LP1 + DA 1

* Title: Iris Flower dataset

* Problem statement:

Download the Iris flower dataset or any other d.s. into a Dataframe. Use Python/R and perform -

→ How many features are there and what are their types?

→ Compute and display summary statistics for each feature available in the dataset.

→ Data Visualization - Create a histogram for each feature in the dataset to illustrate the feature distributions.

→ Create a boxplot for each feature in the dataset

→ Should be combined into a single plot.

* Software requirements :

Jupyter Notebook, Aca Anaconda navigator

* H/W requirements :

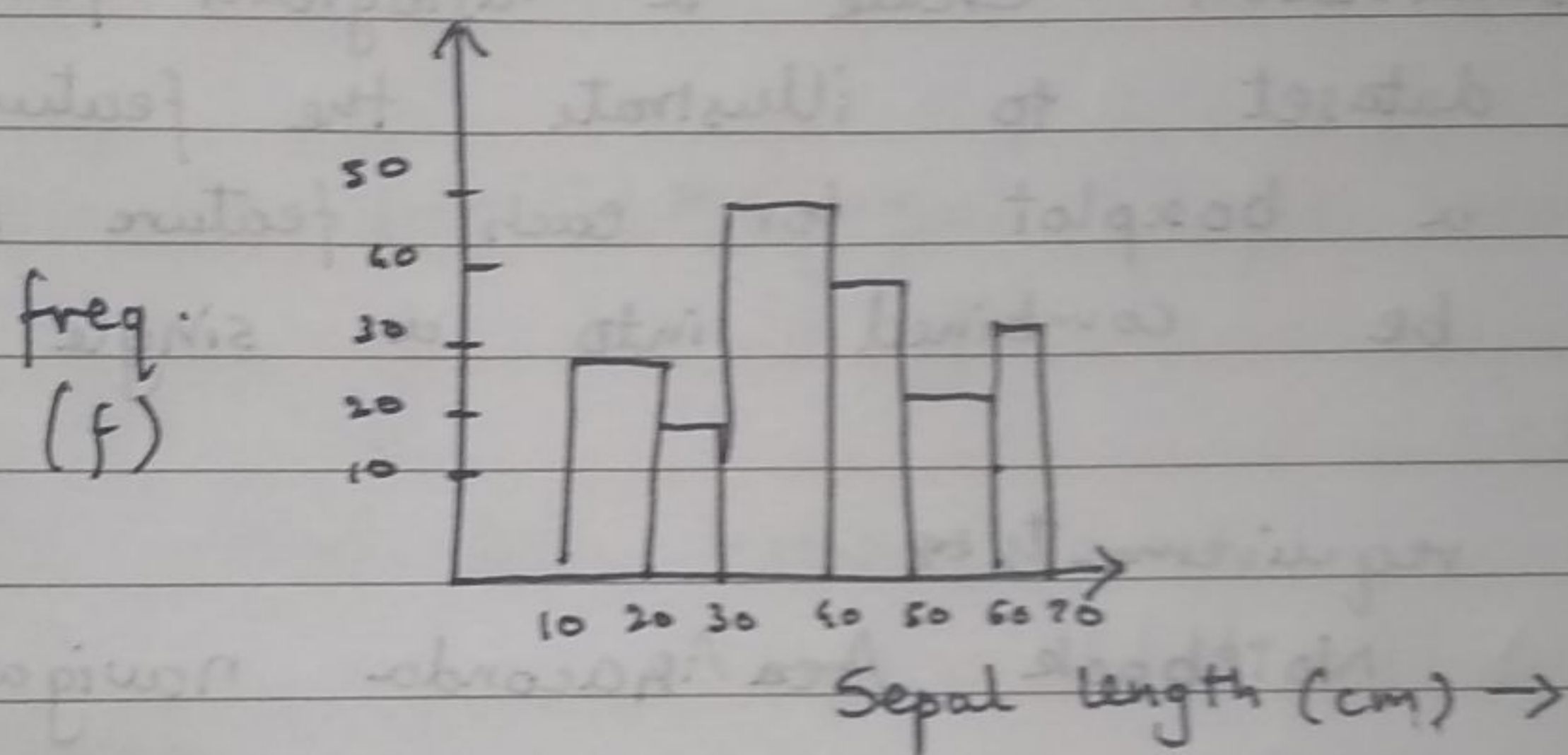8 GB RAM, 1 TB HDD

* Theory :

① Histogram :

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

b) while representing the tabulated data into the histogram, the tabulated frequency at every interval /instance is represented by every bar in a histogram and the total area of a histogram is equal to the no. of data.

