

ASSESSMENT 3 BRIEF	
Subject Code and Title	DMV302 Data Mining and Visualisation
Assessment	Data visualisation presentation
Individual/Group	Individual
Length	4–6 minute presentation plus source code
Learning Outcomes	<p>The Subject Learning Outcomes demonstrated by the successful completion of the task below include:</p> <ul style="list-style-type: none"> c) Identify key components of the computing environment for developing data mining models. d) Apply data mining techniques to generate statistical predictions. e) Articulate the outcomes of data mining models to a diverse range of stakeholders.
Submission	Due by 11.55 pm AEST/AEDT on the Sunday at the end of Module 12.
Weighting	30%
Total Marks	100 marks

Context

In Weeks 10 and 11, you explored a number of data visualisation techniques and learnt to programmatically create diagrams, plots and graphs that depict high-level information about and provide insights into data sets. Appropriate visualisation helps data analysts, miners and decision makers to identify patterns or trends that are not apparently manifested in raw data sets. In this assessment, you will demonstrate your ability to:

- Visualise complex data sets using programmatically generated plots and diagrams;
- Choose the best visualisation for a certain data set given its characteristics; and

Integrate data visualisation techniques with data mining algorithms.

Task Summary

You are to individually complete two data visualisation tasks. The first requires you to programmatically create two specific diagrams for a three dimensional (3D) data set. The second requires you to be creative and programmatically visualise a data set that contains historical weather data. You are also required to deliver a 4–6 minute presentation explaining your choice of diagram, discussing the patterns, insights and information that you have drawn from your visualisation and reflecting on your experience in this assessment.

Please refer to the Task Instructions for details on how to complete this task.

Task Instructions

This assessment has two tasks. The first task requires you to visualise a specific data set in two designated ways, while the second task requires you to be as creative as possible and choose the best way to visualise a specific data set.

Task 1 (12 marks)

Download a comma separated values (csv) file titled `DMVA3T1.csv`. Each row of this file contains a 3D data point in $[-100,100]^3$.

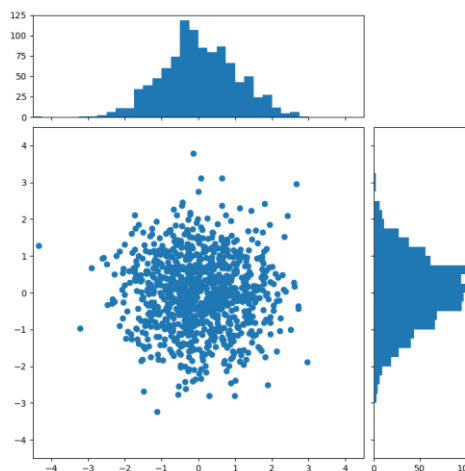


Figure 1. An example of a scatter plot with two histograms

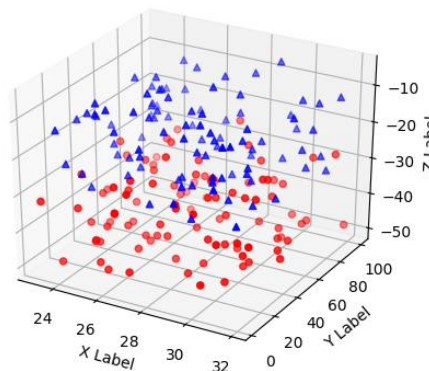


Figure 2. An example of a 3D scatter diagram with clusters

(a) Implement a Python program that reads in the `DMVA3T1.csv` file, and programmatically create a scatter plot with two histograms (see Figure 1). **(6 marks)**

- Project the 3D data (x,y,z) onto a two-dimensional (2D) plane. If your student ID ends with 0, 3, 6 or 9, then project each data point onto a x-y plane; if your student ID ends with 1, 4 or 7, then project it onto a x-z plane; otherwise, project it onto a y-z plane.



- Plot a scatter diagram, then project the 2D data onto each dimension and create a histogram for each dimension.
- All three plots are in ONE graph.
- Label all three plots appropriately.
- You can import `numpy`, `scipy`, `csv`, `pandas`, `matplotlib` or any other library that your facilitator approves.
- Save the source code as, `'DMV302T1a.py'`, and save the graph as a PNG file `'DMV302T1a.png.'`

(b) Implement a Python clustering program that reads in the `DMVA3DS1.csv` file, and programmatically create a 3D scatter plot with the clusters (see Figure 2). **(6 marks)**.

