UNIVERSITY OF LONDON

CM2015

BSc EXAMINATION

COMPUTER SCIENCE

Programming with Data

Release date: Monday 13 September 2021 at 12:00 midday British Summer Time

Submission date: Tuesday 14 September 2021 by 12:00 midday British Summer Time

Time allowed: 24 hours to submit

INSTRUCTIONS TO CANDIDATES:

Section A of this assessment paper consists of a set of **10** 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. 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) Web pages are often made up utilising a range of technologies. Explain how a dynamic webpage might present issues when web scraping. [2]

(b) Web scraping is a contentious topic in data science, with different perspectives about legality and appropriateness. Describe two ethical issues associated with web scraping. Provide examples for each. [4]

(c) Write a program using Pandas to extract the first 9 records from a dataframe 'cars'. [2]

(d) Write an SQL query to extract the first 9 records from the table 'CARS'. [2]

(e) Assume we take some user input e.g.

monthName = input("Select a month name from the 12 months of the year e.g. December.")

Write a program that calculates the number of days in the month selected by the user. [10]

(f) Write a Python function to find and show the length of values contained in a dictionary. [10]

Question 2

(a) Write a Python program to check that a password p meets the following rules:

- Contains both lower and uppercase letters [4]
- Contains a number [2]
- Is between 4 and ten characters [4]

(b) Describe a viable technique for scraping content from dynamic webpages and explain how it works. [2]

(c) Provide two examples of things we cannot test using unit tests. [4]

(d) Assume we have a dataset with missing or erroneous data. Describe three strategies to identify erroneous data. Provide an example for each strategy. [6]

(e) Give three examples of ways that we might represent location-based data. [3]

(f) Write a python function that can add up the five lowest (positive) values from a list. Assume the list only contains unique elements. Provide an example of utilising the function on a list of numbers that contains more than five elements. [5]

Question 3

(a) Provide two examples of how you might remove whitespaces from a string. [4]

(b) Describe what stopwords are. Provide two examples of stopwords. [2]

(c) The daily temperatures for the week are defined as follows:
temps = [5, 18, 34, 32, 20]
Write some Python code to print out the second hottest daily temperature. [2]

(d) Unit tests enable us to test lots of functional aspects of our code. Provide two examples of things we can test using unit tests. [4]

(e) What sort of error might we produce if we tried to change an immutable object in Python? [2]

(f) Describe three strategies to identify data duplication. For each strategy, provide an example of how this problematic data might manifest in a dataset. [6]

(g) Describe three issues with 3d visualisations. [3]

(h) Describe and provide an example of a representation format that may overcome some presentational challenges of 3d data. [3]

(i) Describe what the following code is doing. [4]

```
def test(my_word):
   return any(c1 == c2 for c1, c2 in zip(my_word, my_word[1:]))
```

END OF PAPER

.