Software Engineering Tools Lab

Assignment 1

PRN No: 2019BTECS00084

Full name: Pravin Santosh Lokhande

Batch: T2

Title of Experiment: Introduction to OSS

Q.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 pre-processing, data classification (use any appropriate ML algorithm) and data visualization efficiently on given dataset.

Use the Iris dataset given

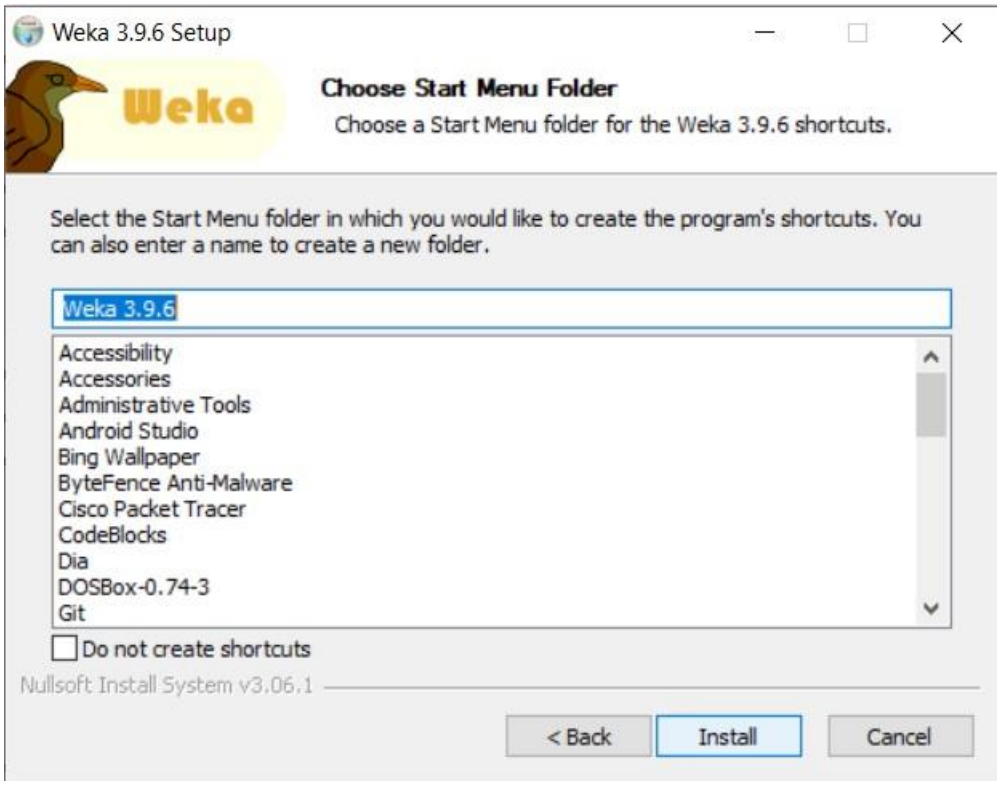
https://drive.google.com/file/d/1A3Fxsfzm6BSfhFZGDrjI47RTe45bSgYP/view

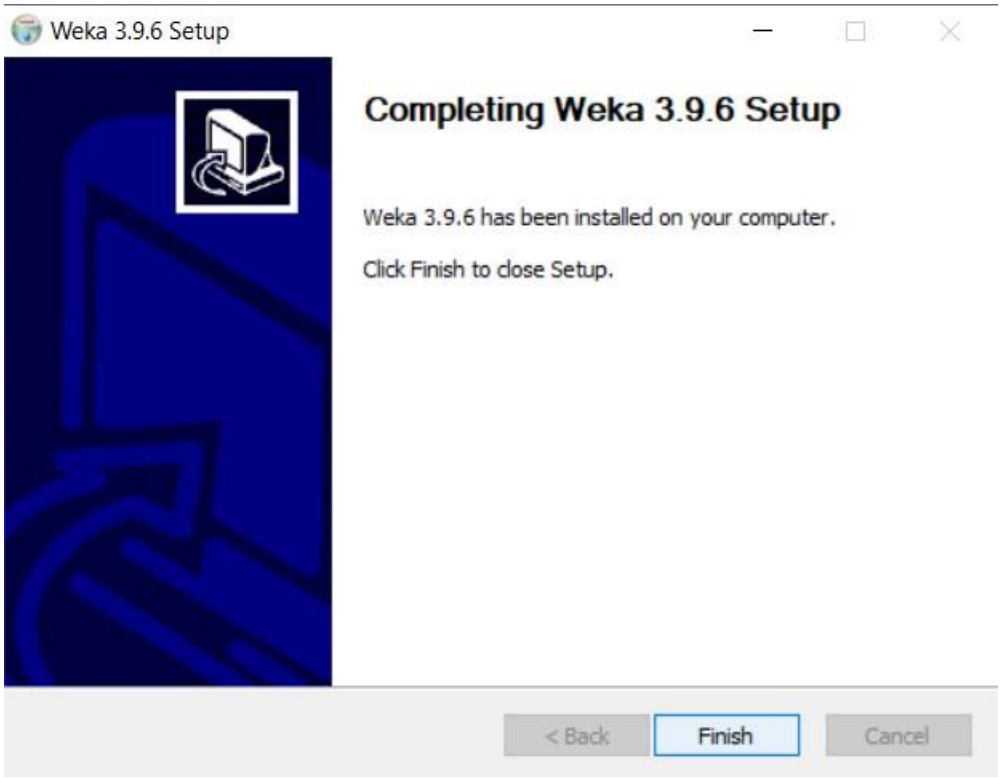
Note-provide screen shots for every task Create a report which will illustrate the details of tasks performed (for e.g., to perform pre-processing of data provide details of navigation and selection of appropriate parameters)

Ans:

