DA- Assignment 1.

Date : 23/07/2021

Title :

Download the Iris flower dataset or any other dataset into a DataFrame.

(eg. https:// archieve. ics, uci. edu/ml/datasets/ Iris)

Use Python/R and Perform following:

1. How many features are there and what are their types (e.g., numeric, nominal)?
2. Compute and display summary statistics for each feature available in the dataset (eg. mininum value, maximum value, mean, range, standard deviation, variance, and percentiles).
3. Data Visualization - Create a histogram for each feature in the dataset to illustrate the feature distributions. Plot each histogram.
4. Create a boxplot for each feature in the dataset. All of the boxplots should be combined into a single plot. Compare distributions and identify outliers.

Objectives :

1. To identify and understand R/Python commands
2. To understand Data Visualization.

Outcomes:

1. Understand the data visualization and perform the operations from minimum, maximum, mean, range values.

Software Requirements:

Jupyter Notebook, Anaconda Navigator

Hardware Requirements:

8GB RAM, 1TB HDD

Theory :

1. Histogram:

a. A vertical bar chart is used to draw a histogram which represents the distribution of a set of data over a continuous interva of certain time period and relationships of a single variable over set of classes.

b. While representing the tabulated data into histogram, the tabulated frequency at every interval/bin/instance is represented by every bar in a histogram. And the total area of a histogram is equal to the number of data.

c. The one of the most commonly used graphical presentation of data is Histogran

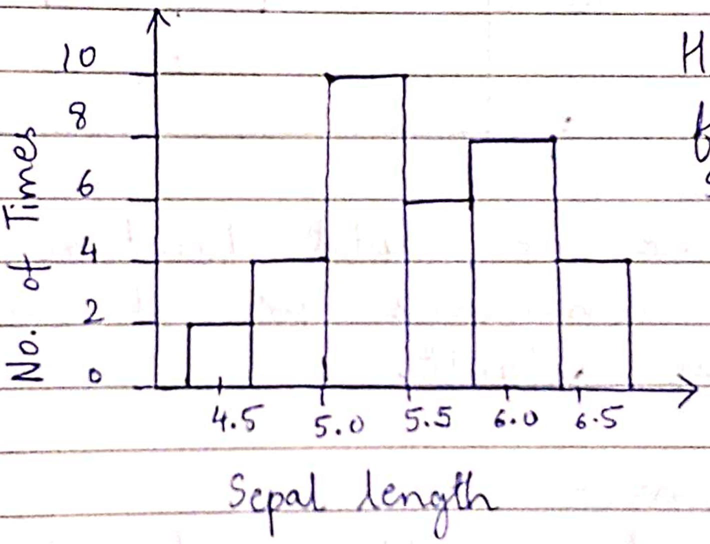
Date:

d. Histogram is used to graphically reprepent the huge amount of area / measurements/ dimensions contained by table.

e. A histogram origanizes and displays the table data in user-friendly format.

f. That means the histogram constructed to visualize the data will make that data easy to understand by representing the number of points.

Example:



Histogram of feature : Sepal length

Box- Plots :

1. A boxplot or box and whiskers plot is a graphically summary of a distributions.

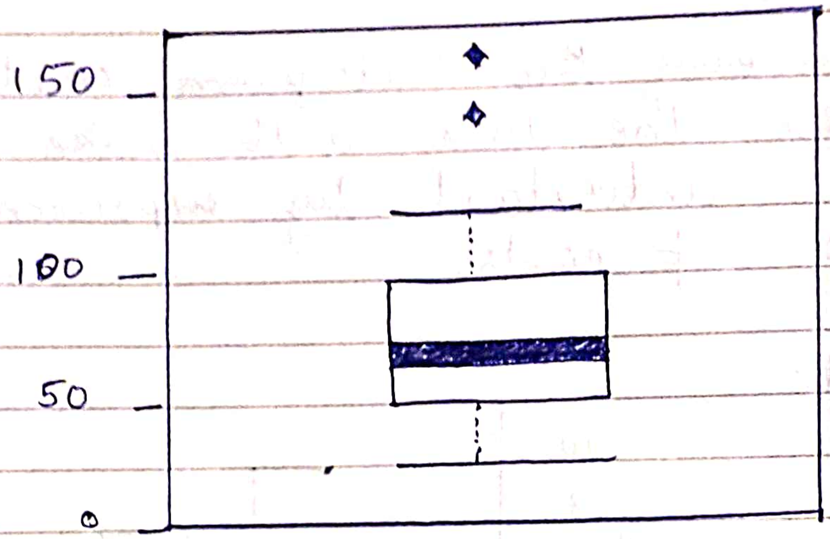
2. The box in the middle indicates hinges (close to the first and third quartiles) and median

3. The lines shows the largest and smallest observations that falls within the distance.

4. A boxplot can often given a good idea

of the data distribution and is often more useful to compare distributione side-by-side as it is more compact than a histogram.

Example of a boxplot:



5. Thus use of boxplot function to calculate quick summaries for all variables in our set by default.

Conclusion:-
I studied installing and setup of Anaconda Navigator and Jupyter Notebook and how to perform basic data analysis on the Iris dataset.