



**BSc EXAMINATION**

**COMPUTER SCIENCE**

**Programming with Data**

**Release date:** Wednesday 8 March 2023 at 12:00 midday Greenwich Mean Time

**Submission date:** Thursday 9 March 2023 by 12:00 midday Greenwich Mean 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) In Video 2.401 we relied on a NumPy a related module that offered us some additional utility. Describe and provide three examples where utilising these might create undesirable outcomes. [9]
- (b) In Lab 2.405 we explored a few libraries and techniques to help us conceptualise progression of events from some point in the past. Describe three challenges of working with this type of data and three strategies to overcome said challenges. [12]
- (c) In Lab 3.13, Activity three, we used a technique to quantify the value of words according to one of three pre-defined categorical representations. Describe these three categorical representations of the data that we saw in the example. Provide example data that pertains to each category. [6]
- (d) In Lab 3.13 we experienced challenges in trying to quantitatively evaluate data. Describe three challenges of analysing the type of data that we explored in this lab activity. [3]

### Question 3

- (a) In Lab 3.13, activity two, we removed some stopwords and then plotted some metrics for the remaining data. What did this allow us to discover? [3]
- (b) Which of the things that we visualised in Lab 3.13 do not belong in the analysis? Explain your decision. [3]
- (c) In videos 7.002, 7.004, and 7.006 we explore some approaches to problem solving with challenging datasets.  
Describe three techniques that we used to deal with values that were perceived to be either invalid or not fit for purpose. [6]
- (d) Describe three examples where the strategies shown in topic 7 for dealing with invalid or unfit for purpose data might pose problems. [6]
- (e) In the data plotting topic we explored a variety of different formats for visualisation. Select three of these visualisations and describe a problem associated with each. [6]
- (f) In each of the previous examples that you have provided exhibiting problems associated with visualisations, explain a strategy for solving the problem that you described. [6]

#### Question 4

- (a) In the DataFrames and exploratory analysis lecture video we looked at high level features of our dataset, including things such as mean (average) values. Suggest five other things that we could explore under this analytical lens. [6]
- (b) In the exploratory data analysis video series, it was suggested that you do not rely entirely on your own analysis. Instead, it was suggested that you engage with a wide range of sources of information to help to inform your analysis. Provide three examples of resources you might use and describe the utility that they provide for an exploratory data analysis. [6]
- (c) Find an example of error handling from the course. Present the code and explain how it works. [3]
- (d) Describe an alternative approach for dealing with the error handling scenario you addressed in part (c) of this question. [3]
- (e) In Video 3.07 we took a verbose approach to counting words. Draw a diagram to explain the logic of this process. [6]
- (f) Provide an example of counting words by utilising the power of an external library. [3]
- (g) Provide three examples of how you might standardise the sort of data you see in Video 3.07. [3]

END OF PAPER