**PRN – 2019BTECS00039**

**Name – Vaishnavi Santosh Bhajibhakare**

**Date – 04/02/2022**

**Software Engineering Tools Lab**

**Assignment 1**

**(Module 1- Introduction to OSS)**

1. Weka is a GUI workbench that empowers data wranglers to assemble machine learning pipelines, train models, and run predictions without having to write code. Using Weka tool perform below tasks such as data preprocessing, data classification (use any appropriate ML algorithm) and data visualization efficiently on given dataset. Use the Iris dataset given

<https://drive.google.com/file/d/1A3Fxsfzm6BSfhFZGDrjI47RTe45bSgYP/view> Note-provide screen shots for every task Create a report which will illustrate the details of tasks performed (for e.g to perform preprocessing of data provide details of navigation and selection of appropriate parameters)

STEPS –

* Installing weka from the link –

<https://sourceforge.net/projects/weka/>

* GUI Screen of wekasoftware :

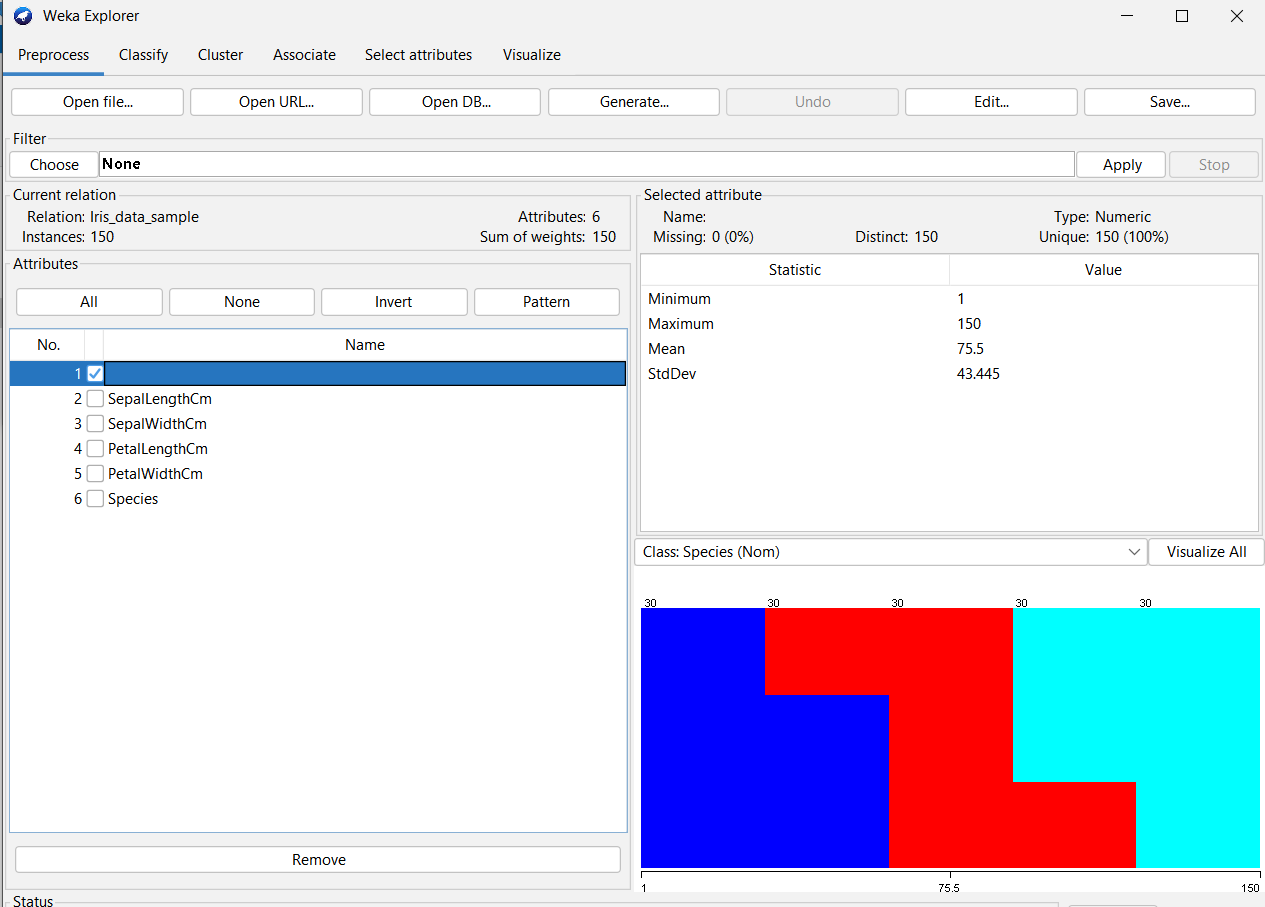
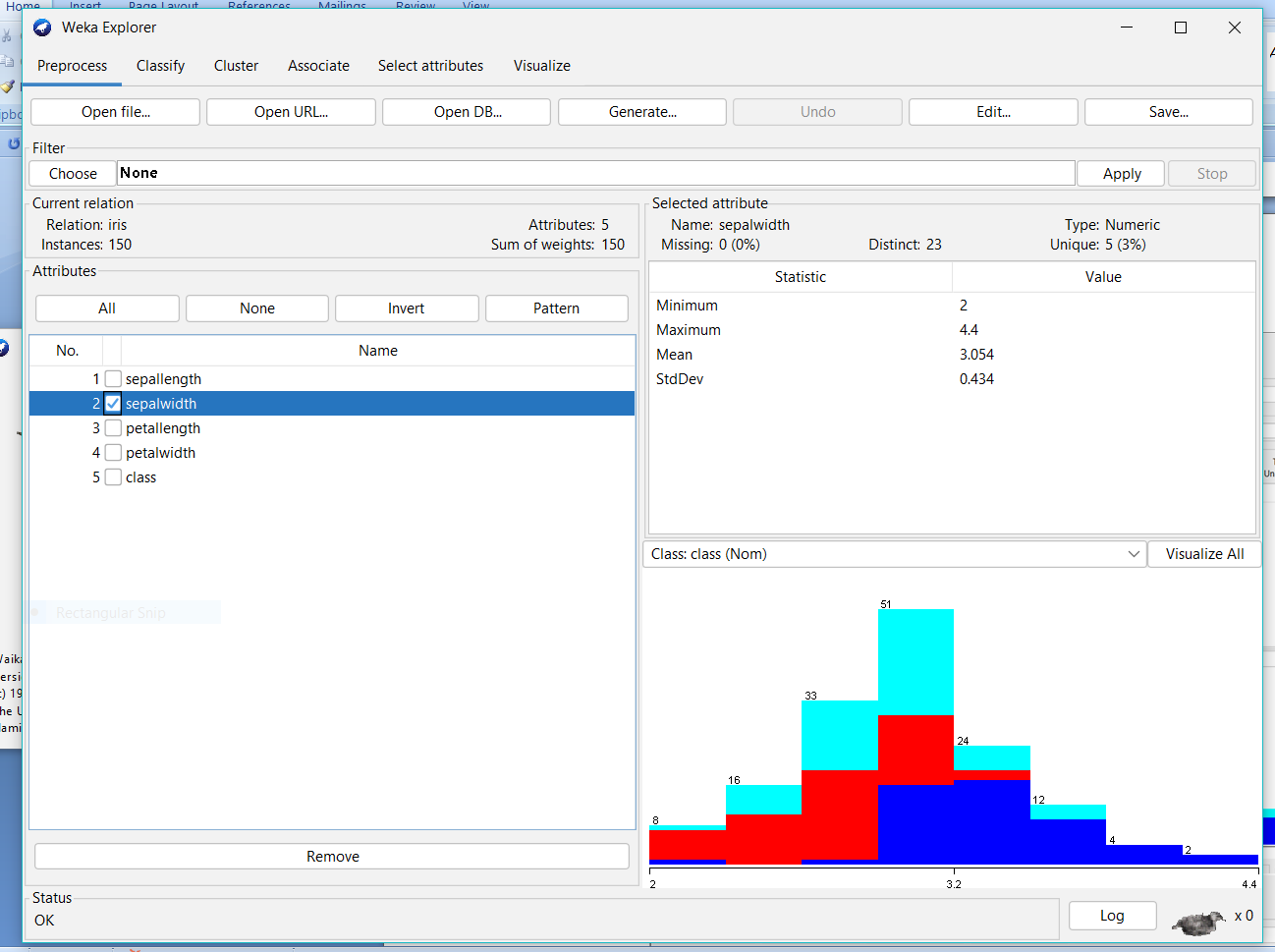


