Rohit Kulkarni 41346

M T W T F S S
Page No.:
Date:
YOUVA

DA ASSIGNMENT 1

TITLE:
Download the Iris dataset or any other dataset into a dataframe. Use Python/R & perform the following:

1] How many features are there & their types.
2] Compute & display summary statistics for each feature available in the dataset (min value, max mean, standard deviation, variance, percentiles)
3] Data visualization. Create a histogram for each feature in the dataset to illustrate the feature distribution. Plot each histogram.
4] Create a boxplot for each feature in the dataset. All of the boxplots should be combined into a single box plot. Compare distributions & identify outliers.

OBJECTIVES:
1] To identify & understand python commands.
2] To understand data visualization.

OUTCOMES:
1] Understand the data visualization & perform the operations from minimum, maximum, mean, range values.

S/W & H/W REQUIREMENTS:
Jupiter Notebook, Anaconda Environment, & GB RAM PC, Monitor, Keyboard, mouse.

Page No.:
Date:
YOUVA

THEORY:

① Histogram:

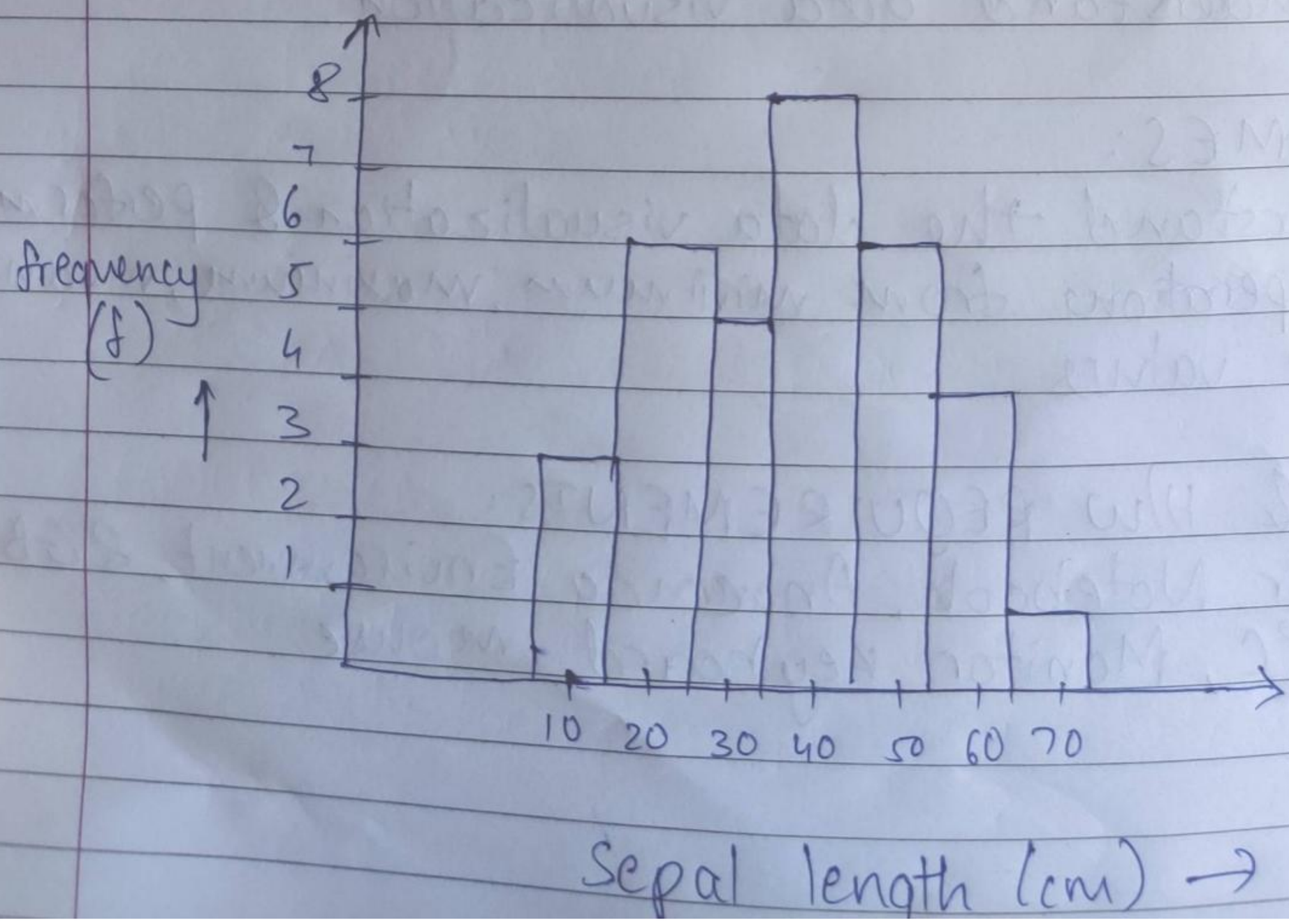
i) A vertical bar chart is used to draw a histogram which represents the distribution of a set of data over a continuous intervals or contain time period.

