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	*Peter Chapman*
5. Assessment	*Practical coursework part I*
6. Weighting	*20% of module assessment*
7. Size and/or time limits for assessment	*None*
8. Deadline of submission	Your attention is drawn to the penalties for late submissions *24th October 2018 before 2355*
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	*See following document*
12. Special instructions	*None*
13. Return of work and feedback	*General feedback will be provided by the end of week 9.*
14. Assessment criteria	*See following document.*

Coursework - Data Exploration

Peter Chapman

SET09120

Abstract. In this coursework, you will have to explore a data set using visualisation, and produce an optimised representation of one relationship in the dataset. There are many relationships in the dataset, but some are harder to find than others. A report, containing your findings and visualisations, must be produced that is no longer than 6-pages in the LNCS style template: up to 3-pages for your written findings, and up to 3-pages for your charts. The templates for the LNCS style are made availabe for you on Moodle.

1 Overview

Athletes from 5 locations have been competing in a multi-discipline competition. The disciplines are categorised as *sprints*, *jumps*, *throws* and *distance*. In each discipline, athletes have been assigned a score out of 100. In addition, they have been assigned an overall score, again out of 100. Their age, as one of four bands, has been recorded, along with their gender. You have the dataset containing all of this information, and you have been asked to identify any trends, relationships, or other interesting information in it. In summary, each entry in the dataset contains:

- Gender (can be *M* or *F*)
- Location (can be *A*, *B*, *C*, *D* or *E*)
- Age. The age is in four bands, **T**eenager, **U**nder 25 (but over 20), **L**ate 20s, and **O**ver 30.
- Sprints. Score out of 100.
- Jumps. Score out of 100.
- Throws. Score out of 100.
- Distance. Score out of 100.
- Overall. Score out of 100.

2 Report

Your report *must* use Springer's Lecture Notes in Computer Science (LNCS) template. Templates for LaTeX and Word are provided on the Moodle page for the module. The page limit is **6 pages**, of which (the first) 3 contain only text, and 3 which can contain only images. 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, if brief, picture of what you have done and found.
2. *Description of the relationships found* - Through your exploration of the data, you should find some interesting links between some of the attributes of the dataset. In this section, you should say what you found, and how you found them. Include the visualisations as part of your 3 page image allowance at the end of the document, and use appropriate internal referencing to make them easy to locate. Note: there are 7 relationships in the data.
3. *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 3 page image allowance, with appropriate internal referencing.

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

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.

4 Mark Scheme

The coursework will be marked out of 50.

> **Any deviation from the template or page limits will limit your mark to 32. Note that all text must come before all of the images.**

The breakdown of marks will be as follows:

- *Relationships found* **20 marks** - have you found all 7 relationships that were in the dataset? Not all relationships are worth the same number of marks: those that involve more variables are worth more. Do your included visualisations provide evidence that you have found the relationships?
- *Optimised visualisation* **20 marks** - is your chosen visualisation optimisation effective? Have you justified why you have optimised in the way you have?
- *Quality of written report* **10 marks**- adherence to the template, quality of written English, quality of any displayed images, etc.