latitude)
LONGITUDE	Geographical Location (longitude)

Table 1: Fields of the "Victoria_Accident_Data_FIT5147S12024PE1" data set

For you, the most crucial change to the data is that **at least three types of data irregularities or errors have also been included in the data (you will only get marks for three)**. One of the requirements of this assignment is for you to find (using data visualisation), describe and handle them. This modified dataset can be found on Moodle in the Assessments section under the Programming Exercises heading.

The task has two components: **data exploration using Tableau**, and a short **written report**.

Data Exploration using Tableau
You are expected to:

1. Load the dataset in Tableau Public/Desktop

2. Use data visualisation to **check for and find** the three aforementioned irregularities in the dataset. Each type of irregularity may occur multiple times in the data.

3. Amend the data to **correct** these errors using any tool of your choice (e.g., Excel, Python, R, Tableau)

4. **Using visual analysis**, answer the following questions:

   o Q1: Compare and contrast when and what type of accidents occur over time (2012-2024). Consider different time periods: hour of day, day of week, month of year, year in the dataset. What do you notice about the pattern over time? What do you know about that time period in order to provide some explanation of increases, decreases, and similar numbers of accidents.

   o Q2: Compare and contrast where accidents occur geographically. Consider in particular the geometry, speed zone of the road and the urban/rural aspect of the location. How does the spatial mapping of the data and additional information about the road help support, challenge or change your conclusions to the first question?

5. Write a report that **describes** your data exploration process and visual analysis (see below for details).

Note: UNKNOWN/unknown, NULL/null, N/A, tba values should not be regarded as irregularities for this assignment. If Tableau has any issues automatically recognising any date or time information, this is not to be regarded as an irregularity.

Written Report
Once you have finished your data exploration, write a report that contains the following information:

1. Data checking and cleaning (i.e., Steps 1 to 3)

   o An **image** to show what your data looks like after loading it in Tableau

   o A brief explanation (maximum of **one paragraph per error**) and an accompanying **image of each of the errors** or irregularities that you have found, showing how you found them **using Tableau**, and **explaining & justifying** how you resolved them. The image must show a **relevant visualisation**, not just the data or a table.

2. Data exploration (i.e., Step 4)

   o Your answers to Q1 and Q2 with accompanying **images of the data visualisations** that you used to support your analysis, and a brief **explanation** of why you have used those type of visualisations.

The report should also:

- Be submitted as a **PDF file**
- Be **no more than 5 pages** in length, including figures, with a minimum font size of 10 (title page and any table of contents are excluded from the page limit)
- Be **properly structured** with headings, subheadings, figure captions (in-text referencing of captions), page numbers, and references (where appropriate)
- Have **high quality images** of your visualisations with clearly readable and legible text/labels (presume that it is read as part of an A4 document with no zooming).
- You must use proper **academic referencing** for all reports in this unit. This should follow either the APA or IEEE structure as recommended by the Faculty. Use the library referencing guide for support.
- **Not include any code snippets**
- **No Generative AI** software or system may be used to complete this assessment task. This includes using any software that paraphrases, translates or rewrites your text.

2. Assessment Resources

- **Victoria_Accident_Data_FIT5147S12024PE1.csv** (Available on Moodle)

3. Assessment Criteria

The following outlines the criteria which you will be assessed against.

- Demonstrated ability to check and clean data and read into Tableau [1%]

- Demonstrated ability to visually explore data using Tableau [2%]

- Demonstrated ability to see trends/patterns in data [1%]

- Quality of report [1%]

As part of the grading process, mandatory interviews to discuss your submission will occur during the Applied Session in Week 5. If the interview is not successfully attended, you will not receive any mark or feedback for this assessment.

Whilst the assignment can be done using Tableau Public, we recommend you download Tableau Desktop and get a student license in order to be able to save your project.

4. How to Submit

Once you have completed your work, take the following steps to submit your work.

1. Save your report as .pdf.

2. Name your file using the following structure **PE1_Surname_StudentID**

3. Click the **Add Submission** button below to submit and upload your report

Please note that your assignment MUST show a status of "Submitted for grading" before it can be marked. Any submission left in draft mode will not be marked. We recommend always double checking your submission has been completed and that you have uploaded the correct file. Penalties will apply to any submission which needs amendment after the deadline.

5. Word Count & Late penalty

The report must not be more than 5 pages of graded material including figures (min. font size 10). Up to 2 additional pages may be used if you wish, but restricted to:

- 1 page prior to the report as a title page with a table of content.
- 1 page after the report only for references.

1 mark will be deducted if the report does not meet these requirements.

The penalty of 0.5 marks (out of a total of 5 marks) per day will be applied for all late submissions, including weekends.