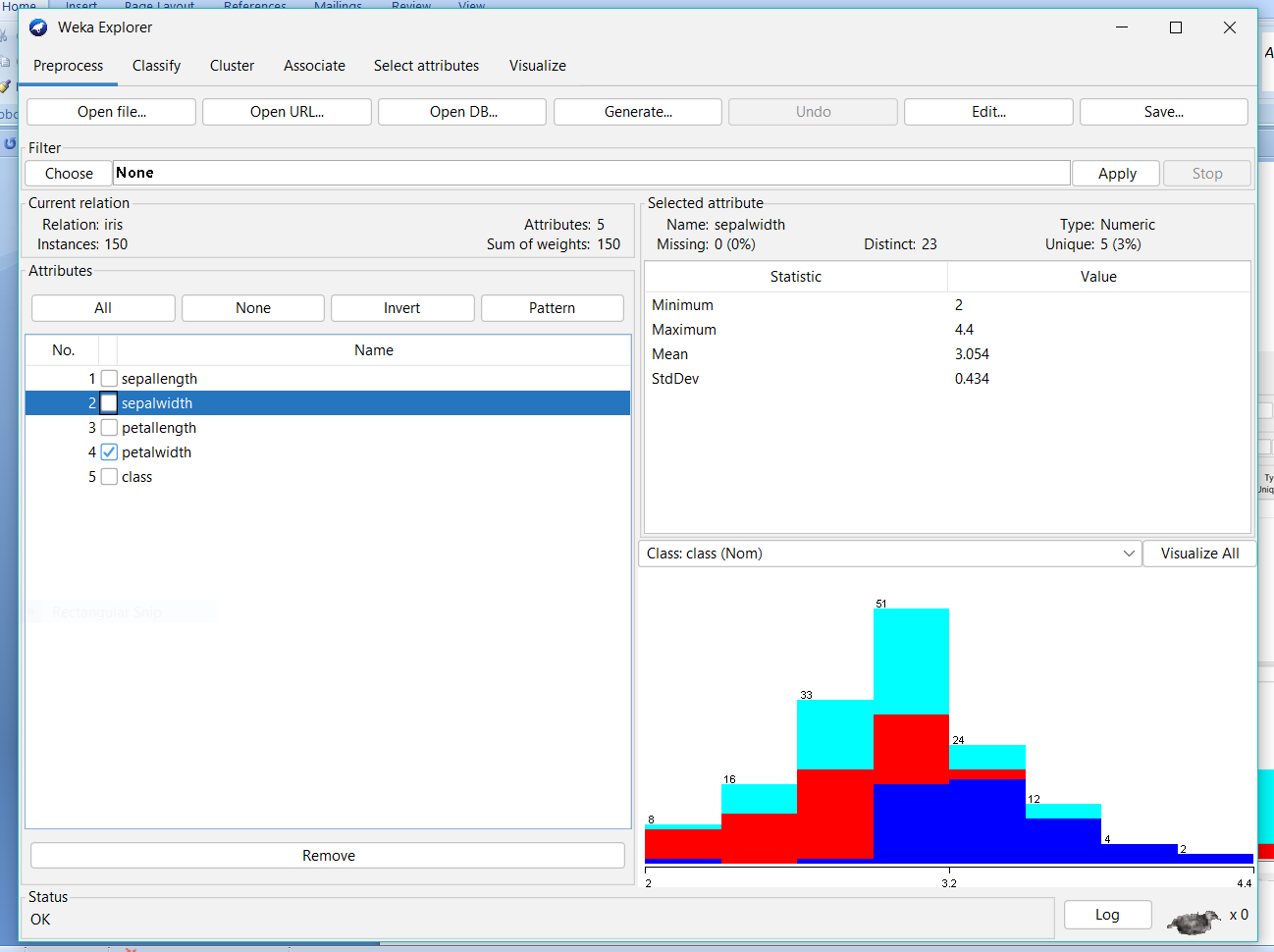
* Data Preprocessing –

At first we need to open the .csv file that is our dataset.

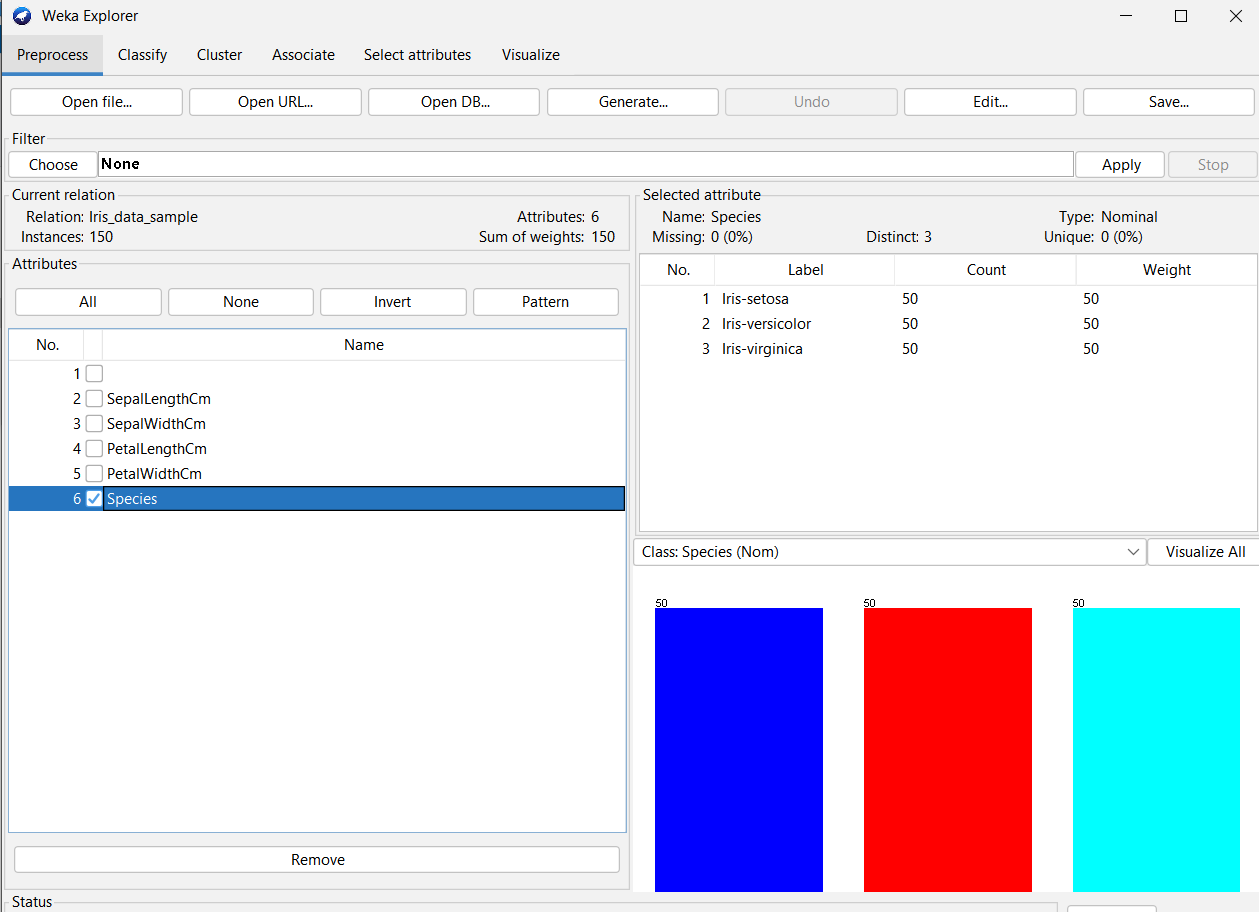
Click on Open file & import the Iris\_dataset file.

Goto the Preprocessing section of the Explorer part & there we can see attribute names on the left side.

