# Assignment 01 - Test your Python skills

This set of questions can help you understand your programming skills in Python. Remember that this is a simple test, and we will work with more complex programming problems during the semester.

# Dataset Information

For this problem, you will be working with COVID-19 sequence processing data from Kaggle. The dataset contains data about the processing of COVID-19 sequences by different countries over time. It comes as a Comma-Separated Value (CSV) file. It includes the following six columns:

1. **location**: the country for which the information is provided
2. **date**: the date of the data entry
3. **variant**: the COVID-19 variant for the data entry
4. **num\_sequences**: the number of sequences *processed* (for the country, variant, and date)
5. **num\_sequences\_total**: the total number of sequences *available* (for the country, variant, and date)
6. **perc\_sequences**: the percentage of the available number of sequences that were processed (*Note: this value is out of 100*)

Each row (or data entry) in the dataset represents the processing of *one* variant by *one* country on *one* day.

A copy of this dataset is provided to you. However, if you want to, you can also find the dataset here: <https://www.kaggle.com/yamqwe/omicron-covid19-variant-daily-cases?select=covid-variants.csv>.

# Problem 1

The three main variants of COVID-19 that we’ve experienced in the United States are:

1. Alpha
2. Delta
3. Omicron

However, there are many other variants recognized by the WHO.

For this problem, determine which other variants are included in the dataset. Additionally, sort the variant names alphanumerically.

*Note: the* ***variants*** *column contains 2 “catch-all” categories called “non\_who” and “others.” Do* ***NOT*** *include these categories in the list.*

# Problem 2

Determine which variant of COVID-19 has the most sequences processed across the entire dataset.

# Problem 3

Determine which country processed sequences across **all** variants (including the “catch-all” categories). The output should be the name of a single country.

# Problem 4

Problem 4 has two parts.

## Part A

Determine which country did the best at processing sequences across the Alpha, Delta, and Omicron variants **only**. The output should be the name of a single country.

## Part B

Determine the ranking of the United States at processing sequences across the Alpha, Delta, and Omicron variants **only**.

*Note: the best country overall should have a ranking of 1, but indexing in Python starts at 0*.

# Problem 5

Determine each country’s total number of **processed** sequences for the Omicron variant on December 27, 2021. Sort the output from the highest number of processed sequences to the smallest number of processed sequences. Each element in the output should include the country's name and the number of processed sequences.

# Problem 6

Determine the percentage of processed sequences for the Alpha, Delta, and Omicron variants **only** in the United States.

# Implementation Requirements

There are only two simple requirements for your implementation:

1. All code should be written in Python 3. We’ll run your code with a Python 3 interpreter, so Python 2 code will almost certainly fail.
2. All code should be either a single Python script (.py file) or Jupyter Notebook (.ipynb file).

Upload your solution in canvas before 11:59 PM ET today.