PRN: 2019BTECS00105 Batch: T7 Vaishnavi Mahajan

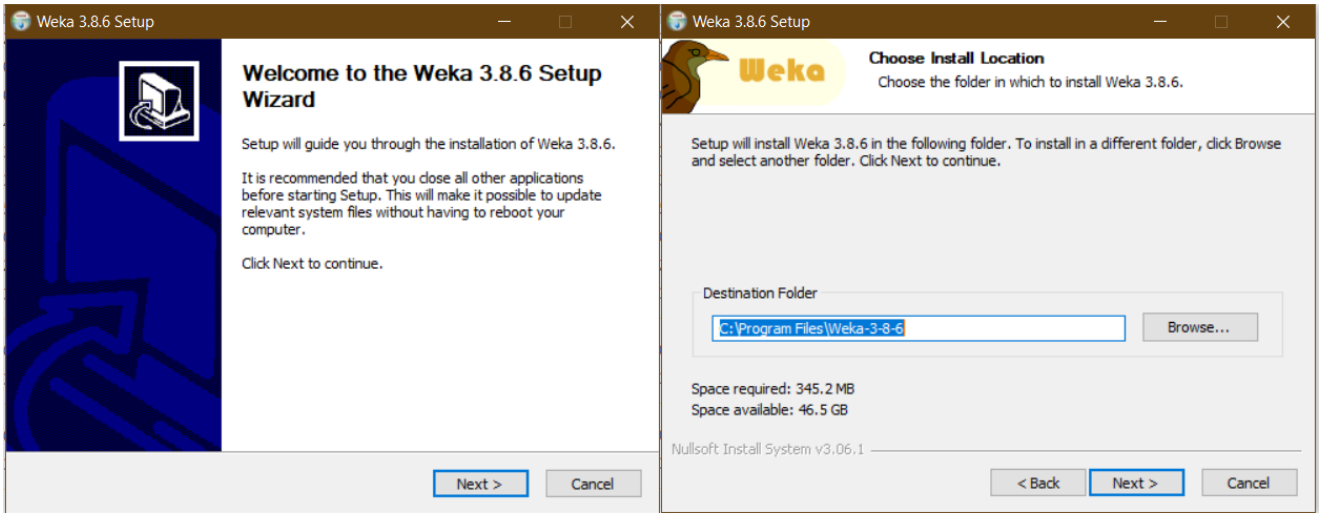
Software Engineering Tools Lab Assignment No-1

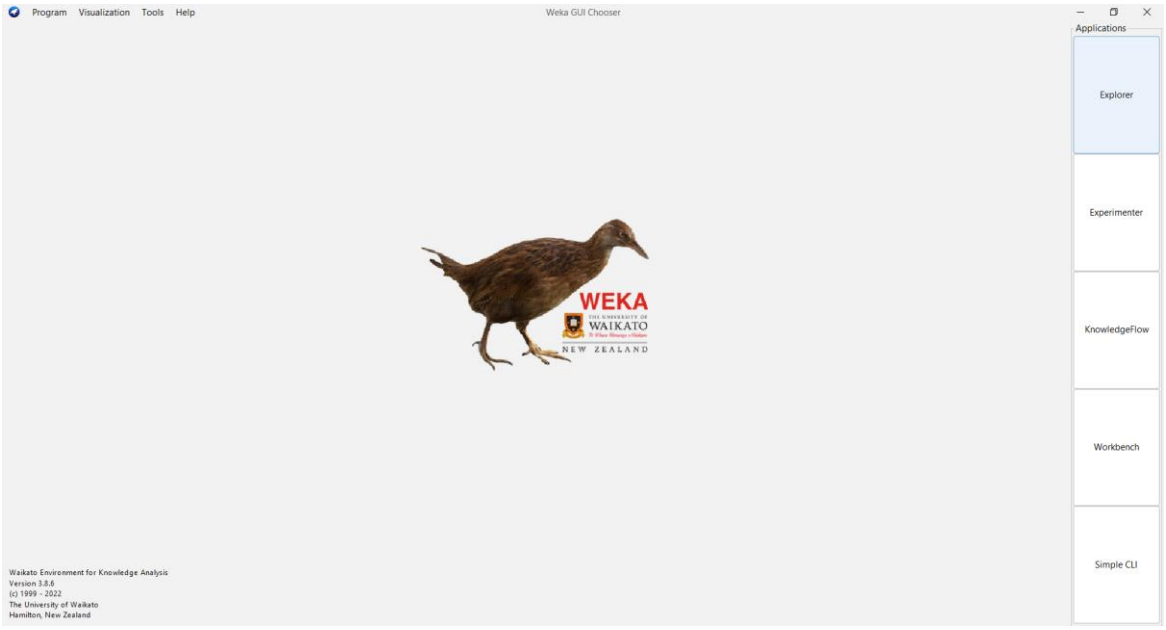
(Module 1- Introduction to OSS)

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)