1. **Installation:**

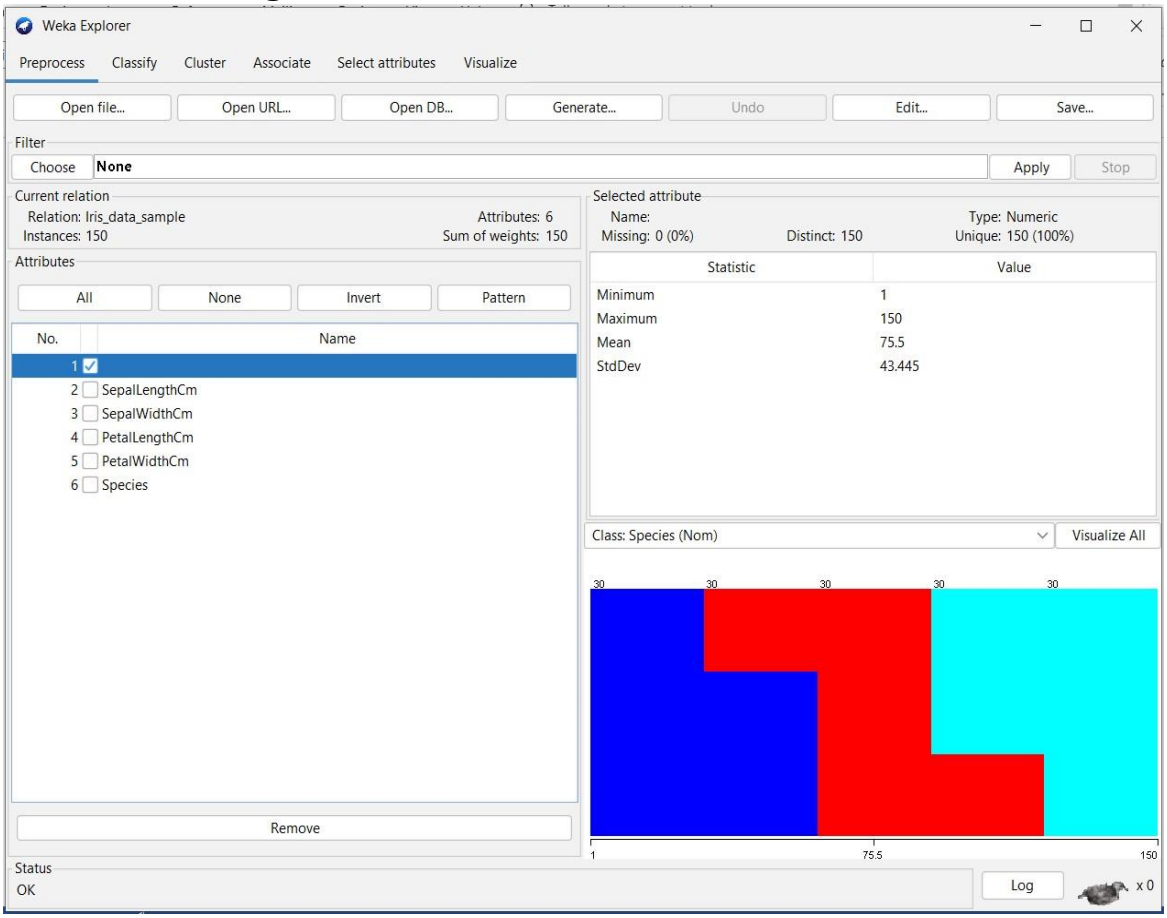
   Download from: https://www.autotechint.com/weka


Weka 3.9.6 Setup
Weka
Choose Start Menu Folder
Choose a Start Menu folder for the Weka 3.9.6 shortcuts.
Select the Start Menu folder in which you would like to create the program's shortcuts. You can also enter a name to create a new folder.
Weka 3.9.6
Accessibility
Accessories
Administrative Tools
Android Studio
Bing Wallpaper
ByteFence Anti-Malware
Cisco Packet Tracer
CodeBlocks
Dia
DOSBox-0.74-3
Git
Do not create shortcuts
Nullsoft Install System v3.06.1
< Back
Install
Cancel


Weka 3.9.6 Setup
Completing Weka 3.9.6 Setup
Weka 3.9.6 has been installed on your computer.
Click Finish to close Setup.
< Back
Finish
Cancel

