

Software Engineering Tools Lab

Assignment No - 1

(Module 1- Introduction to OSS)

PRN : 2019BTECS00055

Name : Aditya Rajendra Manapure

Batch : T2

Que 1 : Using **Weka tool** perform below tasks such as data preprocessing, data classification (use any appropriate ML algorithm) and data visualization efficiently on given dataset.

Ans :

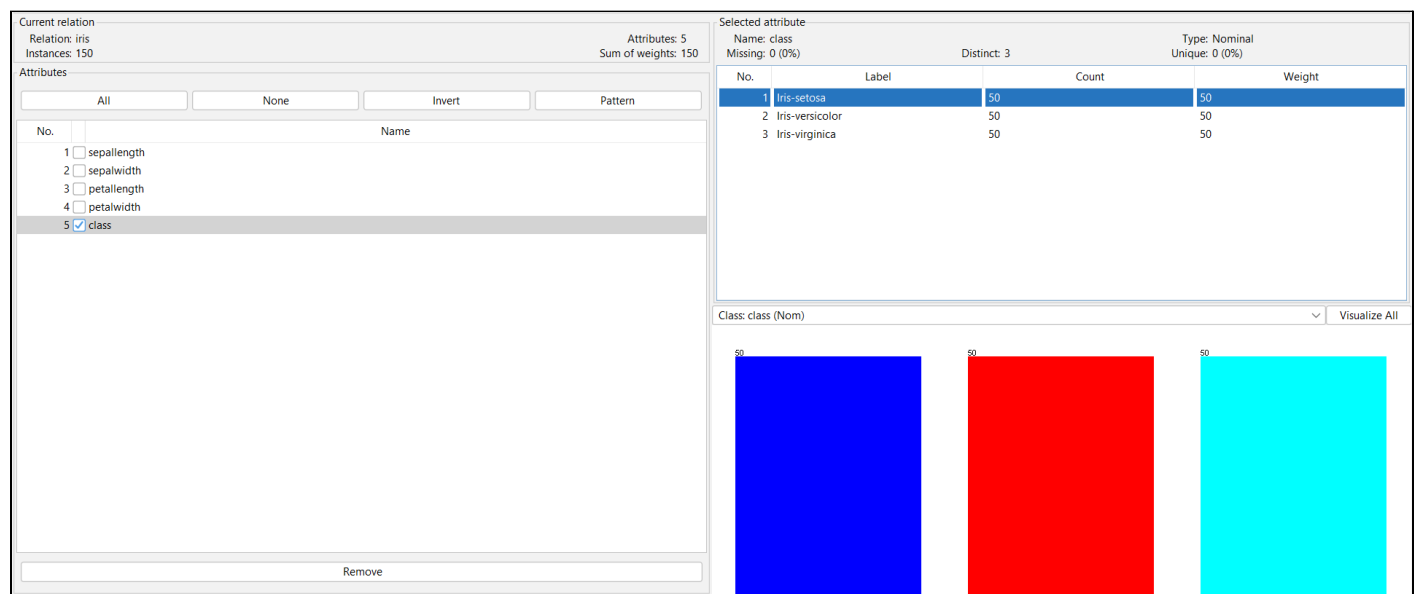
Here we are using IRIS dataset provided by weka tool. The dataset content 3 classes and 4 attributes of of iris flower.

