School of Computer Science Engineering and Technology

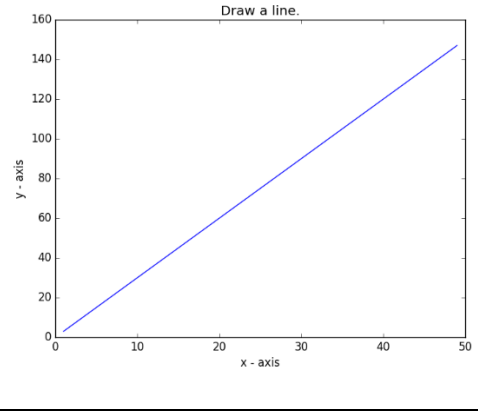
Course- BCA
Course Code- BCA355L
Year- 2022
Date- 02-03-2022

Type- Elective
Course Name - Data Mining and Predictive Analysis
Semester- Even
Batch- NA

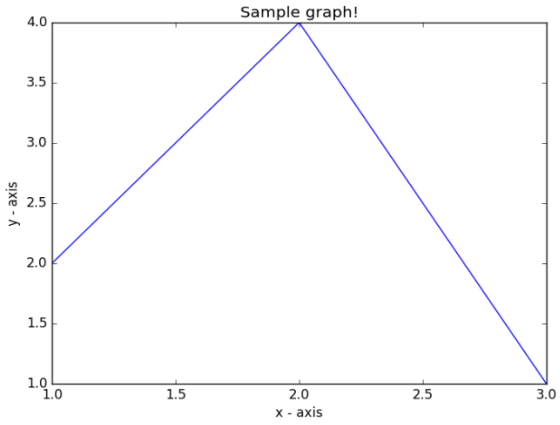
Lab Assignment 5

Objective: Working with Exploratory Data Analysis (EDA).

1. Draw the following plot with the appropriate label in the x-axis, y-axis, and title.
   The output is shown in the following screenshot:(15)



(a)



(b)

2. Draw an appropriate plot/chart to explore the stock market-related data (20)
   Step1: Add the following stock market-related data between October 3, 2021, to October 7, 2021, into the xyz.csv file

```
Date,Open,High,Low,Close
10-03-21,774.25,776.065002,769.5,772.559998
10-04-21,776.030029,778.710022,772.890015,776.429993
10-05-21,779.309998,782.070007,775.650024,776.469971
10-06-21,779,780.47998,775.539978,776.859985
10-07-21,779.659973,779.659973,770.75,775.080017
```

   Step 2: Read the dataset in the form of the data frame
   Step 3: Explore the data with appropriate plots.

3. Draw an appropriate plot to explore the case study: Gold medal achievements of five most successful countries in Olympics. (15)
   Step 1: Store the following data in medal.csv

```
country,gold_medal
India,46
Japan,27
US,26
China,19
Germany,17
```

   Step 2: Read the dataset in the form of the data frame
   Step 3: Explore the data with appropriate plots.

4. Use the following data to create the Data Frame and store it in a variable named "Data" (20)

Year	Unemployment_Rate
1920	9.8
1930	12
1940	8
1950	7.2
1960	6.9
1970	7
1980	
1990	6.2
2000	5.5
2010	6.3

   Step 1: Check the presence of missing values
   Step 2: Use the imputation approach to handle missing values
   Step 3: Explore the above data using the following plots
      a) Line Plot
      b) Scatter Plot
      c) Bar Plot

5. Explore the given dataset using different plots/charts (50)
   1. Download the dataset from the following link https://archive.ics.uci.edu/ml/datasets/iris
   2. Read the dataset in the form of the data frame
   3. Check the shape of the data frame
   4. Check the presence of missing values
   5. Perform exploratory data analysis on this dataset with the help of following data visualization plots like
      a) Line Plot
      b) Scatter Plot
      c) Bar Plot (vertical and Horizontal)
      d) Pie plot
      e) Count Plot
      f) Box Plot

Platform Required: Anaconda, Editor: Jupyter/Spyder/Pycharm/Google Colab

Submission Instructions:

- Submission required 3 things 1) Python file (roll_no.ipynb/.py) 2) Dataset 3) PDF of .ipynb file. All these files are in a single zip folder (Use the naming convention: RollNo_LabNo.docx (Example: 123_Lab1))
- Submission is through LMS only.