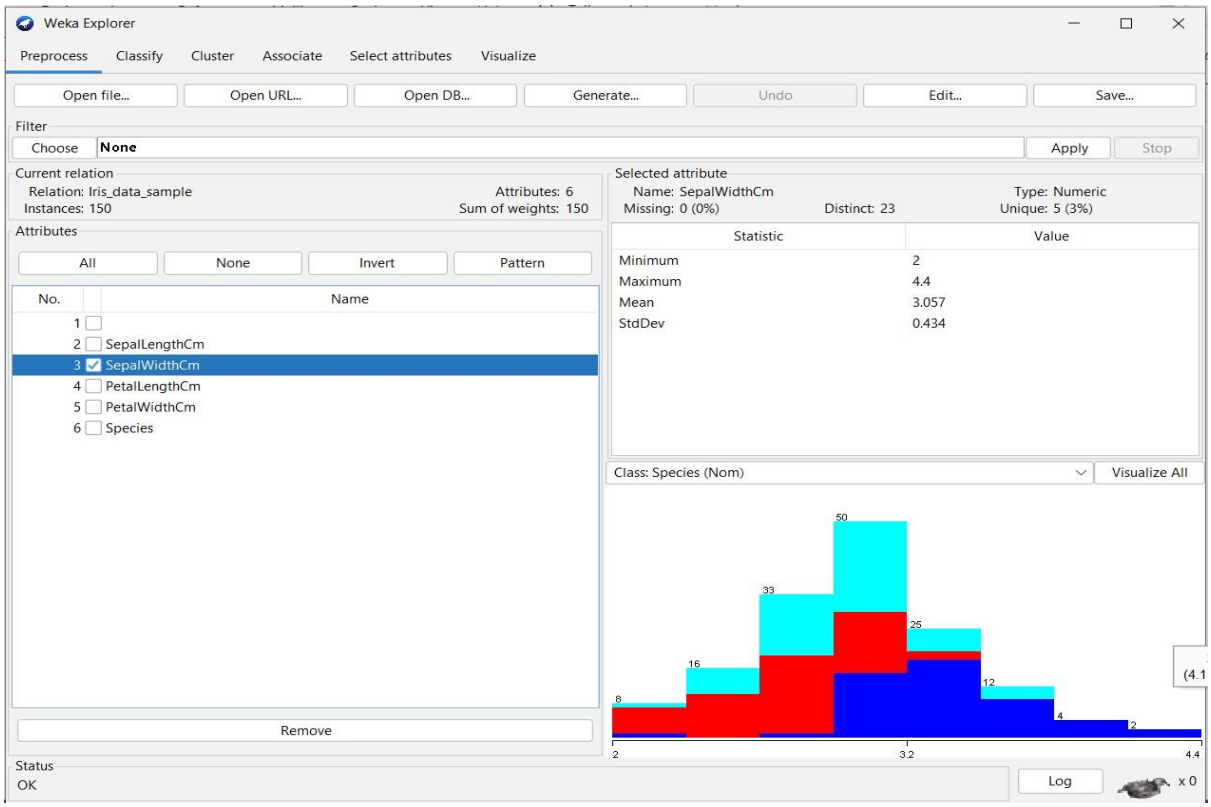
- **Homepage:**

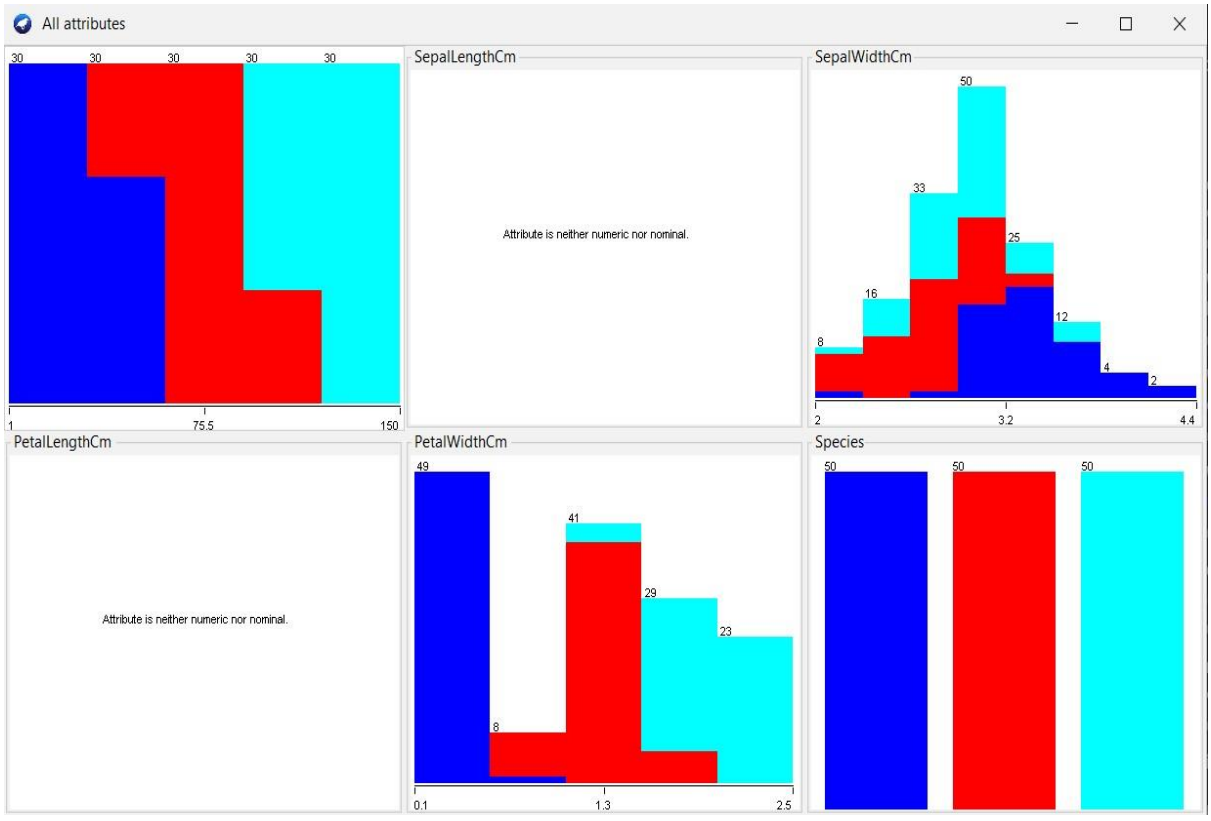
2. Data Processing:



