

Software Engineering Tools Lab

Assignment No-1

(Module 1- Introduction to OSS)

Name: Anchit Gupta

Prn No.2019BTECS00098

Batch:T8

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.

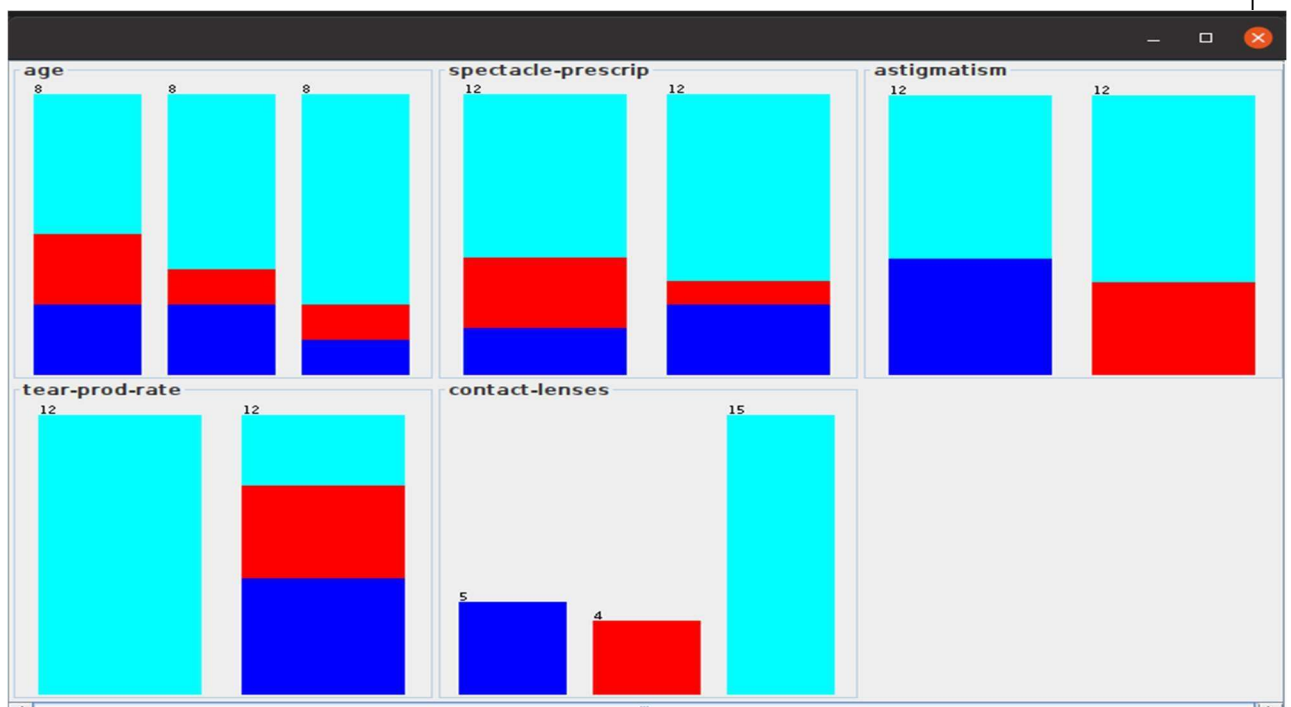
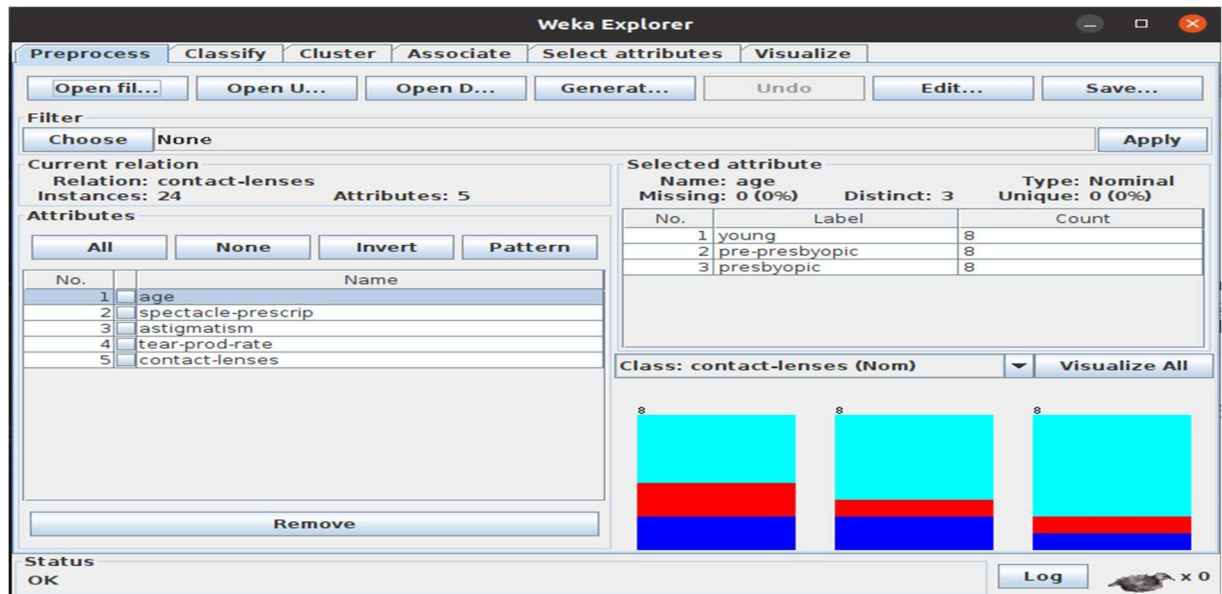
Using Weka tool perform below tasks such as data preprocessing, data classification (use any appropriate ML algorithm) and data visualization efficiently on given dataset.

