

This document is for Coventry University students for their own use in completing their assessed work for this module and should not be passed to third parties or posted on any website. Any infringements of this rule should be reported to [facultyregistry.eec@coventry.ac.uk](mailto:facultyregistry.eec@coventry.ac.uk).

## Faculty of Engineering, Environment and Computing 7143CEM Programming for Data Science



### Assignment Brief

Module Title: <b>Programming for Data Science</b>	Individual /Group: <b>Individual</b>	Cohort (Sept/Jan/May): <b>Sept 2022</b>	Module Code: <b>7143CEM</b>
Coursework Title: <b>Individual Portfolio</b>			Hand out date: <b>19/09/2022</b>
Lecturer: <b>Mark Johnston / Seyed Mousavi</b>			Due date and time: <b>04/11/2022 at 6pm</b>
Estimated Time (hrs): <b>100 hours</b> Word Limit*: <b>6000 words</b>	Coursework type: <b>Portfolio</b>		Credit value assessed: <b>20 credits</b>
Submission arrangement: <b>online via Aula</b> File types: <b>docx, pdf</b>  Mark and Feedback date (DD/MM/YY): <b>18/11/2022</b> Mark and Feedback method (e.g. in lecture, electronic via Aula): <b>Aula</b>			

#### Module Learning Outcomes Assessed:

**ILO2.** Design, build, test, adapt and critique small programs in a high-level programming language.  
**ILO3.** Critically assess, select and apply data science tools, libraries or algorithms appropriate for various phases of the data science project lifecycle.

#### Task and Mark distribution:

The *Individual Portfolio* 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 answer all parts of each task.

Please ensure that your Portfolio is all your own work and you clearly reference any sources you have used. **No collaboration with other students is permitted.**

Please submit one report (e.g. as a single Microsoft Word document) covering all of the tasks below, 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). The word limit is a maximum rather than a target. Concentrate on producing a clear and concise answer to each subtask.

### Task 1. Explain, critique, comment and debug code (ILO2)

Suppose you have found the following (incomplete) Python code on a scrap of paper inside a textbook on *Computer Graphics*. The two Python functions are not quite finished and contain some errors.

```
def liDDA(a,b,c,d)                def liB(a,b,c,d)
    P = []                        P = []
    e = c-a                      e = c-a
    f = d-b                      f = d-b
    s = max(e,f)                 g = 2*f - e
    x,y = a,b                    y = b
    for i in range(s)            for x in range(a,c+1)
        P.append((round(x),round(y)))    P.append((x,y))
        x += e/s                  if (g>0)
        y += f/s                  y += 1
    print(P)                     g = g - 2*e
                                g = g + 2*f
                                print(P)

liDDA(0,1,6,4)
liB(0,1,6,4)
```

- (1) Firstly, give a detailed critique of the incomplete Python code given above. Secondly, fix any syntax errors so that the Python code runs, ensure the code follows good Python coding style, and improve all the variable and function names. *Make sure you give your full critique of the code before fixing anything. Then provide a corrected version of the code. Do not fix any semantic errors. Be very careful about the indenting given.*

[10 marks]

- (2) Add appropriate print statements to your Python code from part (1) to generate (directly from the code) a clear example illustrating the two algorithms running in detail (using the example function calls given). The level of detail should show what comparisons are being made and what decisions are being taken. *Please provide both code (with print statements added) and the output from the code running in a way that illustrates the algorithms in detail but also gives insights into how they work and what they do. You can also add additional explanation to your example.*

[5 marks]

- (3) The two Python functions liDDA and liB are supposed to solve the same task. Firstly, explain the task clearly and then fix any semantic errors in the two algorithms (so that they actually do solve the task correctly). Secondly, compare the two algorithms in terms of accuracy (whether they produce the same answers), efficiency (how quickly they solve the task), and whether there is any strange behaviour of these functions. *You can use further examples to illustrate your answers but do not include much output.*

[10 marks]

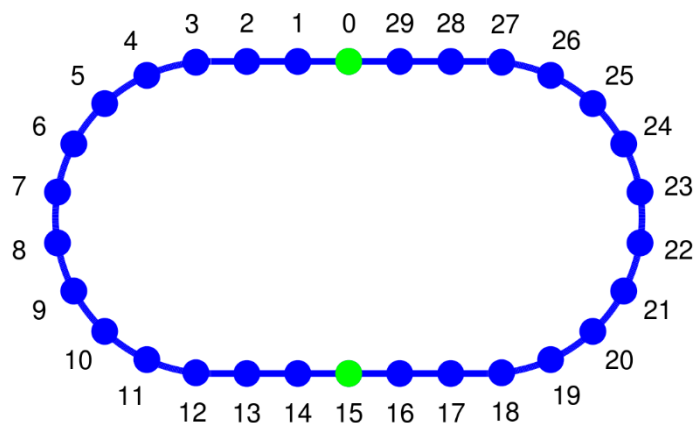
## Task 2. Design, build and test (ILO2)

The image below shows the Velodrome where the track cycling events were held for the 2012 London Olympics. The *Individual Pursuit* is a track cycling race in which two individual cyclists start the race at the same time but from opposite sides of the track. The winner is the cyclist who manages to “catch” (overtake) their opponent or whoever records the fastest time over a given distance (if neither cyclist catches the other).