Installation

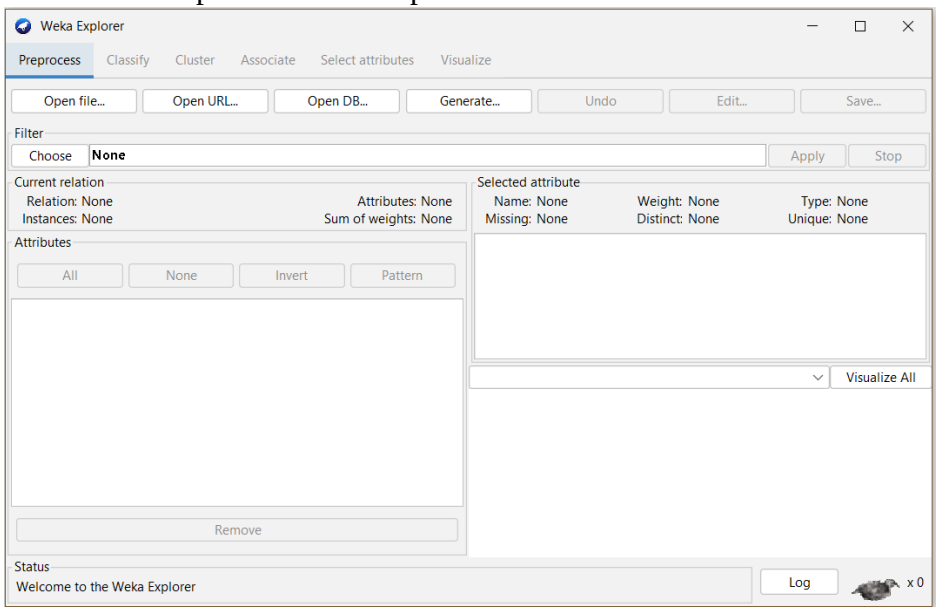




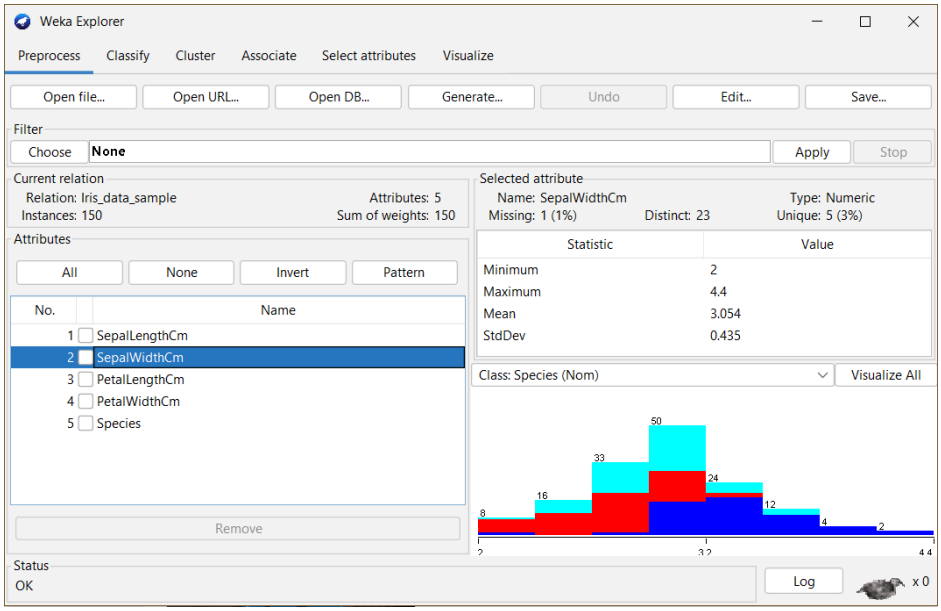
PRN: 2019BTECS00105 Batch: T7 Vaishnavi Mahajan

Preprocessing

1. Click on explorer.
2. Click on Preprocess tab in explorer window.



3. Choose the file

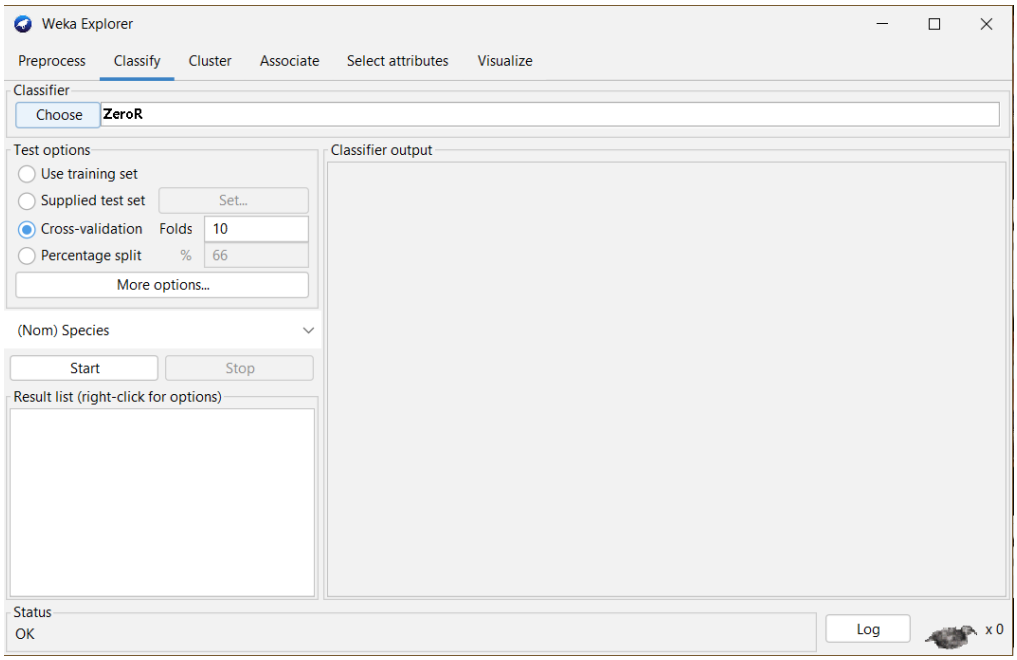


4. Various information about the data and it's attributes is shown.
5. I removed string attributes.
6. Click on save to save the file with .arff type.

