Edinburgh Napier UNIVERSITY

School of Computing, Napier University

Assessment Brief

1. Module number	**SET11121 002**
2. Module title	**Data Wrangling**
3. Module leader	*Yanchao Yu*
4. Tutor with responsibility for this Assessment	Yanchao Yu *(Y.Yu@napier.ac.uk)* *Jeff Mitchell (J.Mitchell@napier.ac.uk)*
5. Assessment	Coursework
6. Weighting	30% of module assessment
7. Size and/or time limits for assessment	up to 500 words **plus** tables with results and developed code. Code and figures are **not** included in the word limit.
8. Deadline of submission Your attention is drawn to the penalties for late submission	04/03/22 at 1500 UK time
9. Arrangements for submission	Your Coursework must be submitted via Moodle. **Further submission instructions are included in the attached specification and on Moodle**
10. Assessment Regulations	All assessments are subject to the University Regulations**.**
11. The requirements for the assessment	See Attached
12. Special instructions	See Attached
13. Return of work	Feedback and marks will be provided **within three weeks** of submission.
14. Assessment criteria	Your coursework will be marked using the marking sheet attached as Appendix A. This specifies the criteria that will be used to mark your work. Further discussion of criteria is also included in the coursework specification attached.

Assessment Brief

The assignment aims to cover the learning outcomes specified for the module:

LO1: Critically evaluate the tools and techniques of the data storage, interfacing, aggregation and processing
LO2: Select and apply a range of specialised data types, tools and techniques for data storage, interfacing, aggregation and processing
LO3: Employ specialised techniques for dealing with complex data sets
LO4: Design, develop and critically evaluate data-driven applications in Python

The goal of this assignment is to extract information from a provided dataset.

Data

For this assignment, you **must** use the dataset provided on Moodle. *The use of other datasets will result in failing the module.*

Deadline: Friday 4th of March at 3 pm (UK time).

Deliverable: You will need to use the dataset provided on Moodle to answer the following questions based on data. For each question, you will need to provide the Python code as well as a short description of why you solved the problem using the techniques/tools/libraries you chose and a general discussion of your results.

1) Create a DataFrame that consists of the following columns: (1) location; (2) average deaths for 2020; (3) average deaths of 2018; (4) the difference between average deaths of 2020 and 2018. The *location* column should include the name of the respective country.
2) For every country/location, find the *top-5 weeks* with the lowest number of excess deaths (Excess deaths are defined as the difference between the deaths that occurred in a specific week in 2020 vs 2018).
3) Find all the countries showing a positive excess (e.g., the diff between 2020 and 2018 greater than 0);
4) For each country/location, find which year had the highest mortality.
5) For every week, show the country/location with the lowest mortality.

Tips and Clarifications

You **must** use Python 3 and its libraries taught in the class to tackle these tasks.

You will submit

1. The code of your solution is a Python notebook, and an up to 500 words **.pdf document** explaining your solution.

Marking:

15% for code, 15% for explanations and general discussion/reporting of results. See Appendix A for more explanations.

Note: regarding the explanation, you are expected to describe which techniques/tools/libraries you chose to solve the problem and why you choose them with a rationale. You are also expected to describe what results you received and discuss what you find out behind the results. for example. what is the meaning of the results? and why do the results behave in a certain way?

Feedback:
Apart from the markings, you will also receive text feedback from Moodle 3 weeks after the submission deadline. The feedback will contain a further explanation about what you have done good and what needs to improve, corresponding to your marks.

Late submission policy
Coursework submitted after the agreed deadline will be marked at a maximum of 40%. Coursework submitted over five working days after the agreed deadline will be given 0%.

Extensions
If you require an extension, please contact the module leader **before** the deadline. Extensions are only provided for exceptional circumstances and evidence may be required. See the Fit to Sit regulations for more details.

Plagiarism
Plagiarised work will be dealt with according to the university's guidelines (Please read - especially if this is the first time in a UK university): http://www2.napier.ac.uk/ed/plagiarism/

Appendix A: Marking Scheme

	No Submission	Very poor	Inadequate	Adequate	Good	Very good	Excellent	Outstanding
A1 Code **15%**	No work submitted	Solution submitted, but contains a lot of errors	Solution submitted but doesn't answer most of the questions	Solution submitted but only answers half of the questions correctly	Solution submitted but only answers most of the questions correctly	The solution answers all questions correctly, but the code is not very readable.	The solution answers all questions correctly.	The solution answers all questions correctly and the code is clean and readable.
A2 Explanations and Report **15%**	No work submitted	Report submitted but doesn't present the results and not explain why a certain approach/library is chosen to solve the task.	Report missing essential discussion and explanation on both results and reasons for choosing certain approaches.	Report missing some essential discussion, rationale for selected tools, and explanations.	Good report and discussion, but not coherent or in-depth discussion.	Very good report and discussion, but not coherent or in-depth discussion, e.g. the results are just stated without an attempt to explain them.	Excellent report with and a good discussion on the methods chosen as well as the results. An excellent submission will attempt to offer insights on the results.	Outstanding report with an in-depth explanation on choices of tools/libraries/approaches and in-depth discussion of results. An outstanding submission will go beyond just the data and offer insights on the results.