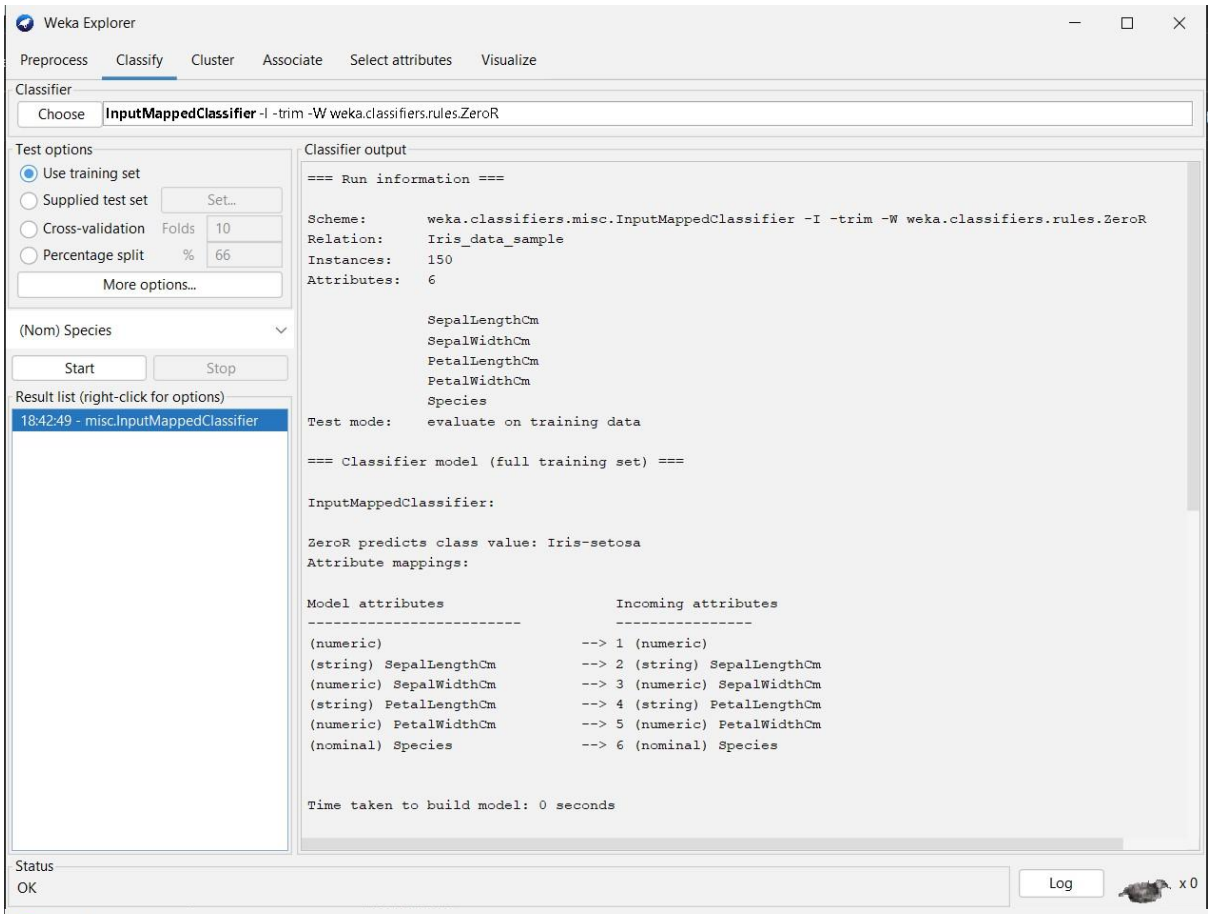
DATASET

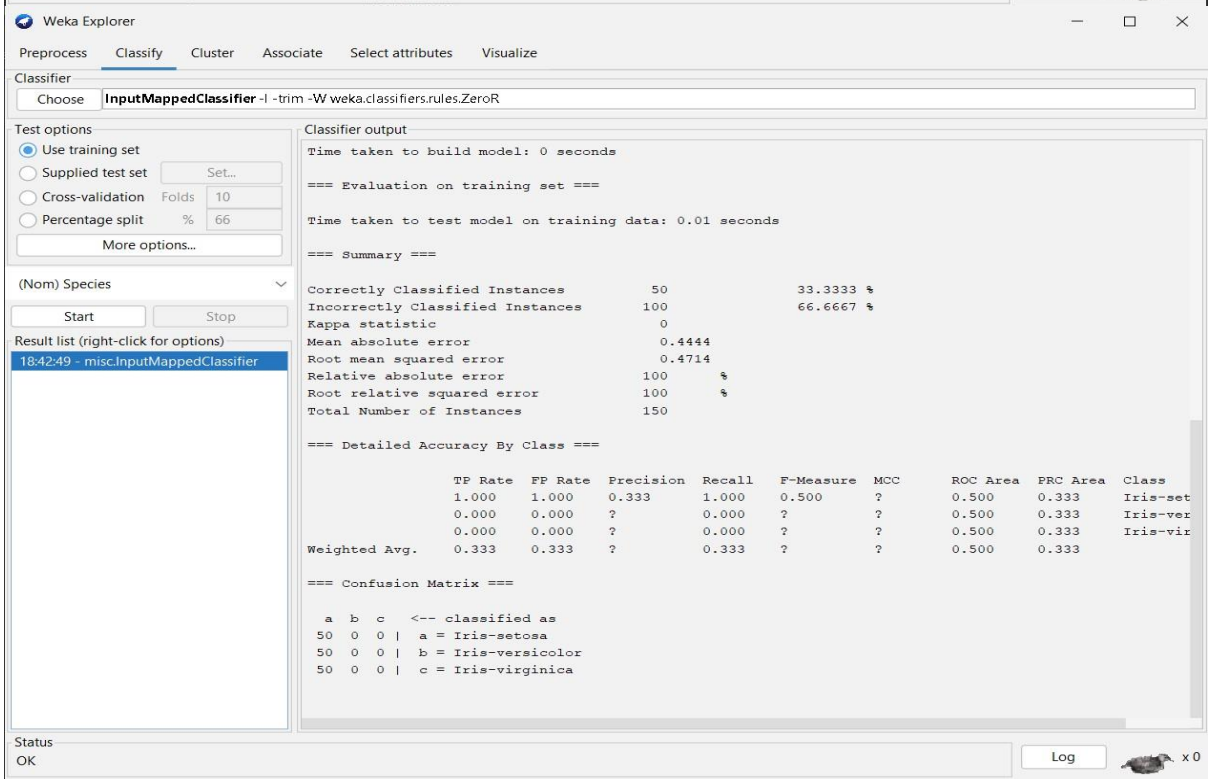
- Attribute : 6
- Total instance : 150





3. Data classification:





Q2.

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

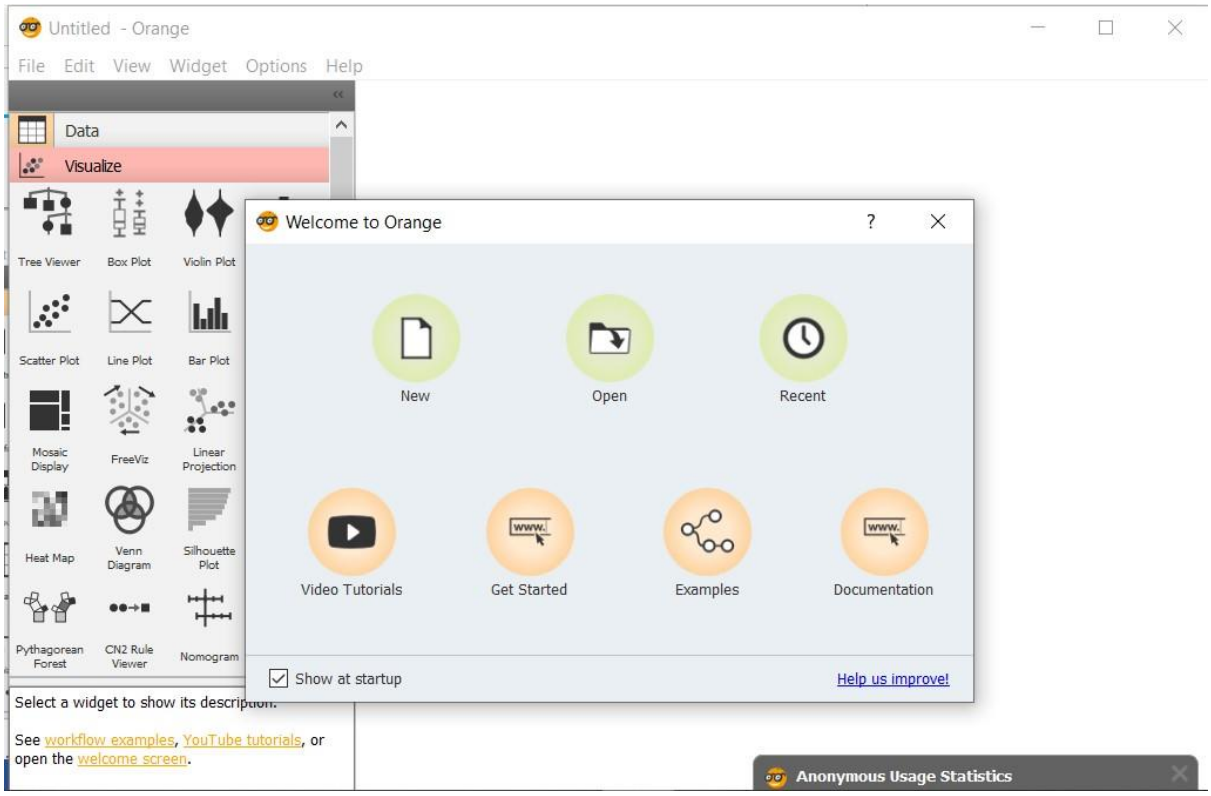
c. Show linear projection

d. Show FreeViz

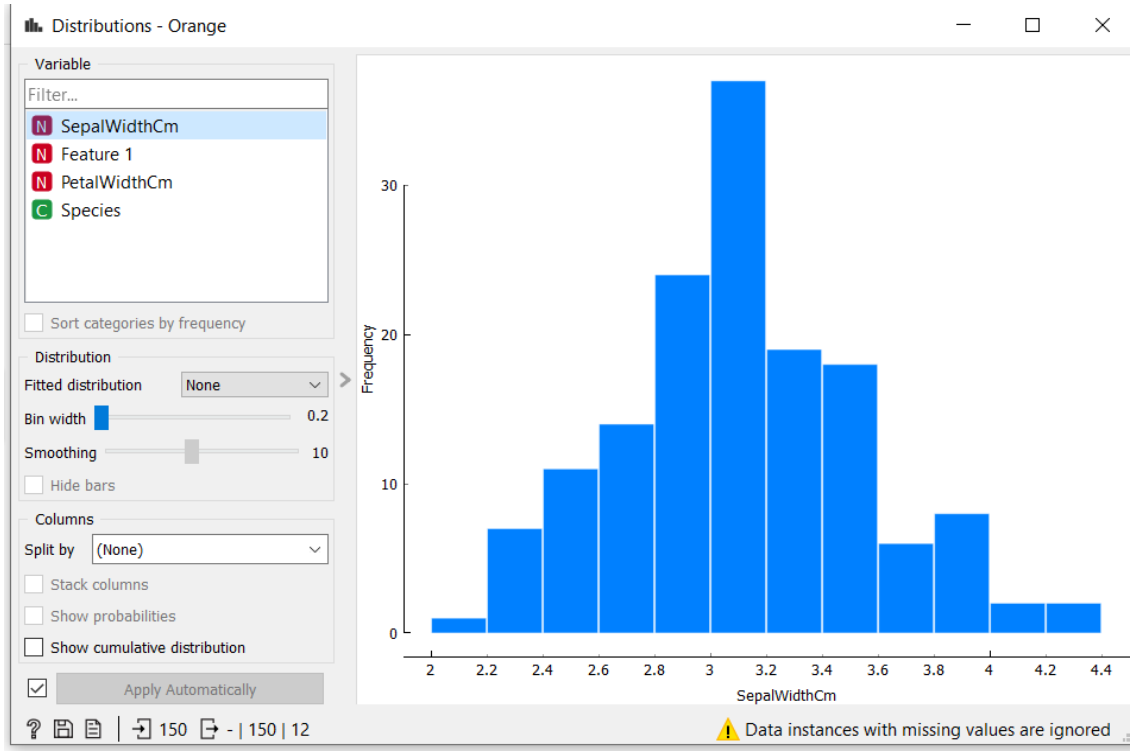
Use dataset
https://drive.google.com/file/d/1m6sKI1Dap0XK6Bw1edUd5PohwpPwXnd9/view