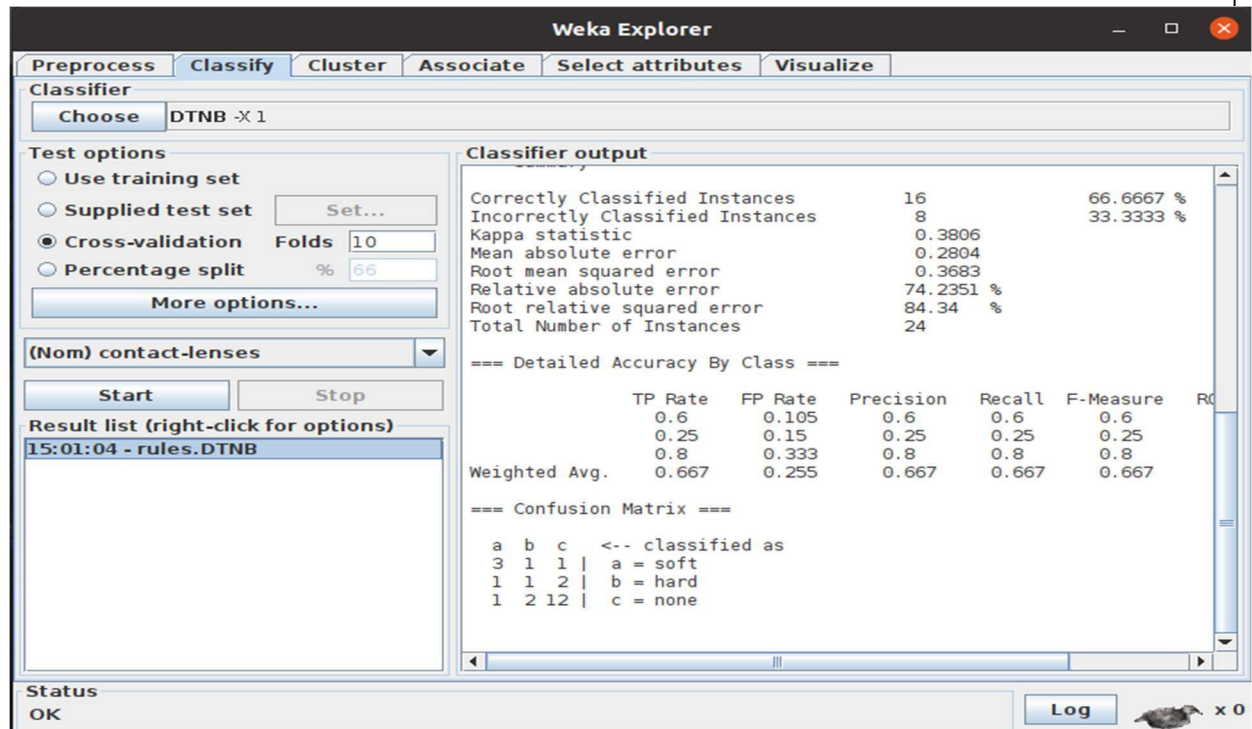
Use the Iris dataset given:

<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)

