

DAV 5400 Fall 2019 Week 4 Assignment (30 Points)

Our Week 4 Assignment makes use of the functionality provided by Python's **NumPy** package.

1. 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. 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. 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**.

4. Suppose you have ridden a bicycle from New York City to Key West, Florida. Your bicycle odometer shows the total miles you have travelled thus far, which you make a note of each day with paper and pencil. Your first two entries might be '55' and '120', indicating that you rode your bike 55 miles on day 1 and 65 miles on day 2. Your task is to create a NumPy array wherein you can record the cumulative miles you recorded each day during your trip. Then use your Python skills to show the total miles that you rode each day. Assume the trip took a total of 35 days and included at least 4 non-consecutive days where no cycling was possible due to either weather conditions or personal fatigue/soreness. For the remaining 31 days, you would have covered the entire distance between NYC and Key West. Complete your work within your Jupyter Notebook for this assignment and be sure to include a writeup explaining your approach to this problem, including the ways in which you decided to make use of NumPy.

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

5. Open the Jupyter Notebook provided to you with this assignment. The first code block within that file will load a small data set from a GitHub repository into a NumPy 2D array. The original source of that data contains an overview of its contents:

<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.

Using the NumPy array that has been loaded for you in the Jupyter Notebook, 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 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 upload it to your online DAV5400 GitHub directory. Be sure to save your Notebook using the nomenclature we've been using, i.e., **first initial_last name_W4_assn**" (e.g., J_Smith_W4_assn_).

As a reminder, this assignment is due no later than 11.59pm on Sunday Sep 22.