Assignment - DA1

* TITLE AND PROBLEM STATEMENT: Download the Iris flower dataset or any other dataset into a DataFrame. Use Python / R & Perform following -
  - How many features are there & what are their types ?
  - Compute & 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. Plot each histogram.
  - create a boxplot for each feature in the dataset. All of the boxplots should be combined into single plot. Compare distributions & identify outliers.

* OBJECTIVES: i) To understand Python commands.
  ii) To understand data visualization.

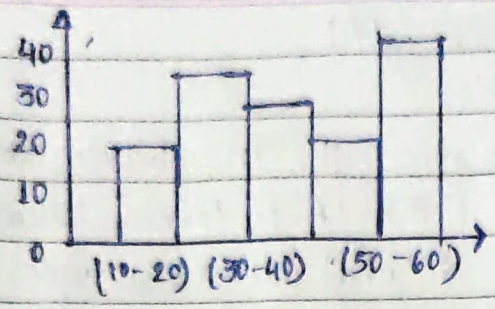
* OUTCOMES: Understood the data visualization and performed operation for summary statistics mentioned above.

* SOFTWARE AND HARDWARE REQUIREMENTS: Jupyter nootebooks, required libraries (pandas, seaborn, scipy, sklearn), Python 3., 8GB RAM, UNIX/LINUX OS.

* THEORY:

Data Visualization:

i) Histogram: vertical bar chart is used to draw a histogram; which represents the distributions of a set of data over a continuous interval or certain time period & relationships of a single variable over set of classes. while representing the tabulated data into an histogram, the tabulated frequency at every interval is represented by every bar in a histogram.



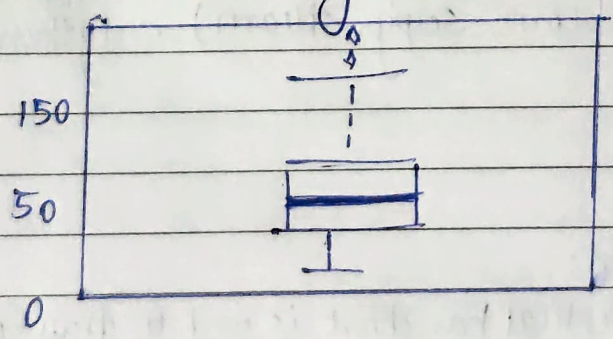
Histogram example

2) Boxplot: a graphical summary of distributions.
- The box in the middle indicates hinges (close to the first & third quantities) and median.
- The lines show largest & smallest observation that falls within dist.
- Boxplots are used to calculate quick summaries for all the variables in our set by default.

- About the Iris dataset:

The IRIS dataset is a multivariate dataset introduced by British statistician, eugenicist and biologist Ronald Fisher in a paper. The dataset contains 50 samples from each of the 3 species of Iris. Four features were measured for each sample. the length & width of the sepals & petals, in centimeters.

This dataset is thus very useful for statistical classification techniques in machine learning as well as a good starter data set.



Boxplot example

* CONCLUSION:

The python commands for basic statistical techniques were understood. and data visualization was performed on the results.