**LABS**

**Lab 1: Setup**

* Install Python and setup environments
* Dowload Dataset:

<https://github.com/hvphi/Data-Analytic>

Google Dataset Search <https://datasetsearch.research.google.com/>

The U.S. Government’s open data <https://data.gov/>

UCI Machine Learning Repository <https://archive.ics.uci.edu/ml/index.php>

**Lab 2: Python programing**

**Problem 1.** Write a program to manage a random integer array A with size of n using pointers (no array indexing). The program consists of the following functions:

a)     Input data (random numbers).

b)     Print out the array.

c)     Count all odd numbers.

d)     Compute sum of the array.

e)     Compute average of the array.

f)      Check a value x if x esists in A.

g)     Count number of elements between a and b.

h)     Insert a new element x into A at position i.

i)       Remove an element at position i

**Problem 2.** Write a program to find GCD (Greatest Common Divisor**)** and LCM (Lowest Common Multiple) of two integer numbers using recursion.

**Lab 3: Numpy Library**

**Problem:** Write a program to manage a random integer array A with size of n using Numpy library. The program consists of the following functions:

a)     Input data (random numbers).

b)     Print out the array.

d)     Compute sum of the array.

e)     Compute average of the array.

f)      Find x.

h)     Insert a new element x into A at position i.

i)       Remove an element at position i

**Lab 4: Read data from files (CSV, Excel)**

Write a program to read file tiny.tsv and Job\_survey.xlsx

**Lab 5: Read data from website/API**

Write a program to read data from website/API:

Google Dataset Search <https://datasetsearch.research.google.com/>

The U.S. Government’s open data <https://data.gov/>

UCI Machine Learning Repository <https://archive.ics.uci.edu/ml/index.php>

**Lab 6: Data Assessment**

* Write a program to read data from files patients.csv, treaments.csv and assess it:

Quick look into data (head()/tail()/sample())

Get data size (shape)

Get information (info())

Get sumary (describe())

Get column name (columns)

Check NaN values (isna())

Check unique values (unique())

Check duplicates (duplicated().sum())

….

**Lab 7: Clean data**

* Write a program to clean data:

Remove rows/columns

Add rows/columns

Missing values dealing

Rename columns

Data type casting

Merge datasets

…

**Lab 8: Visualization**

* Write a program to visualize data:

Line,

Bar,

Piece,

Histogram,

Heat map,

…

**Lab 9: Visualization**

* Write a program to refine charts:

Colour,

Background

Labels,

…

**Lab 10: Present the analysis results**

Observation

Comments

Suggesstions