**BSc EXAMINATION**

**COMPUTER SCIENCE**

**Programming with Data**

**Release date:** Thursday 7 March 2024 at 12:00 midday Greenwich Mean Time

**Close date:** Friday 8 March 2024 by 12:00 midday Greenwich Mean Time

**Time allowed:** 4 hours to submit

**INSTRUCTIONS TO CANDIDATES:**

**Part A** of this assessment consists of a set of **TEN** Multiple Choice Questions (MCQs). You should attempt to answer **ALL** the questions in **Part A.** The maximum mark for Part A is **40**.

Candidates must answer **TWO** out of the **THREE** questions in **Part B**. The maximum mark for Part B is **60**.

**Part A and Part B** will be completed online together on the Inspera exam platform. You may choose to access either part first upon entering the test area but must complete both parts within **4 hours** of doing so.

A handheld non-programmable calculator may be used when answering questions on this paper but it must not be able to display graphics, text or algebraic equations. Please hold your calculator to the camera at the start of the examination to clearly show the make and type.

You may use **ONE** A4 page of previously prepared notes in this examination. Please hold up your notes to the camera at the start of the examination.

File upload is **NOT** permitted.

Do not write your name anywhere in your answers.

**PART A**

Candidates should answer the **TEN** Multiple Choice Questions (MCQs) in Part A.

**PART B**

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

**Question 2**

(a) Describe how to handle conditionals in Python. Provide an example of simple conditional logic. Provide **TWO** examples of more complex logic scenarios in Python.

[5 marks]

(b) What is exception handling in Python, and why is it important? Provide an example of exception handling and provide an example where you might use exception handling over logic controls.

[5 marks]

(c) Describe how a Python program can retrieve data from the Web. Explain how HTTP requests work and how to handle API responses.

[5 marks]

(d) Write an example SQL query to retrieve data from a relational database. Explain the components of a relational database and explain why we might utilise a relational model to store and retrieve data.

[5 marks]

(e) Explain the principles of effective data visualisation and describe how you would justify your choice of visualisation type for **THREE** distinct purposes.

[5 marks]

(f) Explain the importance of exploratory data analysis (EDA) in the context of data programming. Provide a step-by-step example of how you would conduct an EDA.

[5 marks]

**Question 3**

(a) What is a Python lambda function, and how does it differ from a regular function? Provide an example of when you might utilise lambda functions.

[5 marks]

(b) Discuss the significance of data pre-processing in the data programming workflow. Provide a detailed example of data pre-processing steps using Python libraries, highlighting their importance in preparing data for analysis.

[5 marks]

(c) Explain the concept of data ethics and its relevance in data programming. Provide an example of an ethical dilemma related to data collection, analysis, or sharing, and discuss potential solutions.

[5 marks]

(d) Discuss the principles of reproducible research in data programming. Provide an example of a scientific project where reproducibility is essential and explain how you would ensure that the results of your project were reproducible by other data scientists.

[5 marks]

(e) Describe the concept of data-driven decision-making and its impact. Provide an example of how data could be used to reflect upon elements of the degree programme that you are currently studying.

[5 marks]

(f) Explain the concept of a Python dictionary (dict) as a data structure for key-value pairs. Discuss common operations, methods, and use cases for dictionaries in Python programming. Provide an example of converting a list to a dictionary.

[5 marks]

**Question 4**

(a) You are developing a Python application for a student information system used to store, administer, and manage student data at a university. Explain the importance of unit testing in ensuring the reliability and correctness of your code.

[5 marks]

(b) Reflect on the strengths and weaknesses of Matplotlib and Seaborn as Python plotting libraries for data visualisation.

[5 marks]

(c) Imagine you are working with a dataset containing incomplete information about student performance in an educational database. Discuss the potential challenges of making assumptions about missing student data, and explain how these assumptions could affect decision-making or analysis within an educational setting. Provide strategies to handle such situations effectively while maintaining data accuracy.

[10 marks]

(d) Explain the purpose and advantages of using the Pandas library in Python for data analysis and manipulation. Provide a practical example of a data manipulation task that Pandas can efficiently handle, and describe the key features or functions of Pandas that make it suitable for this task.

[5 marks]

(e) Discuss the fundamental differences between text-based data and numerical data in the context of data analysis. Explain the unique challenges and considerations when working with each type of data. Provide examples of situations where text-based data is more appropriate for analysis and situations where numerical data is preferred.

[5 marks]

END OF PAPER