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

iii) The one of the most commonly used graphical presentation of data is histogram.

iv) Histogram is used to graphically represent the huge amount of area / measurements / dimensions contained by table.

v) A histogram organizes & displays the table data in user friendly format.

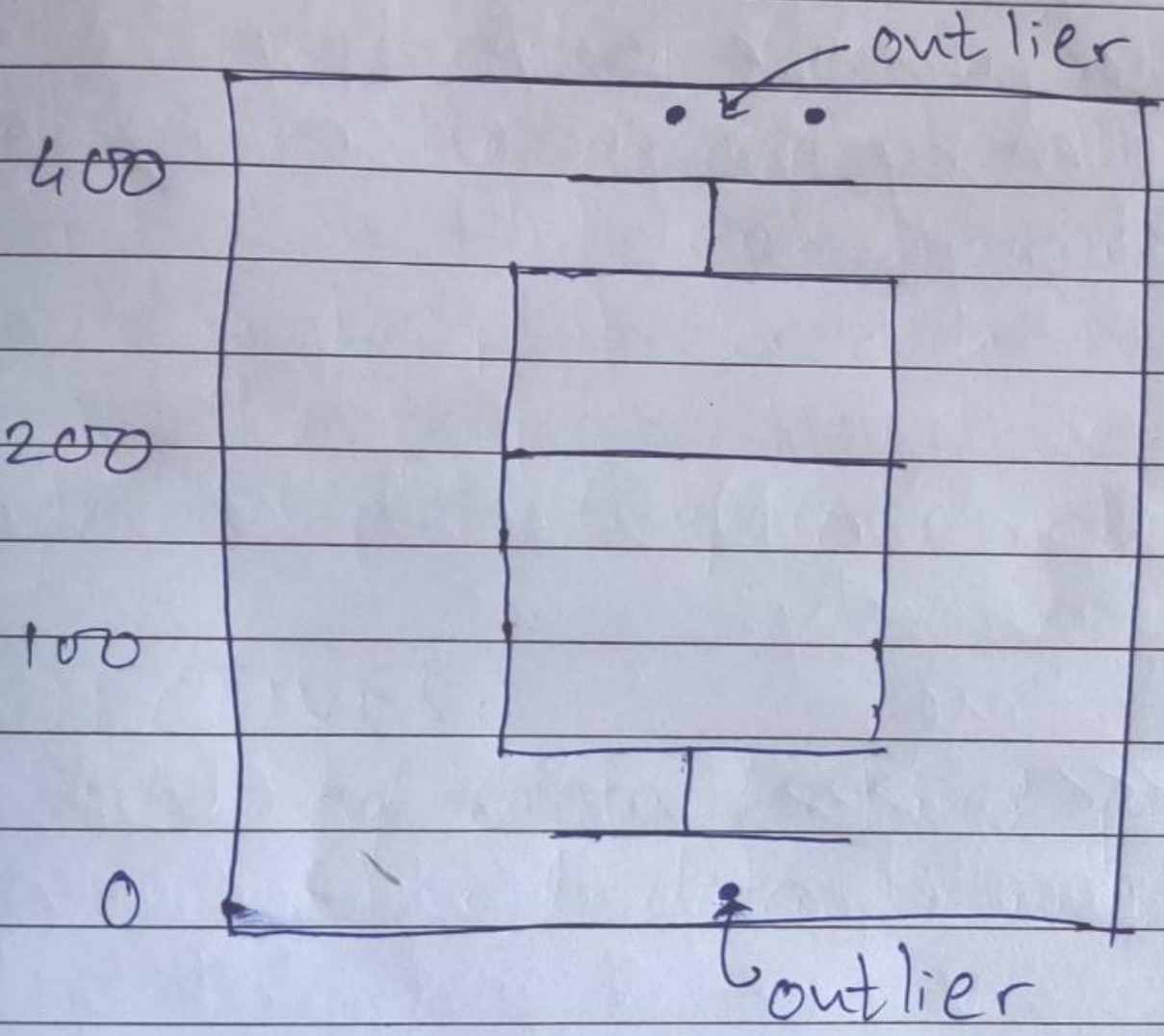


(2) Boxplots:

i] A boxplot or box & whiskers plot is a graphically summary of a distribution.

ii] The box in middle indicates hinges (close to 1st & 3rd quartiles) & median.

iii] A boxplot can often give a good idea of the distribution & is often more useful to compare distributions side-by-side as it is more compact than a histogram.



CONCLUSION:

The Iris dataset was loaded into a dataframe & analysed. Also, required plots were drawn successfully.