1. With all attributes -
2. SepalWidth -
3. Petal width



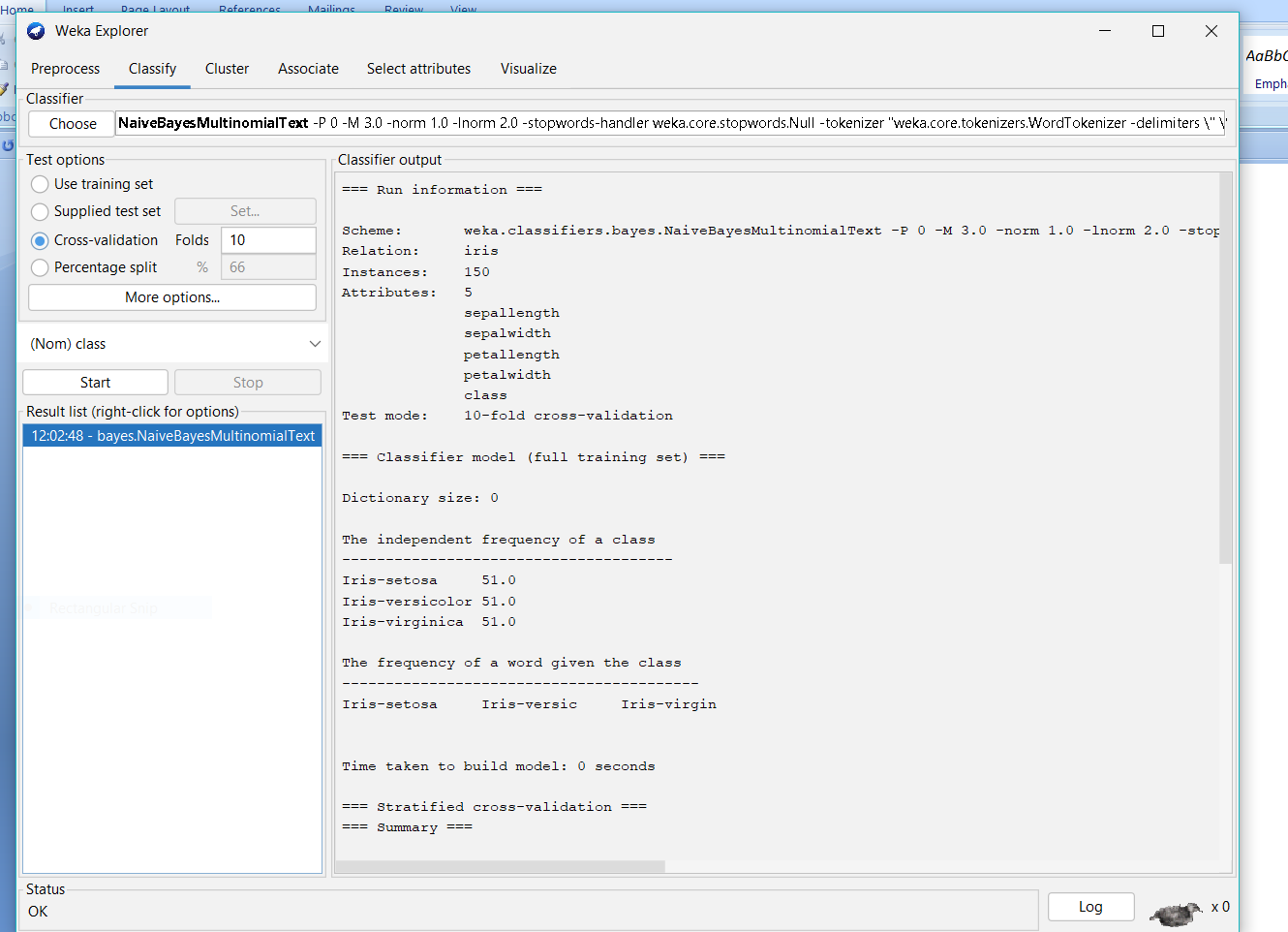
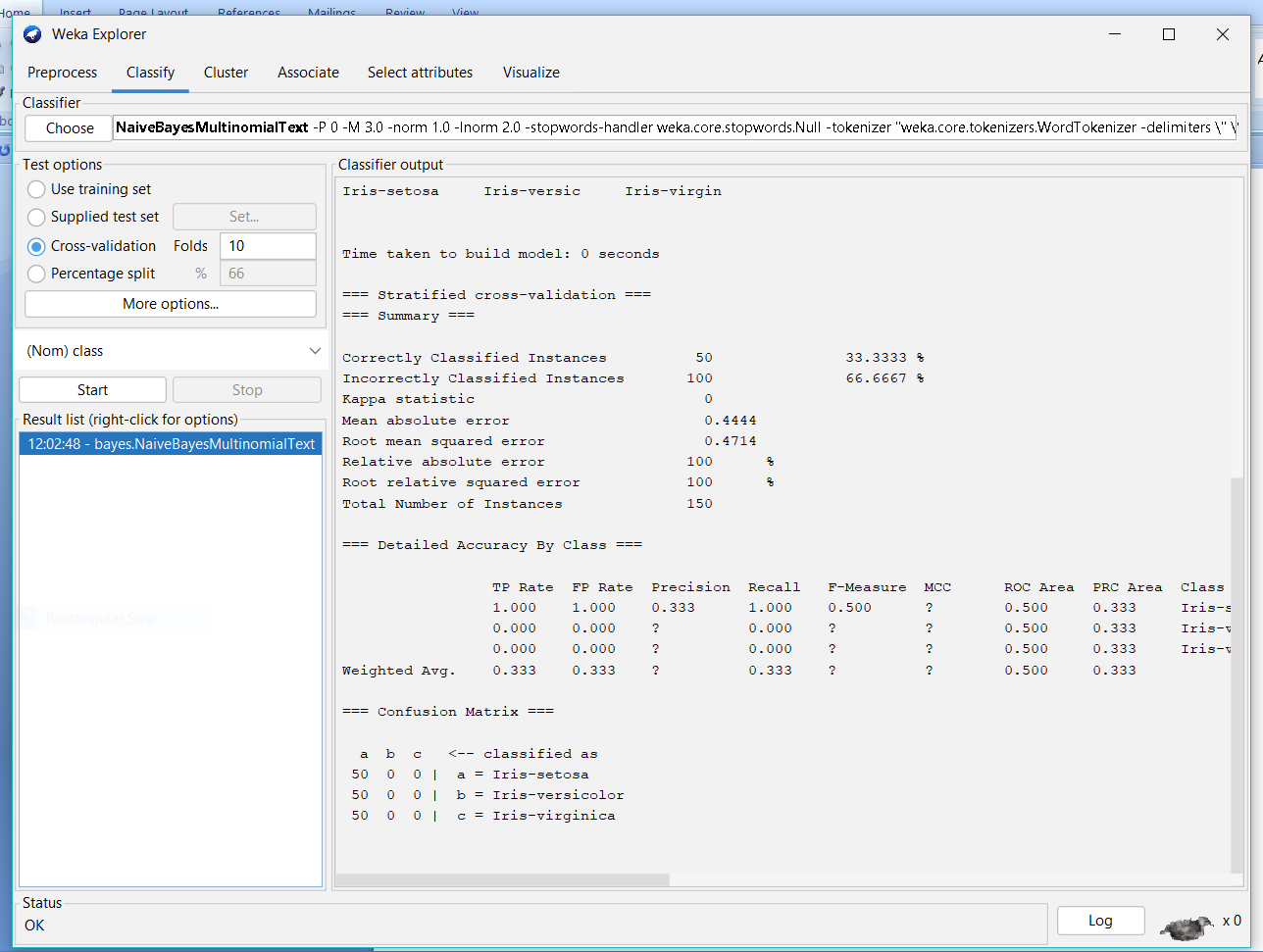
1. Species



* Data Classification –

Chose any ML algorithm from the options. Here I have chosen Naïve Bayes Multimonial Text under bays folder.