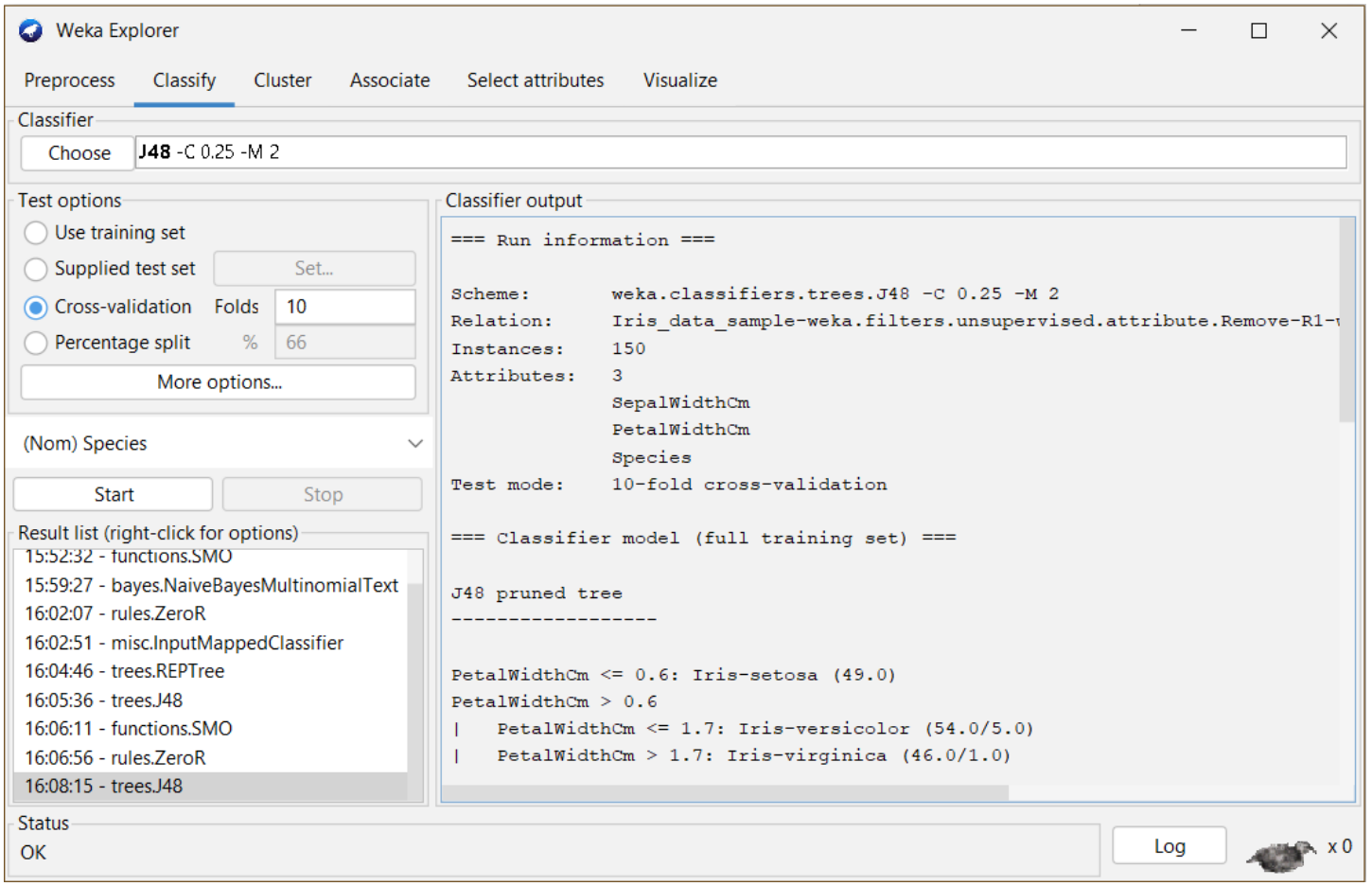
Data classification:

1. Choose the .arff type dataset file.
2. Click on Classify tab.

PRN: 2019BTECS00105 Batch: T7 Vaishnavi Mahajan



3. I selected cross validation with 10 folds for defining training data because the dataset is not very big.
4. Select species for output class.
5. Select J48 as classifier.
6. Click on start.



Result

=== Run information ===

PRN: 2019BTECS00105 Batch: T7 Vaishnavi Mahajan

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: Iris_data_sample-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R2
Instances: 150
Attributes: 3
SepalWidthCm
PetalWidthCm
Species
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

PetalWidthCm <= 0.6: Iris-setosa (49.0)
PetalWidthCm > 0.6
| PetalWidthCm <= 1.7: Iris-versicolor (54.0/5.0)
| PetalWidthCm > 1.7: Iris-virginica (46.0/1.0)

Number of Leaves : 3

Size of the tree : 5

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	141	94.6309 %
Incorrectly Classified Instances	8	5.3691 %
Kappa statistic	0.9195	
Mean absolute error	0.0578	
Root mean squared error	0.1833	
Relative absolute error	13.0031 %	
Root relative squared error	38.8753 %	
Total Number of Instances	149	
Ignored Class Unknown Instances	1	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.980	0.000	1.000	0.980	0.990	0.985	0.985	0.966	Iris-setosa
0.960	0.061	0.889	0.960	0.923	0.883	0.937	0.832	Iris-versicolor