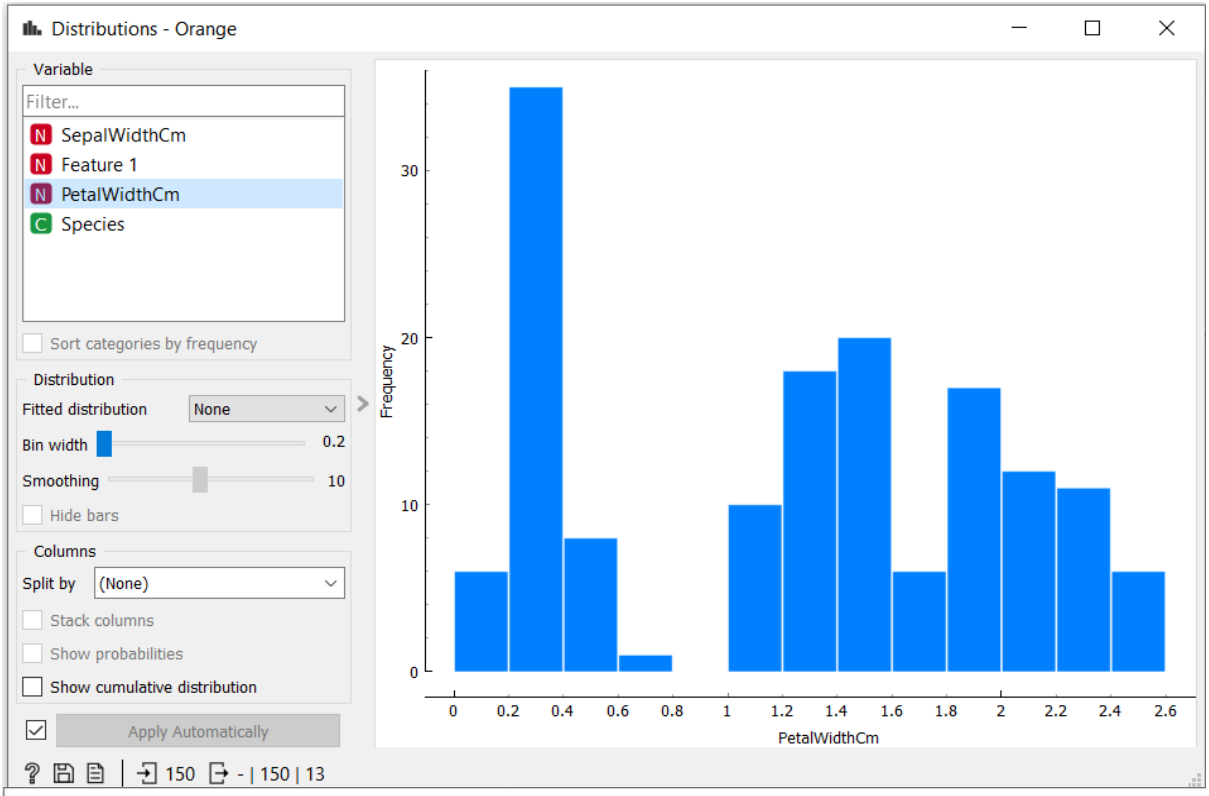
Ans:

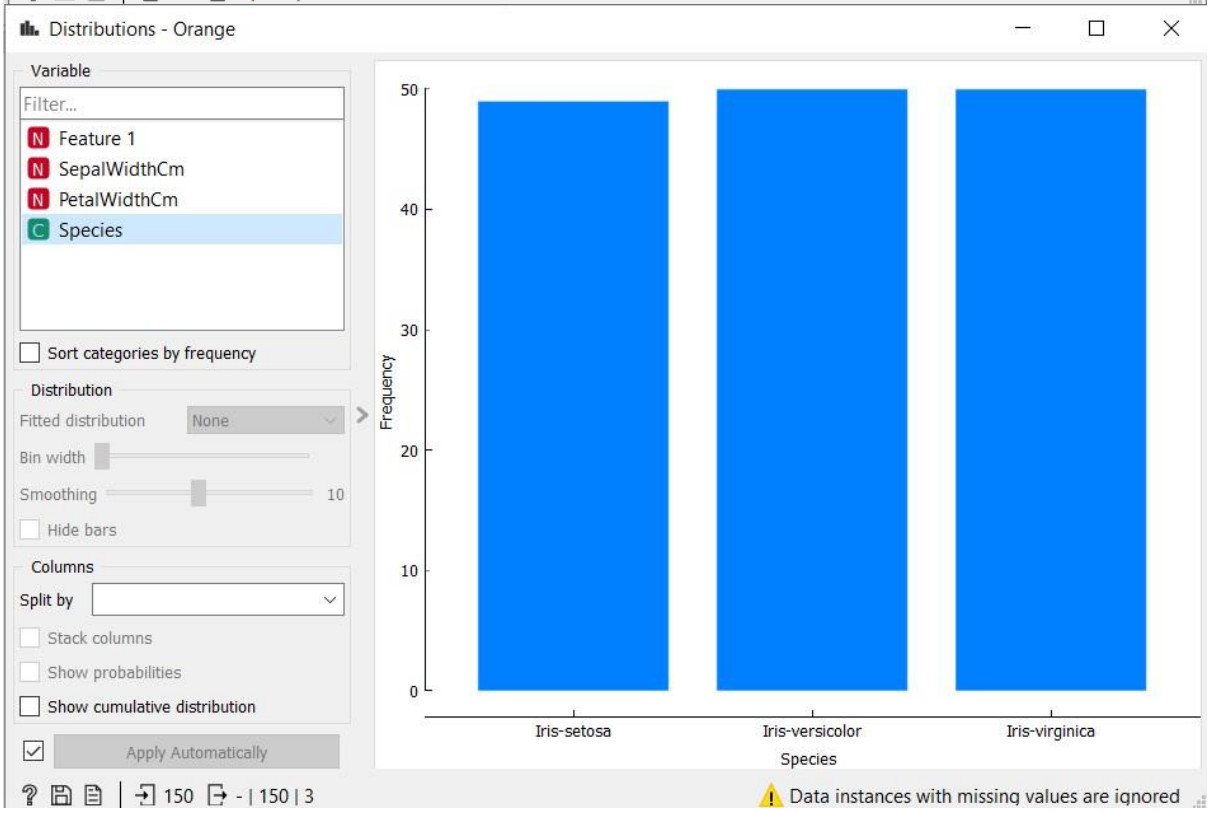
a. **Installation:**
   - **Download from:**
     https://orangedatamining.com/download/#windows