Used Dataset: form weka proived





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

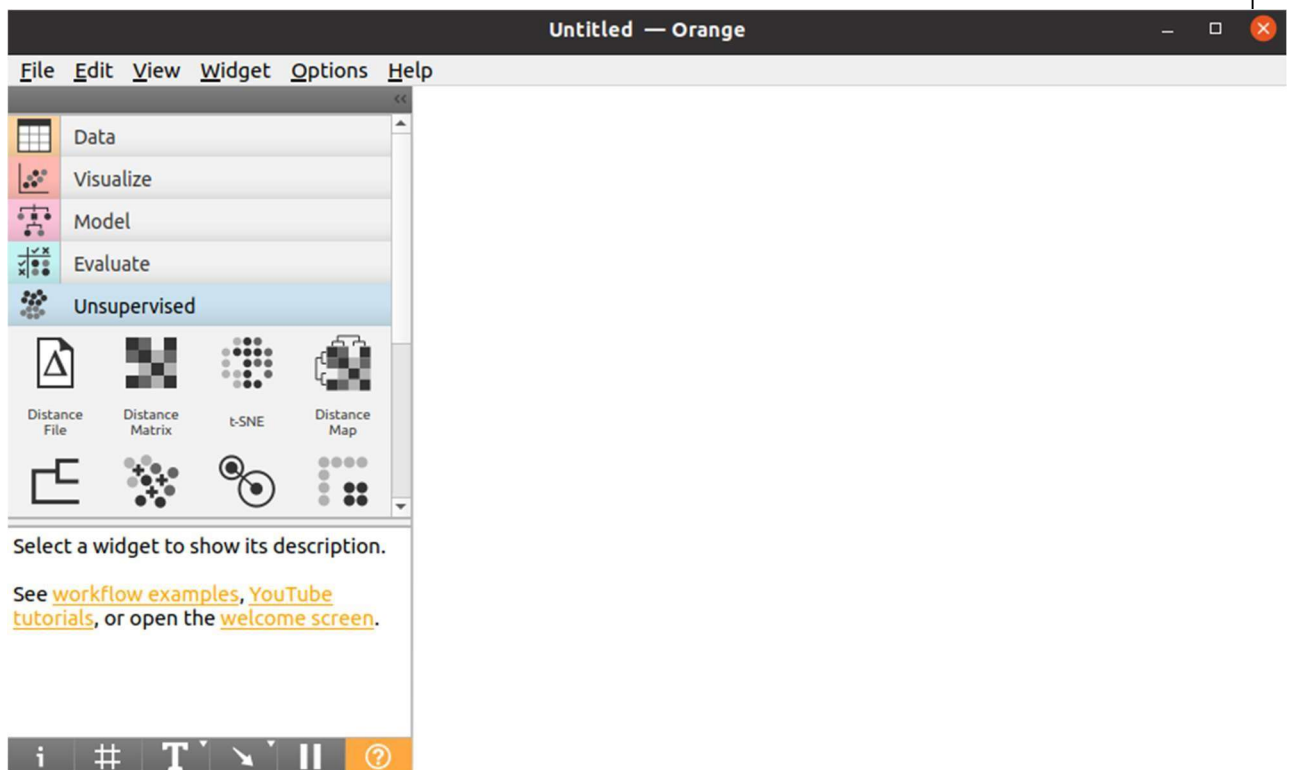
b. Show data distribution

c. Show linear projection

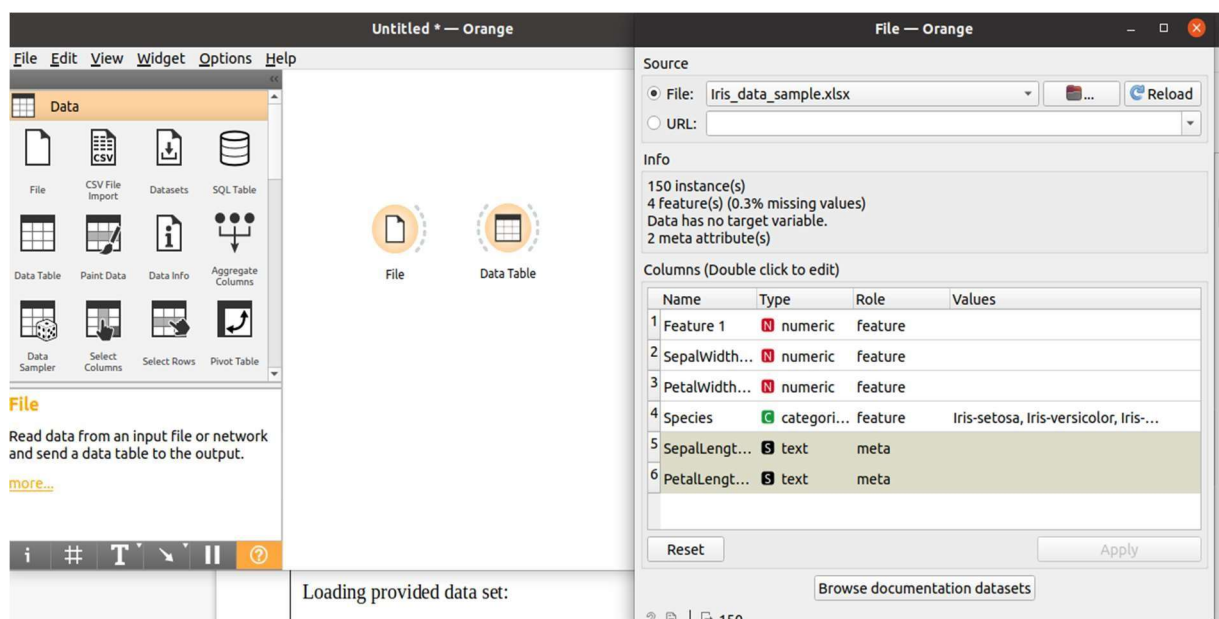
d. Show FreeViz Use dataset:

<https://drive.google.com/file/d/1m6sKl1Dap0XK6Bw1edUd5PohwpPwXnd9/view> Create a report for this task and upload screenshots for the same.