- You can use any clustering algorithm and can make library calls for the clustering algorithm (e.g., `sklearn.cluster.KMeans`)
- Set the number of clusters to 3.
- Draw a 3D scatter diagram illustrating the three clusters. Use different shapes **and** colours to distinguish data points in each cluster.
- Label your diagram appropriately.
- You can import `numpy`, `scipy`, `sklearn`, `csv`, `pandas`, `matplotlib` or any other library that your facilitator approves.
- Save the source code as, `'DMV302T1b.py'`, and save the graph as a PNG file `'DMV302T1b.png'`.

Task 2 (18 marks)

Download a comma separated values (csv) file titled `DMVA3T2.csv`. This file contains weather history data for a city for a whole financial year (1/7 to 30/6). The table below sets out the column numbers and provides a brief description for each column.

Column number	Description
1	The data of the record; formatted dd/mm/yy
2	The actual average temperature for that day
3	The actual minimum temperature for that day
4	The actual maximum temperature for that day
5	The average minimum temperature on that day since 1901
6	The average maximum temperature on that day since 1901
7	The lowest temperature on that day since 1901
8	The highest temperature on that day since 1901
9	The actual amount of rain on that day
10	The average amount of rain on that day since 1901
11	The highest amount of rain on that day since 1901

(a) Be as creative as possible. Write a Python program to programmatically create diagram(s) to illustrate the given data set with the aim of capturing as much **information** and **insights** as possible (not just dumping data on a graph). **(10 marks)**

- You can import `numpy`, `scipy`, `sklearn`, `csv`, `pandas`, `matplotlib` or any other library that your facilitator approves.



- Label your diagram or diagram(s) appropriately.
- There is no requirement as to the number of diagrams you should create. However, you should aim to use as few diagrams as possible, each capturing as much information as possible.
- You may use any data mining algorithms to assist you to visualise this data set.
- Save the source code as, 'DMV302Tb.py' .

(b) Now prepare a 4–6 minute presentation using PowerPoint slides (.ppt or .pptx). Your presentation should contain all the diagrams you created in (a). In your presentation, you should:

- Explain your diagram or diagrams and why you chose to visualise the data set in the way you did. Your discussion should be connected to the characteristics of the data set. **(This should take about 2 minutes; 3 marks).**
- Discuss the insights, knowledge, patterns or information that can be obtained directly from your visualisation of this data set. **(This should take about 2 minutes; 3 marks).**
- Reflect on your experience in this task, and discuss the challenges you faced in the implementation. **(This should take about 1 minute; 2 marks).**

General assessment instructions

You will need to record your presentation and submit the recordings of your presentation. Here are the instructions:

- You may use any software/device to record your presentation. However, it is highly recommended that you use Zoom or an equivalent online conferencing software (e.g., MTeams) to record your presentation. The resulting video file is normally small in size. If you use your mobile phone to record the presentation, the file size may be too large for you to submit it into Moodle.
- If you find that you cannot submit your recording into Blackboard due to file size restrictions, you may upload your recording to YouTube (or another online video platform) or Google Drive (or other cloud storage services) and submit the link to the video/file to Blackboard. You must set up the video or file in a way that allows the tutor to view/download your recording at the time they mark your assessment.
- You must show your face during the presentation for authentication purposes.
- Your presentation must not exceed 6 minutes. The marker will stop watching your recording after 6 minutes.

All Python programs in this assessment should comply with the following requirements:

- Follow Python naming conventions. If in doubt, refer to the *Google Python Style Guide* at <https://google.github.io/styleguide/pyguide.html>;
- Be implemented with Python 3 (**NOT** Python 2.x);
- Be implemented with any IDE of your choice;
- Contain adequate and useful comments; and
- Be thoroughly tested and not contain any syntax or logical errors.

Referencing

It is essential that you use appropriate APA style for citing and referencing research. For more information on referencing, visit: <https://library.torrens.edu.au/academicskills/apa/tool>.

Submission Instructions

Submit this task via the **Assessment 3** link in the main navigation menu in DMV302 Data Mining and Visualisation. The Learning Facilitator will provide feedback via the Grade Centre in the LMS portal. Feedback can be viewed in My Grades.

Academic Integrity Declaration

I declare that except where referenced, the work I am submitting for this assessment task is my own. I have read and am aware of the Torrens University Australia Academic Integrity Policy and Procedure viewable online at <http://www.torrens.edu.au/policies-and-forms>

I am also aware that I need to keep a copy of all submitted material and their drafts, and I agree to do so.

