Walchand College of Engineering, Sangli

Computer Science & Engineering
Third Year

Course: Software Engineering Tools Lab

PRN NO : 2020BTECS00207

Name : Kshirsagar Mayuri Manojkumar

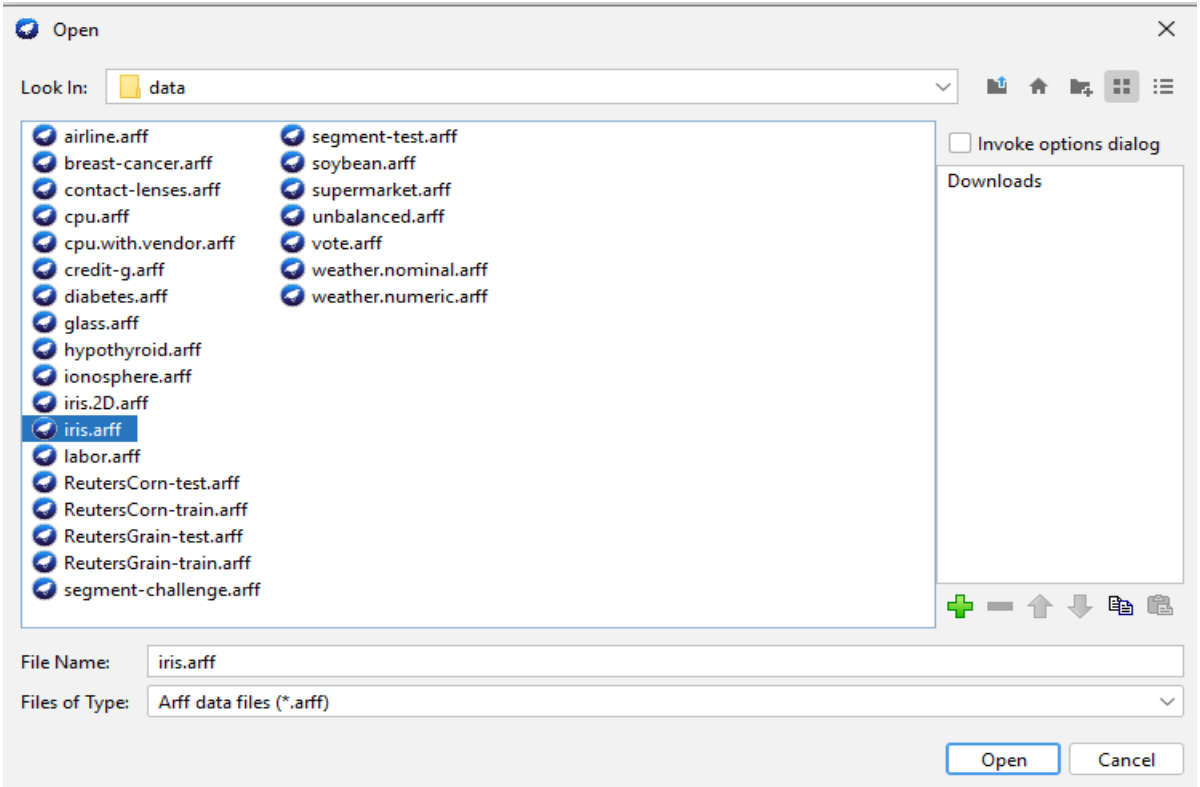
Lab course coordinator : Mr.S .D. Pujari
Batch: - T5

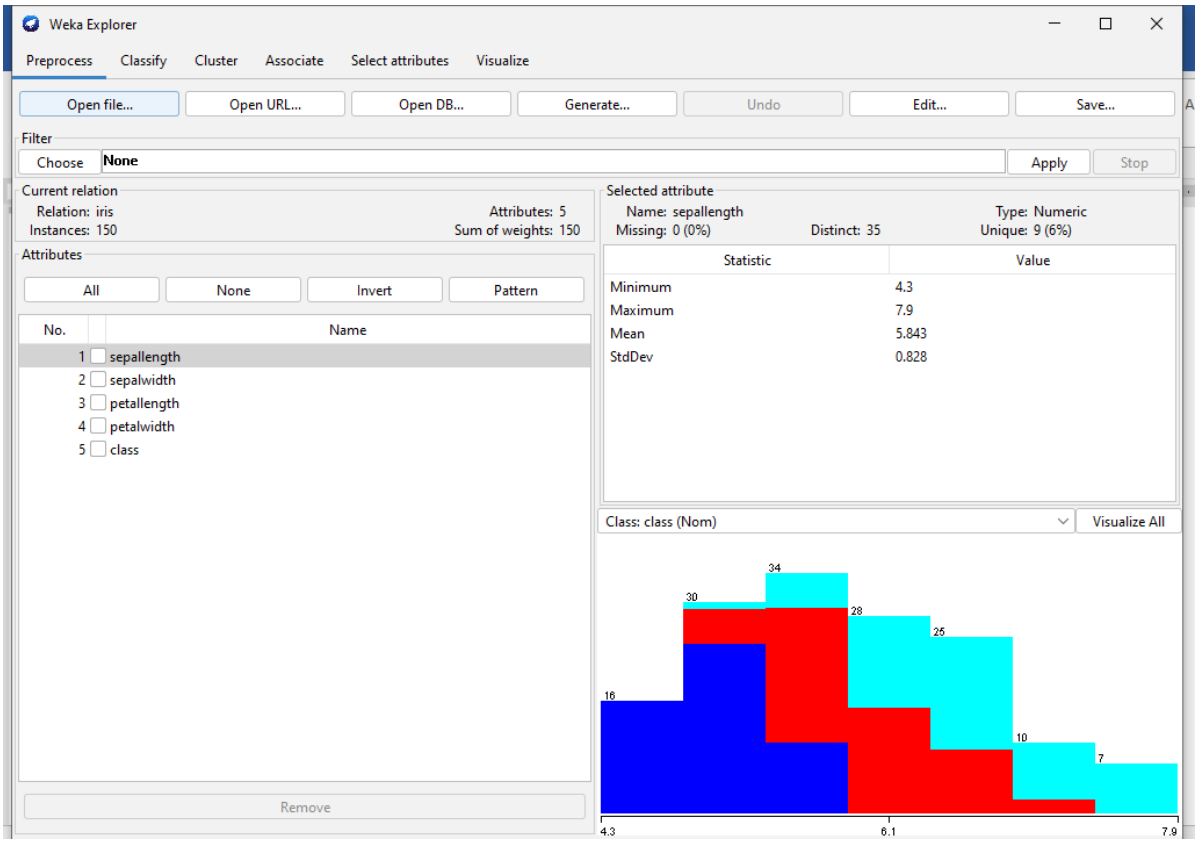
Assignment No-1

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.

