



Assessment Brief Proforma

1. Module number	SET09120
2. Module title	Data Analytics
3. Module leader	Taoxin Peng
4. Tutor with responsibility for this Assessment Student's first point of contact	Thomas Methven
5. Assessment	Practical coursework Part 1
6. Weighting	20% of module assessment
7. Size and/or time limits for assessment	6 Pages (More details in coursework brief)
8. Deadline of submission	Your attention is drawn to the penalties for late submissions Submission: 1 st November 2019 before 1500
9. Arrangements for submission	Upload your PDF (more details given later) to the Moodle submission page.
10. Assessment Regulations	All assessments are subject to the University Regulations

11. The requirements for the assessment	<i>See following document</i>
12. Special instructions	<i>None</i>
13. Return of work and feedback	<i>Marks and general feedback will be provided by the end of week 11</i>
14. Assessment criteria	<i>See following document.</i>

Data Visualisation Coursework (SET09120)

Thomas Methven

School of Computing, Edinburgh Napier University, Edinburgh
t.methven@napier.ac.uk

Abstract. In this coursework, you will have to explore a dataset using visualisation, find relevant patterns, creating visualisations showing these patterns, and then explain what these patterns actually mean. You will also be required to optimise one of these visualisations in order to make the pattern discovered obvious to the reader.

Finally, you will need to submit a report that contains your findings and visualisations. This report will be in the LNCS Style, and will be no longer than six (6) pages, with only three (3) pages of text, and three of images. The template for this is available on your Moodle page.

More information about this coursework is presented in this document, which you should make sure you read in detail before beginning. If you fail to follow the required style, you might lose marks!

Keywords: You won't need these!

1 Overview

Participants from five locations have been asked to complete three different tasks. These tasks have not been explained to you, but you know that each task can return a score from between 0 and 100, as they are measured in percentages. In addition, each participant was timed for how long it took them to complete all three tasks. Finally, some basic demographic information for each of them was recorded. Bear in mind that this information has been anonymised so you, for example, do not know what the tasks are or where the locations are. Instead, you only know there are three different tasks and five different locations where they could have taken place.

You have been provided a dataset which contains 10,000 of these measurements, and have been asked to identify any trends, relationships, or other interesting information or patterns in it. You are also expected to understand what this data means, so you can state these patterns in a sensible way. In other words, rather than saying 'Y is greater than M' you will be expected to say 'Young Adults are, in general, more successful at Part A than Middle Aged participants'. This example is, unfortunately, not one of the expected patterns.

In order to help with this, the following metadata information has been provided:

Table 1. A table showing the metadata for the dataset you have been provided.

Column Name	Data Description
Gender	M or F
Location	A, B, C, D, E
Age.Band	U, Y, M, O (Under 16 (< 16), Young Adult (16 - 39, Middle Aged (40 – 64), Older Adult (65+))
Age	Numeric Value
Part.A.Score	Percentage (0 to 100)
Part.B.Score	Percentage (0 to 100)
Part.C.Score	Percentage (0 to 100)

There are also two classes of outliers in the data that you are expected to find, highlight, and report on. Remember: each class of outlier will have more than one row of data associated with it. In addition, if these outliers are making your other visualisations more difficult to understand, you are expected to remove said data before creating the visualisations.

More information on how to report these is provided in the next section.

2 Report

Your report **must** use the Springer's Lecture Notes in Computer Science (LNCS) template. Templates for Latex and Word and provided on the Moodle page for this module. The page limit is **six (6) pages**, of which the first three contain only text, and the final three will contain your visualisations with appropriate captions. Deviation from either the template or the page limits will result in a penalty on your coursework. For reference, this document is in LNCS format.

