

## Department of Computer Science & Engineering

Final Year B. Tech. (CSE) – I : 2022-23

### 5CS462 : PE5 - Data Mining Lab

#### Assignment No. 1

Design the data analysis tool (GUI) to perform the following task :

1. Data upload ( std. format like .csv, excel etc. ) and view
2. Calculate and show the measures of central tendency for uploaded data : mean , median , mode , midrange , variance and standard deviation
3. Calculate and show the dispersion of data : range , quartiles , interquartile range , five-number summary
4. Graphical display of above calculated statistical description of data (provide the facility - UI form to choose different attributes from uploaded data set) :
  - a. Quantile plot
  - b. Quantile-quantile (q-q) plot
  - c. Histogram
  - d. Scatter plot
  - e. Boxplot

#### Note :

- **Don't use direct / readymade functions from library. Do basic python coding.** You may compare / check the results with std. libraries.
- Use IRIS, Breast Cancer Wisconsin data sets from UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/>).

**Deadline : 01/08/2022**

**Dr. B. F. Momin**  
**Course Teacher**