**CA2 – Data Analysis**

**Value 35%**

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

* **Open book, Individual CA.**
* **120 minutes Duration.**
* **All work to be submitted should include your python code, outputs from the code and any charts created and any explanations you were asked to provide.**
* **Save all work under your Xnumber and upload to the CA2 – Upload link on moodle.**
* **Keep a backup of all your work.**
* **Solutions can use the CSV file reader, pandas or not the choice is yours in how best to solve this.**

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

**Introduction:**

Go to moodle and download the dataset called **COVID\_19\_Data.**This is a .csv file and it contains the following data:

**SNo.:** This is the primary Key

**ObservationDate:**Date of the recording

**Provence/State:**This column is 69% Nulls

**Country/Region:**The name of each country e.g. China

**Last Update**: Time data was recorded.

**PopulationProportionCovidCases:**The proportion of these cases based on the population

**Confirmed:**Number of confirmed cases on the associated date for the given country (provence).

**Deaths:**Number of confirmed deaths on the associated date for the given country (provence)

**Recovered:**Number of confirmed recovered on the associated date for the given county.

**NOTE:**

Marks will **not** be awarded for hard coded lists. The original list of values need to be manipulated to get the lists needed to do the calculations required in this CA.

Ensure all graph layouts are fully readable even if this means making the graphs very small to accommodate the values.

Marks awarded for detailed explanations. Short answers are not sufficient.

**Question:**

You have been asked to analyse the given data in particular you are to do the following:

1. Read in the .CSV file and put it into a suitable 2D list so that you can work on it to get the following answers.

Code:

import statistics as st

import matplotlib.pyplot as plt

#question 1

myfile = open("covid\_19\_data.csv","r")

dataIn = myfile.read()

instances1D = dataIn.split("\n")

output2D =[]

for instance in instances1D[1:]:

templist = instance.replace("\t",",").split(",")

output2D.append(templist)

print(output2D)

output is too LARGE

[6 Marks]

1. Get the total, min and max number of cases of COVID-19 in Ireland from the dataset given.

[6 Marks]

1. Get the total, min and max number of cases of COVID-19 in the US from the dataset given

[6 Marks]

1. Explain what we see and what this might mean from the data results in Question 3 and Question 4 above.

In question 2 and 3, the output shown is the minimum, maximum and total number of cases in Ireland and US. From the data shown, that the US has a much higher amount of covid cases than Ireland which is about 4 times more than the cases in shown in Ireland

[7 Marks]

1. Get the max number of cases on a given day and its full details.

[10 Marks]

1. Get the min number of cases on a given day and its full details.

[10 Marks]

1. Get the Average number of Irish cases and get the average number of US cases.

[8 Marks]

1. What other measures of average could be used to measure the cases? Would either of these other measures have been better? Explain your answer.

[7 Marks]

1. List all countries that have more than 100 deaths on the 12/04/2020. Ensure only one instance of each country appears i.e. remove duplicates from result set. (Note the date is in US format MM/DD/YYYY)

[10 Marks]

1. Display a list of all the countries in the data set. From this list choose any 5 countries and graph the **recovered** for each of these countries on a graph of your choice. Explain your choice of graph and explain what the graph tells.

[15 Marks]

1. Graph the number of Irish cases per day in a line chart along with the number of US cases per day. Make sure Ireland is Blue and US is red. Discuss in detail the chart that is displayed.

[15 Marks]