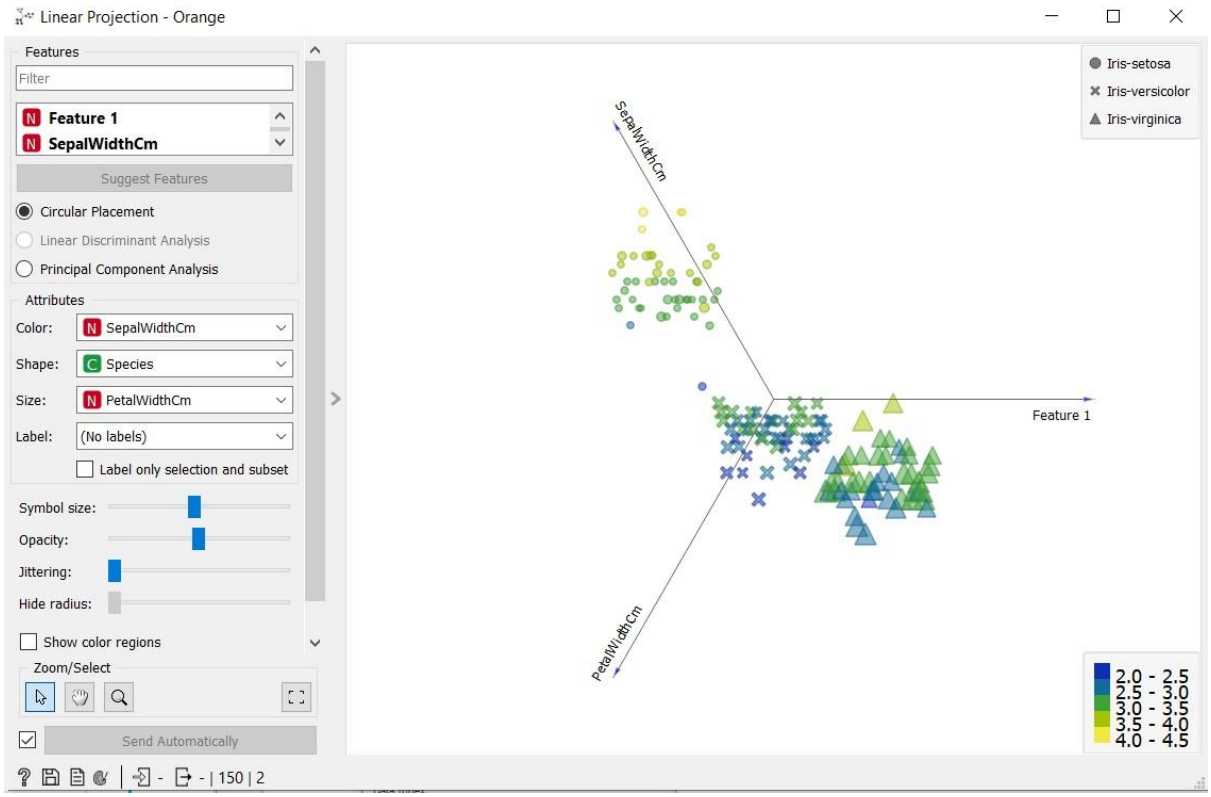
b. Data distribution:

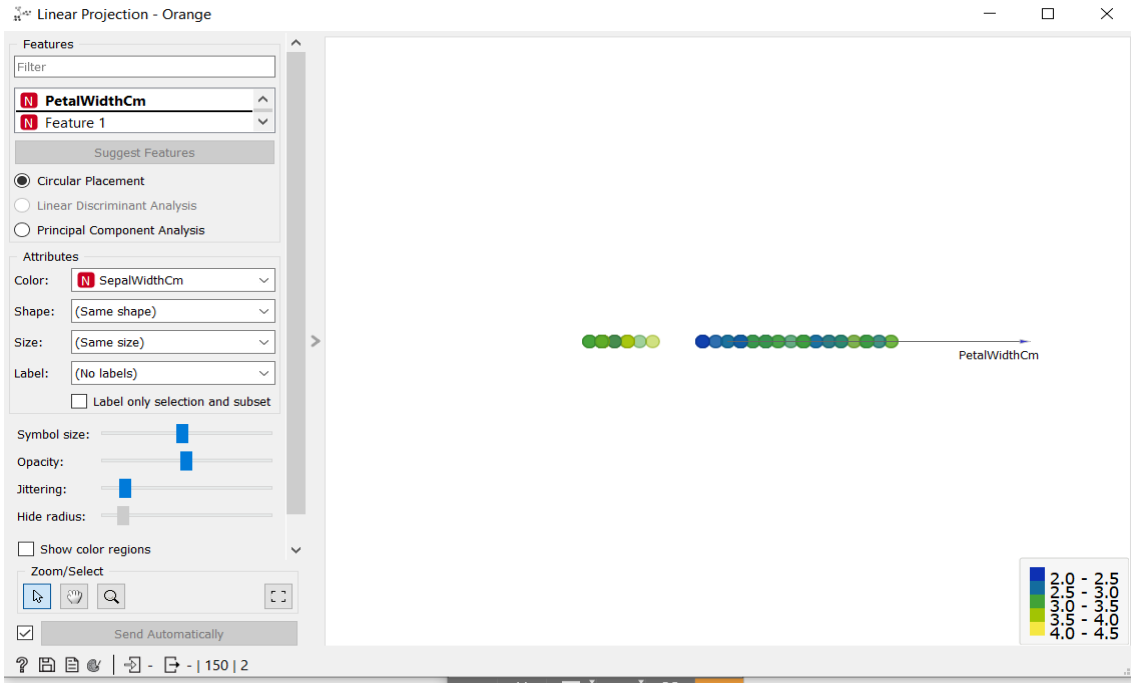



Distributions - Orange
Variable
Filter...
SepalWidthCm
Feature 1
PetalWidthCm
Species
Sort categories by frequency
Distribution
Fitted distribution
None
Bin width
0.2
Smoothing
10
Hide bars
Columns
Split by
(None)
Stack columns
Show probabilities
Show cumulative distribution
Apply Automatically
150 | - | 150 | 13
Frequency
PetalWidthCm


Distributions - Orange
Variable
Filter...
Feature 1
SepalWidthCm
PetalWidthCm
Species
Sort categories by frequency
Distribution
Fitted distribution
None
Bin width
Smoothing
10
Hide bars
Columns
Split by
Stack columns
Show probabilities
Show cumulative distribution
Apply Automatically
150 | - | 150 | 3
Data instances with missing values are ignored
Frequency
Iris-setosa
Iris-versicolor
Iris-virginica
Species

c. Linear projection:





d. FreeViz:


FreeViz - Orange
Optimize
Initialization:
Circular
Start
Attributes
Color:
SepalWidthCm
Shape:
(Same shape)
Size:
(Same size)
Label:
(No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering:
Hide radius:
Show color regions
Show legend
Zoom/Select
Send Automatically
- | 150 | 2
PetalWidthCm
Feature 1
2.0 - 2.5
2.5 - 3.0
3.0 - 3.5
3.5 - 4.0
4.0 - 4.5
Categorical features with more than two values are not shown.