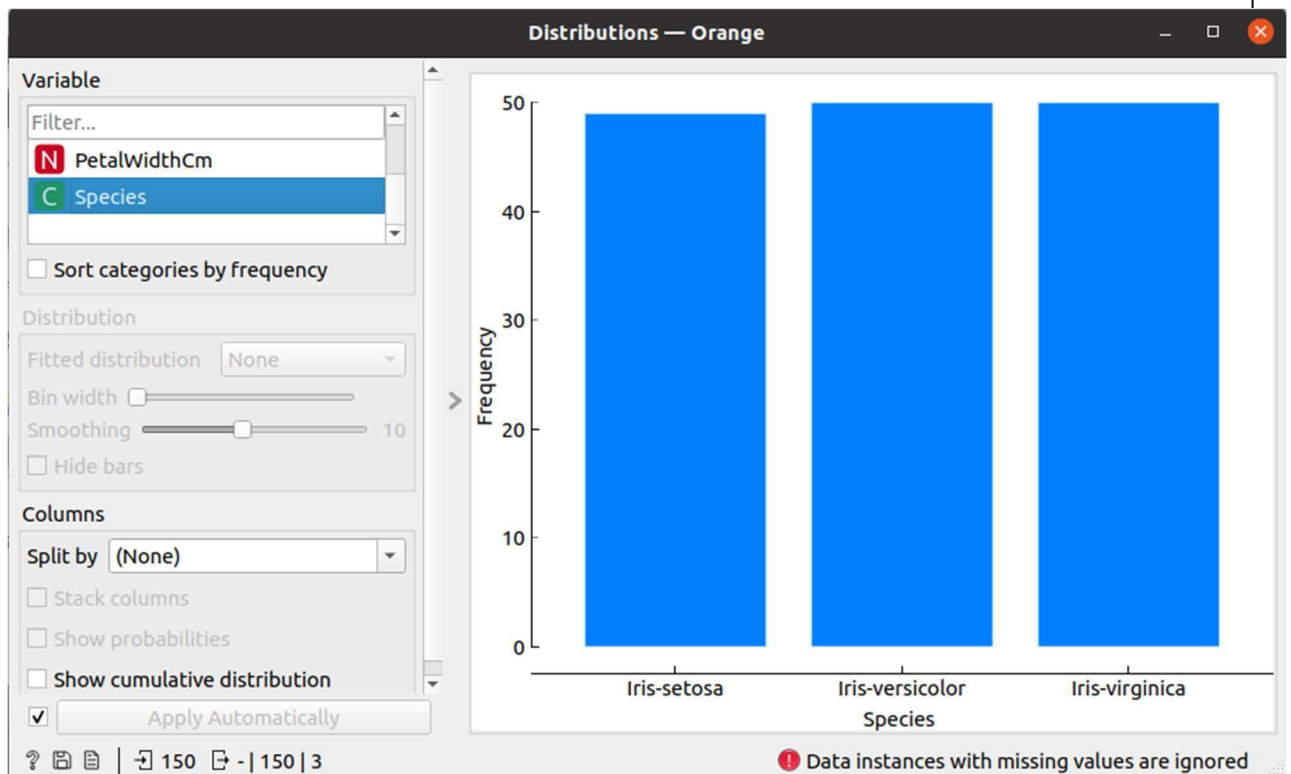
Installation of orange :