ML algorithm used : Naïve Bayes Algorithm

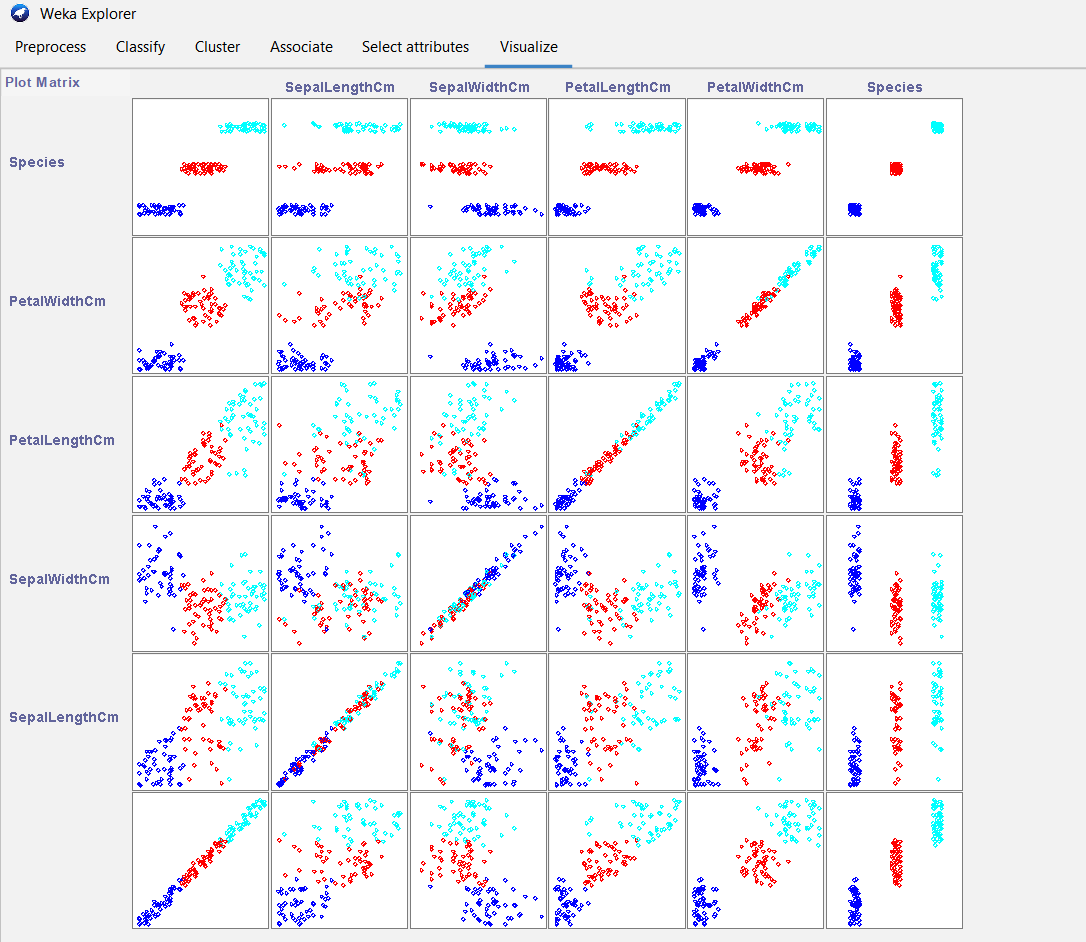
 

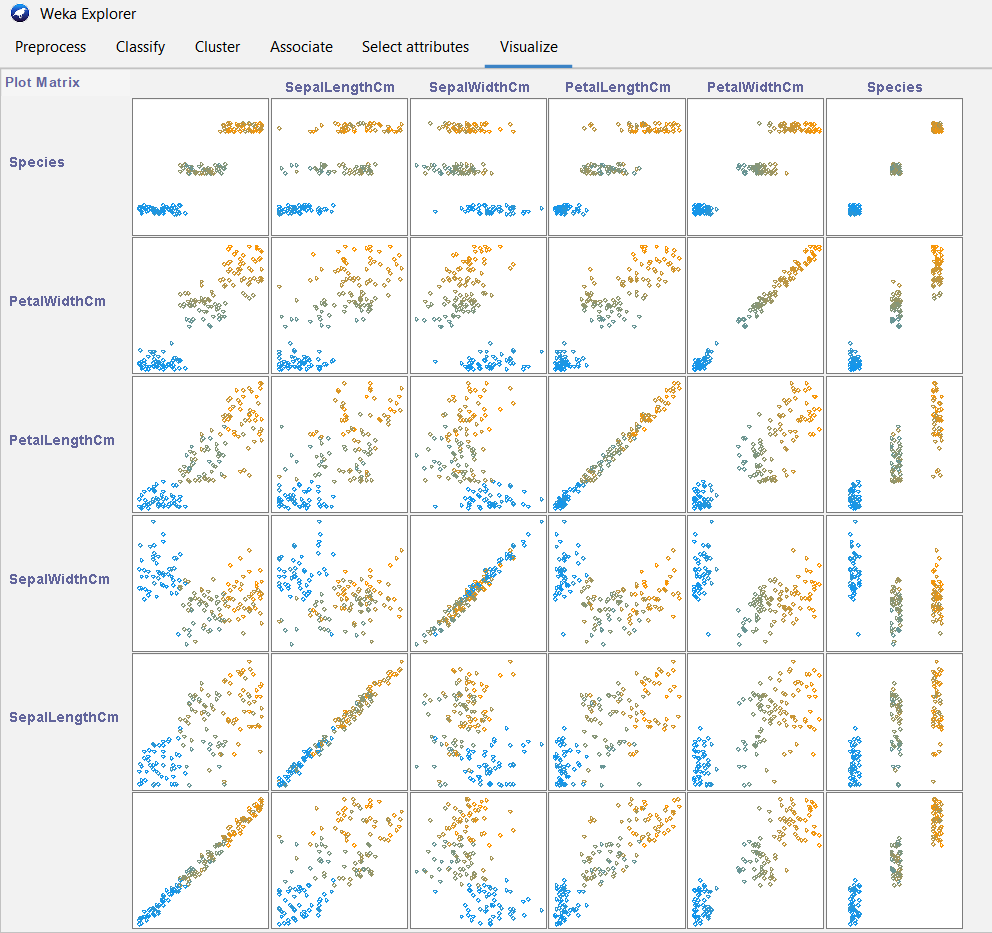
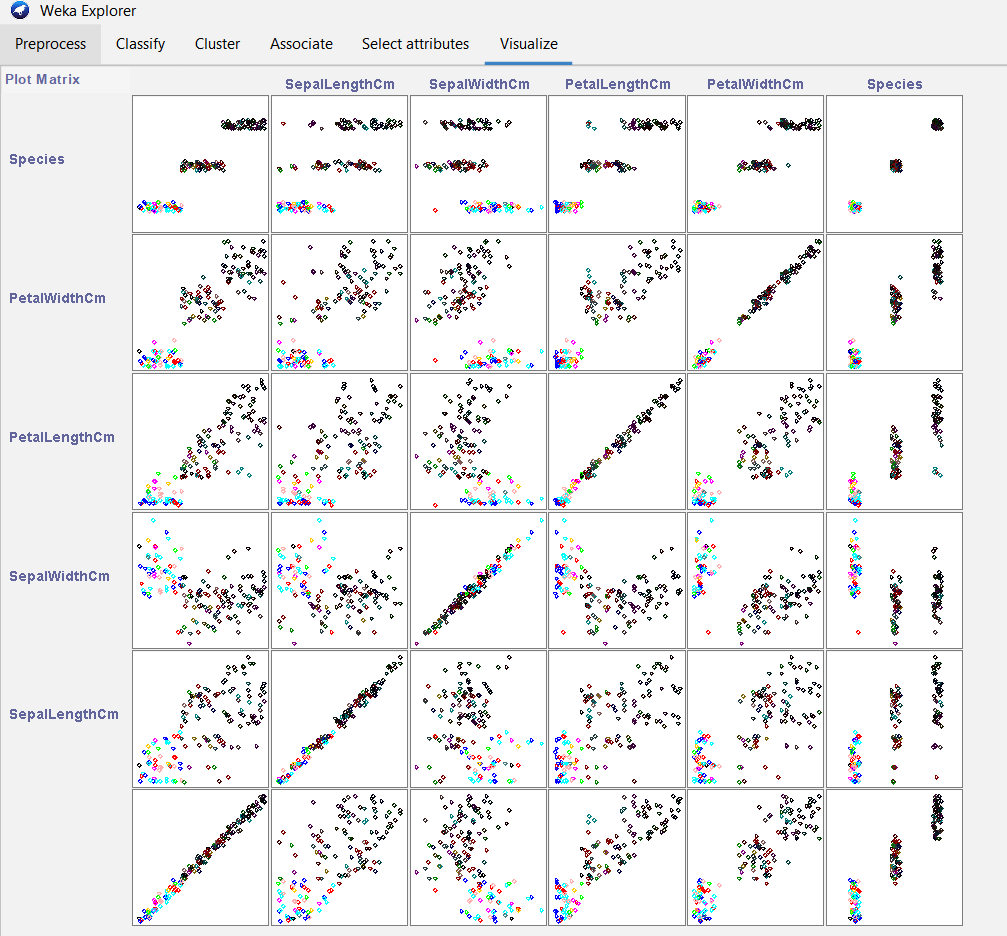
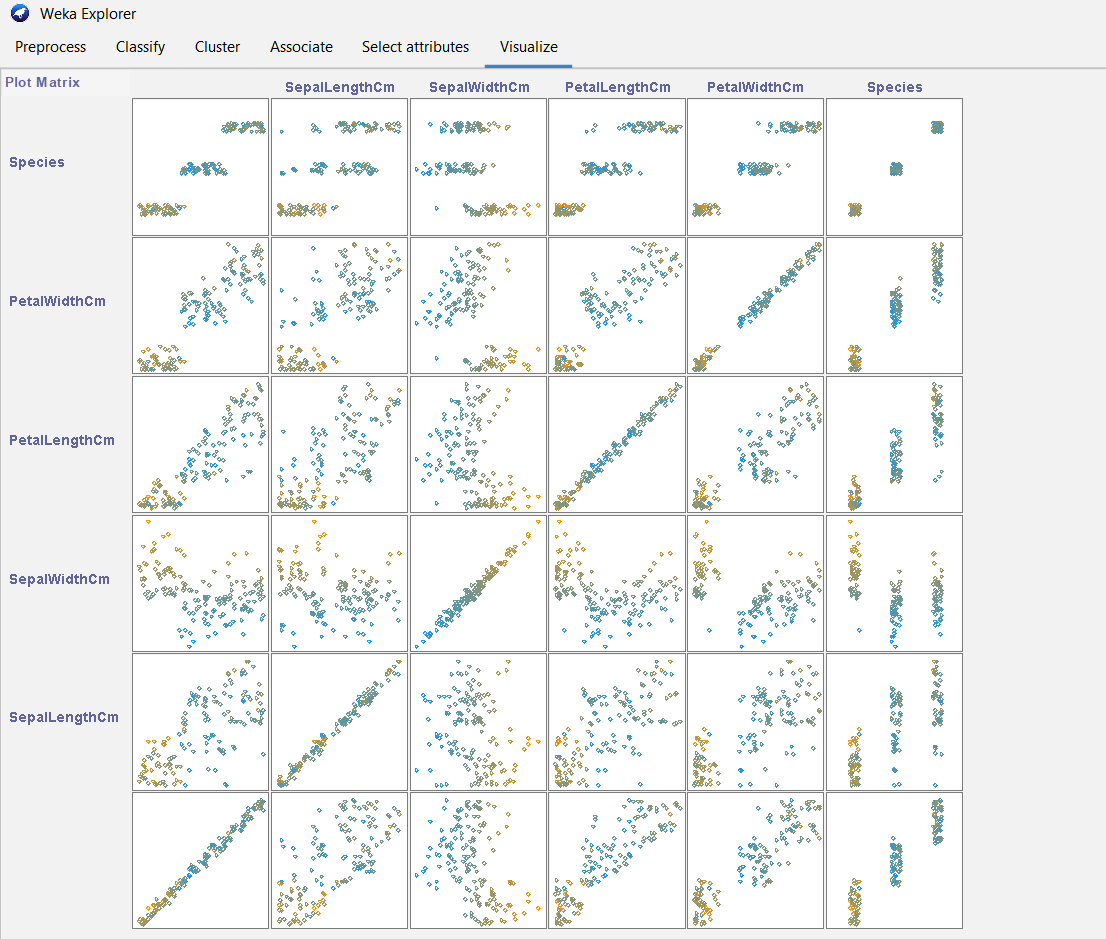
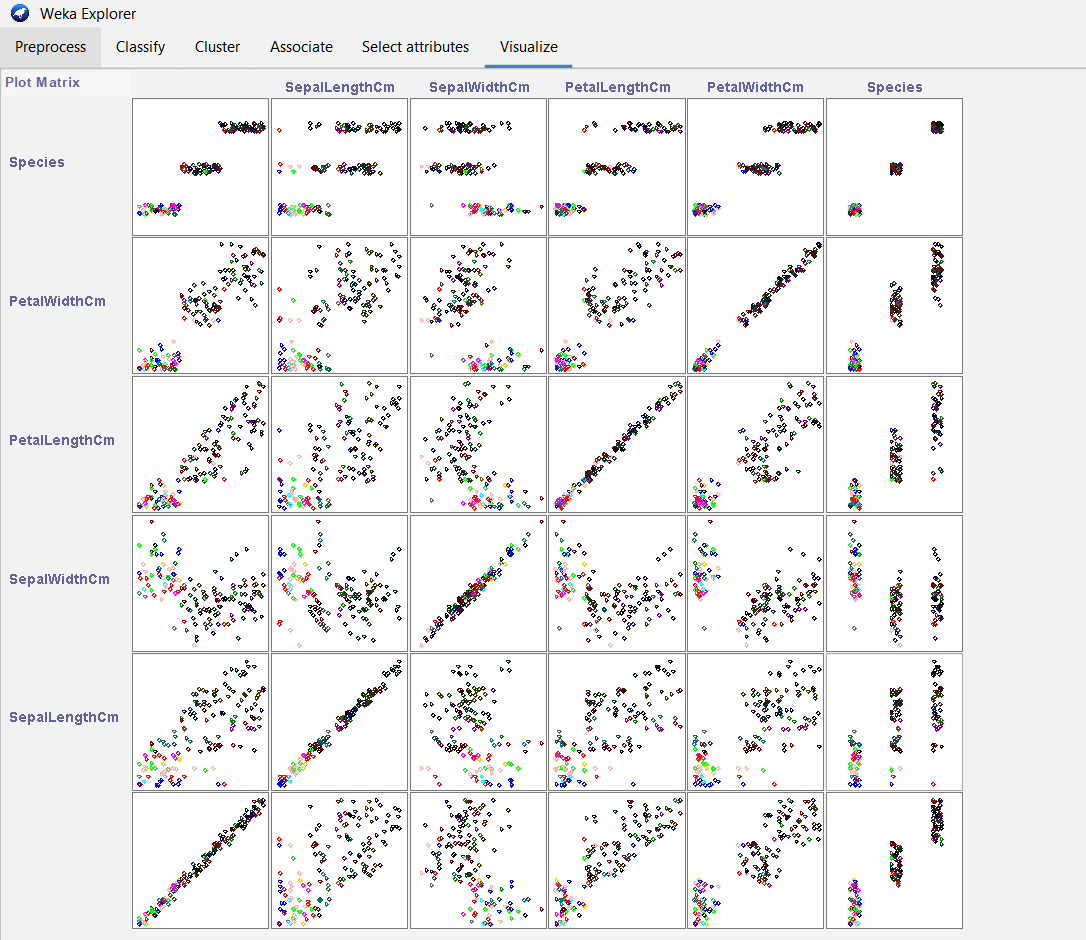
* Data Visualization –