Classes/ type

1. Iris setosa
2. Iris versicolor
3. Iris virginica

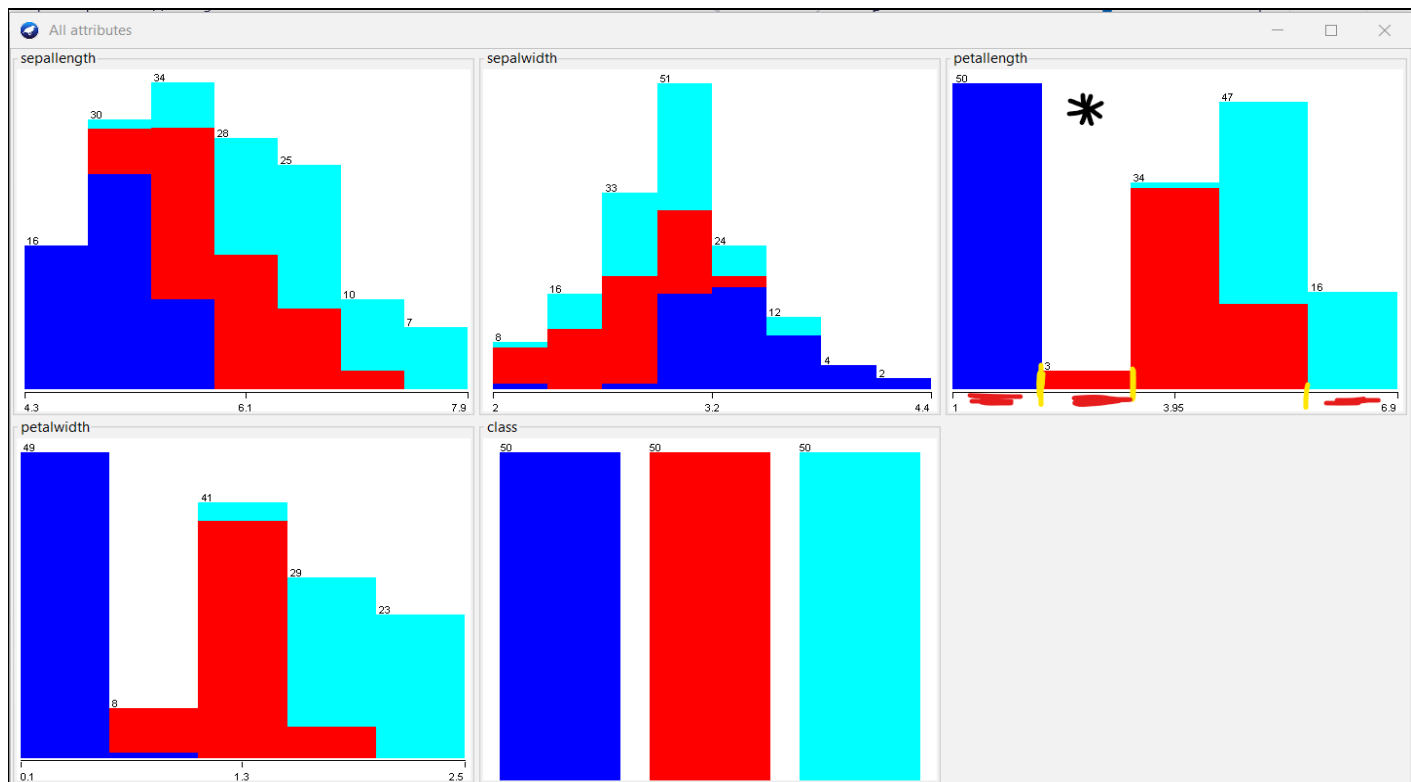
Attributes

1. Sepal Length