Q3.

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

Ans:

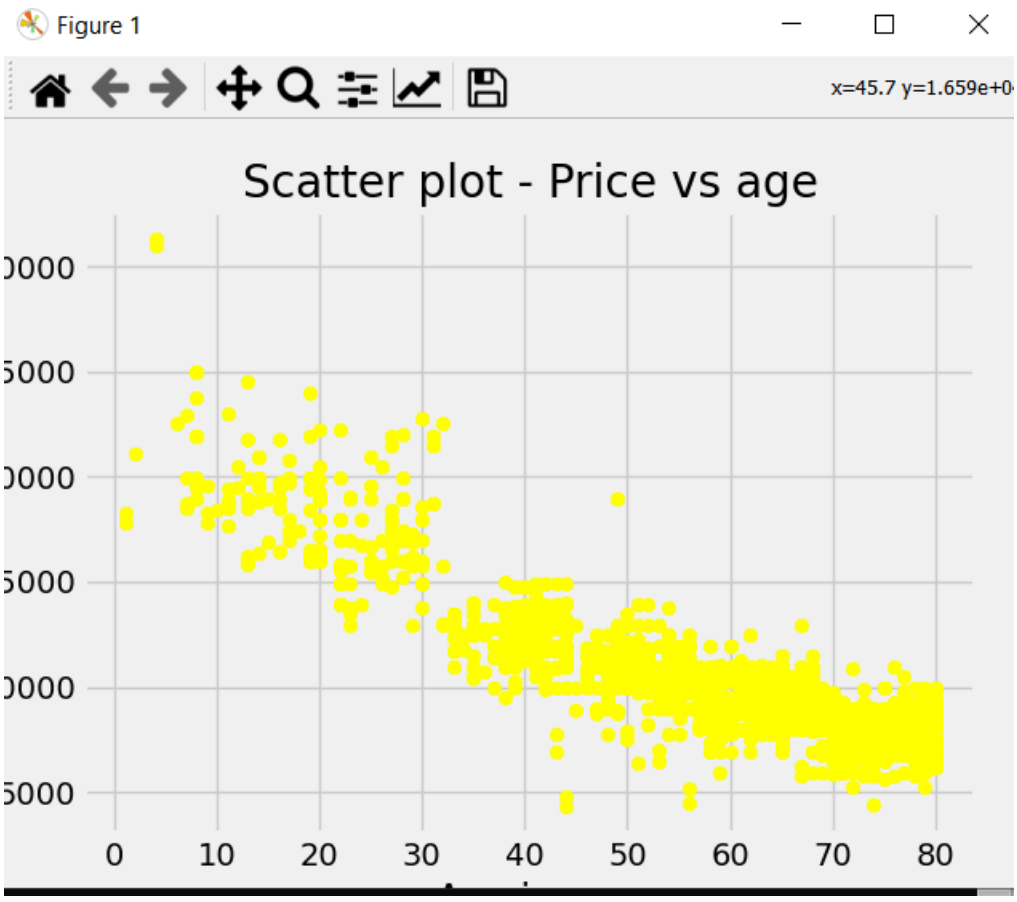
Free Software	Open-Source Software	Proprietary Software
The Source code is not available.	The source code is available and modified.	The source code is highly confidential.
It is free to use.	Except some, all other are free to use.	Requires license for usage.
Owners only can modify source and publish update.	Anyone can modify source code but the version control only publishes the update.	No one can even see the source code, so only owners Modify and publish update.
Very slow update rate.	Very high update rate as anyone can modify.	Very slow update rate.
Less secure than OSS.	As many developers are involved in the development of OSS vulnerability capturing is more, so highly secure.	Only particular group of developers work on these So less secure than os
e. g. valorant	**e. g.** Libre Office	**e. g.** Microsoft office

Q.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
a. Scatter plot- Scatter plot of Price Vs Age
b. Histogram- for Kilometre and CC
c. Bar plot- Bar plot for different fuel types
Ans:

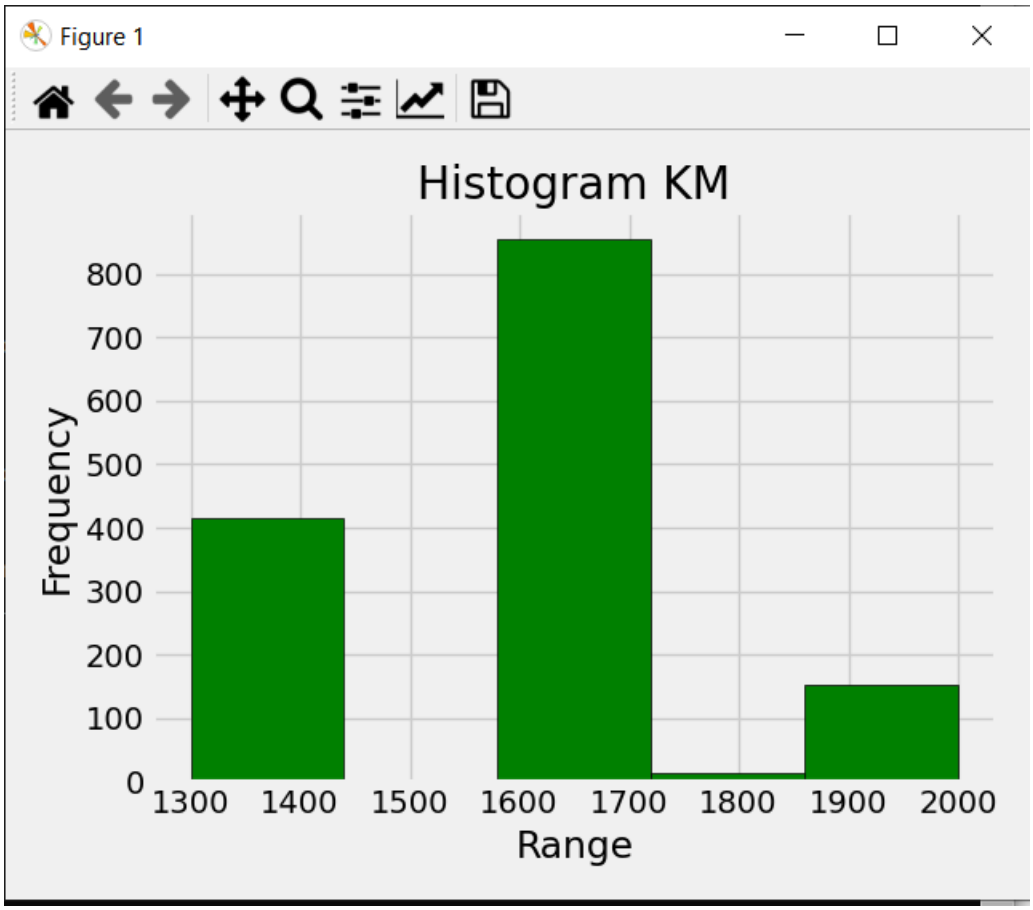
a. **Scatter Plot:**

```
>>> plt.scatter(data['Age'],data['Price'],c="yellow")
<matplotlib.collections.PathCollection object at 0x0000017BBF098910>
>>> plt.title("Scatter plot - Price vs age")
Text(0.5, 1.0, 'Scatter plot - Price vs age')
>>> plt.xlabel("Age in yrs")
Text(0.5, 0, 'Age in yrs')
>>> plt.ylabel("Price")
Text(0, 0.5, 'Price')
>>> plt.show()
```



