Student Assignment Brief

This document is intended for Coventry University Group students for their own use in completing their assessed work for this module. It must not be passed to third parties or posted on any website. If you require this document in an alternative format, please contact your Module Leader.

Contents:

The work you submit for this assignment must be your own independent work, or in the case of a group assignment your own groups' work. More information is available in the 'Assignment Task' section of this assignment brief.

Assignment Information

Module Name: Programming for Data Science

Module Code: 7143CEM

Assignment Title: Individual Portfolio

Assignment Due: Friday 24/11/2023 at 6pm UK time

Assignment Credit: 20 credits

Word Count (or equivalent): 4000 words equivalent (not including reference list or output)

Assignment Type: Written

Percentage Grade (Applied Core Assessment). You will be provided with an overall grade between 0% and 100%. You have one opportunity to pass the assignment at or above 40%.

Assignment Task

This *Individual Portfolio* assessment consists of four tasks (carrying equal marks). You are encouraged to explore these topics, use your initiative, and show some originality, within the time available. Make sure you read each task through carefully, and concentrate on producing a clear and concise answer to each subtask. Aim to demonstrate your understanding of the topics and the module learning outcomes.

Task 1. Explain, critique, comment and debug code (MLO2)

Suppose you have found the following (incomplete) Python code on a scrap of paper inside a textbook on *Algorithms*. It isn't quite finished and contains some errors.

```
def whatdoido(L)
  s = True
  while (s)
    s = False
    for i in range(1,len(L)-3)
      if (L[i] <= L[i+1])
        # now swap values L[i] and L[i+1]
        s = True
    for i in range(len(L)-3,1,-1)
      if (L[i] <= L[i+1])
        # now swap values L[i] and L[i+1]
        s = True
```

(a) Firstly, give a concise critique of the incomplete code given above. Secondly, implement (in Python) the two lines that swap values, fix any errors so that the Python code runs correctly, and ensure the code follows good Python coding style. *Be careful about the indenting already given. Make sure you give your full critique of the code before fixing anything. Include both your critique and your final Python code in your answer (no output is needed).*

[10 marks]

(b) Firstly, modify your Python code from part (a) to count the number of comparisons and swaps, and add appropriate print statements to generate (directly from the code) a clear nontrivial example illustrating the algorithm running in detail. The level of detail should give concise commentary and insight into how the code works and what it actually does, i.e., what comparisons are being made and what decisions are being taken. Secondly, investigate the effect of removing one of the "for loops" from this code. *Please provide both Python code and output from one full example run in your answer. You must not add additional text explanation to your example, i.e., all explanation must be generated through print statements in your Python code.*

[15 marks]

Task 2. Design, build and test (MLO2)

Consider the dice-based cricket match simulation *Owzthat* for one player (this game is at least 90 years old). One dice (called the "batting" dice) is labelled 1, 2, 3, 4, "owzthat" and 6. A second dice (called the "umpire" dice) is labelled "bowled", "stumped", "caught", "not out", "no ball" and "lbw". Before starting the match, the number of balls to be bowled and the number of wickets available are set (so balls and wickets are the resources to be used up over the simulated cricket match). Each time a ball is bowled, the player rolls the batting dice. If the batting dice lands on 1, 2, 3, 4 or 6, that number of runs is added to the total runs scored. However, if the batting dice lands on "owzthat" then the player rolls the umpire dice to determine the outcome of the appeal. If the umpire dice lands on "bowled", "stumped", "caught" or "lbw" then the batter is out (the player loses one wicket) and no runs are scored. If the umpire dice lands on "not out" then no wickets are lost and no runs are scored. If the umpire dice lands on "no ball" then the player adds one run to the total runs scored and the ball must be bowled again. The match finishes when either all the balls have been bowled or all the wickets have been lost (whichever comes first). The aim is to maximize the total number of runs scored.

Image from: https://www.peterberthoud.co.uk/post/a-90-year-old-cricketing-mystery-solved-owzthat

Before you write any Python code, it is worth playing this game using a pencil, paper and a six-sided dice (or use https://rolladie.net/).

