DAV 5400 Module 3 Assignment (30 Points)

Our Module 3 Assignment makes use of the functionality provided by Python's **NumPy** package.

----

1. **(5 Points)** Write a Python/NumPy code block that finds the distinct array items that exist in either array **a** or array **b** but which do not appear within both arrays:

a = np.array([1, 2, 3, 2, 3, 4, 3, 4, 5, 6])

b = np.array([7, 2, 10, 2, 7, 4, 9, 4, 9, 8])

For example, if a 2 is found in both array **a** and array **b**, your output should **not** include a 2. Furthermore, the content of the output should be comprised solely of unique data values.

----

2. **(5 Points)** Create the following 4x5 array using knowledge you have of Python's / NumPy's sequencing functionality so that you do not need to explicitly key in every integer value.

1	5	9	13	17
2	6	10	14	18
3	7	11	15	19
4	8	12	16	20

----

3. **(5 Points)** You are given the following two arrays:

a = np.array([12, 5, 7, 15, 3, 1, 8])

b = np.array([14, 6, 3, 11, 19, 12, 5, 3, 4])

Write a Python/Numpy code block that adds to array **a** any items that are present in array **b** but <u>**not**</u> already present within array **a**.

----

4. **(7 Points)** The process of transforming a multidimensional array into a unidimensional array is referred to as "flattening". Transform the 4x5 array shown above in **Problem 2** into a unidimensional array such that the sequence of values contained within the array is as follows: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20

----

*(** NOTE: The assignment continues on the next page **)*

5. (8 Points) For this problem you will load a small data set to your online GitHub repository and then read it into a NumPy 2D array. You will then use the content that NumPy array to answer a series of analytical questions.

The original source of the data set contains an overview of its content:

https://data.cityofnewyork.us/Environment/Water-Consumption-In-The-New-York-City/ia2d-e54m

Scroll down to the bottom of that page to access an interactive data viewer that includes the column headings for the data. Once you are comfortable in your understanding of the various data attributes, get started on the assignment as follows:

> 1) Load the provided M3_Data.csv file to your online DAV 5400 Github Repository.
>
> 2) Then, from within your M3 Assignment Jupyter Notebook, read the data set from your Github repository and load it into a Pandas dataframe using the code snippet provided below. ** *Please note that you will need to replace the 'link_to_raw_copy of data file in your Github Repository' placeholder shown below with a link to the "raw" version of the file you pushed to your online GitHub repository. ***

Copy the following small Python code snippet into your Jupyter Notebook:

_____

```
import pandas
import numpy as np

# suppress scientific notation
np.set_printoptions(suppress = True)

# read the data file from GitHub and convert to a NumPy ndarray object
# be sure to replace the 'link_to_raw_copy of data file in your Github Repository' placeholder shown below
# with a link to the "raw" version of the file you pushed to your online GitHub repository

nyc_water = pandas.read_csv('link_to_raw_copy of data file in your Github Repository').to_numpy()

# verify you have an ndarray object
type(nyc_water)
```

_____

This code will read the contents of the provided data file into a NumPy ndarray object. Execute the code within your Jupyter Notebook. If successful, the output of the **type(nyc_water)** function should be: **numpy.ndarray**

Using the nyc_water array you've created, answer the following questions using NumPy's indexing, slicing, methods, and functions

- Which year had the maximum per capita daily water consumption?

- Add a new column to the Numpy array containing the **annual change in the population of the city** from year to year. (**HINT**: Your new column should include the calculation of a data value for every year represented within the data set except for the first year, which should contain a data value of 'N/A').

- What is the **median** and **variance** of the population of the city during the time period represented within the data set?

- What was the average annual change in the population of the city?

Please note that it may take a bit of time for Jupyter to load the data set from the GitHub repository. You will know that process has finished once the results of the '**type(nyc_water)**' command get displayed within your copy of the notebook.

Be sure to include some commentary explaining your approach to solving each of the individual problems. Save all of your work for this assignment within **a single Jupyter Notebook** and submit it via the Module 3 Assignment page within Canvas. Be sure to save your Notebook using the nomenclature we've been using, i.e., **first initial_last name_M3_assn**" (e.g., J_Smith_M3_assn_).