c) It is one of the most commonly used graphical representation of data.

d) It displays and organizes the table data into user - friendly format.
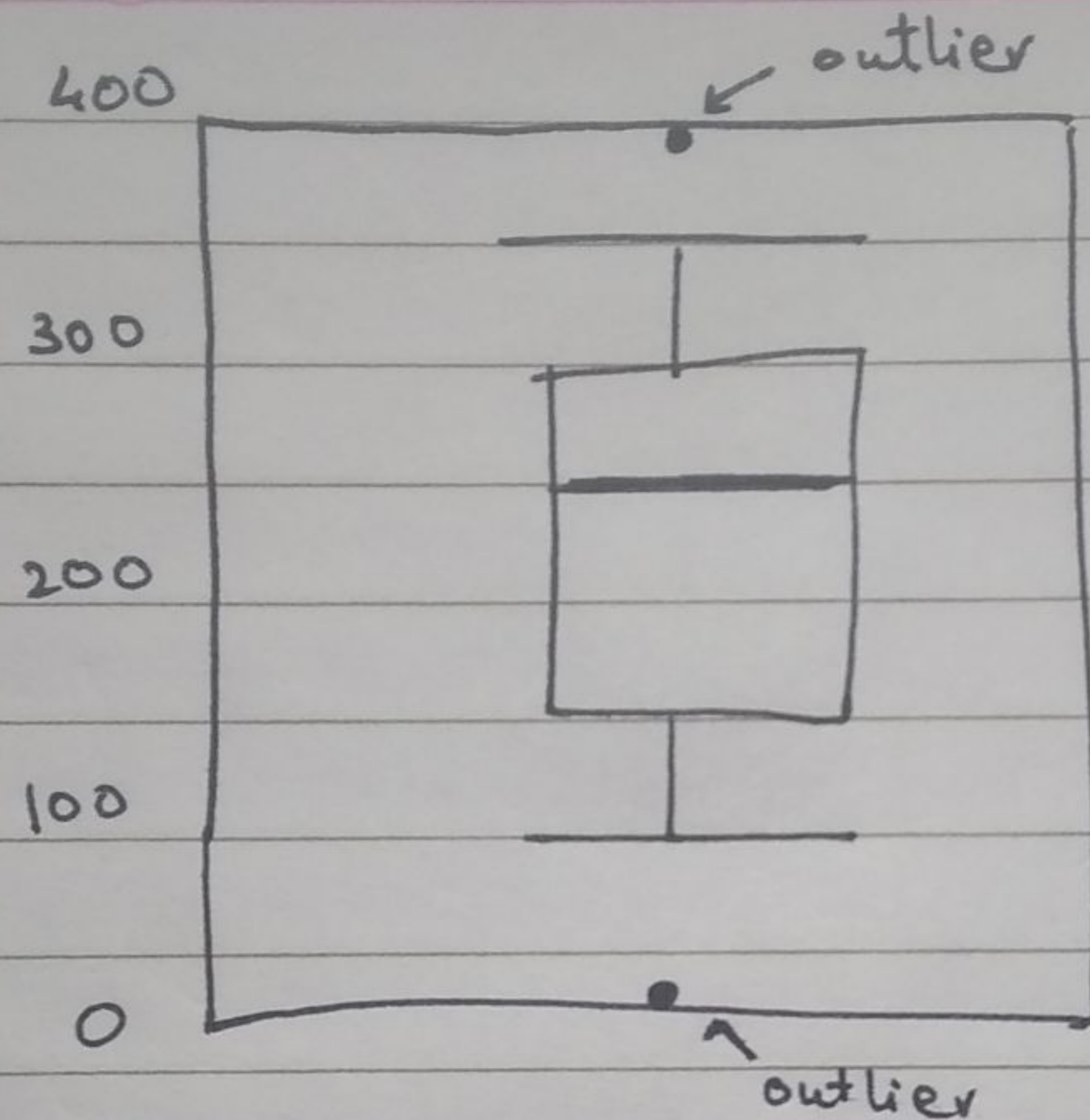
eg.

Sepal length (cm) →

② Box plots:

a) A box-plot or box and whiskers plot is a graphical summary of a distribution.

b) The box in the middle indicates hinges (close to 1st and 3rd quartiles) and median.

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

eg.

The use of boxplot function is to calculate quick summaries for all variables in our set.

* Conclusion:

From this assignment, I was able to understand the basics of data analytics and hence analyze the iris flower dataset.