In the Visulaize section , A plot matrix is prepared & below that there are foloowing options –

Change plot size , point size , jitter etc & the attributes can be selected according to our requirement.

Taking into account different features





1. Orange is an easy to use data visualization tool with a large toolkit. In spite of being a GUI-based beginner-friendly tool, you mustn’t mistake it for a light-weight one. It can do statistical distributions and box plots as well as decision trees, hierarchical clustering and linear projections. a. Install orange b. Show data distribution c. Show linear projection d. Show FreeViz Use dataset

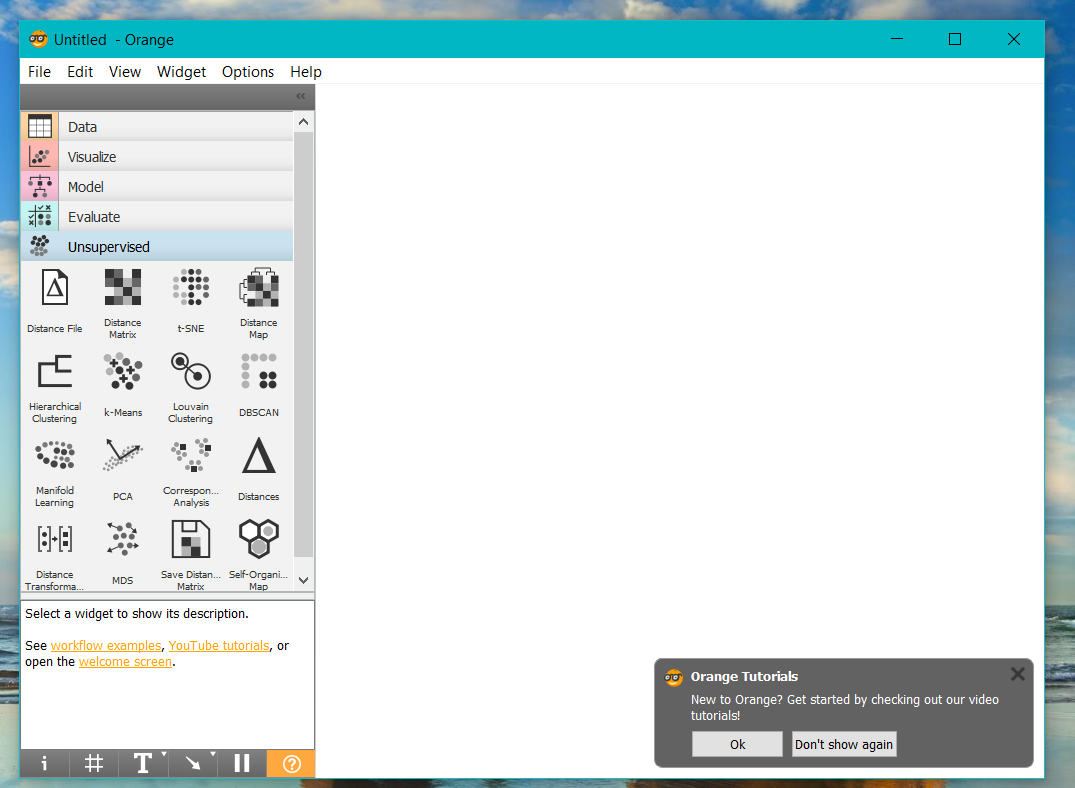
<https://drive.google.com/file/d/1m6sKI1Dap0XK6Bw1edUd5PohwpPwXnd9/view>

Create a report for this task and upload screenshots for the same

* Install Orange from the website –

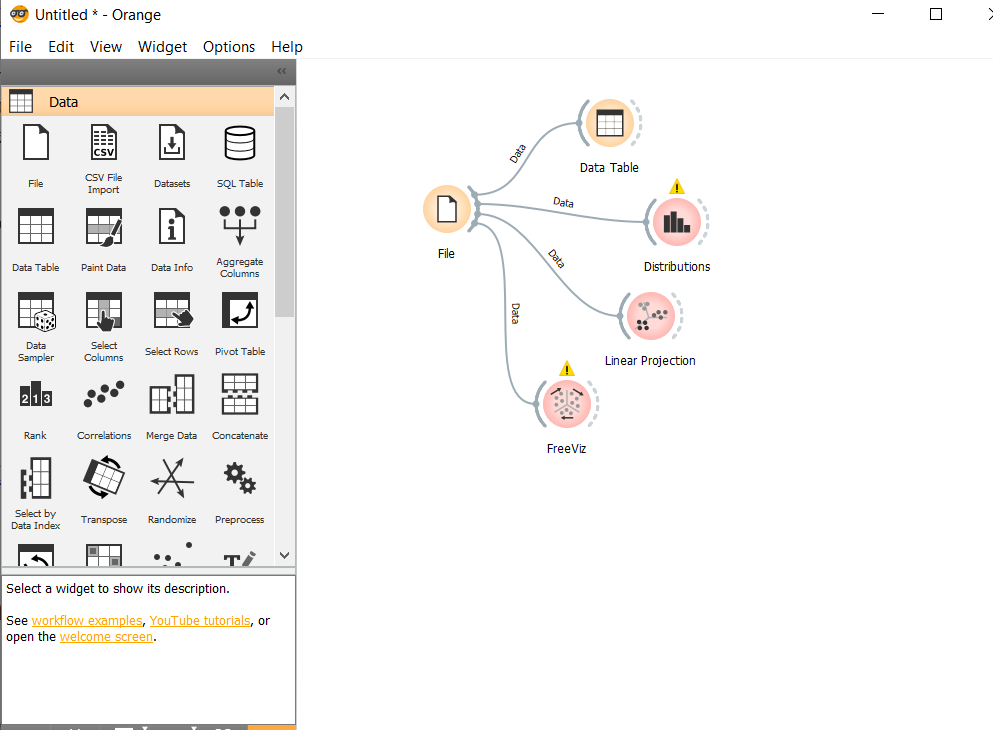
<https://orangedatamining.com/>

* GUI screen of orange –

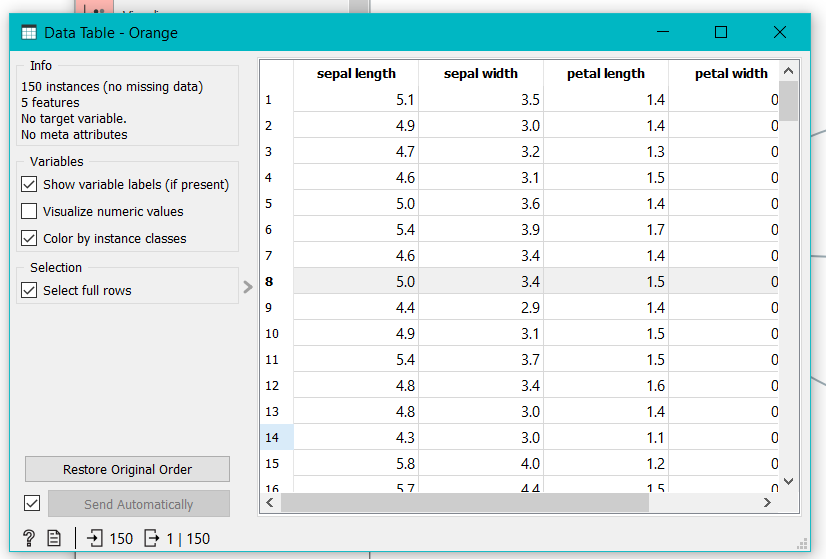


Import the dataset file i.e. , iris\_dataset.xlsx.