(a) Implement this cricket match simulation as one Python function following good Python coding style. Follow the steps and logic you would take in a physical dice game. Your Python code should include enough print statements to generate a clear nontrivial example illustrating a play of the match, telling the story as the dice is rolled and the match score changes. Explain carefully (with clear evidence) how you have used "incremental development" while implementing your Python code. *Please provide both your Python code and the output from two sample matches where there are initially only 10 balls to be bowled and 2 wickets available (to illustrate the two match ending criteria). You must not implement a Python class. Do not attempt to produce a graphical visualisation. The storytelling should be sufficient to be able to test your code by looking only at the output it produces. Marks will be given for both the quality of the code and the quality of the story telling.*

[15 marks]

(b) Modify your Python code to add at least three counts that effectively summarise different aspects of a match. These counts should vary from match to match and you are encouraged to think creatively. Use these summary counts to compare the original Owzthat game with one variant which includes features that more closely resemble a real cricket match. *Provide a summary table of counts for the two game versions, along with a clear and robust conclusion based on your results. Please provide your Python code and a small amount of carefully selected output.*

[10 marks]

Task 3. Critically assess, select and apply data science tools (MLO3)

"No one ever made a decision because of a number. They need a story." – Daniel Kahneman

"Information is giving out; communication is getting through." – Sydney J. Harris

This task involves creating interesting exploratory graphical plots using appropriate Python libraries and then communicating critical insights to decision makers.

(a) "Time series" data is a sequence of data points in time order, usually at equally spaced points in time. Firstly, find an interesting univariate (one variable) time series dataset from the Office for National Statistics (ONS) Time Series Explorer (https://www.ons.gov.uk/timeseriestool) that shows some "seasonality" and build one polished plot of your time series dataset using Python. Secondly, use Python to illustrate (and explain to a relevant decision maker) one time series analysis concept (not previously covered in the module) such as autocorrelation or Winter's method (assuming no "trend") using your time series dataset. *You must clearly indicate exactly where the dataset comes from. The only Python libraries you may use in this part are numpy and matplotlib, except that you may use pandas only to load the dataset and convert the relevant column to a numpy array using* pandas.Series.to_numpy(). *Please provide both Python code and any plots produced.*

[10 marks]

(b) Find an existing graphical plot of a dataset in an article on the BBC News website from 2020 or later. Firstly, reproduce the graphical plot as closely as you can using any two of the seaborn, plotly and plotnine Python plotting libraries separately (but with some help from matplotlib). Briefly assess how well you are able to achieve this. Secondly, use the elements of "data storytelling" to provide a narrative for this graphical plot to communicate insight to a relevant decision maker. *Focus on the content of the plot rather than the exact look and aim to make your two plots as similar as possible. Include the original BBC News image, together with a weblink to the article it is in. Please provide both your Python code and the plots produced. The graphical plot you select must not consist only of text and must not be a pie chart.*

[15 marks]

Task 4. Data protection and data ethics (MLO3)

The General Data Protection Regulation (GDPR) defines a *personal data breach* as "a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data" (Article 4 of GDPR).

(a) Provide details of a significant personal data breach in the UK from 2023 and briefly discuss its impact upon real people.

[10 marks]

(b) The concept of "ownership" of data is different than ownership from physical assets where there is an exclusive right to use, possess and dispose of the asset. There is an argument that since data is not property or a commodity, it therefore cannot be stolen. Critically evaluate how the relevant principles in GDPR and the UK Government Data Ethics Framework address the issue of data ownership in the case described in part (a). Ensure that you consider both the right to control personal data and the right to benefit from personal data.

[15 marks]

Submission Instructions:

Ensure that your Portfolio is all your own work and you clearly cite and reference any sources you have used using APA style referencing. Please include both in-text citations and a list of references for each task (where relevant). **No collaboration with other students is permitted.**

Please submit one report (e.g. as a single Microsoft Word document or a single PDF document) covering all of the tasks above, clearly organised by subtask. Start each task on a new page. Make sure you include your Python code and relevant output/plots directly in the report. You must not submit a zip file. You must not submit a Jupyter notebook (but you can print a Jupyter notebook to a PDF file and submit the PDF file).