2. Sepal Width
3. Petal Length
4. Petal Width



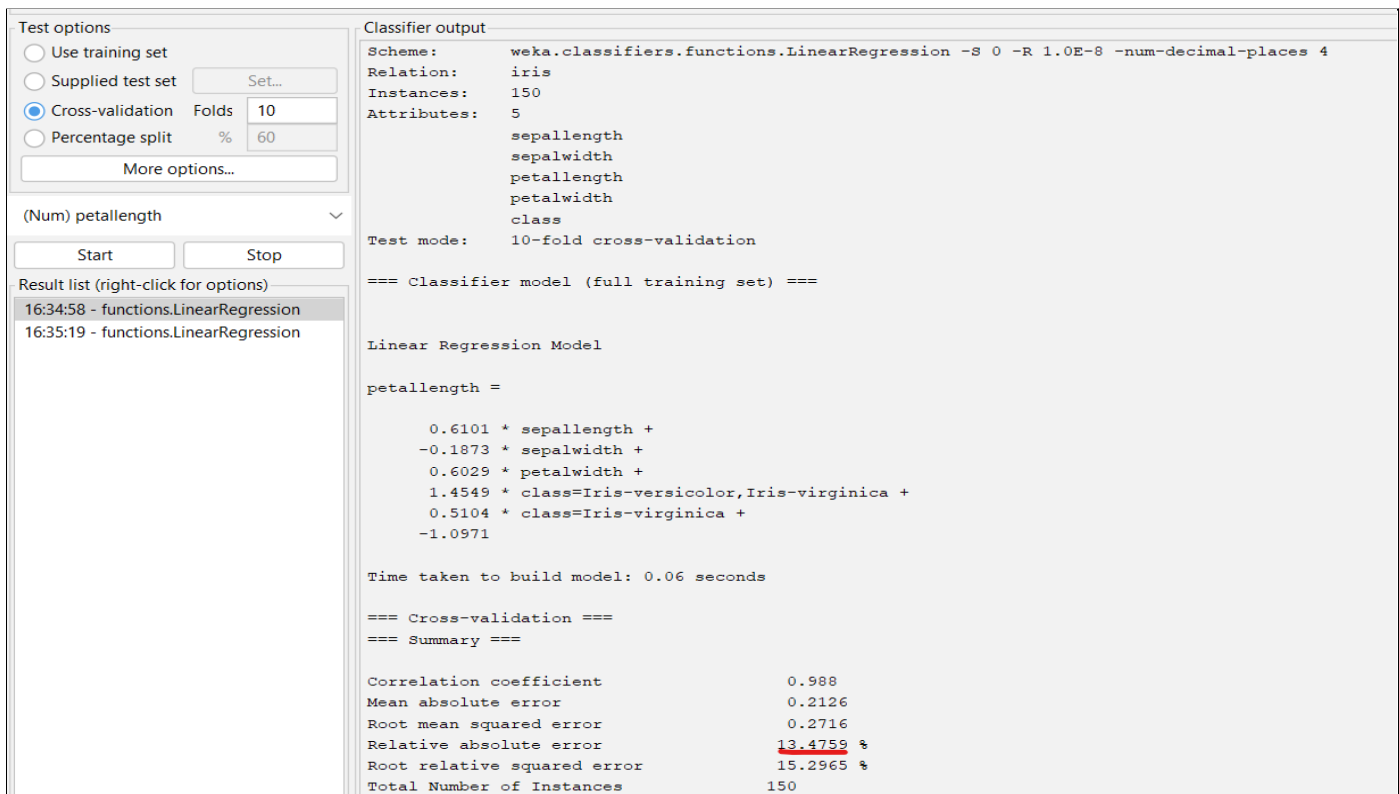
We have 50 samples of each type.

