Assignment No.

Title:- Download the iris flower dataset compute & summarize statistics for the same.

Problem statement:-
Download the iris flower dataset or any other dataset into a dataframe.

1) List down features & their types available in the dataset.
2) Create a histogram for each feature in the dataset to illustrate the feature distributions.
3) Create a boxplot for each feature.

Objectives:- Learn classificatn techniques how to plot histogram & boxplot for given dataset.

Pre-requisites :- R-programming, Rstudio.

Theory :- Let's design a basic data analysis program in R using R studio by utilizing the features of R studio to create some visual representation of that data.

How to install Rstudio ?
We have to follow three basic steps in same order to run R & R studio. on your system.

- Install R
- Install Rstudio
- Install R packages.

Following steps perform to achieve goals.

- Downloading/importing data in R.
- Transforming data/Running queries on data.
- Basic data analysis using statistical averages.
- plotting data distribut^{n}.

— Typical data analysis process:-

Data analysis deals with collecting, inspecting, cleansing, transforming & modeling data to glean valuable insights.

1) Data exploration:-

Having identified the business problem, a data analyst has to go through the data provided by client to analyse root cause of the problem.

2) Data preparation:-

This is most crucial step of the data analysis process wherein any data anomalies with the data have to modelled in the right direct^{n}.

1) Data Modelling:-

The modeling step begins once the data has been prepared. Modeling is an iterative process wherein the model is run repeatedly for improvements.

2) Validation:-

In this step model provided by client & model developed by the analyst are validated against each other to find out if developed model will meet business requirements.

3) Implementation of model & tracking:-

This is the final step of data analysis process wherein the model is implemented in production &

is tested for accuracy & efficiency.

conclusion:-

Thus we have learned how to use features of R studio when learning data visualizatn. How to create histogram for each feature of dataset, boxplot etc.