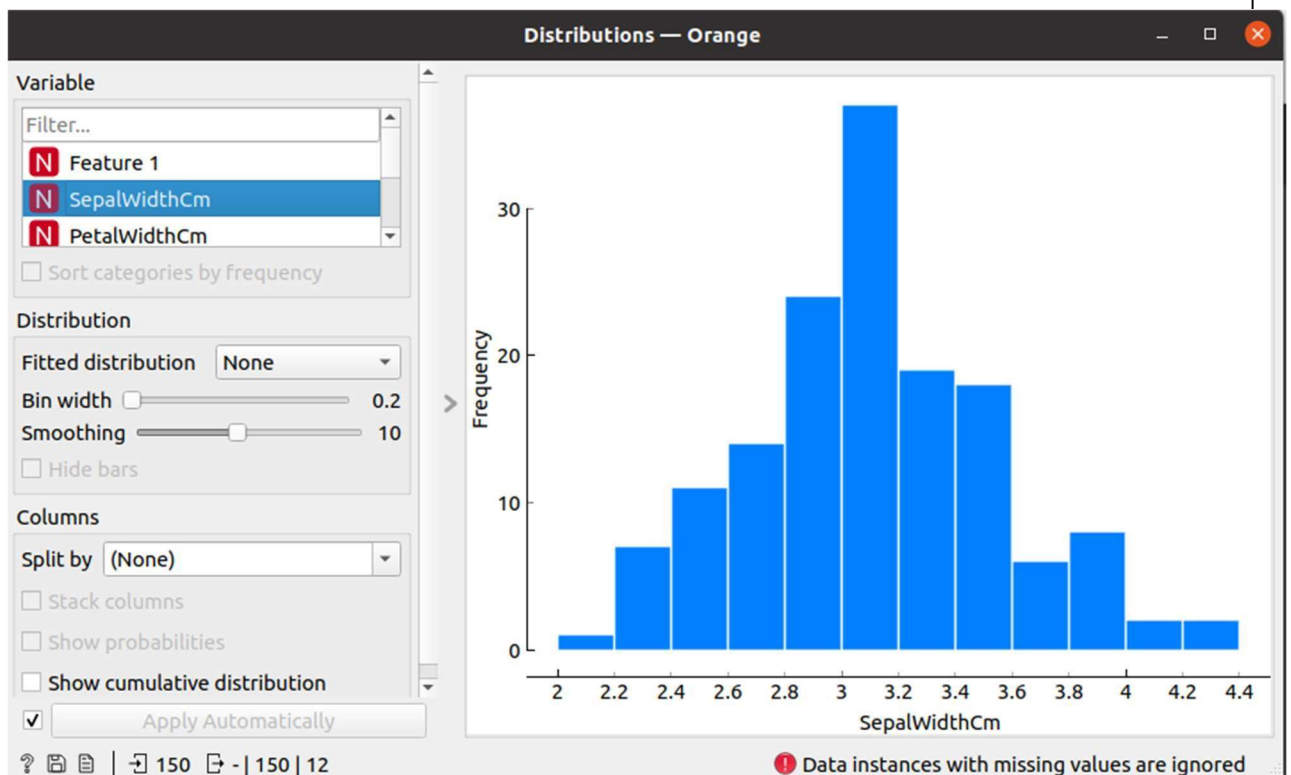
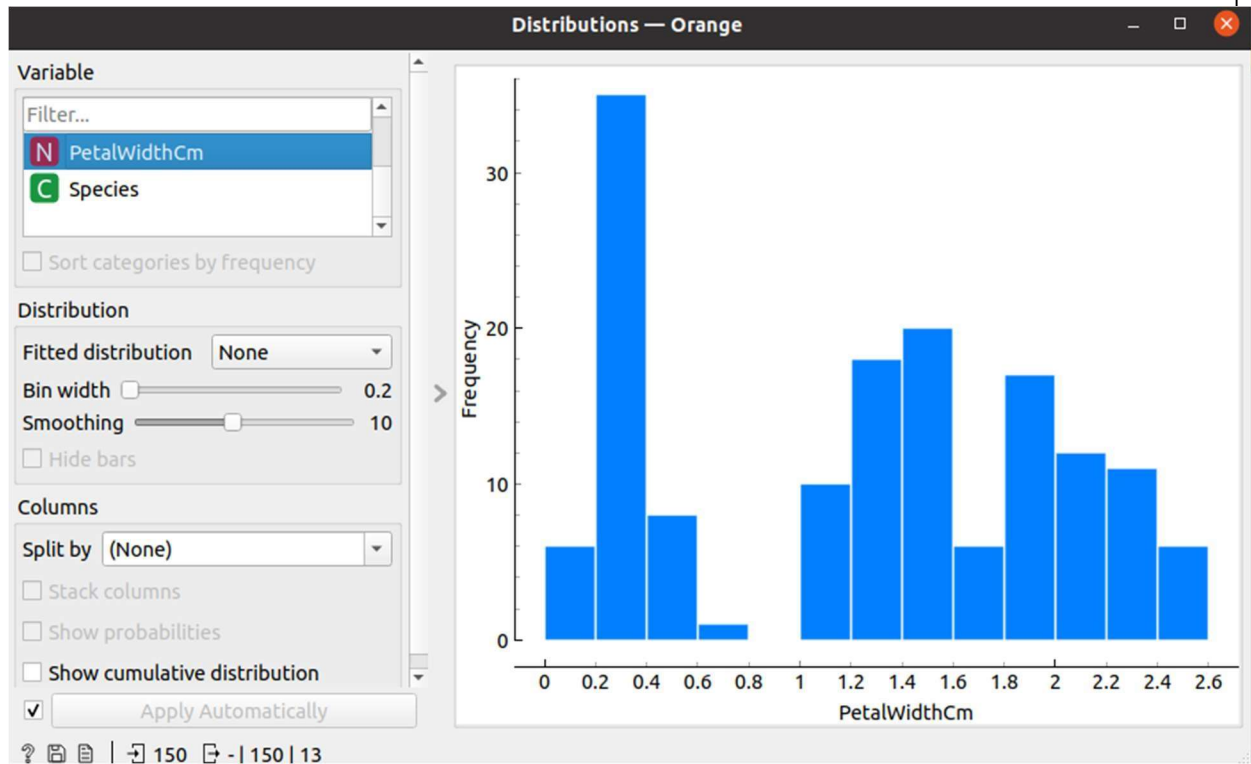


