UNIVERSITY OF LONDON

CM2015

BSc EXAMINATION

COMPUTER SCIENCE

Programming with Data

Release date: Monday 21 March 2022 at 12:00 midday Greenwich Mean Time

Submission date: Tuesday 22 March 2022 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. 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.

UL22/0100

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.

Candidates should provide print outs of code working, where appropriate.

Question 2

(a) Write two functions to sort words alphabetically and remove duplicate words from a sentence. [8]

(b) You have decided to invest in a new cryptocurrency. You have tracked the average monthly price of said cryptocurrency over a six-month period. Write a function that can calculate when you could have bought and sold a single token of the cryptocurrency to maximise your profit margins during this period. [8]

(c) Describe two edge cases where your cryptocurrency calculation algorithm might fail. [4]

(d) Write a function to find the longest word in a sentence. [6]

(e) Define two instances where commenting in Python provides value and provide an example for each case. [4]

Question 3

(a) Explain and provide examples of the steps performed in an exploratory data analysis. [10]

(b) Explain how web scraping works. Describe three ethical considerations you would take into account when web scraping. [7]

(c) Define Test Driven Development and describe two advantages of utilising Test Driven Development in practice. [7]

(d) Provide an example of a type of dataset where standard stopword such as those provided by the nltk.corpus package might not be appropriate. Describe two techniques that you could utilise to identify common stopwords from a corpus of text. [6]

Question 4

(a) Describe three types of visualisations you can use in Python. Provide example codes to produce each. [6]

(b) Identify two cases where a 2D visualisation might be preferable over 3D visualisation and explain why. [4]

(c) Write a program to calculate the number of upper case characters in a text file. [6]

(d) Describe three advantages to a business in using Git for version control. [6]

(e) Missing data can be particularly problematic in the context of a data processing pipeline. Explain three strategies for dealing with missing data and identify in each case a potential issue with your approach. [6]

(f) Provide an example of some code that encapsulates error handling. Explain the conditions and inherent logic contained therein. [2]

END OF PAPER