



**BSc EXAMINATION**

**COMPUTER SCIENCE**

**Programming with Data**

**Release date:** Friday 19 March 2021 at 12 midday Greenwich Mean Time

**Submission date:** Saturday 20 March 2021 by 12 midday Greenwich Mean Time

**Time allowed:** 24 hours to submit

**INSTRUCTIONS TO CANDIDATES:**

**Section A** of this assessment 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 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. You are permitted to upload 30 documents. However, we advise you to upload as few documents as possible. 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 B

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

### Question 1

- (a) The open function in python takes two parameters. Describe what these parameters are and give an example for each. [2]
- (b) In Python there are three numeric types. Describe TWO of these types and provide an example of each. [2]
- (c) Define what is meant by mutable and immutable objects. [2]
- (d) Give two examples of mutable built in types. [2]
- (e) Give two examples of immutable built in types. [4]
- (f) The 'Dict' datatype utilises associative mapping. Explain what this is and give your own example of some real data that it would be appropriate to store in a Dict. [6]
- (g) Explain how the sequence of events of a for loop work in relation to an iterator. [2]
- (h) Explain how the sequence of events of a for loop work in relation to an iterable. [2]
- (i) Consider the data processing code you wrote for your mid-term coursework. [2]
  - i. What was the question you were trying to answer with your code?
  - ii. Describe the dataset you were working with. Where was it from? What did it represent? [2]
  - iii. Give a brief, step by step summary of the steps you went through to analyse the data. [4]

## Question 2

- (a) State and describe THREE features of a typical development environment. [3]
- (b) Define the term 'regular expression'. [2]
- (c) Give an example where a regular expression would be appropriate for use in data processing. [2]
- (d) Give an example of some data, a regular expression and the result of applying that regular expression to that data. [4]
- (e) What is the purpose of a metacharacter in a regular expression? [2]
- (f) When we parse web pages, we are working within a particular structure.
  - i. Name the structure - state the acronym and what it stands for. [2]
  - ii. Describe the form of the structure, using diagrams if necessary. [2]
  - iii. Name and describe TWO things one might find in that structure. [2]
- (g) What is the root element of a HTML document? [2]
- (h) What is the purpose of a HTTP status code in the context of a request? [2]
- (i) Name an HTTP status code and explain the circumstances under which it might arise. [2]
- (j) Why are class and id elements useful for navigating HTML using Python? [2]
- (k) What is web scraping? [1]
- (l) Give TWO reasons why many websites do not permit web scraping. [2]

### Question 3

- (a) State and describe FIVE characteristics or properties of a Pandas dataframe. [10]
- (b) Assuming a variable called df is a Pandas dataframe, what is the function of df.dropna()? [2]
- (c) Explain why it is common to use df.dropna when processing real world data. [2]
- (d) Give an example of some data that would be affected by the df.dropna function. [2]
- (e) Which function would show the last 5 rows of a dataframe? Give some example code which calls the function. [2]
- (f) Which function would replace all NA/NaN values of a series with the mean? [4]
- (g) Which function would produce summary statistics for numerical columns? [4]
- (h) What is the property DataFrame.iloc used to select? [4]

END OF PAPER