

AIM 5001 Module 6 Assignment (100 Points)

Our Module 6 Assignment makes use of the functionality provided by Python's **NumPy** package.

1. (10 Points) Write a Python/NumPy code block that finds the distinct/unique common items between these two 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])
```

Your output should contain only the distinct overlapping values. For example, if a 2 is found in both array **a** and array **b**, your output should contain only one 2 even if array **a** contains more than one 2 within it.

2. (10 Points) Create the following 5x3 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	6	11
2	7	12
3	8	13
4	9	14
5	10	15

3. (10 Points) The process of transforming a multidimensional array into a unidimensional array is referred to as "flattening". Transform the 5x3 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

4. (15 Points) Transform the 2-D array shown in Problem 2 into a 3 dimensional array such that the first column becomes the first dimension of the 3-D array, the second column becomes the second dimension of the 3-D array, and the third column becomes the third dimension of the 3-D array.

5. (15 Points) Transform the 3-D array you created in Problem 4 back to the 2-dimensional format shown in Problem 2.

6. (10 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])
```

Write a Python/NumPy code block that removes from array **a** any items that are also present in array **b**.

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

7. (30 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 the 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 M6_Data.csv file to your online AIM 5001 Github Repository.
- 2) Then, from within your M6 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 slicing, methods, and functions

- What is the maximum yearly NYC consumption of water in millions of gallons per day?
- How many calendar years are represented within this data set? NumPy's **shape** command is one way to find out.

- What is the mean and the standard deviation of the per capita daily water consumption?
- What is the increase or decrease in population from year to year? Use NumPy's `diff` function to create an array of differences and save that to a variable called `pop_diff`, then print that variable to the screen.

Please note that it may take a bit of time for Jupyter to load the data set from your 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 upload / submit it within the provided M6 Assignment Canvas submission portal. Be sure to save your Notebook using the nomenclature we've been using, i.e., **first initial_last name_M6_assn**" (e.g., J_Smith_M6_assn_).