1. Data Preprocessing :



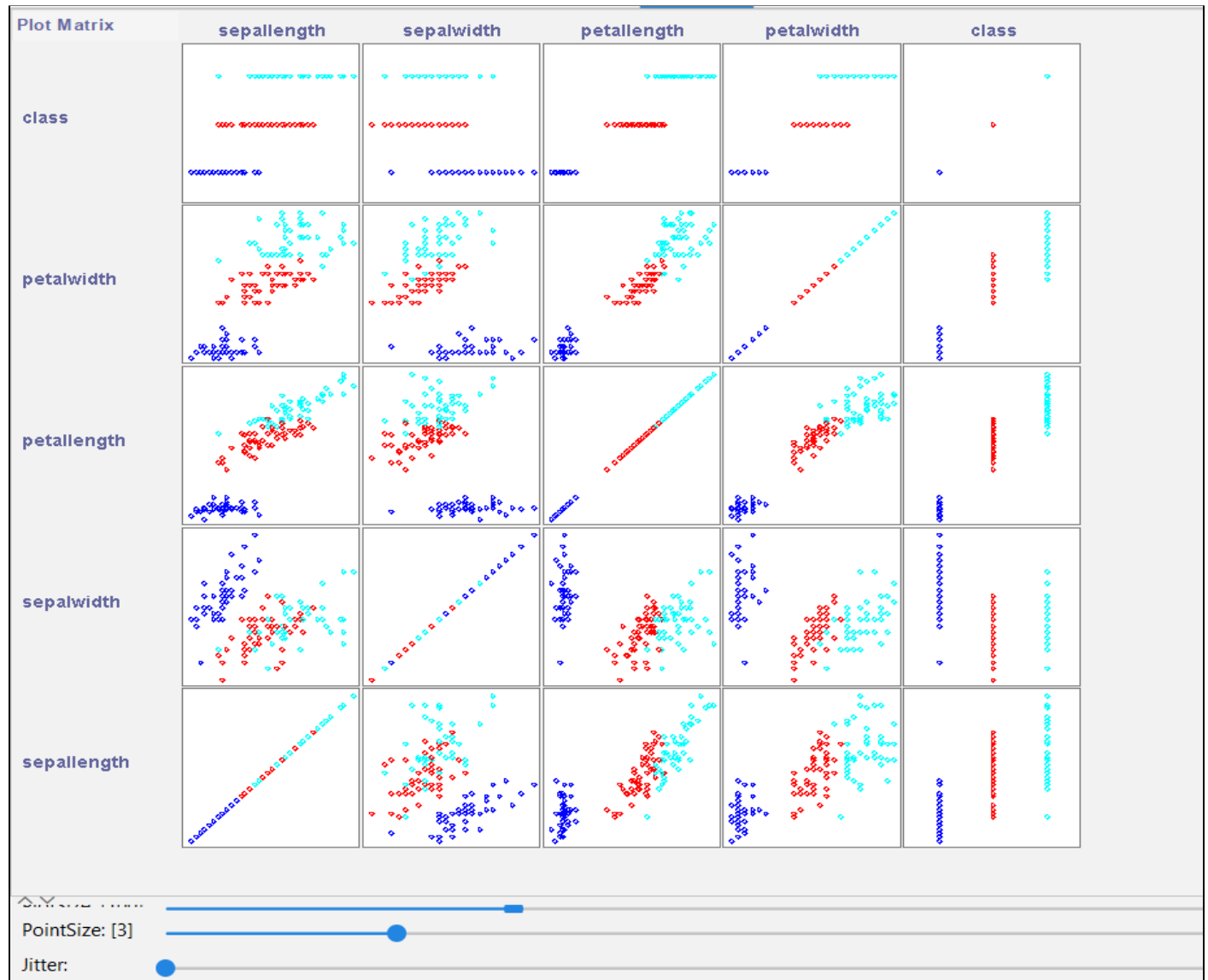
From above fig. We can observe that for attribute **petal length** the data of each type is well separated. Hence we take petal length as our parameter. We will be using **Linear Regression**, so we need only one parameter.

2. Data Classification :



We get a decent **RMS error = 13.47%**. If we select any parameter other than **petallength**, our error increases. This proves that our selection of parameters based on the visualizations was correct.

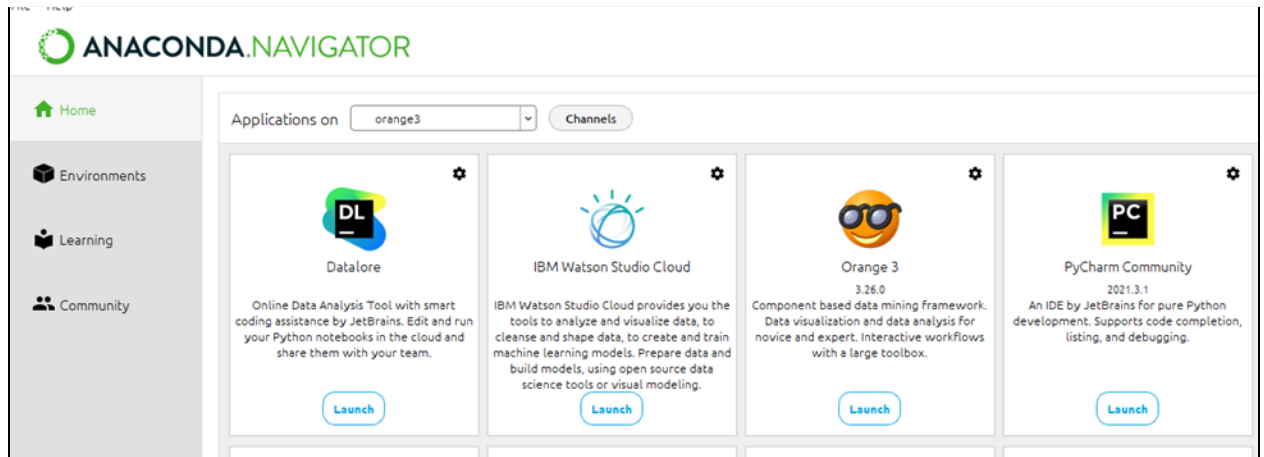
3. Visualization :