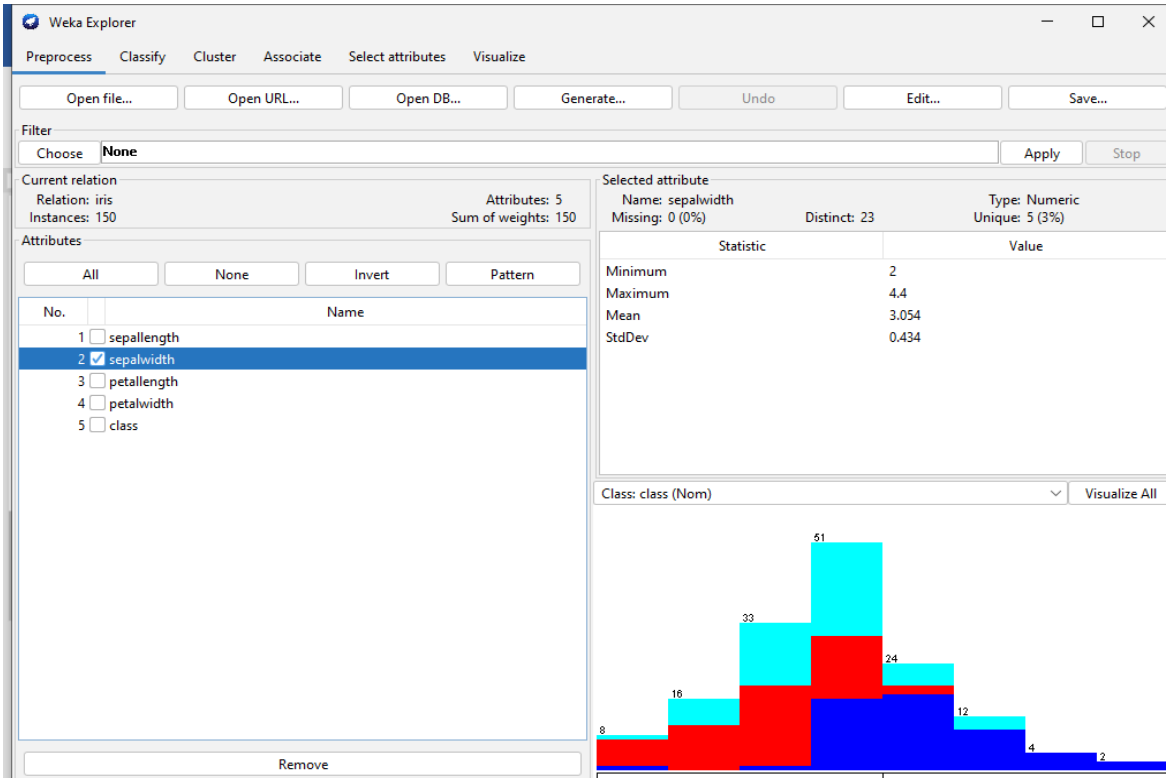
PRN:2020BTECS00207

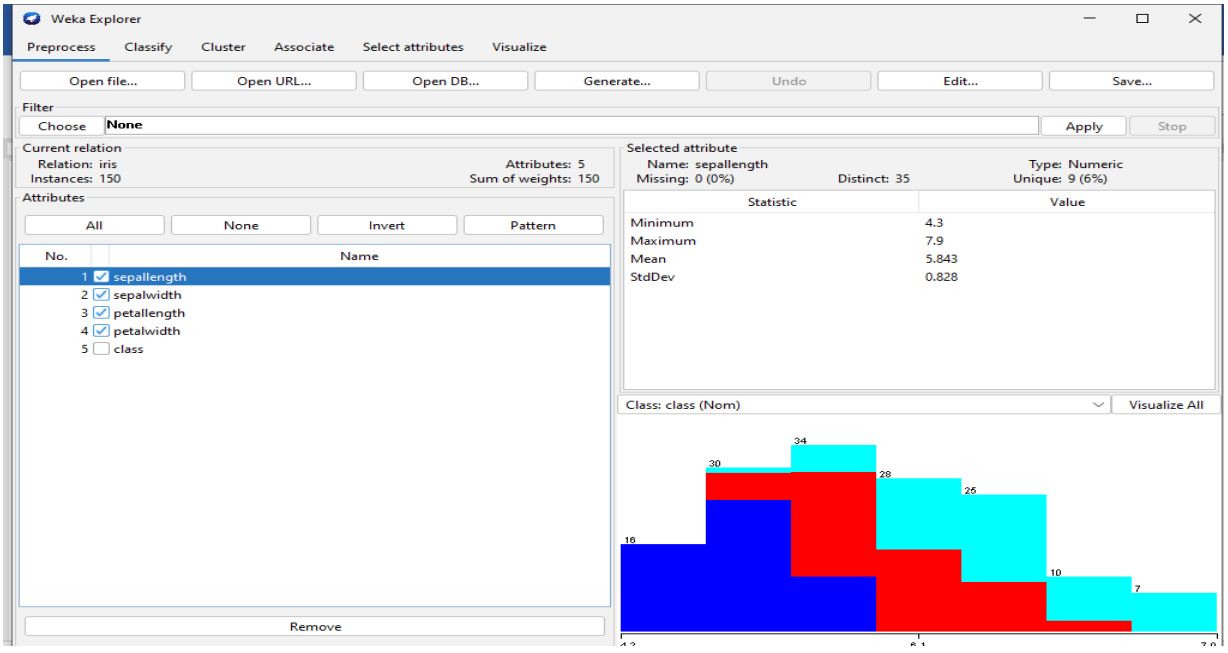
PRN:2020BTECS00207



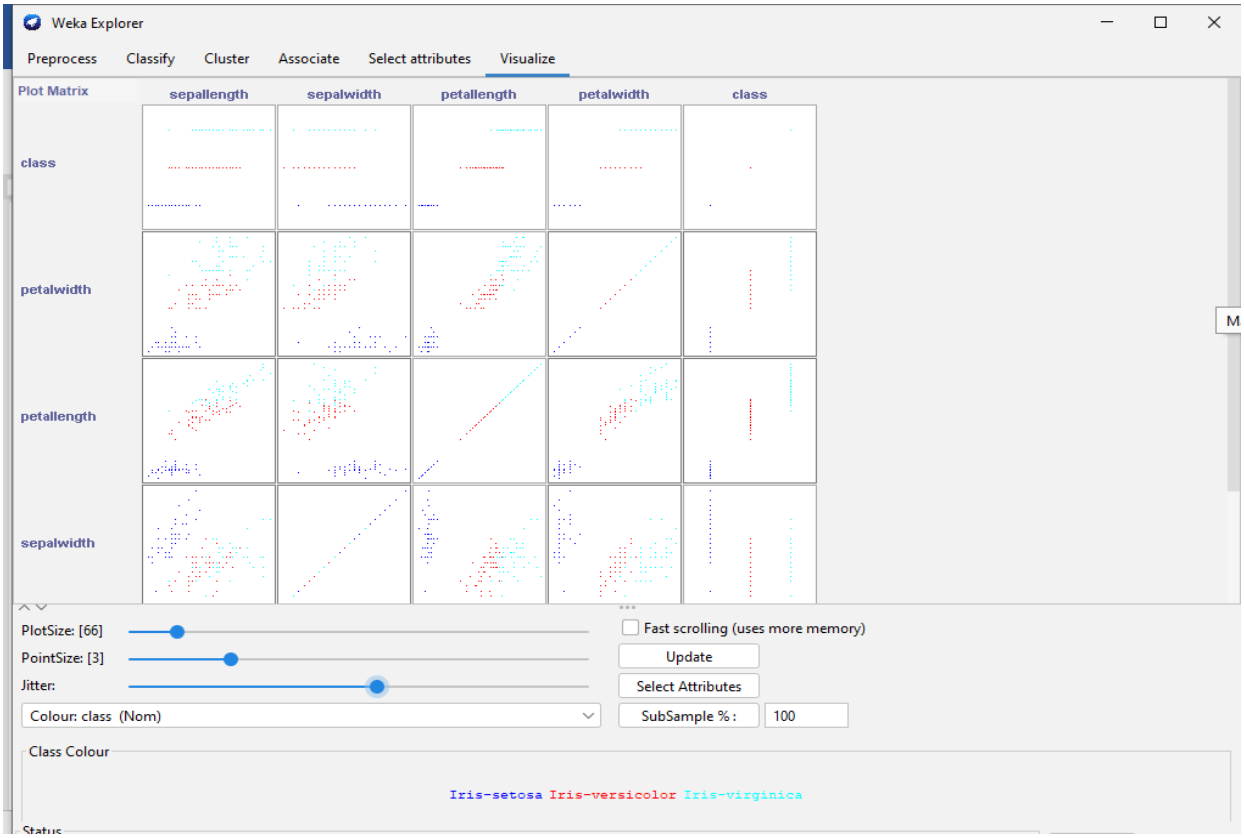