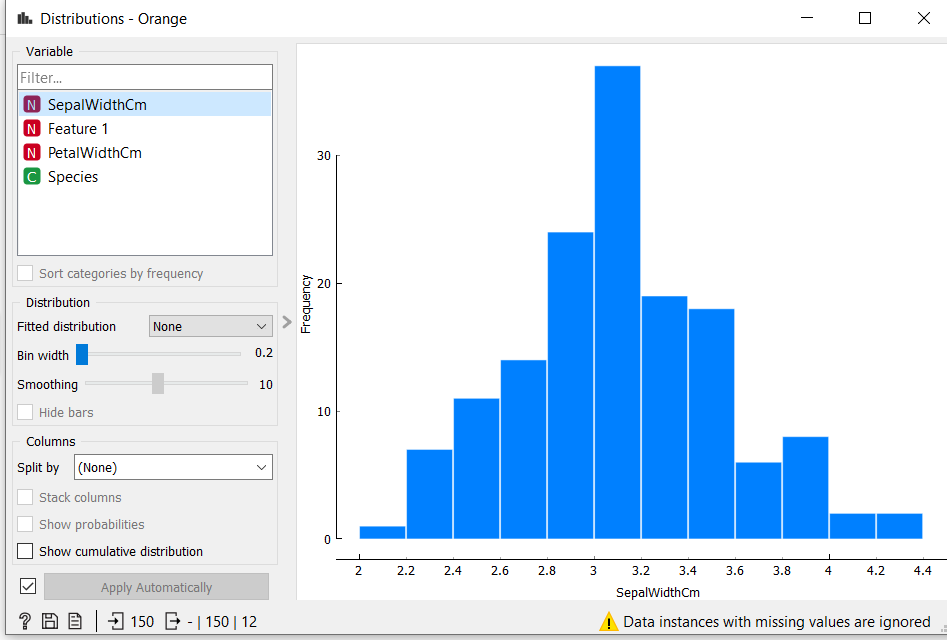
Choose the options according to requiremen . For example – Data distribution , Freewiz , linear regression etc

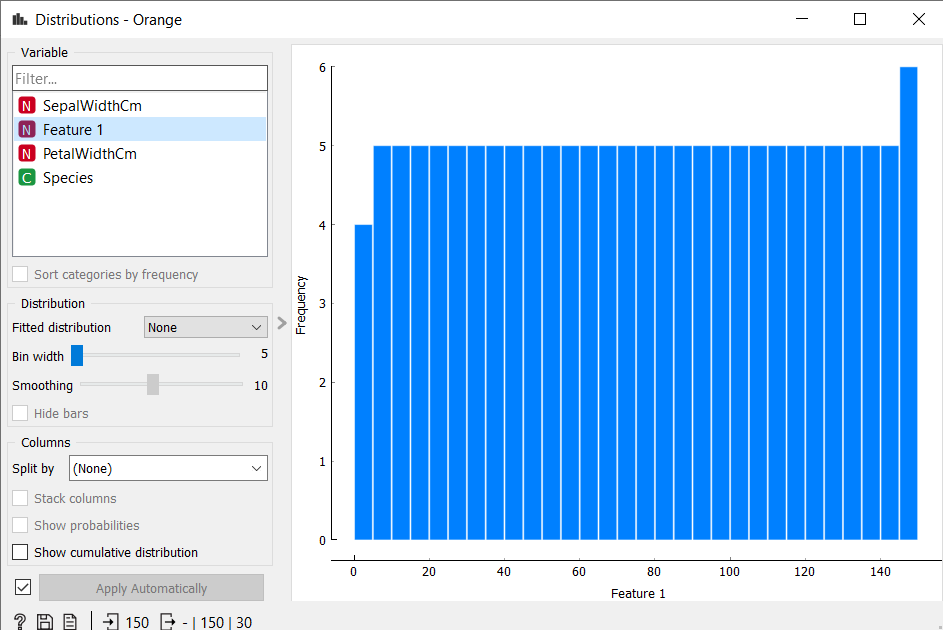
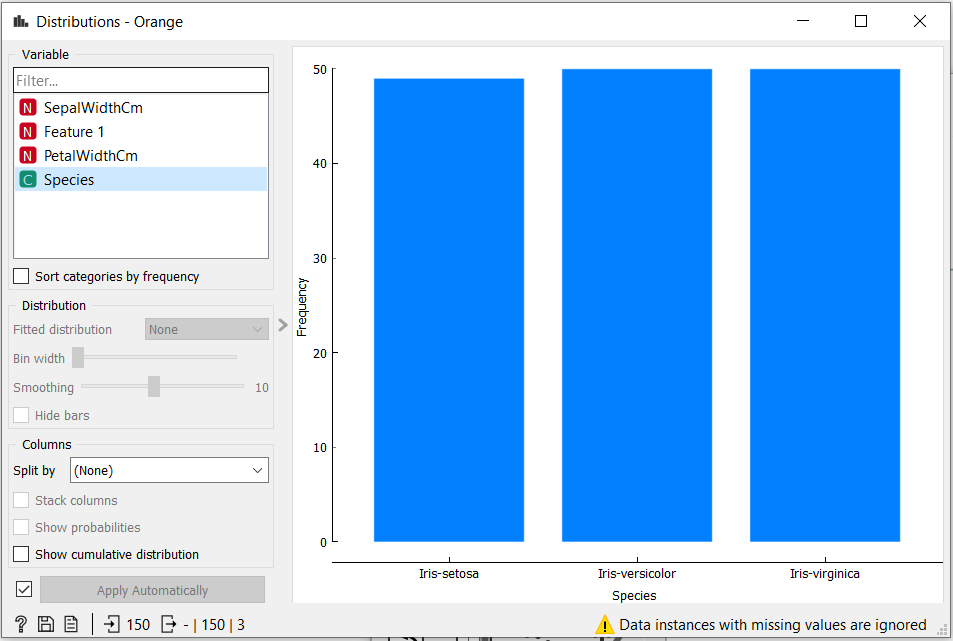
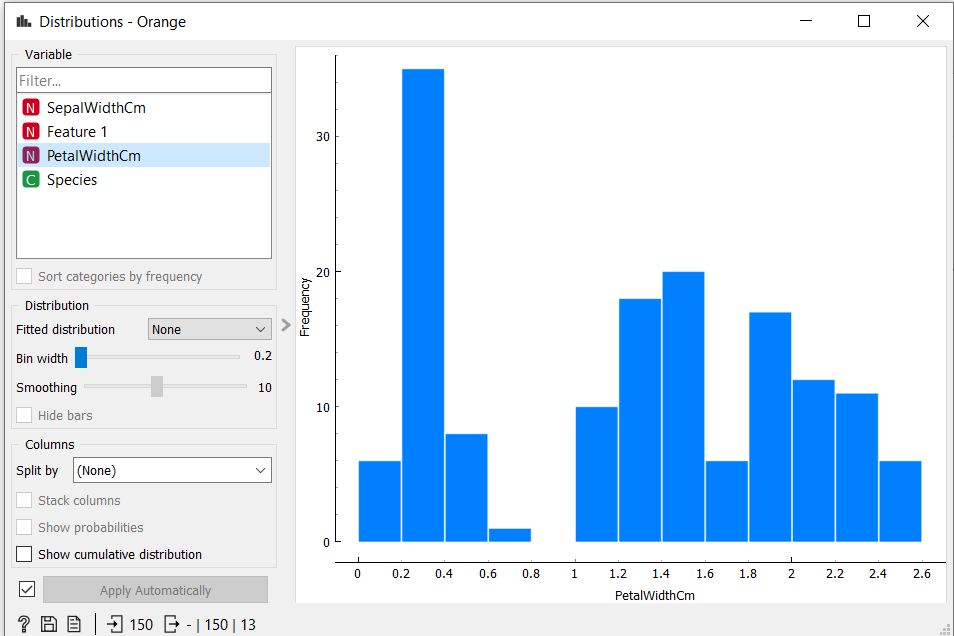