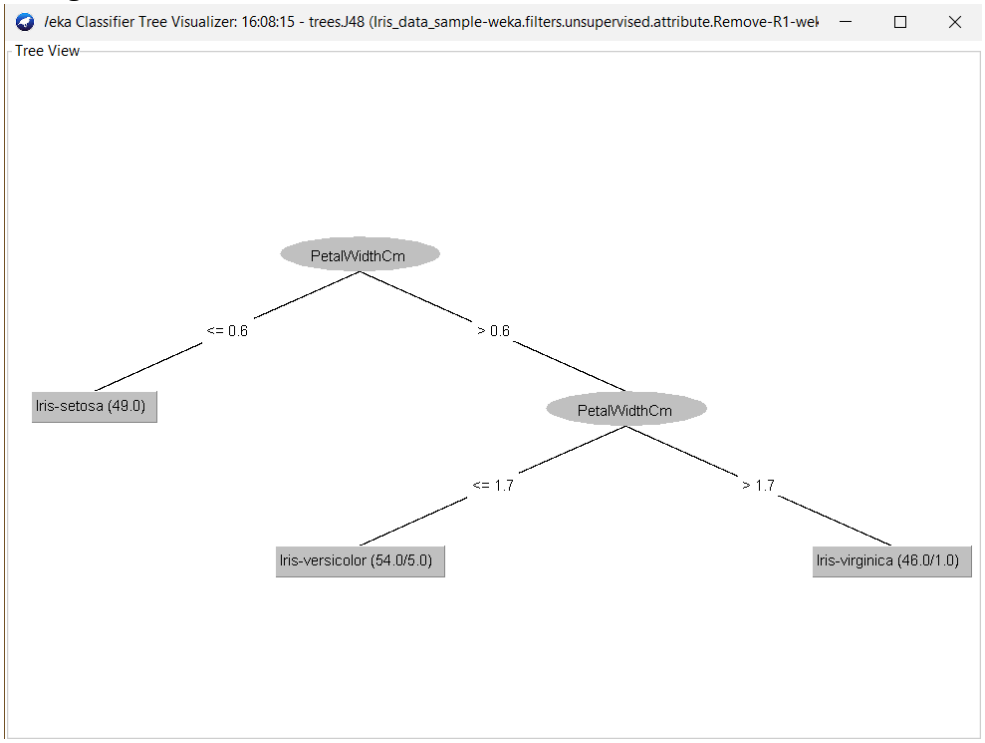
	0.900	0.020	0.957	0.900	0.928	0.894	0.948	0.887	Iris-virginica
Weighted Avg.	0.946	0.027	0.948	0.946	0.947	0.920	0.956	0.894	

=== Confusion Matrix ===

```
 a  b  c   <-- classified as
48  1  0 |  a = Iris-setosa
 0 48  2 |  b = Iris-versicolor
 0  5 45 |  c = Iris-virginica
```

Data Visualization:

1. Right click on result list and select visualize tree.

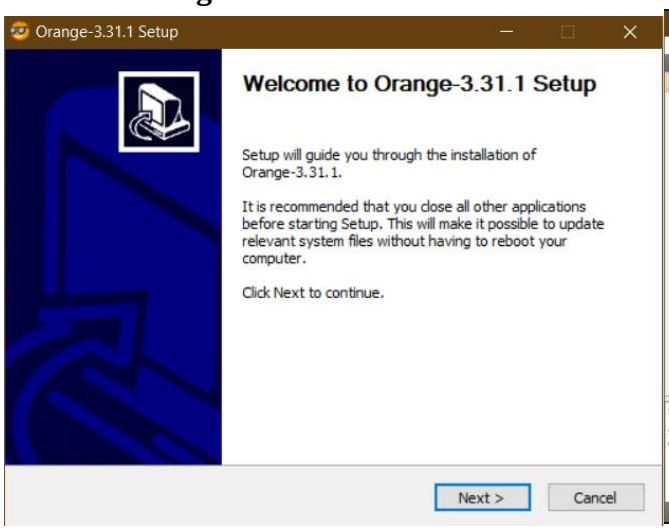


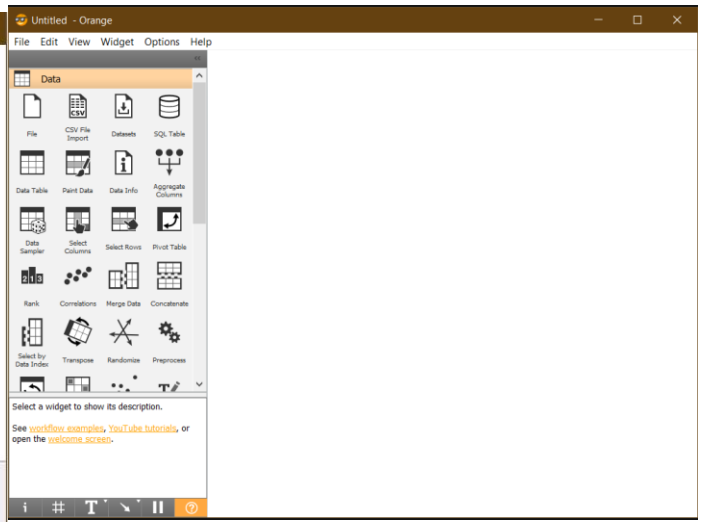
PRN: 2019BTECS00105 Batch: T7 Vaishnavi Mahajan

