

BSc EXAMINATION**COMPUTER SCIENCE****Programming with Data**

Release date: Monday 16 September 2024 at 12:00 midday British Summer Time

Close date: Tuesday 17 September 2024 by 12:00 midday British Summer 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 Inspira 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.

Calculators are **NOT** permitted in this examination.

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 in this examination.

Do not write your name anywhere in your answers.

© University of London 2024

Part A

Question 1

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 the primary differences between lists, dictionaries, and NumPy arrays in Python. How is each used to manage data effectively?
[8 marks]
- (b) Explain the client-server model in the context of retrieving data from the web. What role does HTTP play in this model?
[8 marks]
- (c) You have been provided with a dataset containing various user interactions on a website, `user_data.json`. The dataset includes fields such as user_id, timestamp, page_visited, and action_taken. Write a Python script that reads this file and counts the number of times each page was visited by users.
[8 marks]
- (d) Critically evaluate **TWO** approaches to handling missing data in a dataset. Discuss the advantages and disadvantages of each approach.
[6 marks]

Question 3

(a) Provide **THREE** examples of how you can handle errors in your Python code.

[6 marks]

(b) Define what an SQL query is and explain its role in retrieving data from a relational database.

[6 marks]

(c) Describe the JSON data format and explain why it is commonly used in web applications.

[6 marks]

(d) Write a Python script that connects to a database, retrieves the sales data, and calculates the total sales per month.

[6 marks]

(e) Describe how you engaged in an exploratory data analysis for your coursework assignment. Describe how you were able to verify and validate these against some pre-defined expectations.

[6 marks]

Question 4

- (a) Assume you are working with a dataset of survey responses that has not been verified, optimised or validated, `survey.csv`. The dataset includes binary and alphanumeric values. The file may also include missing and erroneous data. Write a Python script to handle this complex logic scenario. Include comments to describe your logic.

[8 marks]

- (b) Explain why data visualization is important in data analysis. Next, using Python, write a script to create a simple bar chart from a given dataset. Finally, discuss how to choose the appropriate type of visualization for different data types and analysis goals.

[8 marks]

- (c) Describe how a version control system like Git works. What benefits do version control systems provide? Describe **FOUR** use cases for version control systems in regard to reviewing project history.

[8 marks]

- (d) Describe the process of Test-Driven Development (TDD) and how it integrates into the software development lifecycle. What are the primary benefits of adopting a Test-Driven Development approach in software projects?

[6 marks]

END OF PAPER