**UCT31021 – PRACTICAL FOR ARTIFICIAL INTELLIGENCE**
**DEPARTMENT OF ICT**
**FACULTY OF TECHNOLOGY**
**SOUTH EASTERN UNIVERSITY OF SRILANKA**

---

**Lab sheet:** 08

**Date:** 15-05-2025

**Title: Python Libraries**

**Data Processing and Visualization using Python**

**Aims:**

- Work with Python Library
- Work with Data Frames
- Explore the Graphics in Python

**Task:**

1. Read data Using pandas
2. Exploring data frames
3. Data frame Data Types

| Pandas Type | Native Python Type | Description |
|---|---|---|
| object | string | The most general dtype. Will be assigned to your column if column has mixed types (numbers and strings). |
| int64 | int | Numeric characters. 64 refers to the memory allocated to hold this character. |
| float64 | float | Numeric characters with decimals. If a column contains numbers and NaNs, pandas will default to float64, in case your missing value has a decimal. |
| datetime64, timedelta[ns] | N/A (but see the datetime module in Python's standard library) | Values meant to hold time data. Look into these for time series experiments. |

4. Data Frames attributes

| df.attribute | description |
|---|---|
| dtypes | list the types of the columns |
| columns | list the column names |
| axes | list the row labels and column names |
| ndim | number of dimensions |
| size | number of elements |
| shape | return a tuple representing the dimensionality |
| values | numpy representation of the data |

5. Data frame methods

| df.method() | description |
|---|---|
| head( [n] ), tail( [n] ) | first/last n rows |
| describe() | generate descriptive statistics (for numeric columns only) |
| max(), min() | return max/min values for all numeric columns |
| mean(), median() | return mean/median values for all numeric columns |
| std() | standard deviation |
| sample([n]) | returns a random sample of the data frame |
| dropna() | drop all the records with missing values |

6. Select a column in a data Frame
7. Data Frame GroupBy method
8. Data Frame: Filtering
9. Data Frame: Slicing
10. Data Frame: Selecting rows
11. Data Frame: method Loc
12. Data Frame: Sorting
13. Aggregation Function in Pandas
14. Descriptive Statistics

| df.method() | description |
|---|---|
| describe | Basic statistics (count, mean, std, min, quantiles, max) |
| min, max | Minimum and maximum values |
| mean, median, mode | Arithmetic average, median and mode |
| var, std | Variance and standard deviation |
| sem | Standard error of mean |
| skew | Sample skewness |
| kurt | kurtosis |

15. Graphics: -

| | description |
|---|---|
| histplot | histogram |
| barplot | estimate of central tendency for a numeric variable |
| violinplot | similar to boxplot, also shows the probability density of the data |
| jointplot | Scatterplot |
| regplot | Regression plot |
| pairplot | Pairplot |
| boxplot | boxplot |
| swarmplot | categorical scatterplot |
| factorplot | General categorical plot |