Name:- Mangesh T. Gite
Page No. 1 | BE(A) Roll no.:- 46 | Date: / /20

AIR Assignment - 05

Title :- Download the iris flower dataset into a Dataframe using Python & perform following-
(i) How many features are there & what are their types (e.g. numeric, nomical)
(ii) Compute & display summary statistics for each feature available in the dataset (e.g. minimum values, maximum values, mean, range, standard deviation, variance, percentiles)
(iii) Data visualization - create a histogram for each feature distribution plot the histogram.
(iv) Create a boxplot for each features in the dataset. All of the boxplots should be combined into a single plot compare distributions & identify outliers.

Objectives :- (i) To understand python commands
(ii) To understand Data visualizations

Outcomes :- (i) Understand the data visualization & perform the operations for minimum, maximum, mean, Range values.

Software :- Jupiter Notebook [Web application]

Hardware :- 512 MB RAM, 500 GB HDD

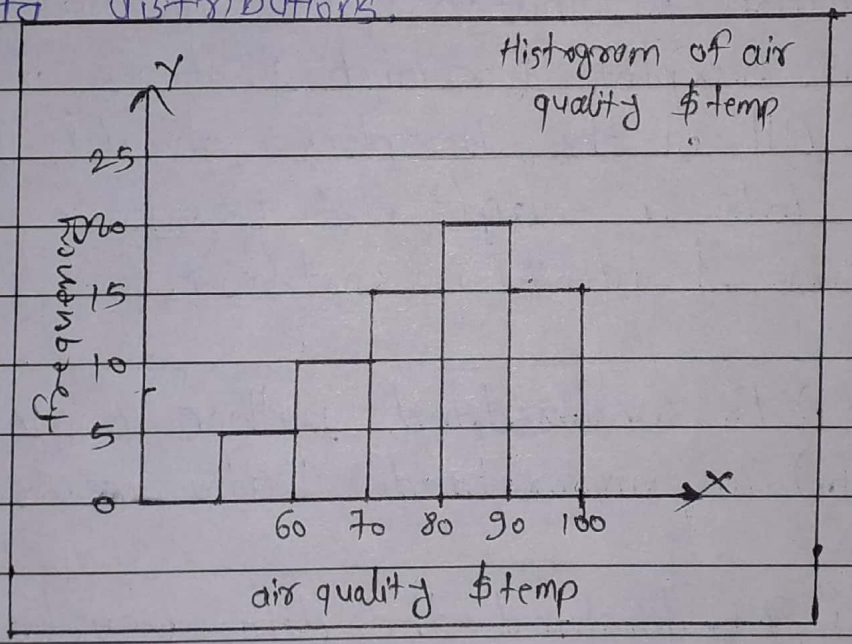
OPTIMA-58

* Theory Concepts :-

Data Visualization :-

(i) **Histogram** :- A histogram is a graphical representation that organizes a group of data points into user-specified range.

- A histogram is a bar graph-like representation of data that buckets a range of outcomes into columns along the x-axis.
- The y-axis represents the no. count or percentage of occurrences in the data for each column & can be used to visualize data distributions.

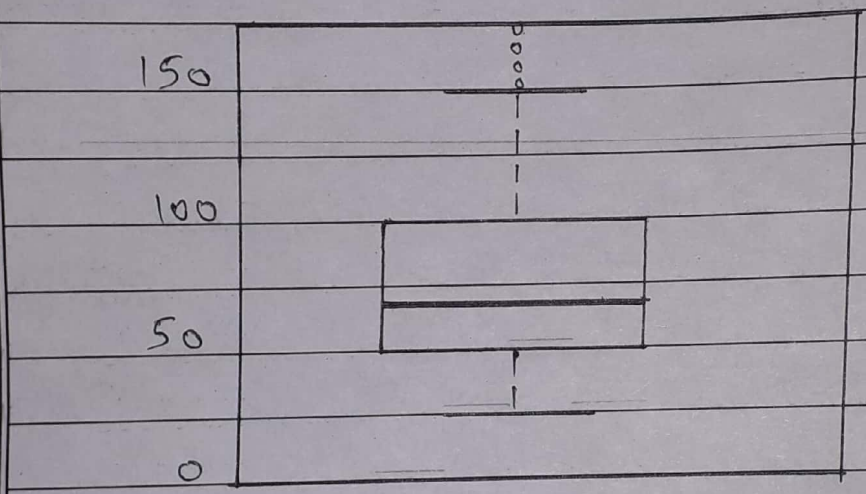


(ii) **Box Plots** :-

A boxplot or box & whiskers plot is a graphical summary of a distributions.

- The box in the middle indicates hinges (close to the first & third qualities) & median.

- The lines shows largest & smallest observation that falls within the distance.
- A boxplot can often give a good ided of the data distribution & is often more useful to compare distributions side by side as it is more compact than a histogram.
- Thus use of boxplot function to calculate quick summary for all the variables in our set by default.



* Commands :-

(i) Dataframe - head () function → used to get first n rows.

(ii) Dataframe - info () function → Used to print concise summary of a dataframe.

(iii) Data frame - shape → stores no. of rows & columns

(iv) Data frame - describe → Used for calculating some statistical data like percentile, mean & std for numerical values of the series.

(v) Dataframe – dtype → describe how the bytes in the fixed size block of memory corresponding to an array item should be interpreted.

Conclusion :-

From this assignment, we understand python commands & data visualization.