* Data Table

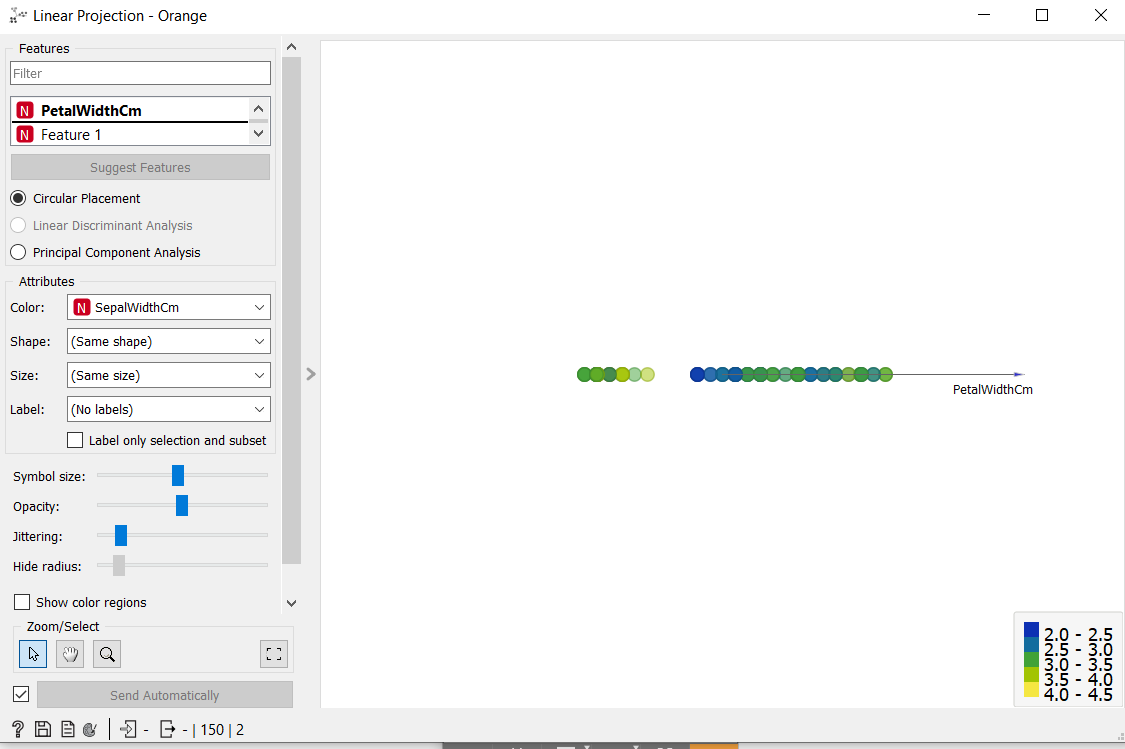


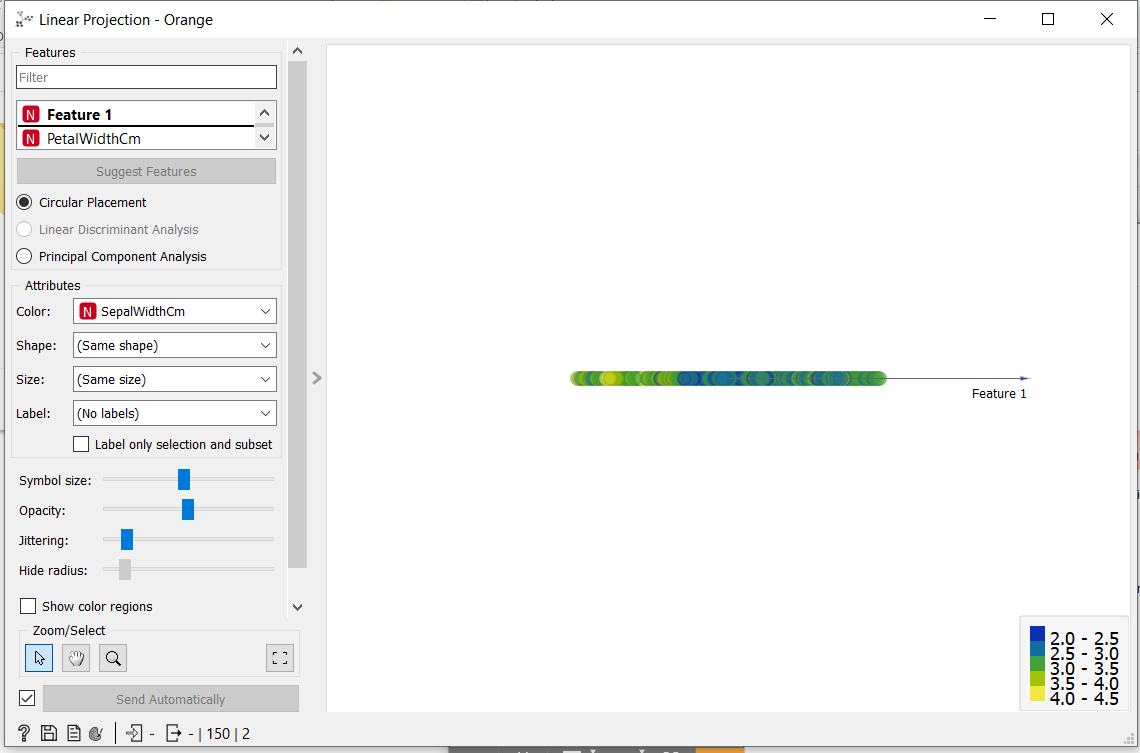
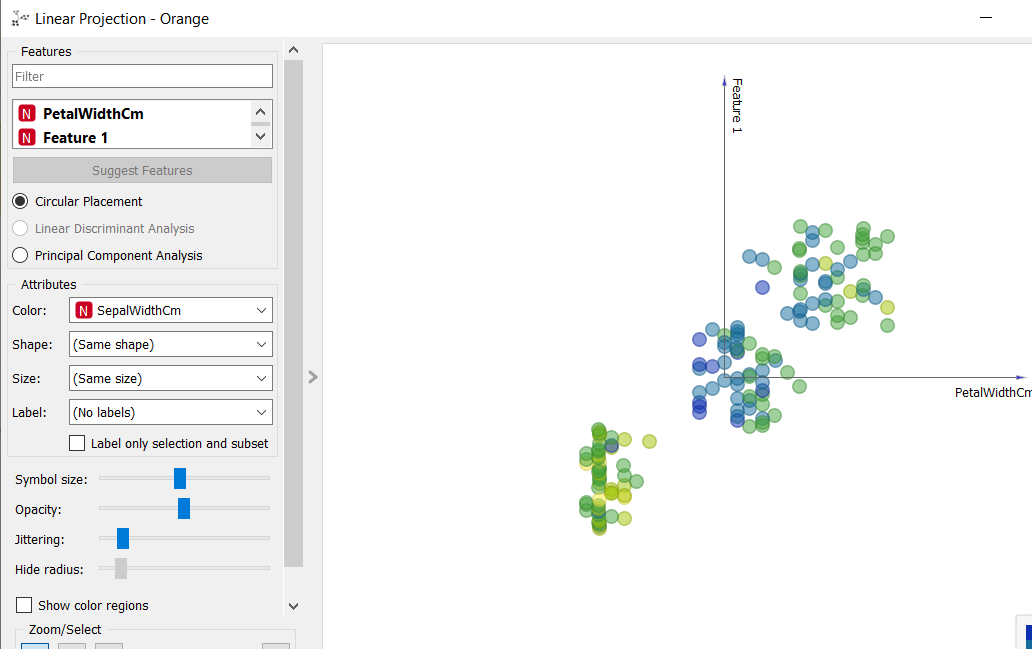
1. Data distribution :



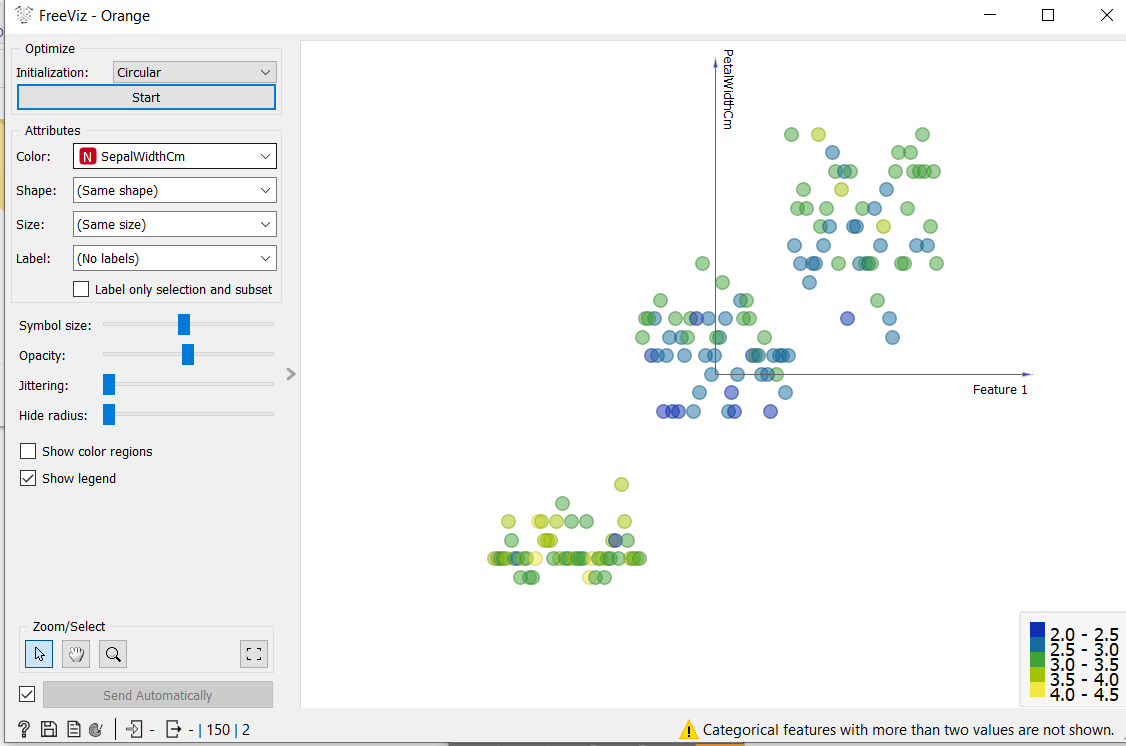


1. Linear projection –





C ) Freeviz



1. Differentiate in between free software, Open source software and proprietary software with respect to its properties.

-> 1. Open source Software :

Open source software is a computer software whose source code is available openly in internet and programmers can modify it to add new features and capabilities without any cost. Here the software is developed and tested through open collaboration. This software is managed by an open source community of developers. It provides community support as well as commercial support if available for maintenance. We can get it for free of cost. This software also sometimes comes with license and sometimes does not. This license provides some rights to users like.

* Software can be used for any purpose
* Allows to study how software works
* Freedom to modify and improve the program
* No restrictions on redistributions

Some examples of Open source software includes Android, Ubuntu, Firefox, Open Office etc.

2. Proprietary Software :

Proprietary software is a computer software where the source codes are not publicly not available only the company which has created can modify it. Here the software is developed and tested by the individual or organization by which it is owned not by public. This software is managed by an closed team of individuals or groups that developed it. We have to pay to get this software and it commercial support if available for maintenance. The company gives a valid and authenticated license to the users to use this software. But this license put some restrictions on users also like.

* Number of installations of this software into computers
* Restrictions on sharing of software illegally
* Time period up to which software will operate
* Number of features allowed to use

Some examples of Proprietary software includes Windows, MacOS, Internet Explorer, Google earth, Microsoft Office etc.

3. Free Software:

“Free software” means software that respects users’ freedom and community. Roughly, it means that the users have the freedom to run, copy, distribute, study, change and improve the software.