PRN:2020BTECS00207

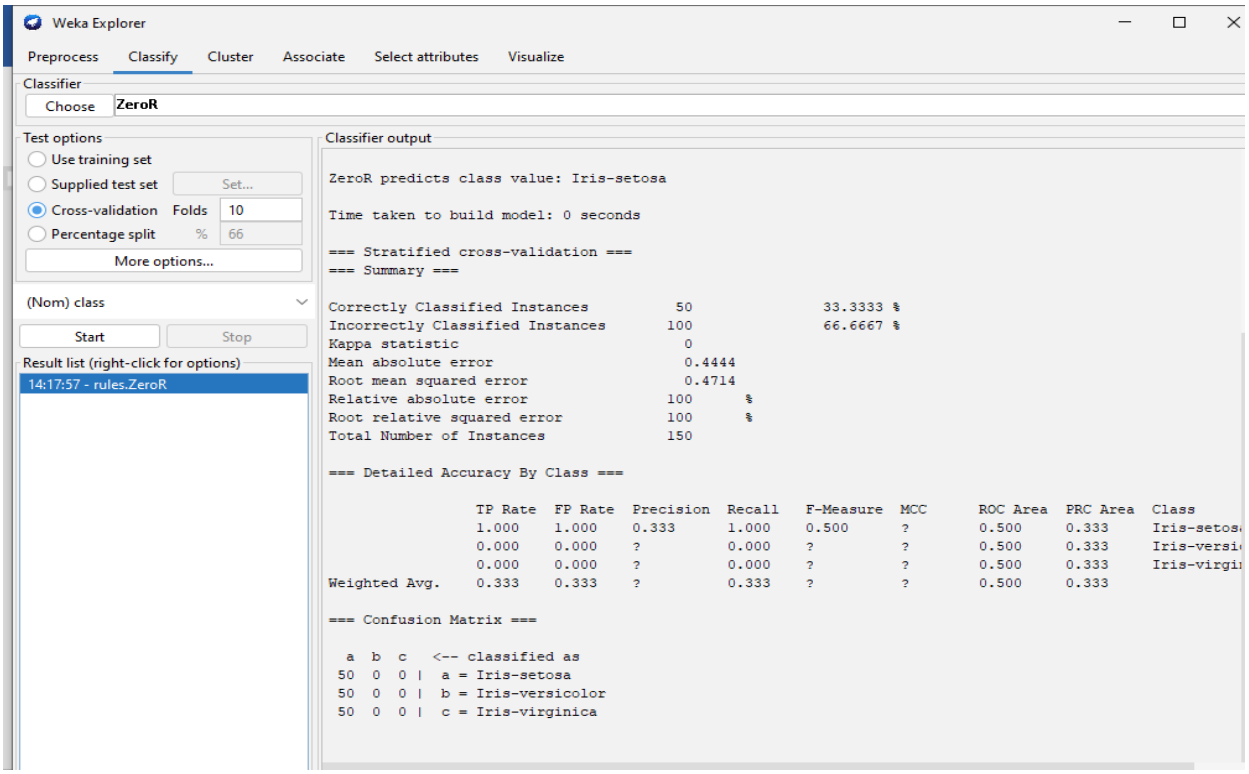


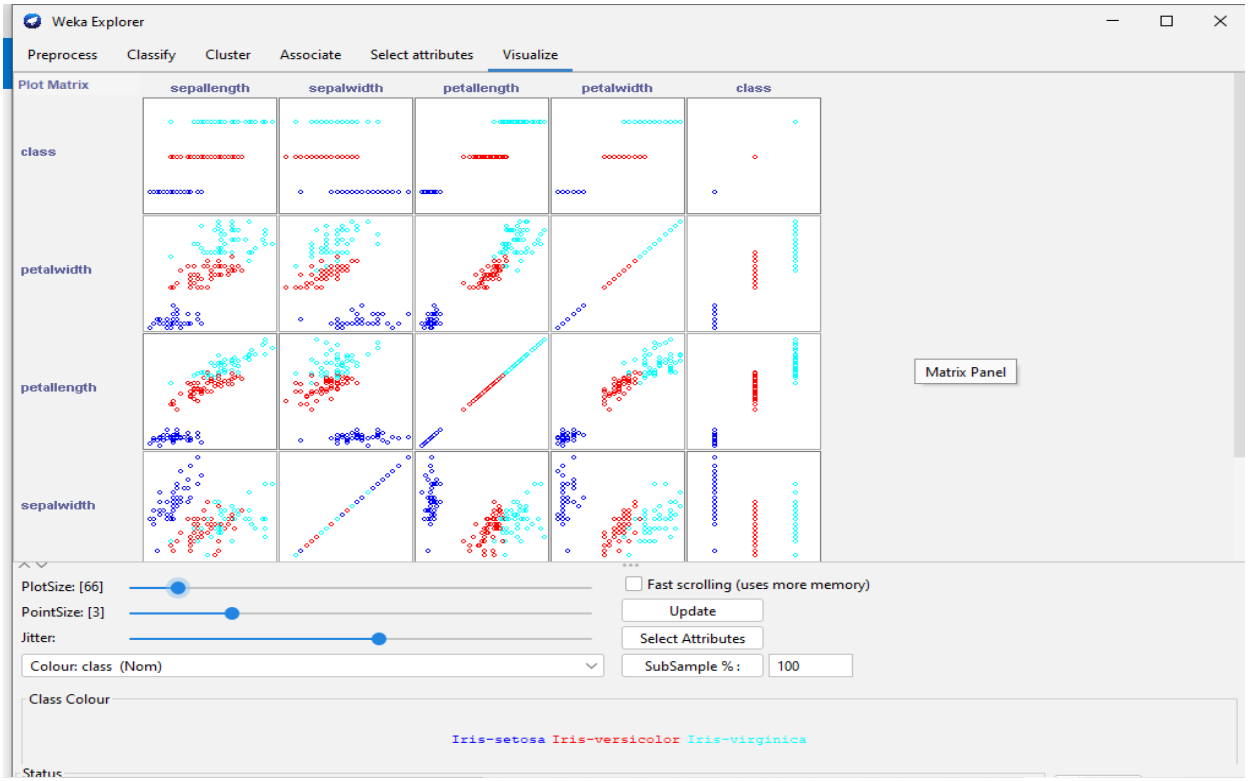


PRN:2020BTECS00207



Data Classification:





2. 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

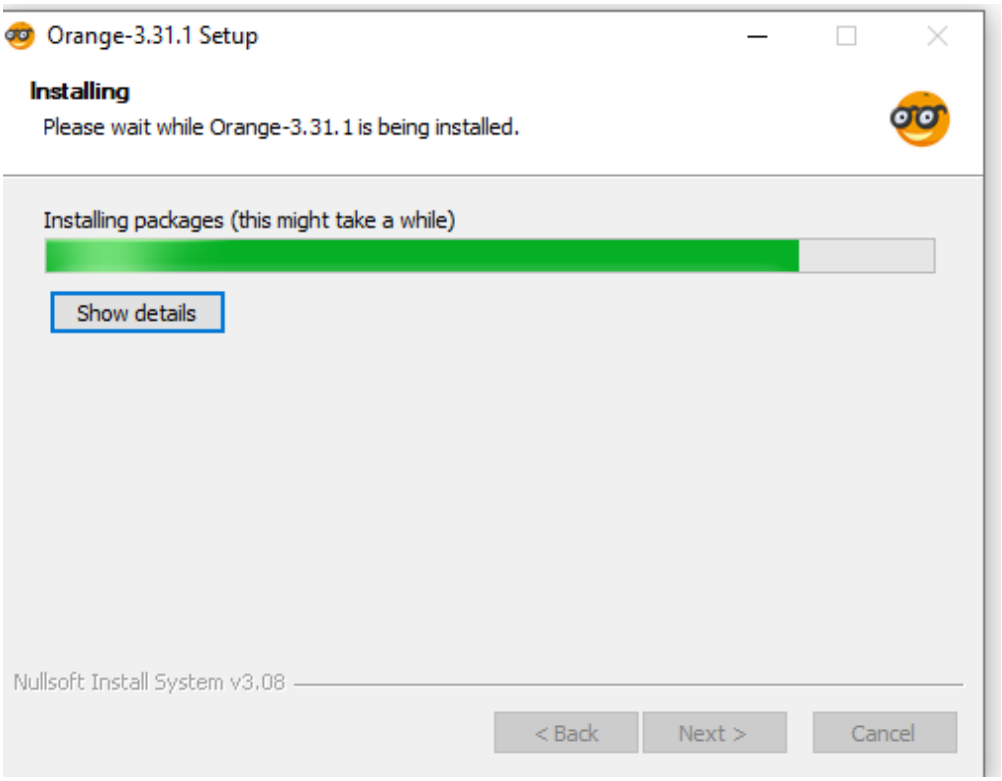
b. Show data distribution

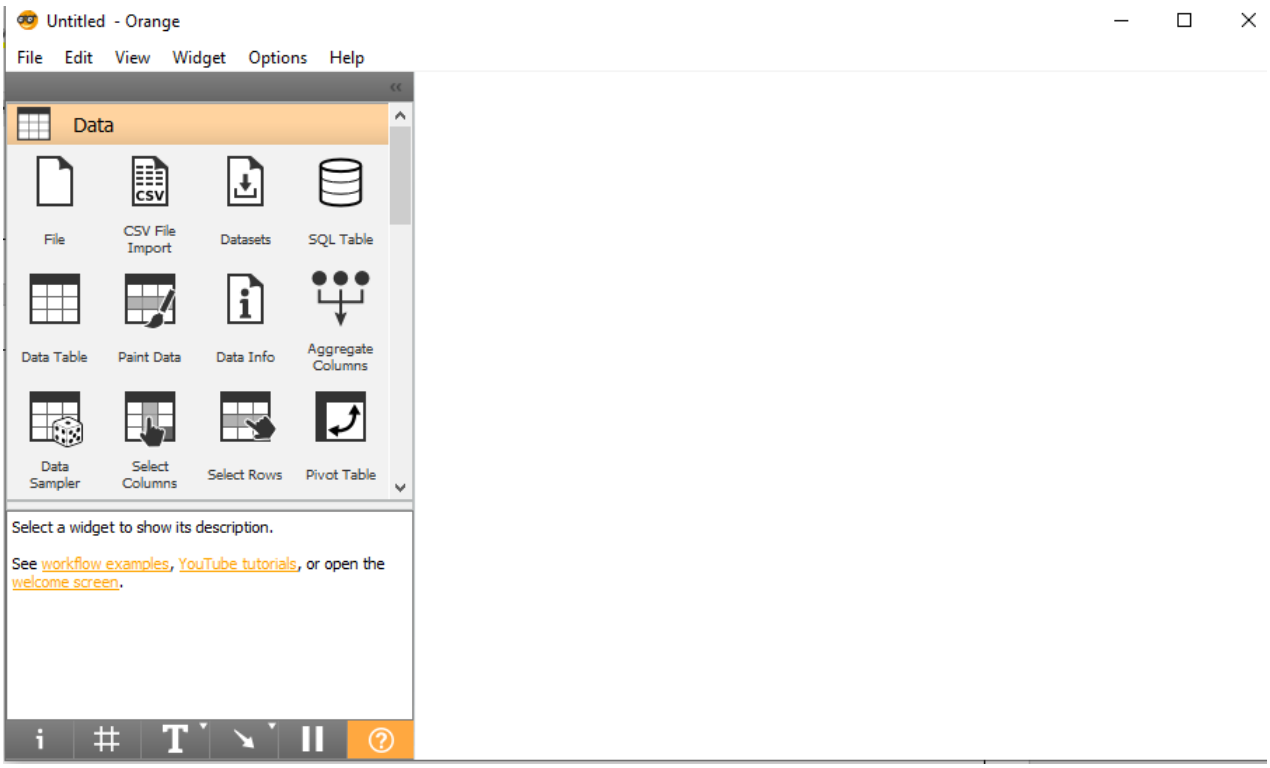
c. Show linear projection

d. Show FreeViz

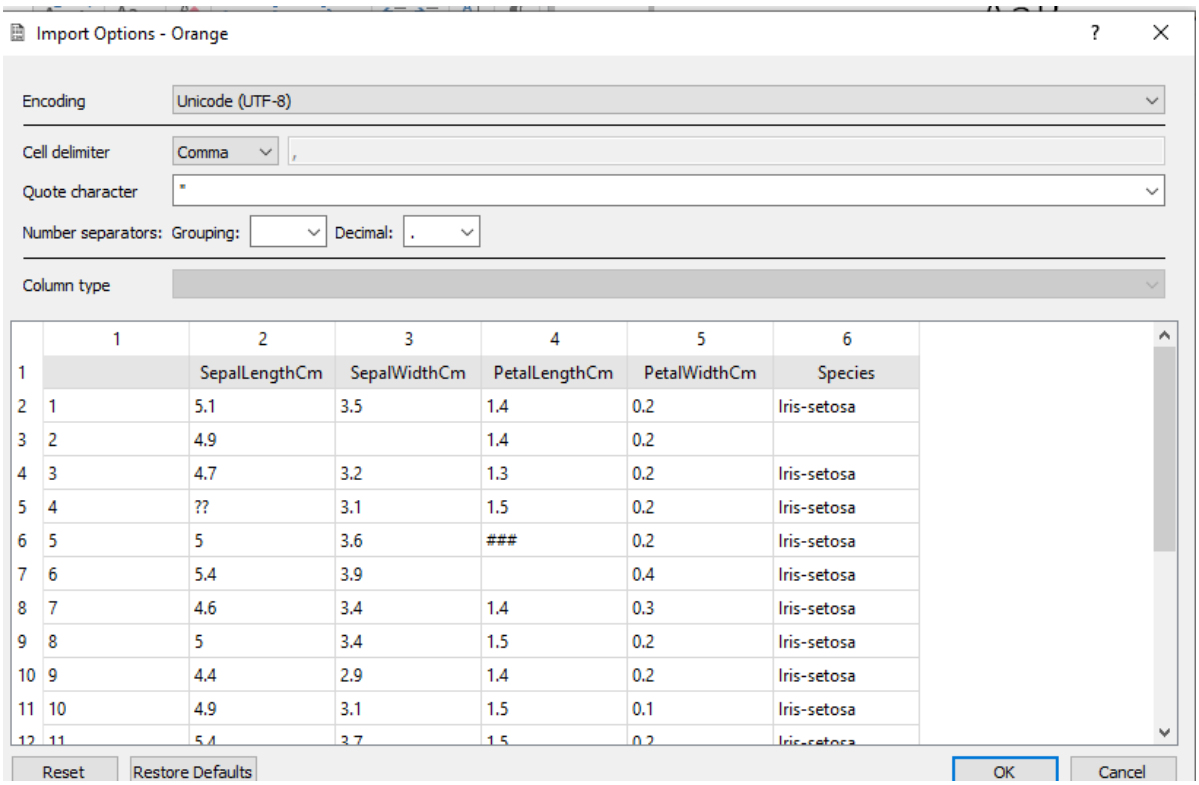
Installation of orange:

PRN:2020BTECS00207

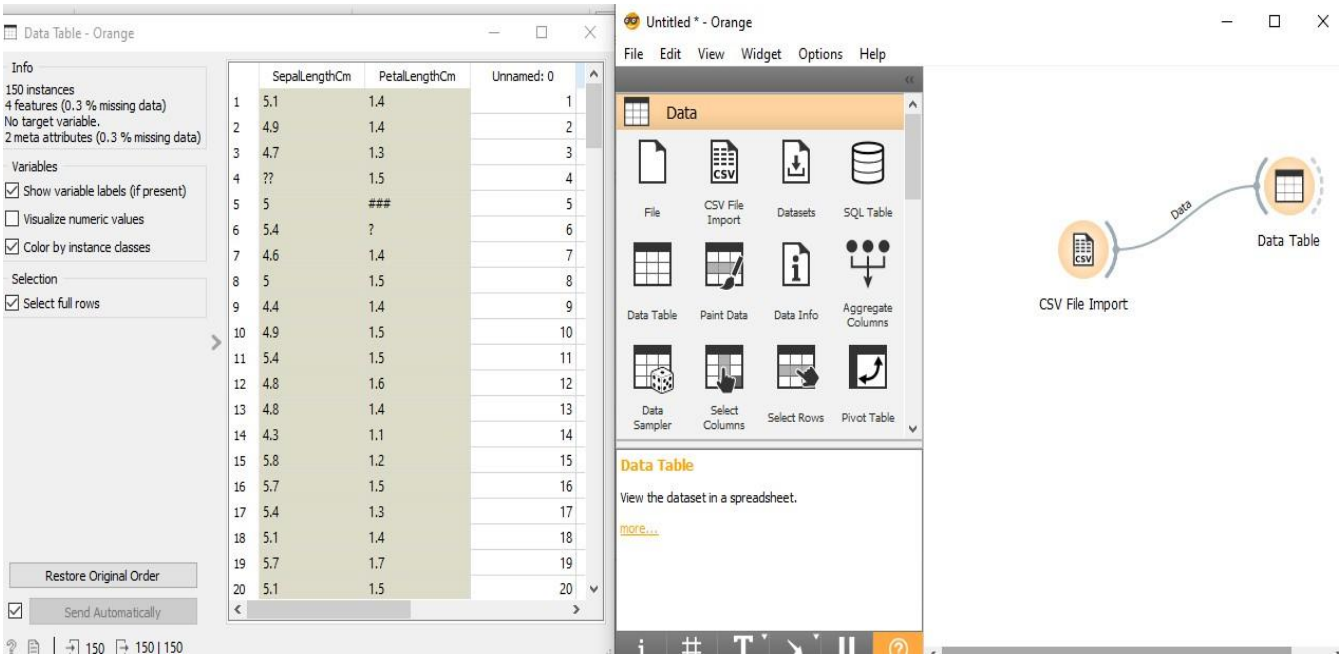




PRN:2020BTECS00207

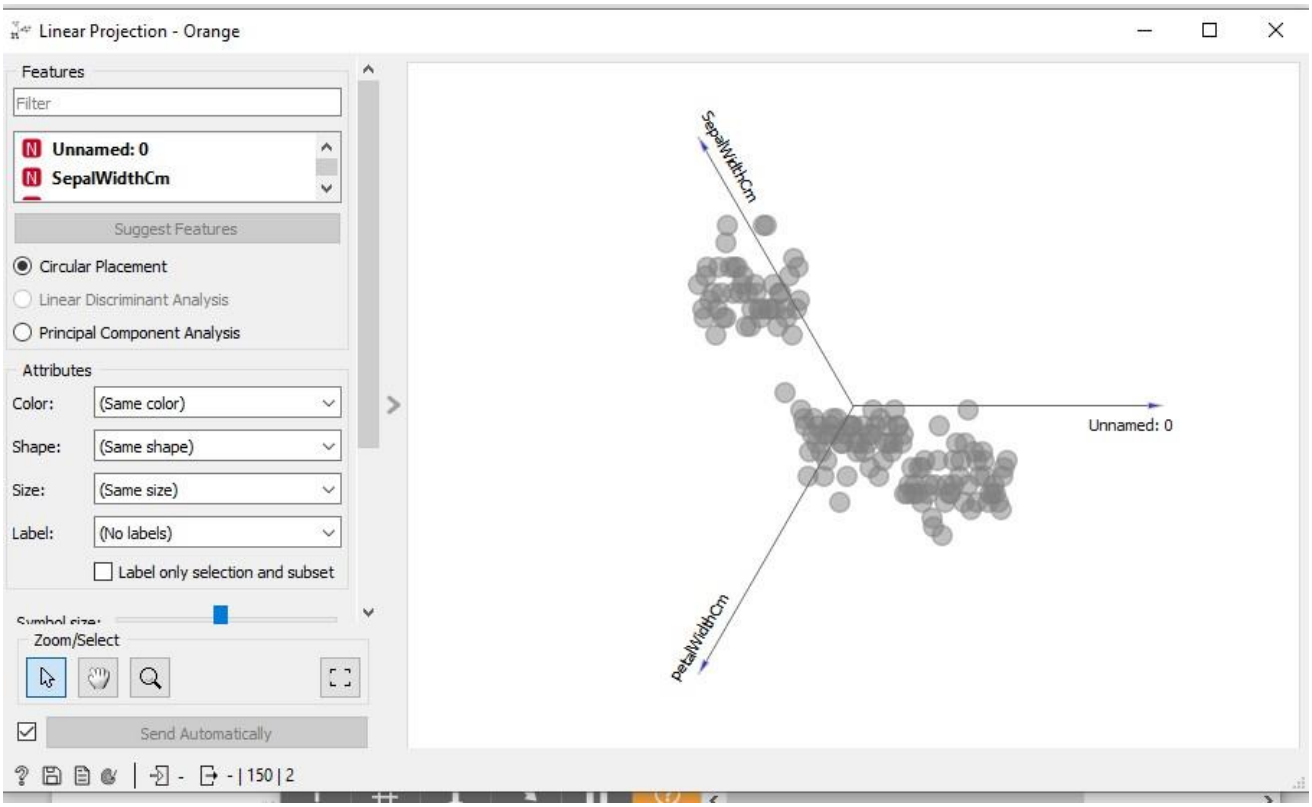


	1	2	3	4	5	6
1		SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
2	1	5.1	3.5	1.4	0.2	Iris-setosa
3	2	4.9		1.4	0.2	
4	3	4.7	3.2	1.3	0.2	Iris-setosa
5	4	??	3.1	1.5	0.2	Iris-setosa
6	5	5	3.6	###	0.2	Iris-setosa
7	6	5.4	3.9		0.4	Iris-setosa
8	7	4.6	3.4	1.4	0.3	Iris-setosa
9	8	5	3.4	1.5	0.2	Iris-setosa
10	9	4.4	2.9	1.4	0.2	Iris-setosa
11	10	4.9	3.1	1.5	0.1	Iris-setosa
12	11	5.4	3.7	1.5	0.2	Iris-setosa