Que 2: ORANGE Data Virtualization Tool (Iris dataset analysis)

Ans :

Download Orange from using Anaconda Navigator

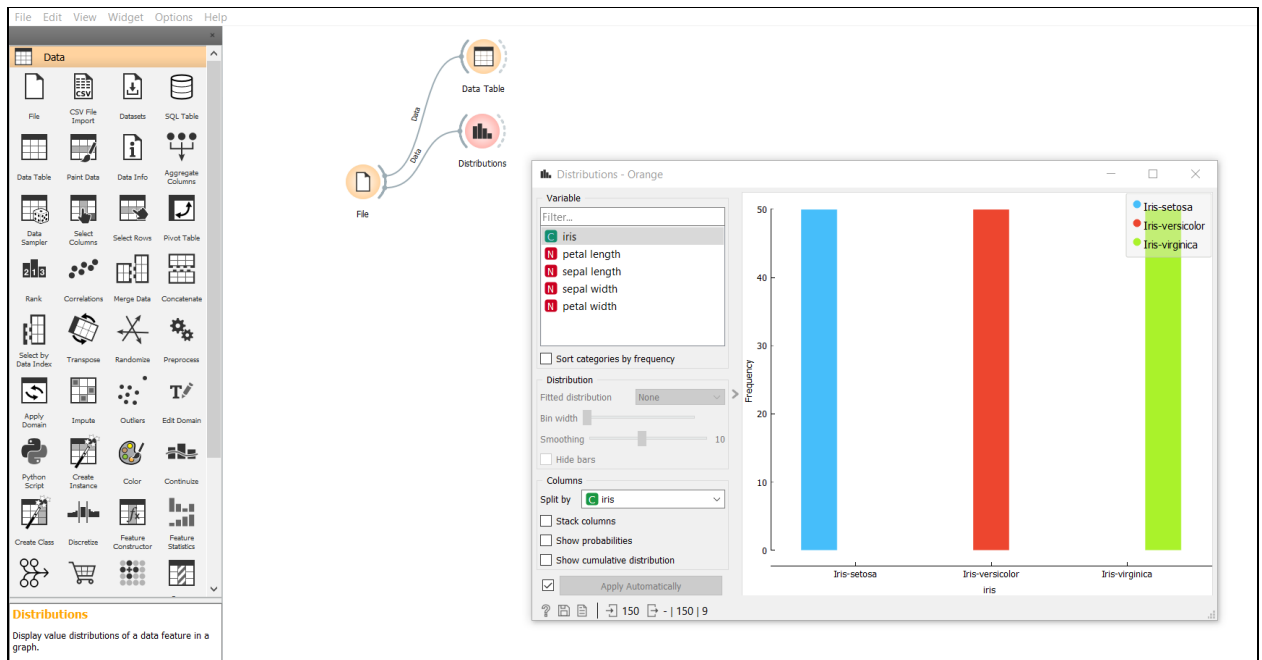


Here I use available dataset iris.tab

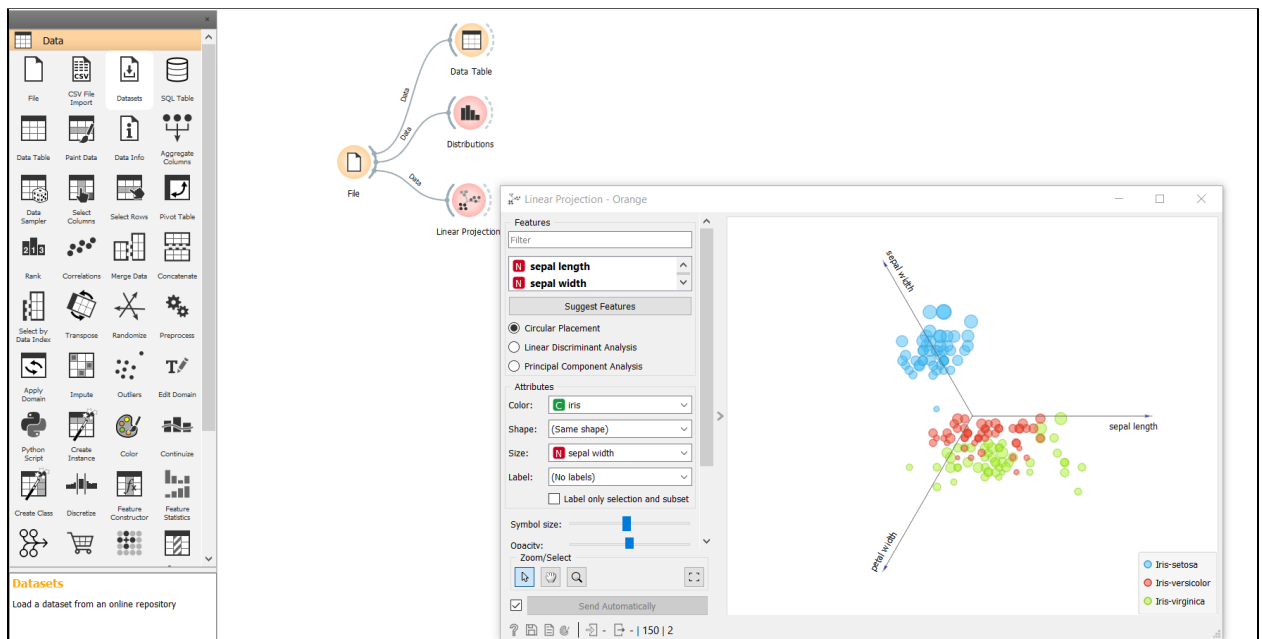
The screenshot shows the Orange Data Mining software interface. On the left is a 'Data' widget palette with various tools like File, CSV File Import, Datasets, SQL Table, Data Table, Pivot Data, Data Info, Aggregate Columns, Data Sampler, Select Columns, Select Rows, Pivot Table, Rank, Correlations, Merge Data, Concatenate, Select by Data Index, Transpose, Randomize, Preprocess, Apply Domain, Impute, Outliers, Edit Domain, Python Script, Create Instance, Color, Continue, Create Class, Discretize, Feature Constructor, and Feature Statistics. The main workspace shows a workflow with a 'File' widget connected to a 'Data Table' widget. The 'Data Table' widget is open, displaying the Iris dataset in a spreadsheet format. The left pane of the 'Data Table' widget