Image from: <https://olympics.com/ioc/legacy/london-2012/lee-valley-velopark>

In this task, we consider a simplified dice-game version of an Individual Pursuit cycling race. One lap of the track will consist of 30 positions labelled 0 to 29 as shown below.



The Olympic competition starts with a *time trial*. For our dice-game version, suppose one cyclist starts from position 0 on the track. A single 6-sided dice is rolled, and the cyclist moves forward the number of positions indicated by the dice rolled. The cyclist continues to roll the dice and move around the track until they have completed six laps of the track. The “time taken” is the number of dice rolls required to complete the six laps. *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/>)*

- (1) Implement the dice-game version of this time trial (for one cyclist) as one Python function. Follow the steps and logic you take in the corresponding physical dice game. Your code should include enough print statements to generate a clear nontrivial example, telling the story as the dice is rolled and the cyclist moves around the track. Indicate clearly in the output when the cyclist completes each lap, and how many dice rolls are needed to complete the time trial. *Please provide both your Python code and the output from a sample run. 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 code and the quality of the storytelling.*

[10 marks]

The Olympic competition continues with the Individual Pursuit races. For our dice-game version, suppose two cyclists race against each other, where one cyclist starts from position 0 on the track, and the other cyclist starts from position 15 on the track. Each cyclist rolls their own 6-sided dice (simultaneously) and moves forward the number of positions indicated by their dice roll. If one cyclist “catches” (overtakes) their opponent, then they are the winner. If neither cyclist catches the other, then the first cyclist to complete six laps of the track is the winner. It is also possible that both cyclists complete six laps of the track on the same dice roll, in which case the race is a draw. Therefore, there are five possible outcomes of the Individual Pursuit race. *Note that this is a simultaneous-move game not an alternating-move game.*

- (2) Modify your Python function from part (1) to implement the dice-game version of this individual pursuit (for two cyclists). The function should return one of the five possible outcomes of the race and the number of dice rolls until a winner is found. You should include enough print statements to generate a clear nontrivial example, telling the story (commentary) as the dice are rolled and the cyclists move around the track. Indicate clearly in the output when each cyclist completes a lap, which cyclist is in the lead, and the “split” (lead) between the player positions (taking account of the different start positions). *Please provide both your Python code and the output from two sample runs (one that lasts for the full six laps and one that involves one cyclist catching the other).*

[10 marks]

- (3) Suppose two cyclists compete over 1000 such dice-game Individual Pursuit races. Firstly, produce a table of results to show how many races finish in each of the five possible outcomes and comment on these results. Secondly, the description of Individual Pursuit dice game assumes that the cyclists have the same ability. Investigate (by example) what happens in your table of results if the cyclists do not have exactly the same ability.

[5 marks]

### Task 3. Critically assess, select and apply data science tools (ILO3)

This task involves creating interesting exploratory graphical plots of a dataset using appropriate Python libraries and then critically assessing/comparing the capabilities of those plotting libraries. *You must include at least two of seaborn, plotly and plotnine plotting libraries.*

Firstly, you will need to select a publicly available dataset about something topical (in the news during 2022) or related to the UN sustainable development goals (<https://sdgs.un.org/goals>). *Do not use data sourced from Kaggle. Clearly show where the dataset comes from and give a small snippet to show the structure of the dataset.*

Some interesting UK data are available from the following websites:

<https://www.data.gov.uk/>

<https://www.ons.gov.uk/>

<https://sdgdata.gov.uk/>

<https://geoportal.statistics.gov.uk/>

- (1) Build one polished simple plot that shows something interesting about the dataset you have chosen. You must build the same plot using three different Python plotting libraries separately. Aim to make the plots as similar as possible. For example, you might use pandas to build a stacked barchart, and then build the same stacked barchart using seaborn (using the same variables), and then build the same stacked barchart again using plotnine. Comment on what the plot shows and why it is interesting. *Do not use a pie chart. The Python Graph Gallery (<https://python-graph-gallery.com/>) is a good place to start. Please provide both your Python code and the plot that it creates. You must clearly reference any sources you have used.*

[5 marks]

- (2) Repeat part (1) using the same dataset but producing a plot that involves some complexity, i.e., a plot that consists of multiple nontrivial components working together or that requires some tricky preprocessing to construct. Examples of complex plots would be a Mekko chart (also known as a mosaic plot or Marimekko chart), a circular barchart, or a ternary scatterplot. *Note that you are attempting to assemble some useful experience that you can reflect on in part (3).*