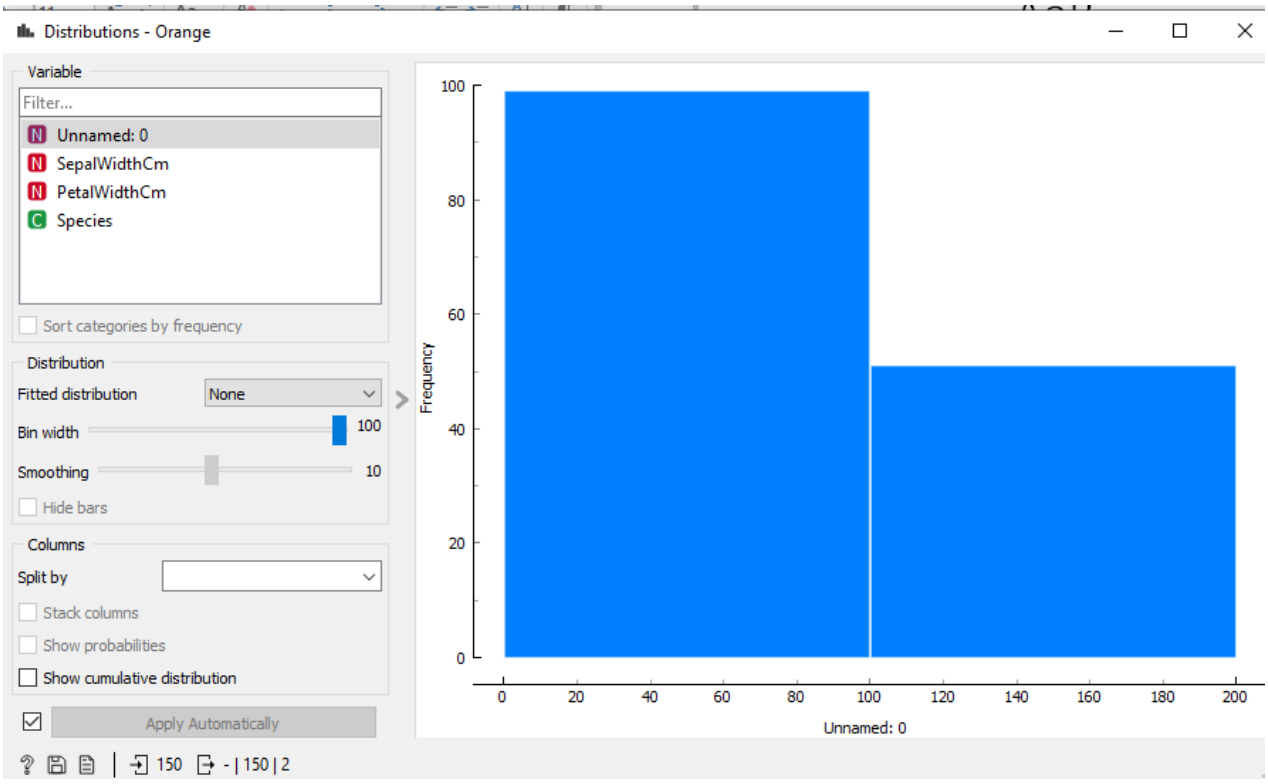


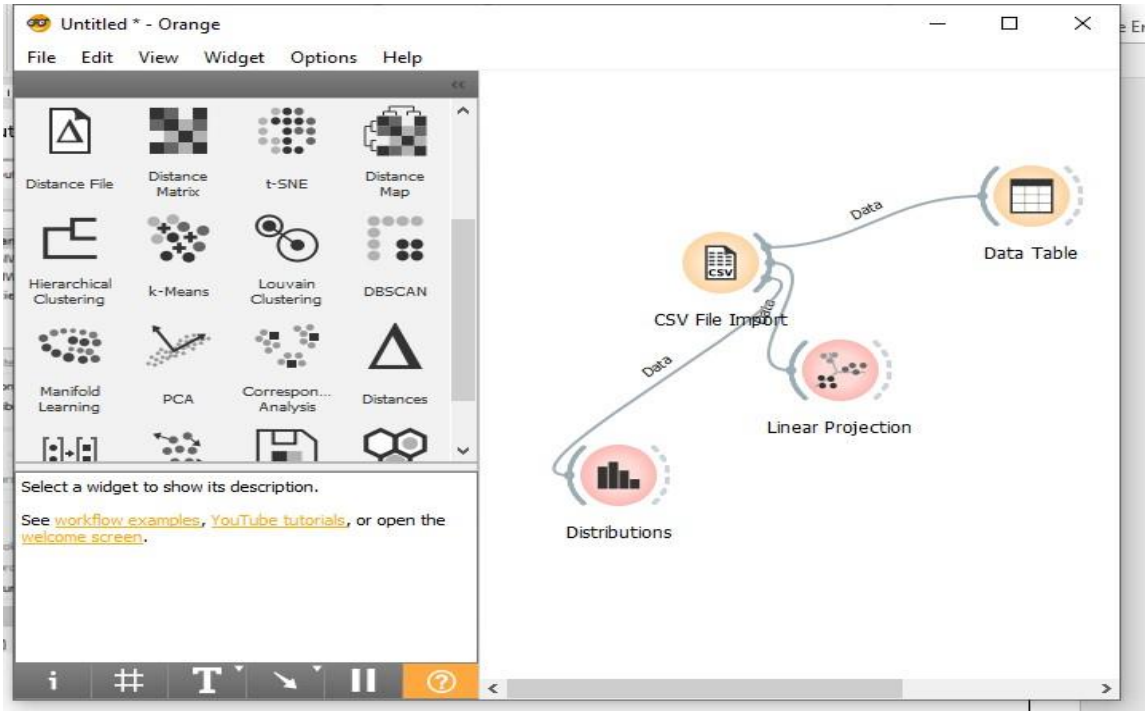
	SepalLengthCm	PetalLengthCm	Unnamed: 0
1	5.1	1.4	1
2	4.9	1.4	2
3	4.7	1.3	3
4	??	1.5	4
5	5	###	5
6	5.4	?	6
7	4.6	1.4	7
8	5	1.5	8
9	4.4	1.4	9
10	4.9	1.5	10
11	5.4	1.5	11
12	4.8	1.6	12
13	4.8	1.4	13
14	4.3	1.1	14
15	5.8	1.2	15
16	5.7	1.5	16
17	5.4	1.3	17
18	5.1	1.4	18
19	5.7	1.7	19
20	5.1	1.5	20

Linear projection:



Distribution:





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

Open Source Software	Proprietary 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.	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 through open collaboration.	Here the software is developed and tested by the individual or organization by which it is owned not by public.
