

Assignment 2

Python Functions

MIS4001/5000

2025-07-06

1 Instruction

File to submit: YourFirstNameLastName.ipynb Due Date: 7/20/2025

You will practice Numpy, Pandas, and Matplotlib in this assignments. All of your code must be in the same file, YourFirstNameLastName.ipynb. In your file, also include a comment at the top with your name, student ID, and GovState Email. All code must be well commented.

2 Part 1 Numpy

Please follow the instruction below to perform the following task.

1. Can you create the following Numpy array and print the exact value, [1, 1.5, 2, 2.5, 3, 3.5]?
2. Please create a Numpy array with shape (5, 10) and fill it with random values between (0, 1).
3. Please create a one dimension Numpy array with length 10 and fill it with random value between (1, 100). (Note: You will need this Numpy array for the following questions. A random seed is recommended but not required.)
4. Can you shuffle the above Numpy array? (Note: this is not covered in the lecture. You need to do a little research and figure this out.)
5. Can you sort this Numpy array to a Non-decreasing order? Note: you can use any method to sort your array.
6. Can you explain how did you sort this Numpy array?

3 Part 2 Pandas

1. Find and download a data on Kaggle: <https://www.kaggle.com/datasets>. Your dataset should at least contains 5 numeric columns and more than 30 rows.
2. Please load the dataset to your programming environment.
3. Please check if there is missing values.
4. Please check the data type of each column.
5. Please calculate the basic statistics for each columns. Note there is a function to help you do this.
6. Can you check if there is linear relationship between two random column using visualization? Note you must use pair-wise visualization method we discussed in the lecture.
7. Discuss what kind of information can you get by performing some analysis with the dataset you selected.

4 Part 3 Matplotlib

You should already develop one visualization in Part 2. Now, based on your dataset selection, please develop 3 more visualization and practice at least 3 of the following, line plot, scatter plot, histogram, bar chart. Please note creating visualization means you want to show some information in the dataset. You must be able to explain what information you can get after you create your visualization. Fail to provide the explanation will not be graded.