Loading provided data set:

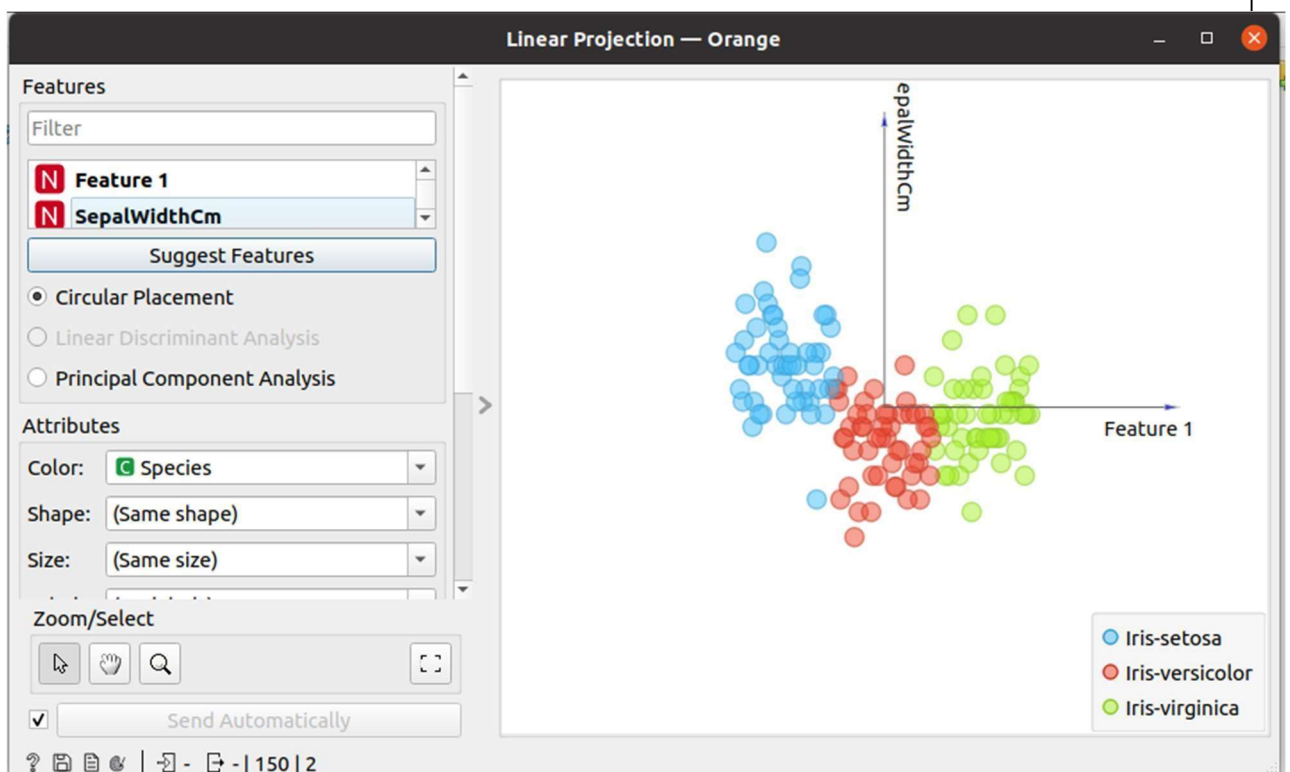
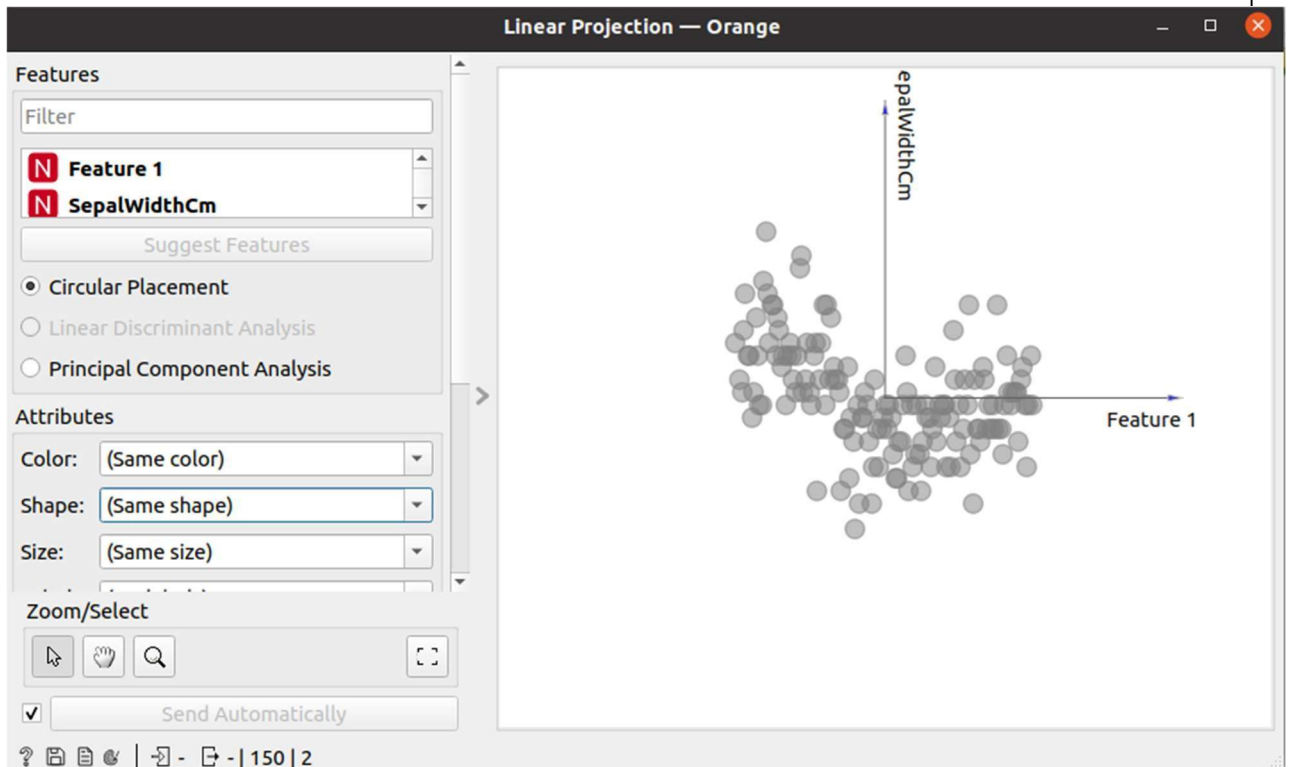