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

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

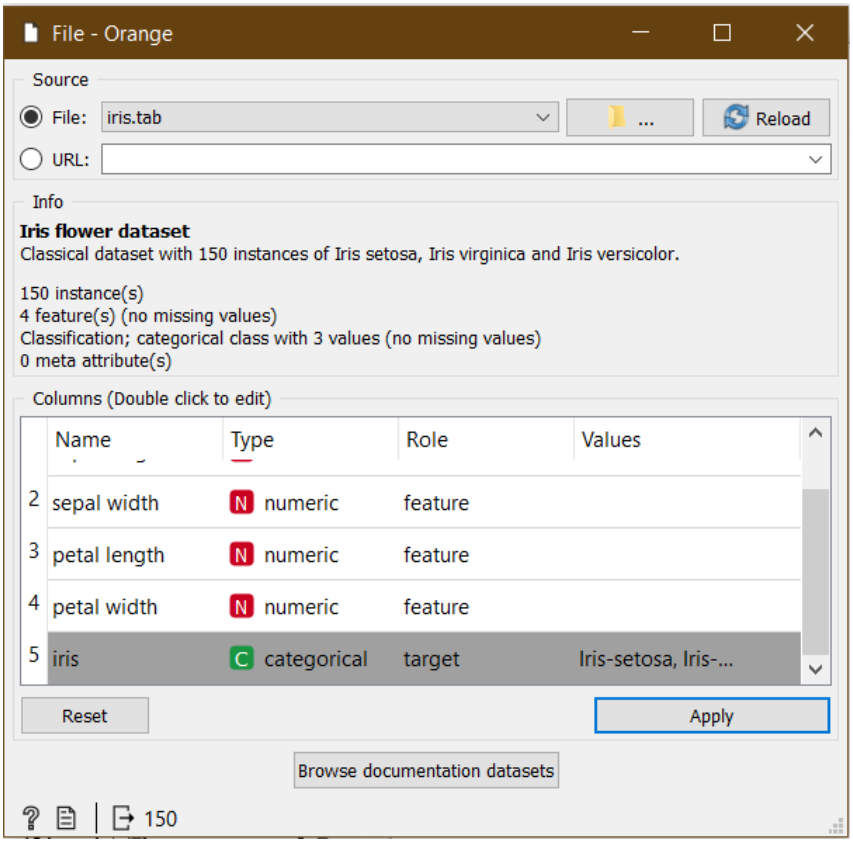
a. Install orange

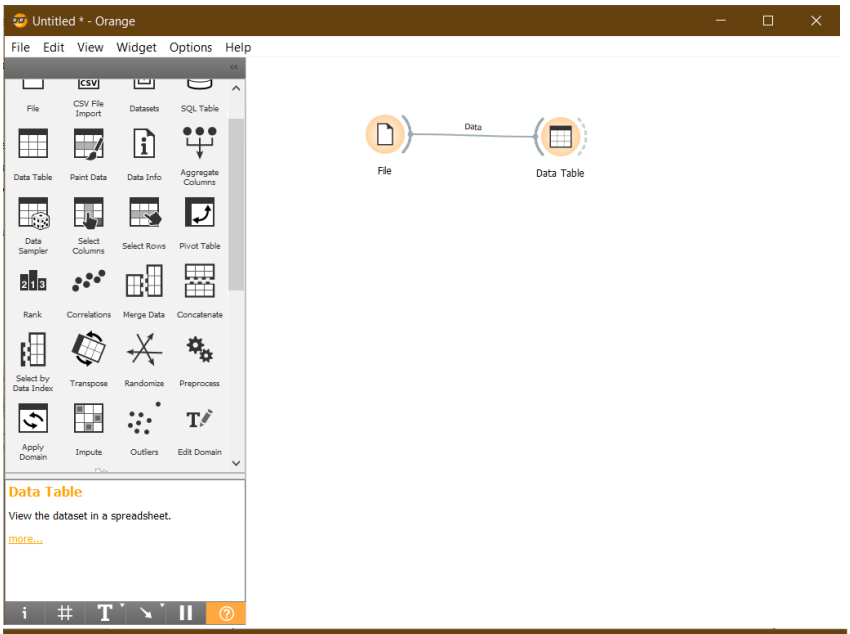


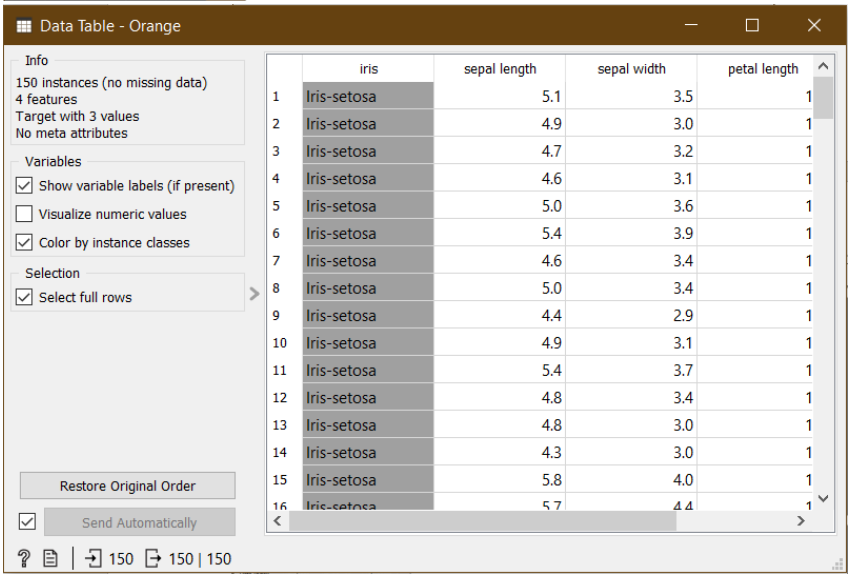


b. Show data distribution

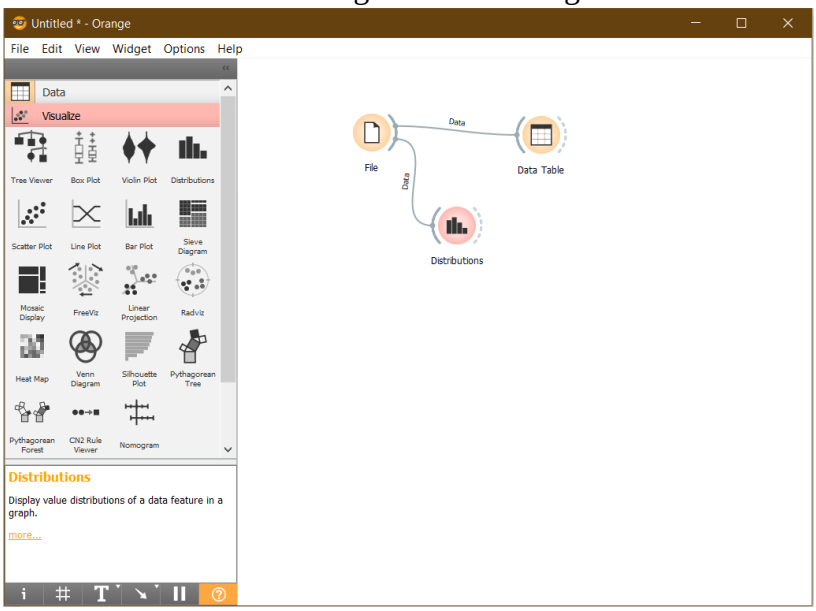
1. Import the data using file widget.
2. Connect it to table widget to view.

