UNIVERSITY OF LONDON

DSM020

MSc EXAMINATION

DATA SCIENCE

MOCK EXAM - Data Programming in Python

Date and Time of exam: TBC

Time allowed: 2 hours

INSTRUCTIONS TO CANDIDATES:

This examination paper is in three parts: Part A, Part B and Part C. You should answer **ALL** of Question 1 in Part A and **ALL** of Question 2 in Part B. You should answer **ALL** parts of **EITHER** Question 3 **OR** Question 4 in Part C. Part A carries 40 marks, Part B and Part C carry 30 marks each. If you answer more than **ONE** question from **Part C** only your first answer will be marked.

The marks for each part of a question are indicated at the end of the part in [.] brackets. There are 100 marks available on this paper.

You may use any calculator for any appropriate calculations, but you may not use computer software to obtain solutions. Credit will only be given if all workings are shown.

PART A

Candidates should answer **ALL** of Question 1 in Part A.

Question 1

(a) To manipulate contents of a cell in Jupyter Notebook you need to
Select ALL statements that apply. [4]

i. Enter edit mode.
ii. Enter command mode.
iii. Press esc.
iv. Open the command palette.

(b) The body of a loop is
Select ALL statements that apply. [4]

i. The iterator.
ii. Indented.
iii. Encapsulated in curly braces.
iv. Encapsulated in brackets.

(c) Exponentiation is performed by
Select ALL statements that apply. [4]

i. a $\wedge$ b
ii. a != b
iii. a * b
iv. a ** b

(d) Mixed arithmetic with operands of different types is possible in Python.
Choose ONE option. [4]

i. True
ii. False

(e) An example of a standard mapping type is.
Select ALL statements that apply. [4]

i. Dictionary
ii. List
iii. Tuple
iv. Set

(f) Functions in Python are
Select ALL statements that apply. [4]

i. Run as and when they are called.
ii. Declared using the def keyword.
iii. Incapable of handling parameters.
iv. Incapable of returning data as a result.

(g) The csv.writer object returns a writer object responsible for which of these?
Select ALL statements that apply. [4]

i. Converting the user's data into delimited strings on the given file-like object
ii. Converting the user's data into non-delimited strings on the given file-like object
iii. Supporting objects without a write() method.
iv. Converting the user's data into non-delimited strings on the given none file-like object

(h) pandas.read_csv supports which of the following parameters for filepath?
Select ALL statements that apply. [4]

i. Only URLs
ii. Only local files
iii. URLs and local files but not FTP
iv. Any valid string path

(i) HTTP handles which of these?
Select ALL statements that apply. [4]

i. Markup
ii. The display of content in the browser
iii. Client side content rendering
iv. Data Communication across the World Wide Web

(j) Consider the following code.

$$http.client.HTTPConnection('www.python.org:80')$$

What does 80 describe in this example? Select ALL statements that apply. [4]

i. The protocol
ii. The port
iii. The time to live
iv. The maximum transmission unit

PART B

Candidates should answer **ALL** of Question 2 in Part B.

Question 2

Everything in Python is an object. With this in mind, consider the following cases.

(a) Define what is meant by mutable and immutable objects. [4]

(b) Give two examples of mutable built in types. [4]

(c) Give two examples of immutable built in types. [4]

(d) Define a method that we could use on a list object. [4]

(e) The 'Dict' datatype utilises associative mapping. Explain what this is and give an example. [6]

(f) Describe Python object serialization (pickle.) [4]

(g) Which two conditions determine iteration steps in a loop? [4]

PART C

Candidates should answer **ALL** parts in **EITHER** Question 3 **OR** Question 4.

Question 3

(a) Assuming we query the table 'records' how would we use SQL to get all of the columns from the table? [4]

(b) Which command do we use in SQL to remove a table? [4]

(c) Describe the nature of an SQL attack. [4]

(d) How does Pandas handle summing NA data? [4]

(e) What would the sum of only NA data equate to in Pandas? [4]

(f) Describe what fillna() does in the context of a dataframe. [2]

(g) What does the assert expression do? [4]

(h) Describe the differences between a syntax error and an exception. [4]

Question 4

(a) Describe a Pandas Series. [4]

(b) Describe two types of test cases we might use in a unit test. [4]

(c) Describe the difference between distributed and centralised version control. [4]

(d) What information does a git log show? [4]

(e) Describe two different libraries for visualisation and how they differ. [10]

(f) What is the utility of hierarchical indexing in Pandas? [4]

END OF PAPER

.