[5 marks]

- (3) Summarise and reflect on your experience (and observations) from using the Python plotting libraries in parts (1) and (2) in terms of the following criteria: difficulty of coding, adaptability of the code, level of control over the plot layout or labelling, and quality of the graphical plot produced. *You must include at least two of seaborn, plotly and plotnine plotting libraries. Present your summary in an appropriate table.*

[5 marks]

- (4) Compare your experience summarised in part (3) with additional literature or sources (e.g. articles, blogs, documentation or tutorials) in order to critically assess the libraries you used to produce the particular plots in parts (1) and (2). *Ensure you clearly reference any sources you use following APA referencing style (both in-text citations and a reference list). Use the criteria given in part (3).*

[5 marks]

- (5) Suppose you wish to visualise geospatial data. Briefly describe the capabilities of one Python library that you would recommend and give an example (code and output) using any UK data. *You may find the Open Geography Portal (<https://geoportal.statistics.gov.uk/>) useful.*

[5 marks]

#### Task 4. Data protection and data ethics (ILO3)

Face recognition technology uses data science and machine learning to find a match for an image of a human face, in a dataset of images of human faces.

- (1) Describe (in detail) a significant example of the use of face recognition technology in the UK from the last 10 years that **had some unintended consequences** and discuss its impact (including benefits, consequences and risks). *Ensure you clearly reference any sources you use.*  
[10 marks]
- (2) A commonly used saying in Ethics is “Just because you can doesn’t mean you should”. Critically evaluate how the relevant principles of the UK Data Ethics Framework (particularly fairness, accountability and transparency) and the General Data Protection Regulations (GDPR) were violated in the case you have described in part (1), and how these principles and regulations could (or should) have prevented these unintended consequences and mitigated these risks.  
[10 marks]
- (3) In 1999, Scott McNealy (CEO of Sun Microsystems) said “You have zero privacy anyway. Get over it.” Briefly discuss to what extent should face recognition technology be restricted (by law) to only particular subsets of organisations or applications. *This part invites you to make and briefly defend a judgement or recommendation.*  
[5 marks]

#### Notes:

1. You are expected to use [Coventry University APA](#) style for referencing. For support and advice on this students can contact [Centre for Academic Writing \(CAW\)](#).
2. Please notify your registry course support team and module leader for disability support.
3. Any student requiring an extension or deferral should follow the university process as outlined [here](#).
4. The University cannot take responsibility for any coursework lost or corrupted on disks, laptops or personal computer. Students should therefore regularly back-up any work and are advised to save it on the University system.
5. If there are technical or performance issues that prevent submitting coursework through the online coursework submission system on the day of a coursework deadline, an appropriate extension to the coursework submission deadline will be agreed. This extension will normally be 24 hours or the next working day if the deadline falls on a Friday or over the weekend period. This will be communicated via your Module Leader.
6. You are encouraged to check the originality of your work by using the draft Turnitin links on Aula.
7. Collusion between students (where sections of your work are similar to the work submitted by other students in this or previous module cohorts) is taken extremely seriously and will be reported to the academic conduct panel. This applies to both coursework and exam answers.
8. A marked difference between your writing style, knowledge and skill level demonstrated in class discussion, any test conditions and that demonstrated in a coursework assignment may result in you having to undertake a Viva Voce in order to prove the coursework assignment is entirely your own work.

This document is for Coventry University students for their own use in completing their assessed work for this module and should not be passed to third parties or posted on any website. Any infringements of this rule should be reported to **facultyregistry.eec@coventry.ac.uk**.

9. If you make use of the services of a proofreader in your work you must keep your original version and make it available as a demonstration of your written efforts. Also, please read the university [Proof Reading Policy](#).
10. You must not submit work for assessment that you have already submitted (partially or in full), either for your current course or for another qualification of this university, with the exception of resits, where for the coursework, you may be asked to rework and improve a previous attempt. This requirement will be specifically detailed in your assignment brief or specific course or module information. Where earlier work by you is citable, i.e., it has already been published/submitted, you must reference it clearly. Identical pieces of work submitted concurrently may also be considered to be self-plagiarism.

This document is for Coventry University students for their own use in completing their assessed work for this module and should not be passed to third parties or posted on any website. Any infringements of this rule should be reported to [facultyregistry.eec@coventry.ac.uk](mailto:facultyregistry.eec@coventry.ac.uk).

### Marking Rubric (PG)

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.



This document is for Coventry University students for their own use in completing their assessed work for this module and should not be passed to third parties or posted on any website. Any infringements of this rule should be reported to **facultyregistry.eec@coventry.ac.uk**.

60-69%		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 - 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 - 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%	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.

This document is for Coventry University students for their own use in completing their assessed work for this module and should not be passed to third parties or posted on any website. Any infringements of this rule should be reported to **facultyregistry.eec@coventry.ac.uk**.

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.