In open source software the source code is public.	In proprietary software the source code is protected
Open source software can be installed into any computer.	Proprietary software can be installed into any computer without valid license.

Users do not need to have any authenticated license to use this software.	Users need to have a valid and authenticated license to use this software.
Users can get open software for free of charge.	Users must have to pay to get the proprietary software.
In open source software faster fixes of bugs and better security is availed due to the community.	In proprietary software the vendor is completely responsible for fixing of malfunctions.

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

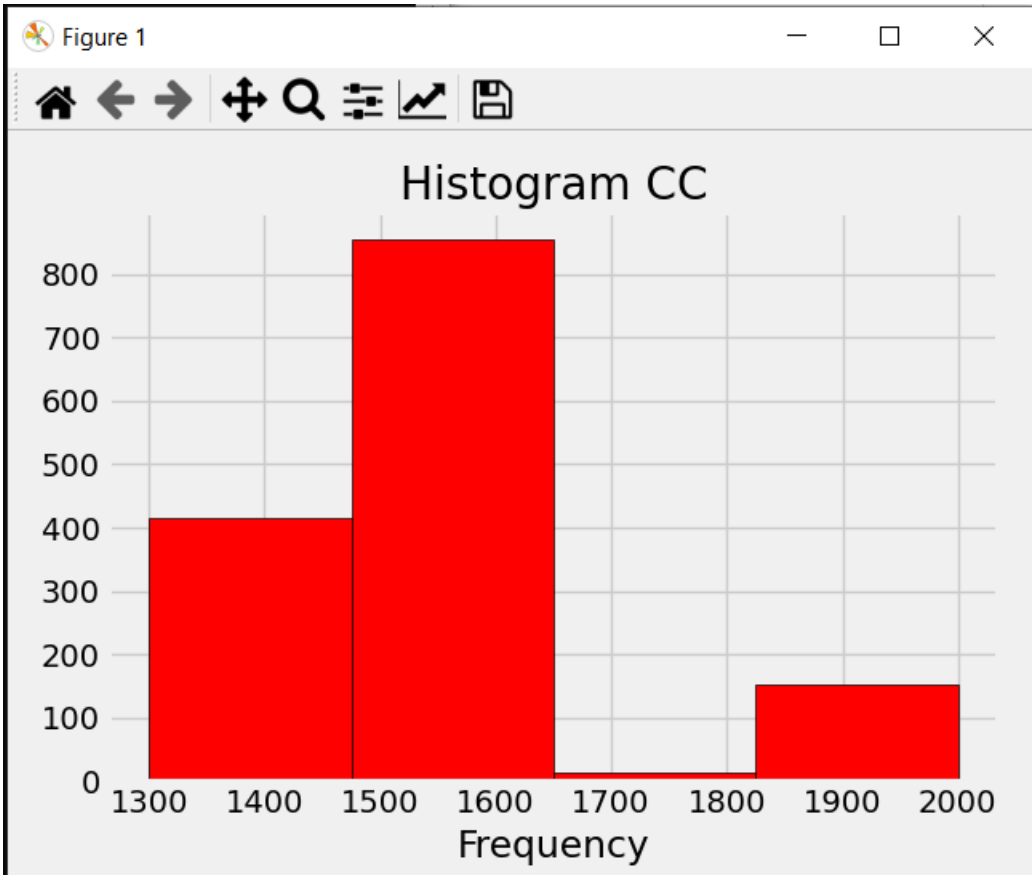
a. Scatter plot- Scatter plot of Price Vs Age