Data Distribution

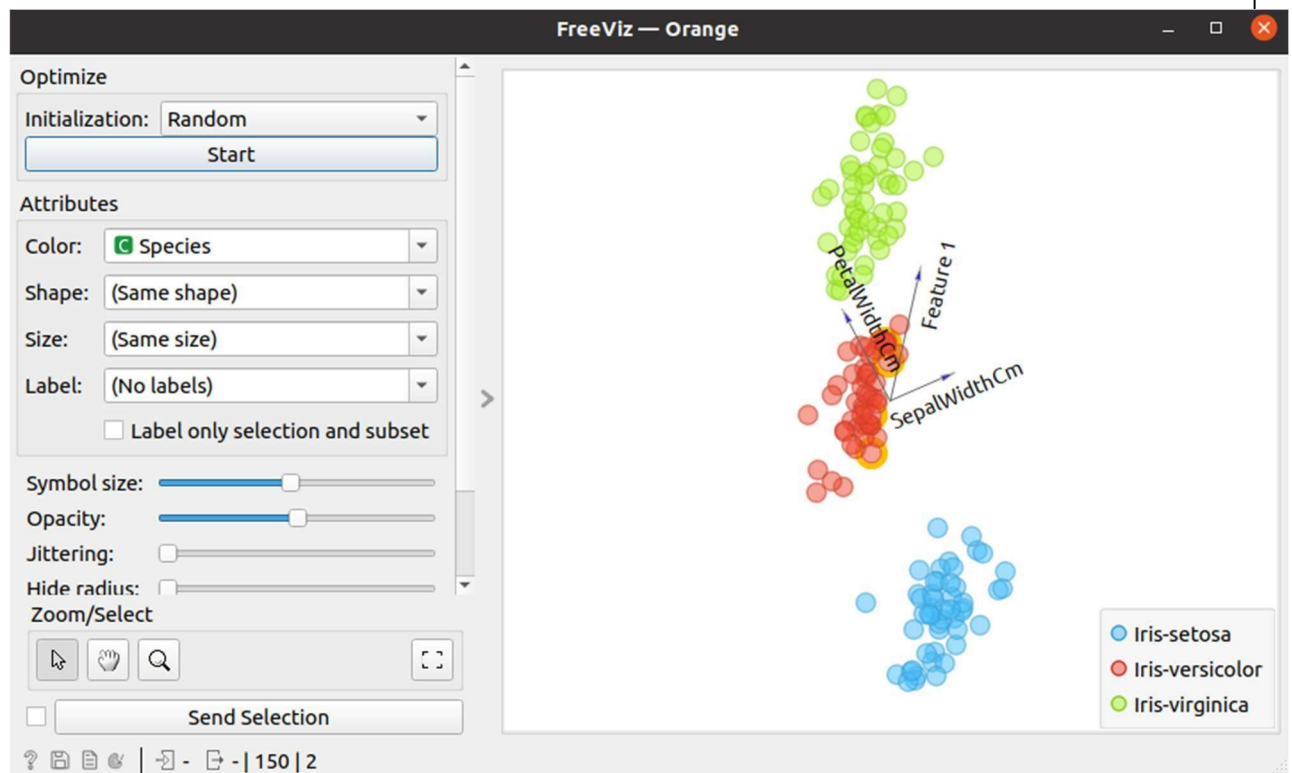
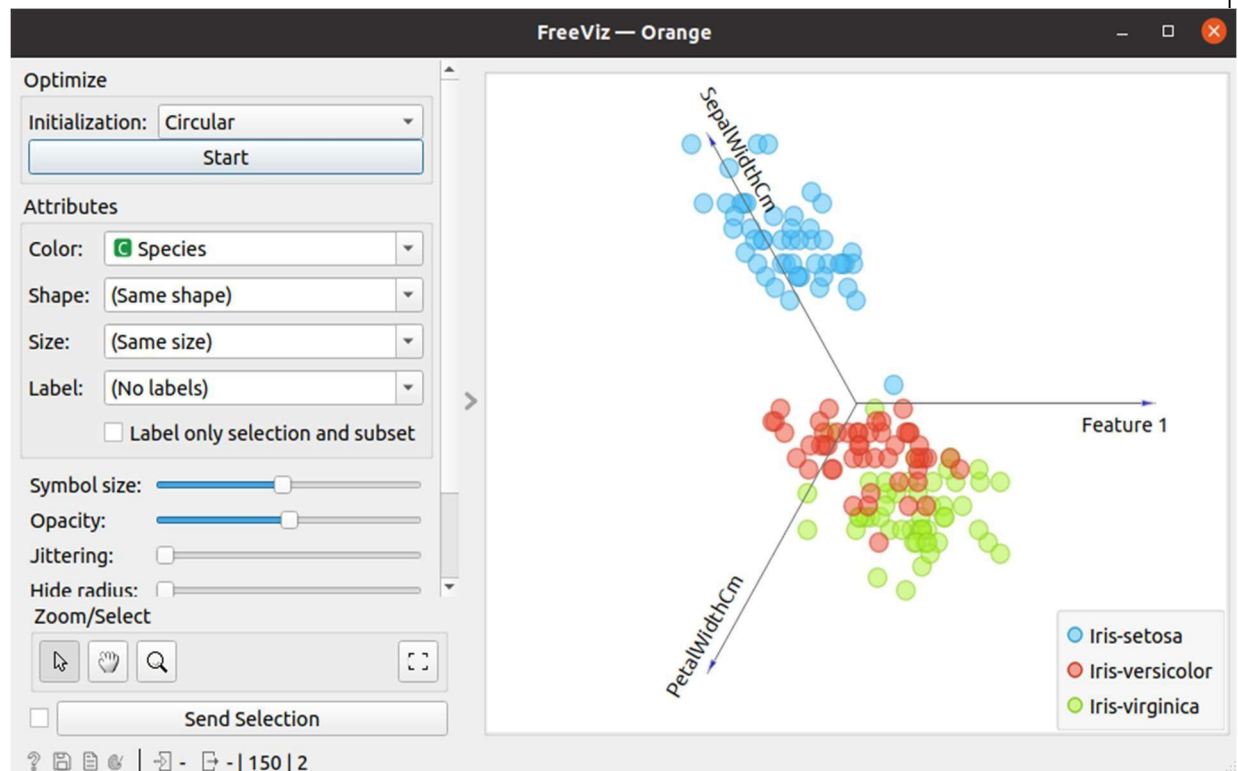




Linear Projection:



Freeviz



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

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.