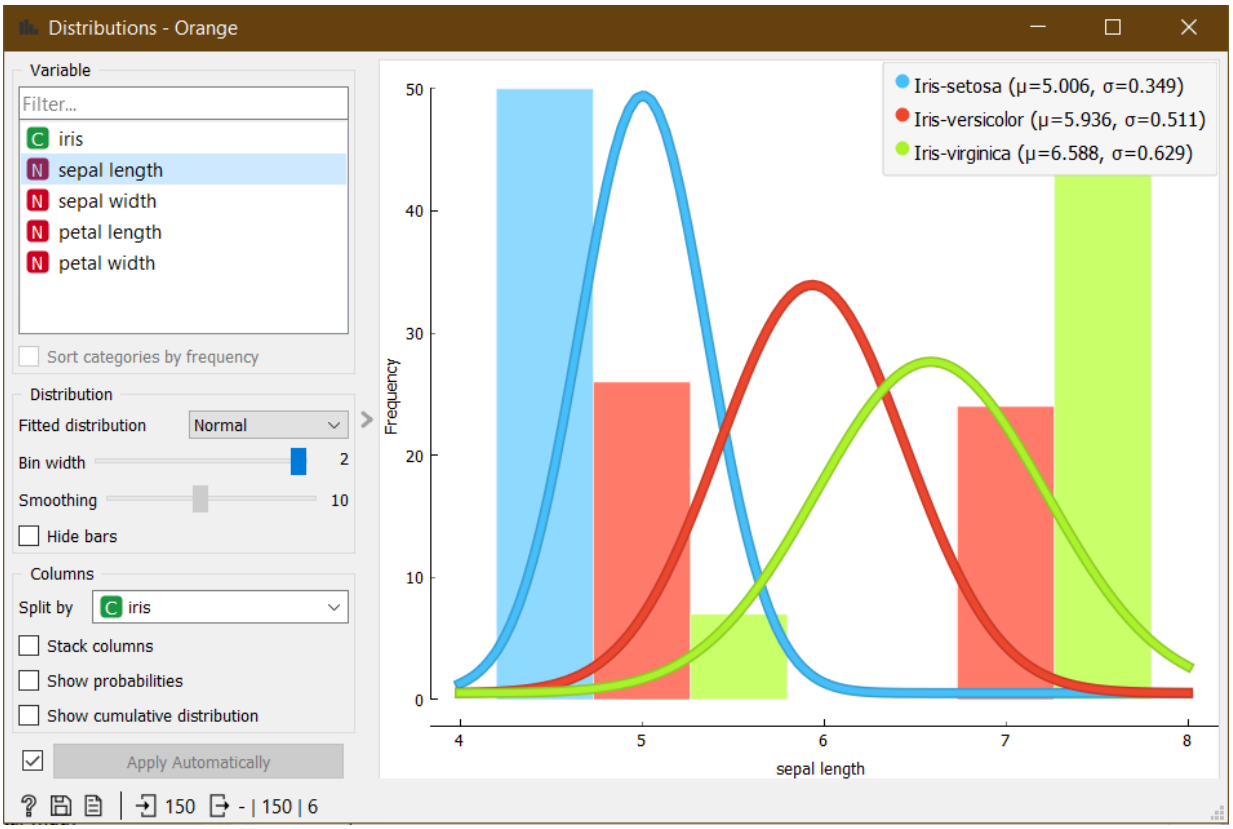






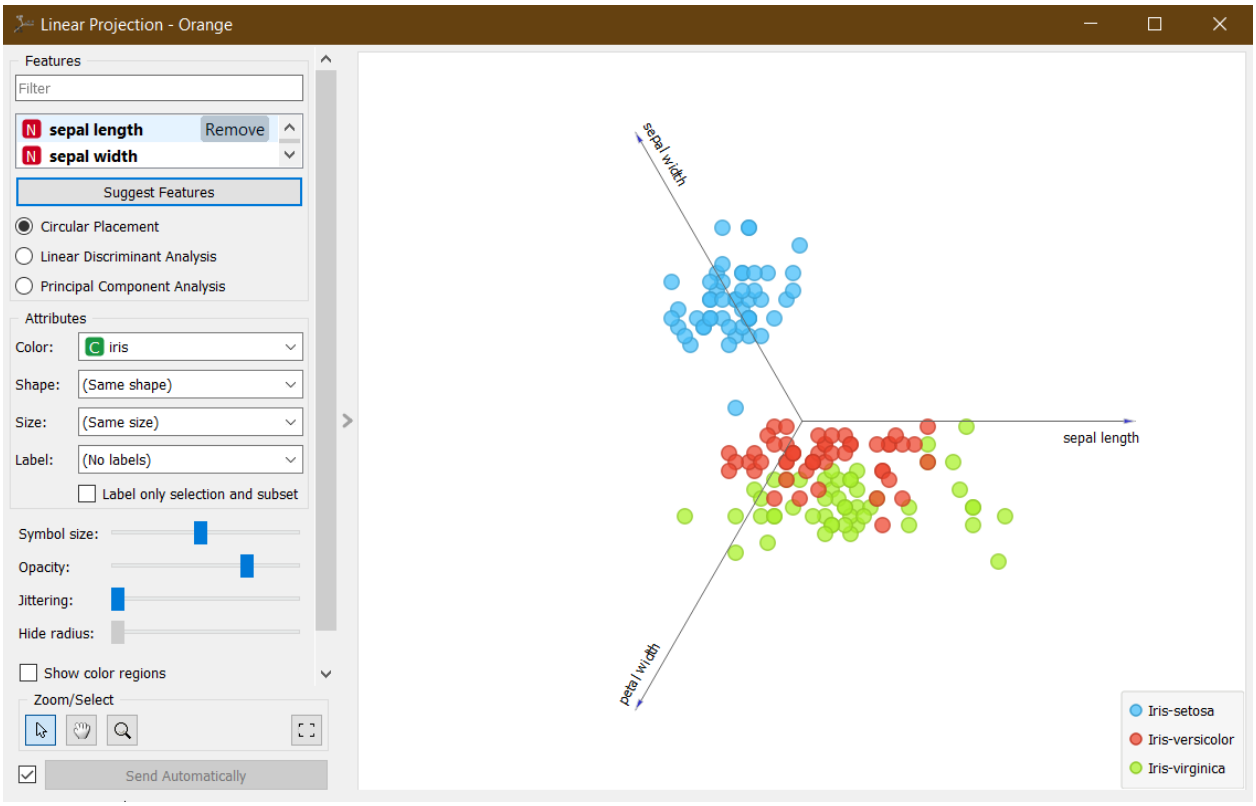
3. Connect distribution widget with file widget.