b. Histogram- for Kilometer and CC

c. Bar plot- Bar plot for different fuel types

a. Scatter plot- Scatter plot of Price Vs Age

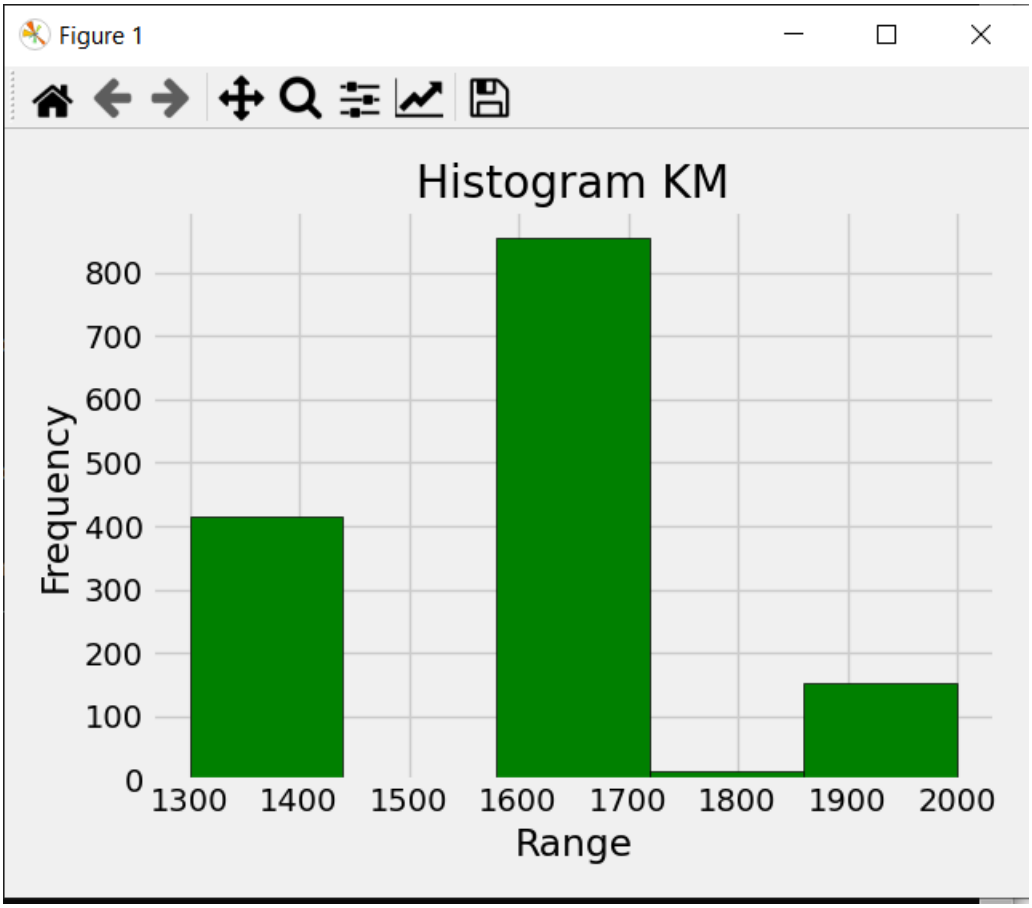
```
>>> import pandas as pd
>>> import numpy as np
>>> from matplotlib import pyplot as plt
>>> plt.style.use('fivethirtyeight')
>>> data=pd.read_csv('Downloads/Toyota.csv')
>>> cc=data['CC']
>>> data.head(1)
   Unnamed: 0  Price   Age     KM FuelType  HP  MetColor  Automatic    CC  Doors  Weight
0           0  13500  23.0  46986   Diesel  90       1.0          0  2000  three    1165
>>> plt.hist(cc,bins=4,edgecolor="black",color="red")
(array([416., 854.,  14., 152.]), array([1300., 1475., 1650., 1825., 2000.]), <BarContainer object of 4 artists>)
>>> plt.title("Histogram CC")
Text(0.5, 1.0, 'Histogram CC')
>>> plt.xlabel("Frequency")
Text(0.5, 0, 'Frequency')
>>> plt.tight_layout()
>>> plt.show()
```



b. Histogram- for Kilometer and CC

```
>>> 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()
```

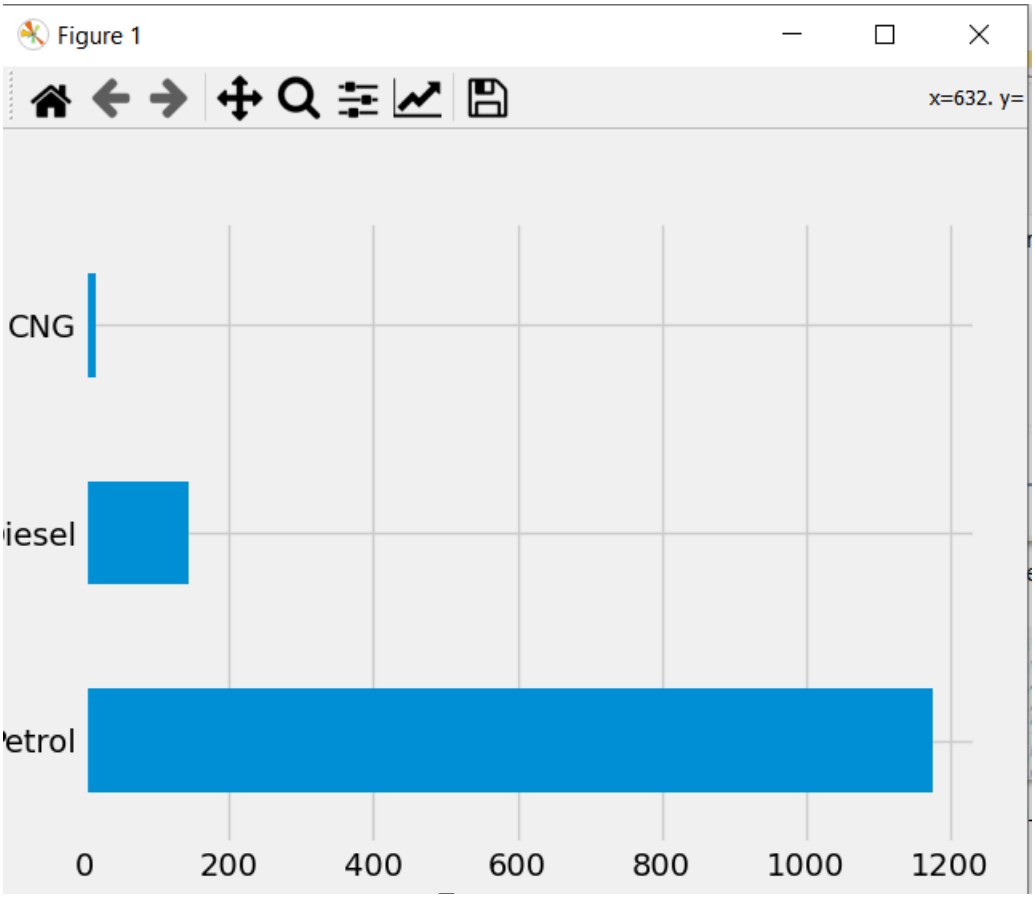
PRN:2020BTECS00207




```
>>> 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()
```

c. Bar plot- Bar plot for different fuel type

```
>>> 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()
```



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

FOSS:

Examples are Android, Linux, Firefox, Open Office, GIMP, VLC Media player etc.

Proprietary Software:

Examples are Windows, MacOS, Internet Explorer, Google earth, Microsoft Office, Adobe Flash Player, Skype etc.