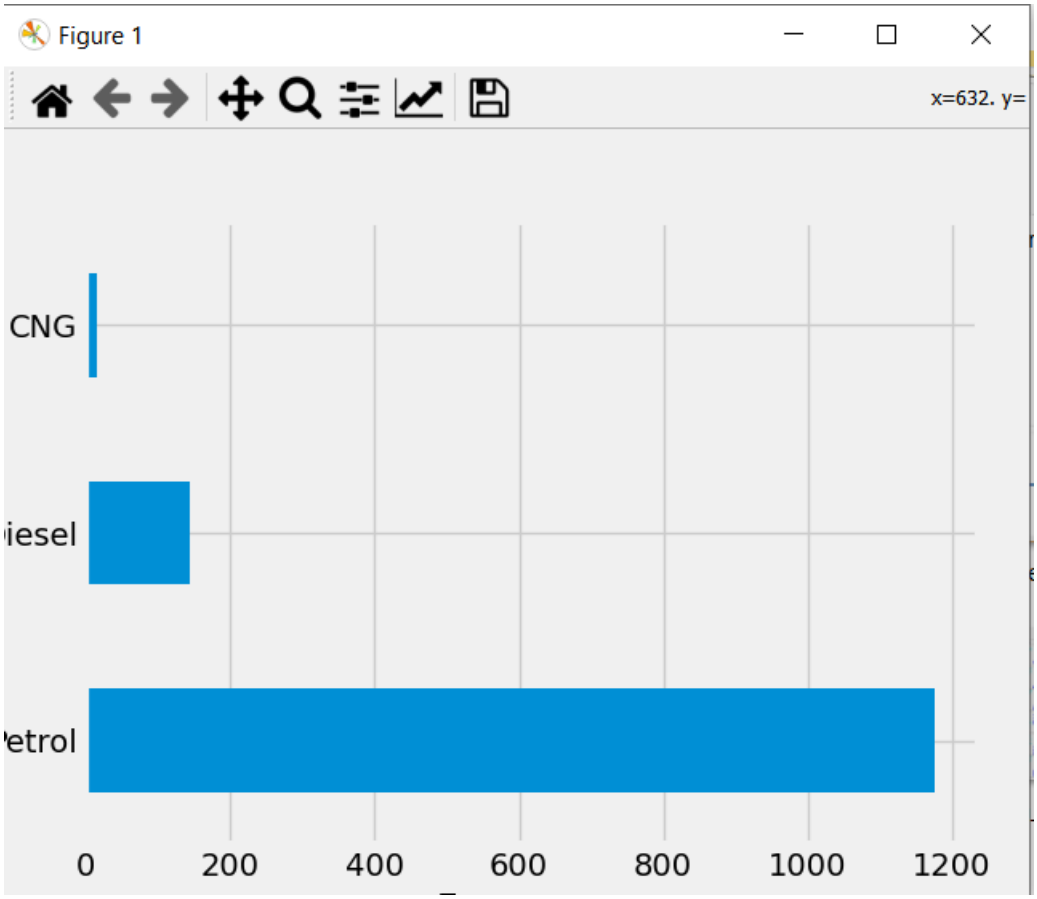
b. Histogram:

```
>>> import numpy as np
>>> import pandas as pd
>>> from matplotlib import pyplot as plt
>>> plt.style.use('fivethirtyeight')
>>> data=pd.read_csv('Downloads/Toyota.csv')
>>> km=data['KM']
>>> data.head(2)
   Unnamed: 0  Price   Age     KM FuelType  HP  MetColor  Automatic    CC  Doors  Weight
0           0  13500  23.0  46986   Diesel  90       1.0          0  2000  three    1165
1           1  13750  23.0  72937   Diesel  90       1.0          0  2000      3    1165
>>> data.head(2)
   Unnamed: 0  Price   Age     KM FuelType  HP  MetColor  Automatic    CC  Doors  Weight
0           0  13500  23.0  46986   Diesel  90       1.0          0  2000  three    1165
1           1  13750  23.0  72937   Diesel  90       1.0          0  2000      3    1165
>>> plt.hist(cc,bins=5,edgecolor="black",color="green")
(array([416.,   0., 854.,  14., 152.]), array([1300., 1440., 1580., 1720., 1860., 2000.]), <BarContainer object of 5 art
ists>)
>>> plt.title("Histogram KM")
Text(0.5, 1.0, 'Histogram KM')
>>> plt.xlabel("Range")
Text(0.5, 0, 'Range')
>>> plt.ylabel("Frequency")
Text(0, 0.5, 'Frequency')
>>> plt.tight_layout()
>>> plt.show()
```



c. Bar Plot:

```
>>> fuel=pd.value_counts(data['FuelType'].values,sort=True)
>>> plt.xlabel("Frequency")
Text(0.5, 0, 'Frequency')
>>> plt.ylabel("Fuel type")
Text(0, 0.5, 'Fuel type')
>>> plt.ylabel("Fuel types Bar plot")
Text(0, 0.5, 'Fuel types Bar plot')
>>> fuel.plot.barh()
<AxesSubplot:xlabel='Frequency', ylabel='Fuel types Bar plot'>
>>> plt.show()
```



Q5.

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

Ans:

Ans :-

FOSS :-

1. Linux Distributions :-
   - all linux distributions are open source operating system.
   - e.g. Kali, ubuntu etc.

2. Android :-
   - The kernal used for android is linux kernal's modified version.
   - This is most papular mobile OS.

3. Brave Browser :-
   - We can surf any website using it.
   - It has an ad-blocking features.

4. Libre office :-
   - We use this to create word file, ppt etc
   - It is replacement of MS office.

5. Openshot :-
   - This oss is used for video editing purpose.

Proprietary Software :-

1. Microsoft office :-
   - It is powered by Microsoft Azure

2. Adobe creative suite :-
   - It is powerd by AWS cloud.

3. Avast security :-
   - It is an antivirus & security software powered by avast cloud.

4. Asset Manager :-
   - It is used for assel managing/tracking.
   - It is powered by liberty street software

5. Asset cloud :-
   - It is used to implement comprehesive assel management system.
   - It is powered by Wasp Barcode.