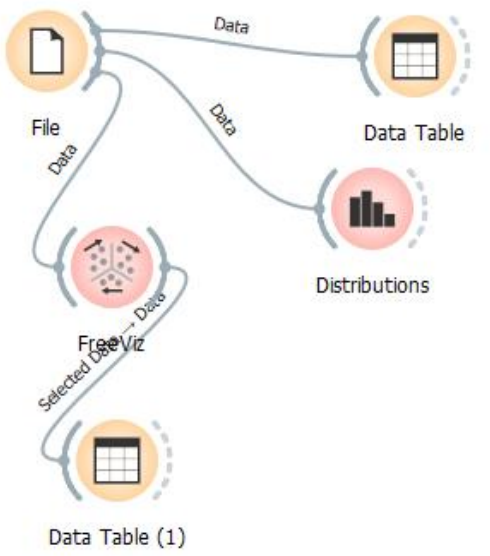
c. Show linear projection

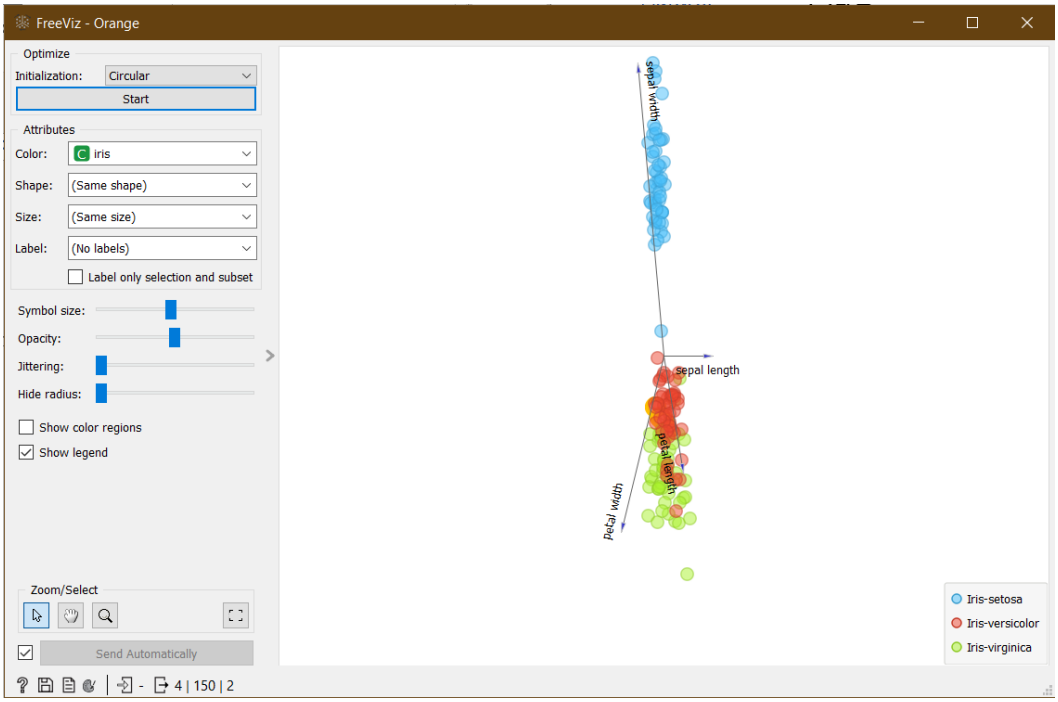
1. Add linear projection widget.
2. Connect the file widget to it.