Do not use screenshots to include Python code or text output into your report. If your report is a Microsoft Word file, then please use a syntax highlighter (such as http://hilite.me/) and copy-and-paste the coloured Python code into your report. Also simply copy-and-paste text output into your report.

Submission is online via Aula using the submission box provided. *Do not leave uploading too late.*

Marking and Feedback

How will my assignment be marked?

Your assignment will be marked by the module team.

How will I receive my grades and feedback?

Provisional marks will be released once internally moderated. Feedback will be provided by the module team alongside grades release. Students will be able to access their feedback via Aula/Turnitin. Your provisional marks and feedback should be available within 2 weeks (10 working days).

What will I be marked against?

Details of the marking criteria for this task can be found at the bottom of this assignment brief.

Assessed Module Learning Outcomes

The Learning Outcomes for this module align to the marking criteria which can be found at the end of this brief. Ensure you understand the marking criteria to ensure successful achievement of the assessment task. The following module learning outcomes are assessed in this task:

MLO2. Design, build, test, adapt and critique small programs in a high-level programming language. *This MLO is assessed in Task 1 and Task 2.*

MLO3. Critically assess, select and apply data science tools, libraries or algorithms appropriate for various phases of the data science project lifecycle. *This MLO is assessed in Task 3 and Task 4.*

Assignment Support and Academic Integrity

If you have any questions about this assignment please see the Student Guidance on Coursework for more information.

Spelling, Punctuation, and Grammar:

You are expected to use effective, accurate, and appropriate language within this assessment task.

Academic Integrity:

The work you submit must be your own, or in the case of groupwork, that of your group. All sources of information need to be acknowledged and attributed; therefore, you must provide references for all sources of information and acknowledge any tools used in the production of your work, including Artificial Intelligence (AI). We use detection software and make routine checks for evidence of academic misconduct.

Definitions of academic misconduct, including plagiarism, self-plagiarism, and collusion can be found on the Student Portal. All cases of suspected academic misconduct are referred for investigation, the outcomes of which can have profound consequences to your studies. For more information on academic integrity please visit the Academic and Research Integrity section of the Student Portal.

Support for Students with Disabilities or Additional Needs:

If you have a disability, long-term health condition, specific learning difference, mental health diagnosis or symptoms and have discussed your support needs with health and wellbeing you may be able to access support that will help with your studies.

If you feel you may benefit from additional support, but have not disclosed a disability to the University, or have disclosed but are yet to discuss your support needs it is important to let us know so we can provide the right support for your circumstances. Visit the Student Portal to find out more.

Unable to Submit on Time?

The University wants you to do your best. However, we know that sometimes events happen which mean that you cannot submit your assessment by the deadline or sit a scheduled exam. If you think this might be the case, guidance on understanding what counts as an extenuating circumstance, and how to apply is available on the Student Portal.

