**Title:** Design Data analysis tool (GUI).

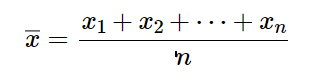
**Objective/Aim:** Performing following tasks

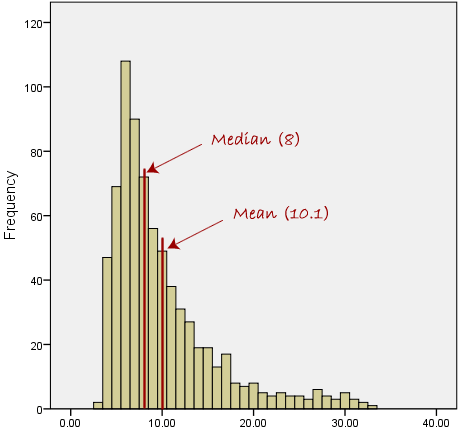
1. Data upload and view.
2. Calculate and show the measures of central tendency for uploaded data: mean, median, mode, midrange, variance and standard deviation.
3. Calculate and show the dispersion of data: range, quartiles, interquartile range, five-number summary.
4. Graphical Display of statistical description of data.

**Introduction:**

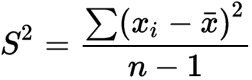
In this assignment our team has explored ways to extract data from conventional.csv/ excel files. Understood and implemented various measures of central tendency calculated from the given data. Also observed the dispersion of data. To enhance comprehensibility we have implement GUI interface to interact with the program application.

**Theory/Diagrams:**

1. Measures of central tendency: A measure of central tendency is a value that attempts to describe a set of data by identifying the central position of the set. The most commonly used measures of central tendency are the mean, median and mode.
   1. Mean: The mean (or average) is the most popular and well known measure of central tendency. It can be used with both discrete and continuous data, although its use is most often with continuous data (see our Types of Variable guide for data types). The mean is equal to the sum of all the values in the data set divided by the number of values in the data set. So, if we have values in a data set and they have values …, the sample mean, usually denoted by (pronounced "x bar"), is: 
   2. Median: The median is the middle score for a set of data that has been arranged in order of magnitude. The median is less affected by outliers and skewed data.
   3. Mode: The mode is the most frequent score in our data set. On a histogram it represents the highest bar in a bar chart or histogram. You can, therefore, sometimes consider the mode as being the most popular option.



* 1. Midrange: In statistics, the mid-range or mid-extreme is a measure of central tendency of a sample defined as the arithmetic mean of the maximum and minimum values of the data set.
  2. Standard Deviation:
  3. Variance: variance is the expectation of the squared deviation of a random variable from its mean.



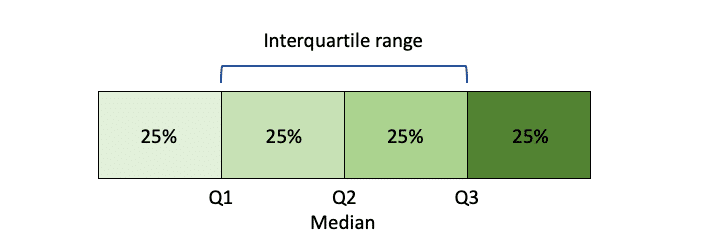
S^2 = sample variance

xi = the value of the one observation

\bar{x} = the mean value of all observations

n = the number of observations

1. Dispersion of data: Statistical dispersion means the extent to which a numerical data is likely to vary about an average value. In other words, dispersion helps to understand the distribution of the data.
   1. Five Number summery: The five-number summary is a set of descriptive statistics that provides information about a dataset. It consists of the five most important sample percentiles:
      1. the sample minimum (smallest observation)
      2. the lower quartile or first quartile
      3. the median (the middle value)
      4. the upper quartile or third quartile
      5. the sample maximum (largest observation)
   2. Interquartile range: the interquartile range, also called the midspread, middle 50%, or H‑spread, is a measure of statistical dispersion, being equal to the difference between 75th and 25th percentiles



**Procedure/ Algorithm:**

Python Formulae.

**Actual Experimentation/ simulation/ result/ Observation:**

Output Screenshots.

**Conclusion:** By performing the given tasks we strengthened out fundamentals of python and data science.

1. Fetch and view data from .csv / excel files.
2. Visual representation and measures of central tendency calculated from uploaded data.

**References:**

1. [https://pythonguides.com/upload-a-file-in-python-tkinter/#:~:text=The%20interface%20has%203%20buttons,button%20to%20upload%20the%20documents.&text=Once%20the%20user%20can%20selected,the%20file%20on%20the%20server.](https://pythonguides.com/upload-a-file-in-python-tkinter/%23:~:text=The%20interface%20has%203%20buttons,button%20to%20upload%20the%20documents.&text=Once%20the%20user%20can%20selected,the%20file%20on%20the%20server.)
2. <https://realpython.com/python-csv/>
3. <https://www.tutorialspoint.com/python_data_science/python_measuring_central_tendency.htm>