



BSc EXAMINATION

COMPUTER SCIENCE

Programming with Data

Release date: Thursday 8 September 2022 at 12:00 midday British Summer Time

Submission date: Friday 9 September 2022 by 12:00 midday British Summer Time

Time allowed: 24 hours to submit

INSTRUCTIONS TO CANDIDATES:

Section A of this assessment paper consists of a set of **TEN** Multiple Choice Questions (MCQs) which you will take separately from this paper. You should attempt to answer **ALL** the questions in Section A. The maximum mark for Section A is **40**.

Section A will be completed online on the VLE. You may choose to access the MCQs at any time following the release of the paper, but once you have accessed the MCQs you must submit your answers before the deadline or within **4 hours** of starting whichever occurs first.

Section B of this assessment paper is an online assessment to be completed within the same 24-hour window as Section A. We anticipate that approximately **1 hour** is sufficient for you to answer Section B. Candidates must answer **TWO** out of the **THREE** questions in Section B. The maximum mark for Section B is **60**.

Calculators are not permitted in this examination. Credit will only be given if all workings are shown.

You should complete **Section B** of this paper and submit your answers as **one document**, if possible, in Microsoft Word or a PDF to the appropriate area on the VLE. Each file uploaded must be accompanied by a coversheet containing your **candidate number**. In addition, your answers must have your candidate number written clearly at the top of the page before you upload your work. Do not write your name anywhere in your answers.

SECTION A

Candidates should answer the **TEN** Multiple Choice Questions (MCQs) quiz, **Question 1** in Section A on the VLE.

SECTION B

Candidates should answer any **TWO** questions from Section B.

Question 2

- (a) Write a Python program that modifies the following set so that it only includes valid types of animals. Provide the code and the output of the program running as text.

```
animals = {"cat", "degree", "dog"}
```

[2]

- (b) Assuming that you could not see the contents of the variable 'animals' from the previous question, how would you know that there are not two items with the same value?

[2]

- (c) Describe two challenges you may face when web scraping a dynamic webpage versus a static webpage.

[2]

- (d) Web scraping is a contentious topic in data science. Describe two ethical issues associated with web scraping. Provide two examples of how such ethical issues have the potential to cause harm.

[4]

- (e) Write a Python program to check that a password has a capital letter, a special character (: ; !) and a number.

[4]

- (f) Describe two advantages and two disadvantage of utilising test-driven development processes for data processing tasks.

[4]

- (g) Provide an example of where you have found test-driven development useful in your own work.

[2]

- (h) Provide an example of where you have found test-driven development challenging in your own work.

[2]

- (i) I record the number of hours that I study each day of the week.

```
study = [2,5,9,7,4]
```

Without importing external modules, write a Python program to calculate my average study time per week based on this data. Provide the code and the output of the program running as text.

[2]

- (j) Write a Python program to check if a given string contains two of the same consecutive letters. Provide the code and the output of the program running as text.

[6]

Question 3

- (a) List three features of datasets you have personally worked with which meant that you had to carry out significant pre-processing or refactoring of the data prior to analysing it. For example, briefly describe the dataset and explain what you had to do in order to be able to work effectively with it. [6]
- (b) Describe three approaches that you might take in an exploratory data analysis of sports performance. Assume that you have a dataset that relates to FA Women's Super League football matches. This data might include things such as shots on target, pass success rate and passes attempted. Provide three examples of appropriate visualisations for these types of data. [6]
- (c) You are working in a team on a data analysis project. You are trying to convince members of your team to work with Jupyter notebooks instead of regular text editors such as vim, Emacs or notepad++. Provide three reasons why data science projects utilise computing platforms like Jupyter over standard text editors. [6]
- (d) Describe one advantage and one disadvantage of relying on modules (code libraries) in Python. [4]
- (e) There are a handful of words in the English language that read the same backwards as they do forwards. Examples of these words include 'civic' and 'refer'. Write a program in Python to evaluate whether words in a list are the same backwards as forwards. Include your home City and at least one valid case of a word that meets this condition of reading identically in reverse. Provide the code and the output of the program running as text. [6]
- (f) Provide an example of a key-value pair relationship from everyday life. Provide an example of this being represented in Python. [2]

Question 4

- (a) Write a program in Python to get the minimum and maximum values held in a dictionary, assuming that the dictionary contains numerical values. Provide the code and the output of the program running as text. [2]
- (b) Write a program in Python to detect if a dictionary contains two values that are the same. Provide the code and the output of the program running as text. [4]
- (c) There is a dictionary containing all student grades for a course. The dictionary contains a unique student ID and corresponding grade for each student as follows:

`{12324:45, 4324324:90}`

Create a new dictionary mapping for degree classifications with specific grade bounds. For reference, a 1st is a grade over 69. A 2:1 is a grade in the range 60-69. A 2:2 is between 50-59. A 3rd class degree is between 40-49 and everything below 40 is a fail. [6]
- (d) Write a function in Python that is capable of returning all words from a list which are longer than six characters. Write a list comprehension and a loop based version showing words greater than six characters. Run it on your machine with a large random test data array. Can you identify any significant differences in speed? [6]
- (e) Provide three reasons why time data is difficult to process. [6]
- (f) Provide two examples of Python code to handle HTTP status codes. [6]

END OF PAPER