d. Show FreeViz

1. Add freeviz widget.
2. Connect the file widget to it.
3. Connect table widget to it.





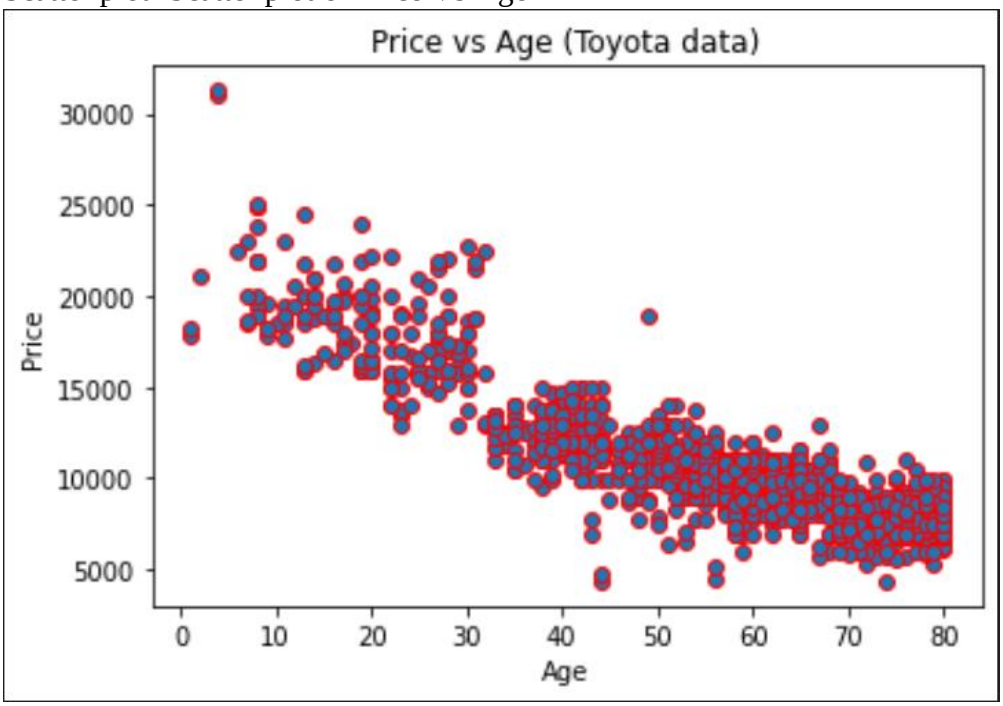
PRN: 2019BTECS00105 Batch: T7 Vaishnavi Mahajan

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

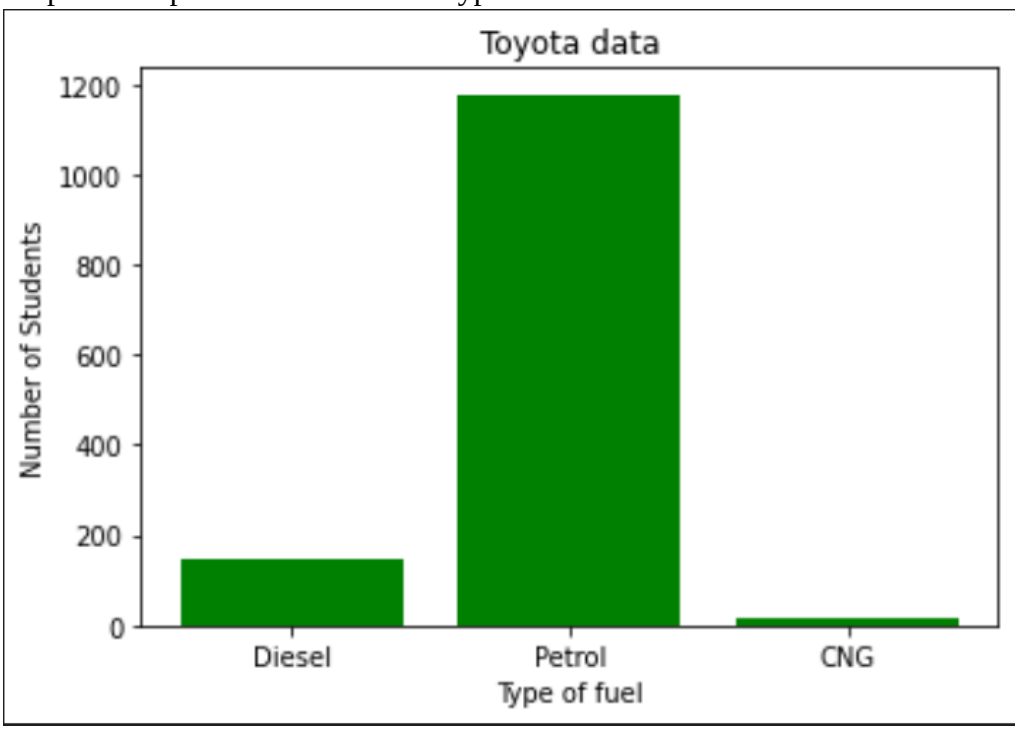
Free Software	Open Source Software	Proprietary Software
Free software means freedom to run, copy, change, study, improve any software.	Open source software's code is freely available to everybody.	Proprietary software code is not available outside the developers and company that owns it.
Software can be developed, tested and improved by open communities.	Developed, tested and improved by open communities.	Developed by a company.
Software should be free of cost accessible.	Users need not pay for using it.	Users have to pay for using it.
Example: C libraries, MySql relational database	Example: Linux, Firefox, VLC media player	Example: Microsoft office, Adobe Flash Player

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. Bar plot- Bar plot for different fuel types



5. Enlist some examples along with its purpose and properties (at least 10) of FOSS and proprietary software with respect to database.
   Free software means freedom to run, copy, change, study, improve any software.
   FOSS can be developed, tested and improved by open communities.
   FOSS is free of cost accessible.
   Example: C libraries, MySql relational database, GNU compiler collection, Linux, Apache web server.