HOMEWORK 6

(This HW covers Numpy and Pandas)

(A single Jupyter Notebook or .PY to be submitted to Wolfware. Do not zip the files before submission. All of the solutions below must use the numpy package)

1) Create one random integer array, aVec, containing 20 values ranging from 0 to 100. Generate a random float value, b, between 0 and 100. Find the index in aVec whose value is closest to b. Print also the value within aVec through the index.

   Example

   A = [46, 76, 74, 27, 6, 74, 53, 63, 86, 7, 77, 17, 72, 16, 61, 77, 25, 67, 18, 34]

   b = 90.07

   Index of A closest to b = 8

   A[8] = 86

2) Write a Python program to create random vector with values between 0 and 1 of size 15 and replace the maximum value within it by -1.

3) Replace all the positions at which non-zeros exist in the given matrix, A to -1. The given 2D array is

   A = [1 0 1 0 1 1 1 0 1 0 0

   0 1 0 1 1 0 1 0 1 1 1

   0 0 0 0 0 0 0 0 1 1 1]

4) Write a numpy program that creates a 1D vector with random integer values ranging from 10 to 100 with 20 elements. Further, for each value that is below the mean of all the values in the vector, replace those values with 0.

5) Consider an elementary system of three linear equations: x + 2y + z = 2, 2x + 6y + z = 7, x + y + 4z = 3, Solve for x, y and z using a python program using the numpy library.

6) Randomly generate two objects. The first object is a 1D array, a vector 'v' of 10 integer elements. Randomly generate a second 2D array, B, a matrix of 10 x 5 elements. Subtract the 1d array, 'v' from the 2d matrix B, such that each item of v subtracts from respective column of A. The resulting matrix must still be of shape 10 x 5. The values in the vector randomly range from (1, 100)

7) A cleaning services company compiled the following data related to the annual profit of the firm to its annual Facebook advertising campaign (measured in thousands) as shown in the table below

Advertising Expenditure	12	14	17	21	26	30
Profit	60	70	90	100	100	120

a) Find the best least squares fit to the data in the form of a straight line given by y = mx + c by writing a numpy program.

b) Plot the points and least square fit line using matplotlib.

c) Calculate the profit if the company allocates in its next FB campaign with a $50,000 budget allocation. Report the value in $ currency.

8) The US Social Security Administration releases baby names given by parents from the Years 1880 up until 2022. The dataset zip file can be downloaded from : Baby Names from Social Security Card Applications - National Data - CKAN (https://catalog.data.gov/dataset/baby-names-from-social-security-card-applications-national-data)

Answer the following questions with regards to the Baby Names Dataset (20).

a. Total number of records across all years (1880 – 2021).

b. Compute the number of times that each name was used, separately for boys and girls across all years. For example, the output should be

Sex	Name	Total
F	Aabha	21

c. Display a single plot showing the popularity of the name 'Harper', 'Evelynn', 'Evan', 'Ethan' across the years.

d. Compute the 10 most popular names for all Male 'M' names and Female – 'F' for every year.