

PGD Data Analytics (PGDDA_SEP)
Database and Analytics Programming (H9DAP)
Continuous Assessment 1: In-class Test

This test contributes towards a **maximum** of 30% of the marks for the module.

Instructions:

- You may consult any of the notes and code from the classes or the labs.
- **All code must be fully commented**
- Upload your code to Moodle
- This is an individual assessment – there should be no communication (including verbal and electronic) between any students during the assessment.

Question 1:

On 12 February 2020, the novel coronavirus was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) while the disease associated with it is now referred to as COVID-19. Since the beginning of the coronavirus pandemic, the European Centre for Disease Prevention and Control (ECDC)'s Epidemic Intelligence team has been collecting, on daily basis, the number of COVID-19 cases and deaths, based on reports from health authorities worldwide.

The file **opendata.ecdc.europa.eu.xml** has been downloaded from [ECDC's website](#), and contains the latest available public data on COVID-19 including a daily situation update, the epidemiological curve and the global geographical distribution (EU/EEA and the UK, worldwide).

Create a function to import this XML file using Python's **xml** package and obtain the root of the xml tree. Your function should include exception handling clauses.

[10 marks]

Write code to print a list of all the tags of the first record of the xml tree above.

Expected Output:

```
['dateRep', 'day', 'month', 'year', 'cases', 'deaths', 'countriesAndTerritories', 'geoId', 'countryterritoryCode', 'popData2019', 'continentExp', 'Cumulative_number_for_14_days_of_COVID-19_cases_per_100000']
```

[5 marks]

Write code to print records from the xml including the *day*, *month*, *year*, *cases*, *deaths* and *Cumulative_number_for_14_days_of_COVID-19_cases_per_100000* for Ireland for October 2020.

[10 marks]

Extract the above XML data and write it to a CSV file. Your file should also contain the column names.

[15 marks]

Question 2:

You are provided with the string below:

```
'H9eoDAPklo-oPGserfDDAoa_SEPiskdsf'
```

Write a Python program to remove lowercase substrings from the string.
[Bonus marks for using a lambda function]

[10 marks]

You are provided with the following text:

```
<p>College Website :</p><a href="https://www.ncirl.ie/">National College of Ireland</a><p>Course Website :</p><a href="https://mymoodle.ncirl.ie/course/view.php?id=633">PGGDA_SEP H9DAP</a>
```

Write a Python program that uses regular expression to identify all the URLs in the text, and returns them as an array.

[10 marks]

You are provided with the following text:

```
Python is a programming language.
Python lets you work more quickly and integrate your systems more effectively.
Python can be easy to pick up whether you're a first time programmer or you're experienced with other languages.
Python is developed under an OSI-approved open source license, making it freely usable and distributable, even for commercial use.
The Python Package Index (PyPI) hosts thousands of third-party modules for Python.
```

Write a regular function that highlights the occurrence of any of the strings Python, you and program.

Expected output:

```
*Python* is a *program*ming language.
*Python* lets *you* work more quickly and integrate *you*r systems more effectively.
*Python* can be easy to pick up whether *you*'re a first time *program*mer or *you*'re experienced with other languages.
*Python* is developed under an OSI-approved open source license, making it freely usable and distributable, even for commercial use.
The *Python* Package Index (PyPI) hosts thousands of third-party modules for *Python*.
```

[10 marks]

Question 3:

Create a **7x5** integer array from a range between 100 to 450 such that the difference between each element is **10** and then **split** the array into five equal-sized sub-arrays
[15 marks]

You are provided with the array below:

```
[[3 ,6, 9, 12], [15 ,18, 21, 24], [27 ,30, 33, 36], [39 ,42, 45, 48], [51 ,54, 57, 60]]
```

Write a function to return an array of odd rows and even columns.

Expected output is:

```
[[ 6 12],[30 36],[54 60]]
```

[10 marks]

Your are provided with the array below:

```
[[34,43,73],[82,22,12],[53,94,66]]
```

Print the maximum value in each row.

Expected output is:

```
[82 94 73]
```

[5 marks]