**George Brown College**

**Centre for Arts, Design and Information Technology**

**Machine Learning Process – Working on a CSV file (Assignment 01)**

The goal of this assignment is to practice working with CSV files and using NumPy functions.

Please follow the following steps using the Jupyter Notebook or any other environment to implement steps. In the end, you need to submit all the codes in a word document into 03-Assignments à Assignment 01 (Submission) folder.

Please note that this is an individual assignment. You can help your team members to understand the problem statement, but everyone needs to do the assignment individually and submit their tasks individually.

Moreover, as there are five different versions of assignment 01 (Assignment 01-(Ver. 00).docx, Assignment 01-(Ver. 01).docx, Assignment 01-(Ver. 02).docx, Assignment 01-(Ver. 03).docx, Assignment 01-(Ver. 04).docx) everyone needs to pick one version, which is calculated from the following formula:

Please count the number of characters in your first name and calculate the remainder of this number divided by 5. The number would be 01, 02, 03 or 04 and you will work on that.

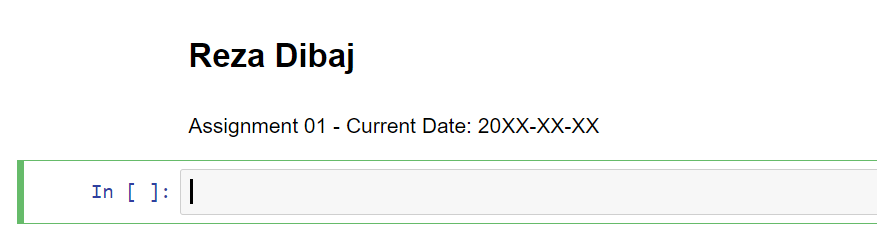
For instance, my name is Reza, 4 characters, 4%5 (4 modulo 5) is equal to 4, so I need to pick Assignment 01-(Ver. 04).docx.

The individual submission is mandatory, while everyone will open and run their assignments in the class to demonstrate what they have learned from it.

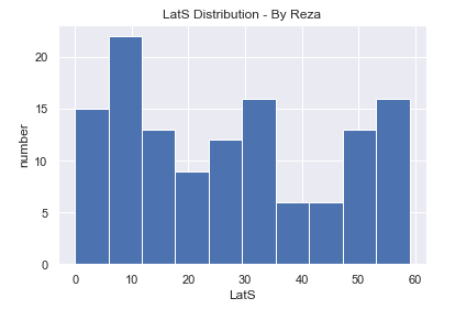
The due for the first assignment is Monday, October 12th, 2020, 11:55 PM. You will present your assignment in a lab session after in your lab section.

**Let’s start the assignment:**

1. Write your name and relevant info in the first cell as follow:



1. Import the **cities.csv** file and place it in an appropriate dataset. (2 points)
2. Fetch the **LatS** column of the dataset and use a variable to hold this array. (1 points)
3. Print the content of the **LatS** column. (1 points)
4. Calculate the average, standard deviation, minimum and maximum of the array and print the result. (2 points)
5. Find the 30th percentile, median and 70th percentile of the **LatS** column. (2 points)
6. Using the Matplotlib library, please draw a histogram chart for the **LatS** column. Please use appropriate title, X-label and Y-label, and include your first name in the title, e.g. ‘**LatS Distribution – By Reza**’. (2 points)



Thank you and good luck,

Reza

You are not a drop in the ocean. You are the entire ocean, in a drop. ~Rumi