Administration of Assessment

Module Leader Name: Dr Mark Johnston

Module Leader Email: ad4039@coventry.ac.uk

Assignment Category: Written

Attempt Type: Standard

Component Code: Por

Assessment Marking Criteria

Mark band	Outcome	Guidelines
90-100% Distinction	Meets learning outcomes	Distinction - Exceptional work with very high degree of rigour, creativity and critical/analytic skills. Mastery of knowledge and subject-specific theories with originality and autonomy. Demonstrates exceptional ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Innovative research with exceptional ability in the utilisation of research methodologies. Demonstrates, creativity, originality and outstanding problem-solving skills. Work completed with very high degree of accuracy, proficiency and autonomy. Exceptional communication and expression demonstrated throughout. Student evidences the full range of technical and/or artistic skills. Work pushes the boundaries of the discipline and may be strongly considered for external publication/dissemination/presentation.
80-89% Distinction		Distinction - Outstanding work with high degree of rigour, creativity and critical/analytic skills. Near mastery of knowledge and subject-specific theories with originality and autonomy. Demonstrates outstanding ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Innovative research with outstanding ability in the utilisation of research methodologies. Work consistently demonstrates creativity, originality and outstanding problem-solving skills. Work completed with high degree of accuracy, proficiency and autonomy. Outstanding communication and expression demonstrated throughout. Student demonstrates a very wide range of technical and/or artistic skills. With some amendments, the work may be considered for external publication/dissemination/presentation
70-79% Distinction		Distinction - Excellent work undertaken with rigour, creativity and critical/analytic skills. Excellent degree of knowledge and subject-specific theories with originality and autonomy demonstrated. The work exhibits excellent ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Innovative research with excellent ability in the utilisation of research methodologies. Work demonstrates creativity, originality and excellent problem-solving skills. Work completed with very consistent levels of accuracy, proficiency and autonomy. Excellent communication and expression demonstrated throughout. Student demonstrates a very wide range of technical and/or artistic skills.

60-69% Merit		Merit - Very good work often undertaken with rigour, creativity and critical/analytic skills. Very good degree of knowledge and subject-specific theories with some originality and autonomy demonstrated. The work often exhibits the ability to fully analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Very good research evidence and shows very good ability in the utilisation of research methodologies. Work demonstrates creativity, originality and problem-solving skills. Work completed with very consistent levels of accuracy, proficiency and autonomy. Very good communication and expression demonstrated throughout. Student demonstrates a wide range of technical and/or artistic skills.
50-59% Pass		Pass - Good work undertaken with some creativity and critical/analytic skills. Demonstrates knowledge and subject-specific theories with some originality and autonomy demonstrated. The work exhibits the ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Good research and shows some ability in the utilisation of research methodologies. Work demonstrates problem-solving skills and is completed with some level of accuracy, proficiency and autonomy. Satisfactory communication and expression demonstrated throughout. Student demonstrates some of the technical and/or artistic skills.
40-49% Pass		Pass - Assessment demonstrates some advanced knowledge and understanding of the subject informed by current practice, scholarship and research. Work may be incomplete with some irrelevant material present. Sometimes demonstrates the ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Acceptable research with evidence of basic ability in the utilisation of research methodologies. Demonstrates some originality, creativity and problem-solving skills but often with inconsistencies. Expression and presentation sufficient for accuracy and proficiency. Sufficient communication and expression with professional skill set. Student demonstrates some technical and/or artistic skills.
30-39% Fail	Fails to achieve learning outcomes	Fail - Very limited understanding of relevant theories, concepts and issues with deficiencies in rigour and analysis. Some relevant material may be present but be informed from very limited sources. Fundamental errors and some misunderstanding likely to be present. Demonstrates limited ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Limited research scope and ability in the utilisation of research methodologies. Limited originality, creativity, and struggles with problem-solving skills. Expression and presentation insufficient for accuracy and proficiency. Insufficient communication and expression and with deficiencies in professional skill set. Student demonstrates deficiencies in the range of technical and/or artistic skills.

20-29% Fail -		Fail - Clear failure demonstrating little understanding of relevant theories, concepts, issues and only a vague knowledge of the area. Little relevant material may be present and informed from very limited sources. Serious and fundamental errors and virtually no evidence of relevant research. Fundamental errors and misunderstandings likely to be present. Little or no research with no evidence of utilisation of research methodologies. No originality, creativity, and struggles with problem-solving skills. Expression and presentation insufficient for accuracy and proficiency. Insufficient communication and expression and with serious deficiencies in professional skill set. Student has clear deficiencies in range of technical and/or artistic skills.
0-19% Fail		Fail - Clear failure demonstrating no understanding of relevant theories, concepts, issues and no understanding of area. Little or no relevant material may be present and informed from minimal sources. No evidence of ability in the utilisation of research methodologies. No evidence of originality, creativity, and problem-solving skills. Expression and presentation deficient for accuracy and proficiency. Insufficient communication and expression and with deficiencies in professional skill set. Student has clear deficiencies in range of technical and/or artistic skills.