shows metadata: 150 instances (no missing data), 3 features, Target with 3 values, and 1 meta attribute. The right pane shows the data table with columns: Iris, petal length, sepal length, sepal width, and petal width. The bottom status bar indicates 150 rows and 150 columns.

	Iris	petal length	sepal length	sepal width	petal width
1	Iris-setosa	1.4	5.1	3.5	0.2
2	Iris-setosa	1.4	4.9	3.0	0.2
3	Iris-setosa	1.3	4.7	3.2	0.2
4	Iris-setosa	1.5	4.6	3.1	0.2
5	Iris-setosa	1.4	5.0	3.6	0.2
6	Iris-setosa	1.7	5.4	3.9	0.4
7	Iris-setosa	1.4	4.6	3.4	0.3
8	Iris-setosa	1.5	5.0	3.4	0.2
9	Iris-setosa	1.4	4.4	2.9	0.2
10	Iris-setosa	1.5	4.9	3.1	0.1
11	Iris-setosa	1.5	5.4	3.7	0.2
12	Iris-setosa	1.6	4.8	3.4	0.2
13	Iris-setosa	1.4	4.8	3.0	0.1
14	Iris-setosa	1.1	4.3	3.0	0.1
15	Iris-setosa	1.2	5.8	4.0	0.2
16	Iris-setosa	1.5	5.7	4.4	0.4
17	Iris-setosa	1.3	5.4	3.9	0.4
18	Iris-setosa	1.4	5.1	3.5	0.3
19	Iris-setosa	1.7	5.7	3.8	0.3
20	Iris-setosa	1.5	5.1	3.8	0.3
21	Iris-setosa	1.7	5.4	3.4	0.2
22	Iris-setosa	1.6	5.1	2.7	0.4

