

# Hands-on Activity 6.1 Introduction to Data Analysis and Tools

## CPE311 Computational Thinking with Python

Name: LastName, FirstName

Section: CPE22Sx

Performed on: MM/DD/YYYY

Submitted on: MM/DD/YYYY

Submitted to: Engr. Roman M. Richard

## 6.1 Intended Learning Outcome

1. Use pandas and numpy data analysis tools.
2. Demonstrate how to analyze data using numpy and pandas

## 6.2 Resources:

- Personal Computer
- Jupyter Notebook
- Internet Connection

## 6.3 Supplementary Activities:

### Exercise 1

Run the given code below for exercises 1 and 2, perform the given tasks without using any Python modules.

```
In [ ]: import random
random.seed(0)
salaries = [round(random.random()*1000000, -3) for _ in range(100)]
```

Using the data generated above, calculate the following statistics without importing anything from the statistics module in the standard library (<https://docs.python.org/3/library/statistics.html>) and then confirm your results match up to those that are obtained when using the statistics module (where possible):

- Mean
- Median
- Mode (hint: check out the Counter in the collections module of the standard library at <https://docs.python.org/3/library/collections.html#collections.Counter>)
- Sample variance
- Sample standard deviation

```
In [ ]: # Write a comment per statistical function
```

### Exercise 2

Using the same data, calculate the following statistics using the functions in the statistics module where appropriate:

- Range
- Coefficient of variation
- Interquartile range
- Quartile coefficient of dispersion

```
In [ ]: # Write a comment per statistical function
```

### Exercise 3: Pandas for Data Analysis

Load the diabetes.csv file. Convert the diabetes.csv into dataframe

Perform the following tasks in the diabetes dataframe:

1. Identify the column names
2. Identify the data types of the data
3. Display the total number of records
4. Display the first 20 records
5. Display the last 20 records
6. Change the Outcome column to Diagnosis

- 7. Create a new column Classification that display "Diabetes" if the value of outcome is 1 , otherwise "No Diabetes"
- 8. Create a new dataframe "withDiabetes" that gathers data with diabetes
- 9. Create a new dataframe "noDiabetes" thats gathers data with no diabetes
- 10. Create a new dataframe "Pedia" that gathers data with age 0 to 19
- 11. Create a new dataframe "Adult" that gathers data with age greater than 19
- 12. Use numpy to get the average age and glucose value.
- 13. Use numpy to get the median age and glucose value.
- 14. Use numpy to get the middle values of glucose and age.
- 15. Use numpy to get the standard deviation of the skinthickness.

In [1]:

# Indicate which item you're answering with a comment

## 6.4 Conclusion

Edit this markdown block.

---

End.