The term “free software” is sometimes misunderstood—it has nothing to do with price. It is about freedom.

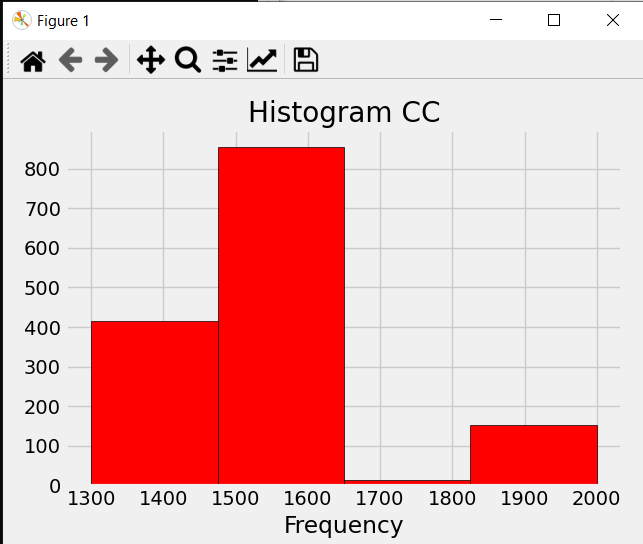
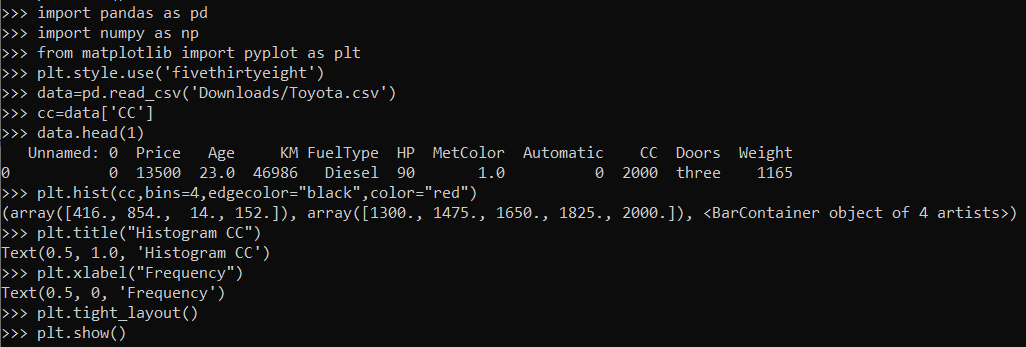
4. Using Anaconda Python create Histogram, Scatter plot and Bar plot for the dataset given below. Dataset- <https://drive.google.com/file/d/1i11BZFe8Xj9kNq7eeE9KOa_Iz1KhEdXJ/view>

Install anaconda from

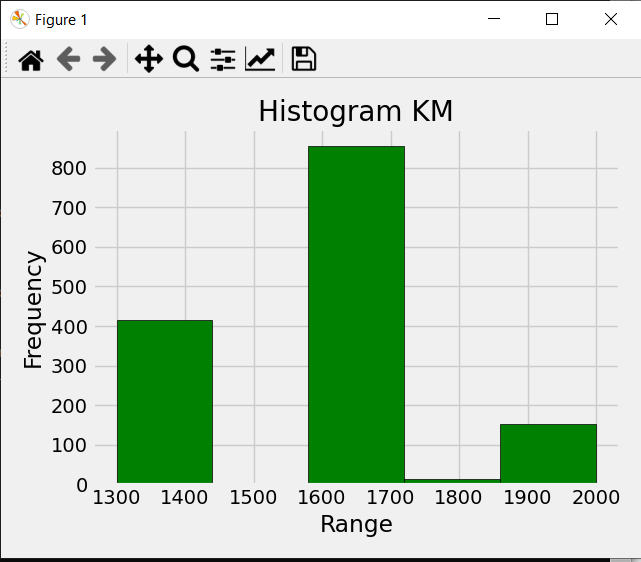
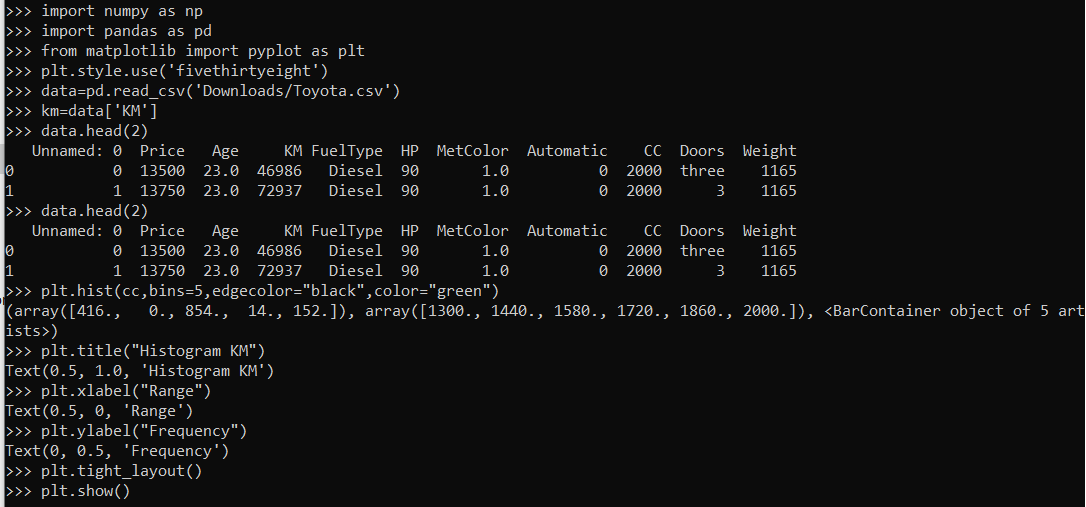
<https://www.anaconda.com/products/individual>

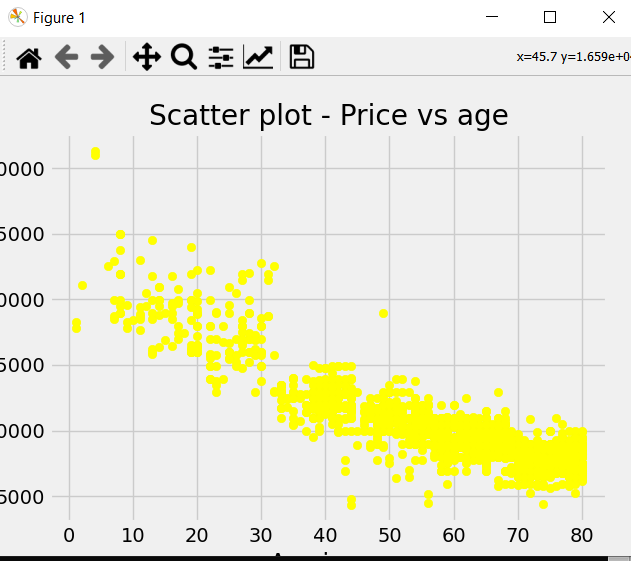
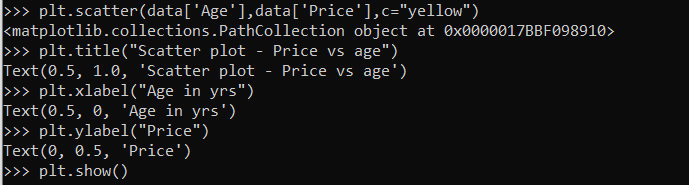
Open anaconda prompt &exexcutefollowing :

a. Scatter plot- Scatter plot of Price Vs Age

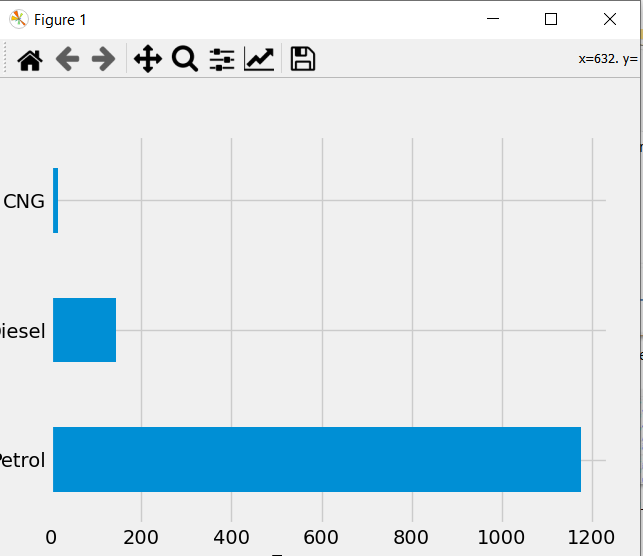
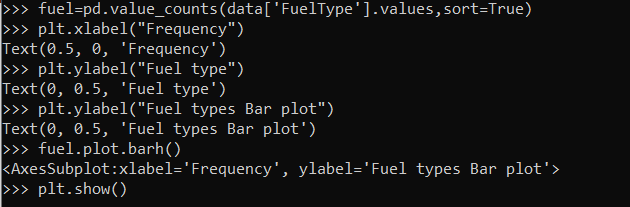


b. Histogram- for Kilometer and CC





c. Bar plot- Bar plot for different fuel type



5. Enlist some examples along with its purpose and properties (at least 10) of FOSS and proprietary software with respect to database

->Examples of proprietary software include:

* Windows and OS X operating systems
* Microsoft Office productivity suite
* Adobe Creative Suite productivity software
* Logic music creation software
* paid-for games for consoles

Examples of open-source software

* LibreOffice.
* GNU/Linux.
* VLC Media Player.
* Mozilla Firefox.
* GIMP.
* VNC.
* Apache web server.
* jQuery.