**East West University**
**Department of Computer Science and Engineering**

**CSE 303: Statistics for Data Science**
**LAB 04**
**Course Instructor: Dr. Mohammad Rezwanul Huq**

---

# Practice Lab on Exploratory Data Analysis using Pandas

## Lab Objective

Practice data analysis using Pandas in Python.

## Lab Outcome

After completing this lab successfully, students will be able to:

1. **Apply** independently the Python Pandas functions and operations for data manipulation, analysis and cleaning.
2. **Use** Pandas functions properly and **Write** appropriate Python programs for data analysis.

## Psychomotor Learning Levels

This lab involves activities that encompass the following learning levels in psychomotor domain.

| Level | Category | Meaning | Keywords |
|---|---|---|---|
| P1 | Imitation | Copy action of another; observe and replicate. | Relate, Repeat, Choose, Copy, Follow, Show, Identify, Isolate. |
| P2 | Manipulation | Reproduce activity from instruction or memory | Copy, response, trace, Show, Start, Perform, Execute, Recreate. |

## Required Applications/Tools

- Anaconda Navigator (Anaconda3)
  - Anaconda is a distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment.
  - Popular Tools/IDEs: Spyder, Jupyter Notebook
- Google Colab: Colaboratory, or "Colab" for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education.

## Lab Activities

### 1. Designing a Python Module

Modules refer to a file containing Python statements and definitions.

A file containing Python code, for example: `example.py`, is called a module, and its module name would be example.

We can define our most used functions in a module and import it, instead of copying their definitions into different programs.

```python
# Python Module example

def add(a, b):
   """This program adds two
   numbers and return the result"""

   result = a + b
   return result
```

*Importing a Python Module:*

We can import the definitions inside a module to another module or the interactive interpreter in Python.

We use the import keyword to do this. To import our previously defined module example, we type the following in the Python prompt.

```
import example
```

We need to make sure that the module resides in the same directory as does the Python Path or in the same directory as the main Python program (`main.py`).

2. **Reading the Dataset and Exploring it**

   Read the given dataset and start exploring it!

**Lab Tasks**

1. How many rows and columns this dataframe has?
2. Describe (numerical summary) the time and amount column.
3. There are 31 columns in the dataset. Compute some statistical measures like mean, median, standard deviation, variance using Pandas Function.
4. Compute the mean of any column using your own module and compare it with the mean value of Pandas.
5. Show the histogram of Time and Amount column.
6. Find the percentage of rows with class value = 0 (Non-Fraudulent) and class value = 1 (Fraudulent).
7. Show the result you have got in 6 using a histogram.
8. Show the histrogram (data distribution) of a few other columns. Differentiate between left-skewed and right-skewed distributions.
9. Find positive correlations among columns.
10. Support your findings in Question 9 using a BoxPlot.
11. Support your findings in Question 9 using a Scatter Plot.
12. Find negative correlations among columns.
13. Support your findings in Question 10 using a BoxPlot.
14. Support your findings in Question 9 using a Scatter Plot.