Your report should include the following sections:

1. *Abstract* – This section, if read alone, should give the reader a clear, albeit brief, picture of what you have done and found.
2. *Description of the outliers found* – Through your exploration of the data, you should find some outliers in the data. In this section you should discuss why you believe these data are outliers, and provide suitable visualisations to show this. These should be included as part of your three page image allowance at the end of the document, and you should use appropriate internal referencing to make them easy to find.
Note: There are two classes of outlier in the data.
3. *Description of the relationships found* – Through your exploration of the data, you should find some interesting relationships between some of the attributes. In this section you should say what you found and how you found them. Include visualisations as part of your three page image allowance at the end of the document, and use appropriate internal referencing to make them easy to find. **Note: There are five relationships in the data**

4. *Most interesting relationship* – Pick one of the relationships from the previous section to focus upon. You are going to optimise the visualisation of this relationship so that a reader can easily identify it. In this section, you have to explain *how* and *why* your optimisations are effective, with reference to the literature. Include your optimised visualisation as part of your three page image allowance, with appropriate internal referencing.

The report must be separated so that all of the written sections come first, followed by all the images. This separation is so the marker can verify you have adhered to the page limits. All included figures must have a caption, and must be referred to at some point in the text, ‘e.g. See Fig. 2’.

If you submit a visualisation as evidence which cannot be read easily due to resolution issues, bad formatting, or any other problem, that visualisation WILL NOT be marked.

3 Submission

The only thing that should be submitted is the pdf version of your report. Do not submit Word or TEX files. The name of the pdf should be your student number. The submission point will be made available on Moodle.

Please note: The relationships have been designed to be different difficulties to find. In other words, some will be more difficult to find than others. Also, when you are selecting the relationship to optimise, you might want to select one of the more complicated relationships, otherwise you will struggle to find much to talk about!

4 Marking Scheme

The coursework will be marked out of 20, and will be worth 20% of your overall grade for the SET09120 Data Analytics module. The breakdown of the marks will be as follows:

1. *Outliers Found (4 marks)* – Have you found the two classes of outliers that were in the dataset? Did you show this with the correct type of visualisation? Did you remove the outliers if appropriate?
2. *Relationships found (10 marks)* – Have you found all five relationships in the dataset? Do your included visualisations provide evidence that you found the relationship? Did you pick the most sensible visualisation as evidence? Bear in mind that some of these relationships are harder to find than others!
3. *Optimised visualisation (6 marks)* – Is your chosen visualisation optimisation effective? Have you justified why you have optimised in the way you have? How much optimisation was even required? Did you pick a sensible relationship to optimise?
4. *Quality of report (-5 marks)* – adherence to the template, quality of written English, quality of images etc. **Please note:** there is a maximum penalty of five marks for issues in this area.

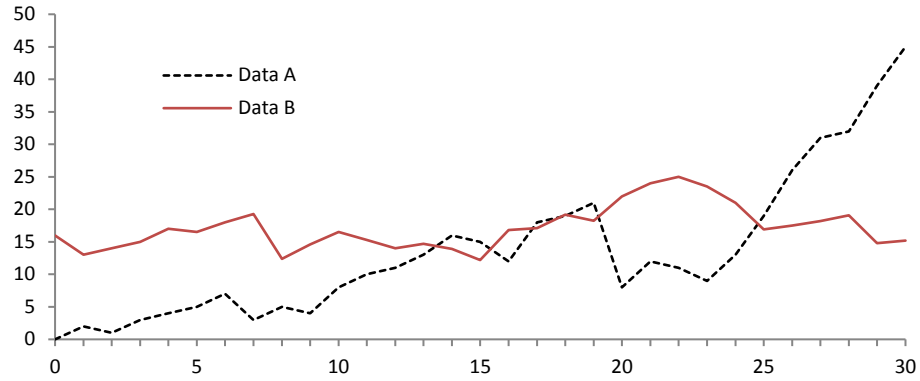


Fig. 1. This is an example visualisation from the template! Remember, any visualisations you create should be on your final three pages only, should have a caption like this, and should be referred from within the text.

References

1. Author, F.: Article title. Journal 2(5), 99–110 (2016).
2. Author, F., Author, S.: Title of a proceedings paper. In: Editor, F., Editor, S. (eds.) CONFERENCE 2016, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016).
3. Author, F., Author, S., Author, T.: Book title. 2nd edn. Publisher, Location (1999).
4. Author, F.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010).
5. LNCS Homepage, <http://www.springer.com/lncs>, last accessed 2016/11/21.