MSc / PGD Artificial Intelligence

Programming for AI (H9PAI)

Continuous Assessment 1

This test contributes towards **20% of the marks for the module**.

Instructions:

- You may consult any of the notes and code from the classes or the labs.
- You may access online documentation.
- All code must be fully commented.
- Upload your code to Moodle via the CA Submission link.
- This is an individual assessment – there should be no communication (verbal, electronic, or otherwise) between students during the assessment.
- Answer all **4** questions.

Question 1: [55 marks]

Data.gov.ie provides access to open datasets. You have been provided with a JSON file on Moodle (**cattle_births_month_county_2016.json**) containing data relating to cattle births by month, county, and type of cattle for 2016.

> The dataset comprises the number of cattle born from 01/01/2016 to 31/12/2016 organised by the cattle breed type, month of the year and county. Only live births were included, stillbirths were excluded from the dataset. Cattle breed types are categorised as either Beef or Dairy. Cross breeds have the suffix _CROSS, only the dam of the cross is identified.

[From https://data.gov.ie/en_GB/dataset/cattle-births-by-month-county-and-type-of-cattle-for-2016?package_type=dataset]

The data consists of an array of JSON objects where each object contains the following pieces of data:

Property Name	Description
BIRTH_YEAR	The year of birth of the cattle.
COUNTY_ORIGIN	The county in which the cattle were born.
BIRTH_MONTH	The month of birth of the cattle.
CALF_BREED_TYPE	The breed of the cattle (e.g., BEEF or DAIRY).
TOTAL_BIRTH	The total number of cattle births associated with the other property values above.

a) Create a function to import the **cattle_births_month_county_2016.json** JSON data file. Your function should include exception handling clauses.

[15 marks]

b) Using a loop structure, write code to

i) calculate the total number of dairy cattle (i.e. CALF_BREED_TYPE == 'DAIRY') that were born in July (i.e., BIRTH_MONTH == 'JUL')

ii) identify the county, and month in which there was the greatest number of dairy cattle births

Print the results from i) and ii) above.

[20 marks]

c) From the JSON data, extract the **BIRTH_YEAR**, **COUNTY_ORIGIN**, **BIRTH_MONTH**, **TOTAL_BIRTH** data for the dairy cattle births and write it to a CSV file. Your file should also contain appropriate column names.

[*Hint:* you might consider accumulating the relevant data for the CSV file as you are looping through the data in part b).]

[20 marks]

Question 2: [15 marks]

a) Create a 2-dimensional NumPy *ndarray* filled with 100 random numbers. Ensure that your array has 20 rows and 5 columns.

[5 marks]

b)

i. Compute the sum total of all entries in the first and last rows.
ii. Compute the sum total of all entries in the second and fourth columns.

[5x2= 10 marks]

Question 3: [20 marks]

Using the comma-separated file *scores.csv* solve the following tasks.

a) Write a function *display_numbers()* which takes one parameter - a file path. The function should read floating point numbers from each line in the specified file, and compute the total for the values on each line. Print each total to 2 decimal places. Use exception handling to deal with the potential case where the input file does not exist.
Apply the function *display_numbers()* to *scores.csv*

b) Write a function *reverse_numbers()* which takes two parameters - an input file path and an output file path. The function should read floating point numbers from each line in the specified input file. The order of the values from in line should then be reversed, and these lines should be written to the specifed output file. Include exception handling code.
Apply the function *reverse_numbers()* to *scores.csv* to create a new file *reversed.csv*.

Question 4: [10 marks]

a) Using pandas data frame load the file 'books-catlogue.csv'

[2 marks]

b) Plot

(i) the number of chapters against the number of characters;
(ii) the number of chapters against the number of words.

[6 marks]

c) Do the plots suggest a correlation in either case?

[2 marks]