Open Source Software :

Open Source Software is something which you can modify as per your needs, share with others without any licensing violation burden. When we say Open Source, source code of software is available publicly with Open Source licenses.

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.

4. Using Anaconda Python create Histogram, Scatter plot and Bar plot for the dataset given below.

Dataset-

https://drive.google.com/file/d/1i11BZFe8Xj9kNq7eeE9KOa_lz1KhEdXJ/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

```
In [10]: details=pd.read_csv("/home/hp/Downloads/Toyota.csv")
print(details)
```

	Unnamed: 0	Price	Age	KM	FuelType	HP	MetColor	Automatic	CC \
0	0	13500	23.0	46986	Diesel	90	1.0	0	2000
1	1	13750	23.0	72937	Diesel	90	1.0	0	2000
2	2	13950	24.0	41711	Diesel	90	NaN	0	2000
3	3	14950	26.0	48000	Diesel	90	0.0	0	2000
4	4	13750	30.0	38500	Diesel	90	0.0	0	2000
...
1431	1431	7500	NaN	20544	Petrol	86	1.0	0	1300
1432	1432	10845	72.0	??	Petrol	86	0.0	0	1300
1433	1433	8500	NaN	17016	Petrol	86	0.0	0	1300
1434	1434	7250	70.0	??	NaN	86	1.0	0	1300
1435	1435	6950	76.0	1	Petrol	110	0.0	0	1600

	Doors	Weight
0	three	1165
1	3	1165
2	3	1165
3	3	1165
4	3	1170
...
1431	3	1025
1432	3	1015
1433	3	1015
1434	3	1015
1435	5	1114

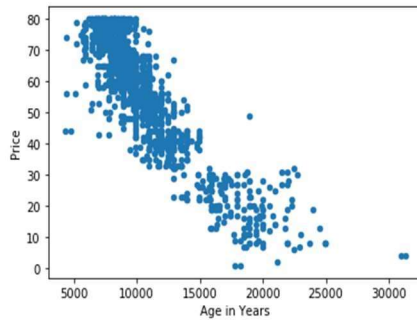
[1436 rows x 11 columns]

a.Scatter plot- Scatter plot of Price Vs Age

```
In [17]: price=[]
```

```
In [20]: details.plot(kind='scatter',x='Price',y='Age')  
plt.xlabel('Age in Years')  
plt.ylabel('Price')
```

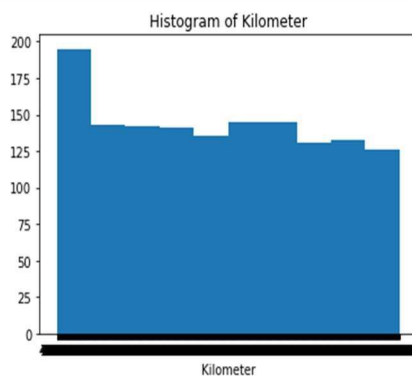
```
Out[20]: Text(0, 0.5, 'Price')
```



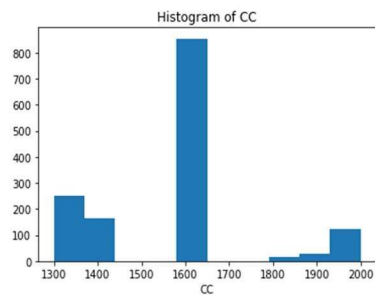
```
In [ ]:
```

b.Histogram- for Kilometer and CC

```
In [40]: plt.hist(details['KM'],)  
plt.title('Histogram of Kilometer')  
plt.xlabel('Kilometer')  
plt.show()
```



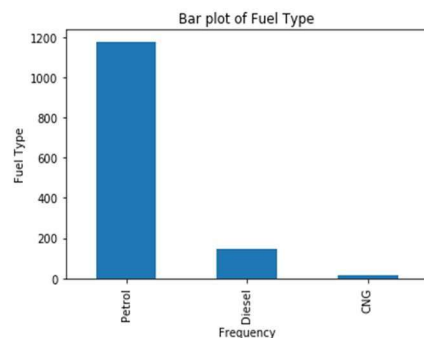
```
In [42]: plt.hist(details['CC'])  
plt.title('Histogram of CC')  
plt.xlabel('CC')  
plt.show()
```



c. Bar plot- Bar plot for different fuel types

```
In [85]: fuel_count = pd.value_counts(details['FuelType'].values, sort = True)
plt.xlabel('Frequency')
plt.ylabel('Fuel Type')
plt.title('Bar plot of Fuel Type')
fuel_count.plot.bar()
```

Out[85]: <matplotlib.axes._subplots.AxesSubplot at 0x7fcb94da6a30>



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

1. Open Source (Database) :

An open source database is a database that anyone can easily view the source code and this is open and free to download. Also for community version some small additional and affordable cost are imposed. Open Source Database provide Limited technical support to end users. Here Installation and updates are administered by user. For examples: MYSQL, PostgreSQL, MongoDB etc.

2. Proprietary (Database) :

Commercial database are that which has been created for Commercial Purpose only.

They are premium and are not free like Open Source Database. In Commercial Database it is guaranteed that technical support is provided. In this Installation and updates are Administrated by software Vendor. For examples: Oracle, IBM DB2 etc.