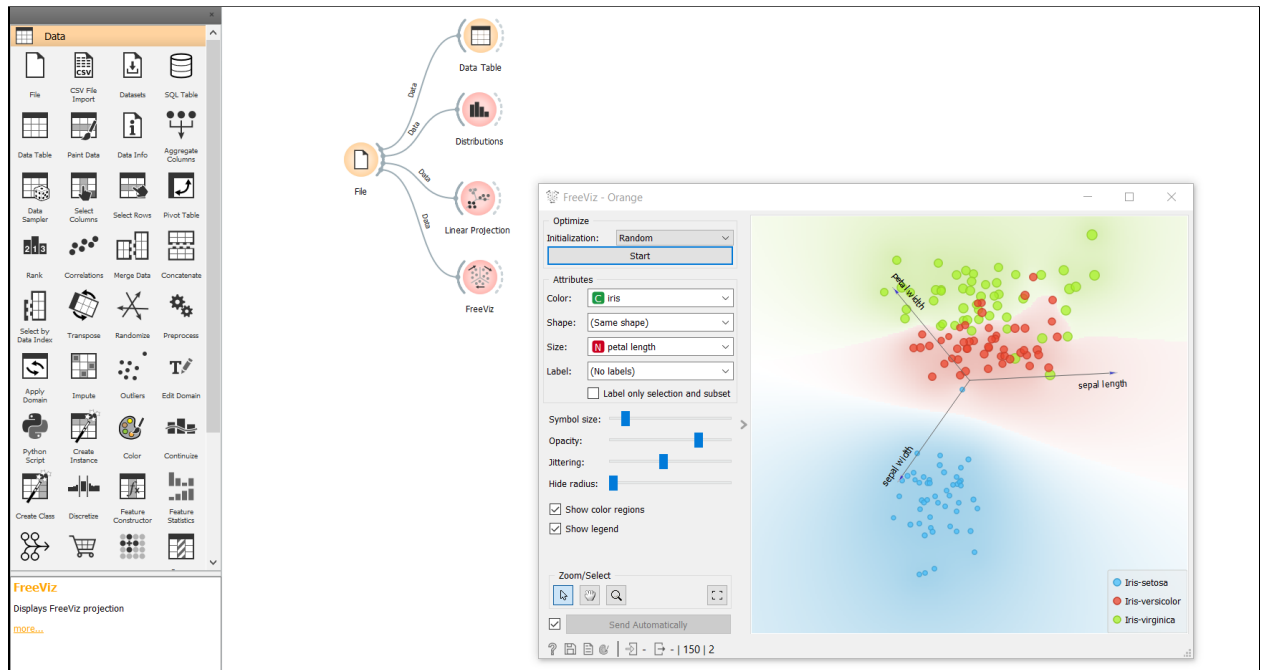
Data Distribution :



Linear Projection :



FreeViz :



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

Ans :

Parameter of Comparison	Open Source Software	Free Software	Proprietary Software
Definition	Open Source software refers to software that contains a source code with license which is open to free use, modification and distribution.	“Free software” means software that respects users’ freedom and community.”	Proprietary software refers to the type of software that contains a licensed source code and is copyrighted for use.
Collaboration	Open Source software is developed for open collaboration.	User contribute but for personal level	Proprietary software is not meant for open collaboration, but only for the creator and users who have paid for it.

Access	Open Source software has open access, that is, can be accessed by anyone.	The users have the freedom to run, copy, distribute, study, change and improve the software.	Proprietary software can be accessed only by those who developed it and those who have paid for it.
Flexibility	Open Source software is flexible, that is, it can be used, modified and distributed by anyone.	It is flexible as user can change and improve software	Proprietary software has restricted flexibility, that is, there are restrictions on its usage.
Example	FreeBSD (Berklee Software Distribution), Android, LibreOffice, Ubuntu are a few examples of Open Source software.	Linux kernel, the BSD and Linux operating systems, the GNU Compiler Collection and C library; the MySQL relational database; the Apache web server; and the Sendmail mail transport agent.	Windows, Microsoft, macOS, Adobe Photoshop, Adobe Flash Player are a few examples of Proprietary software.

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

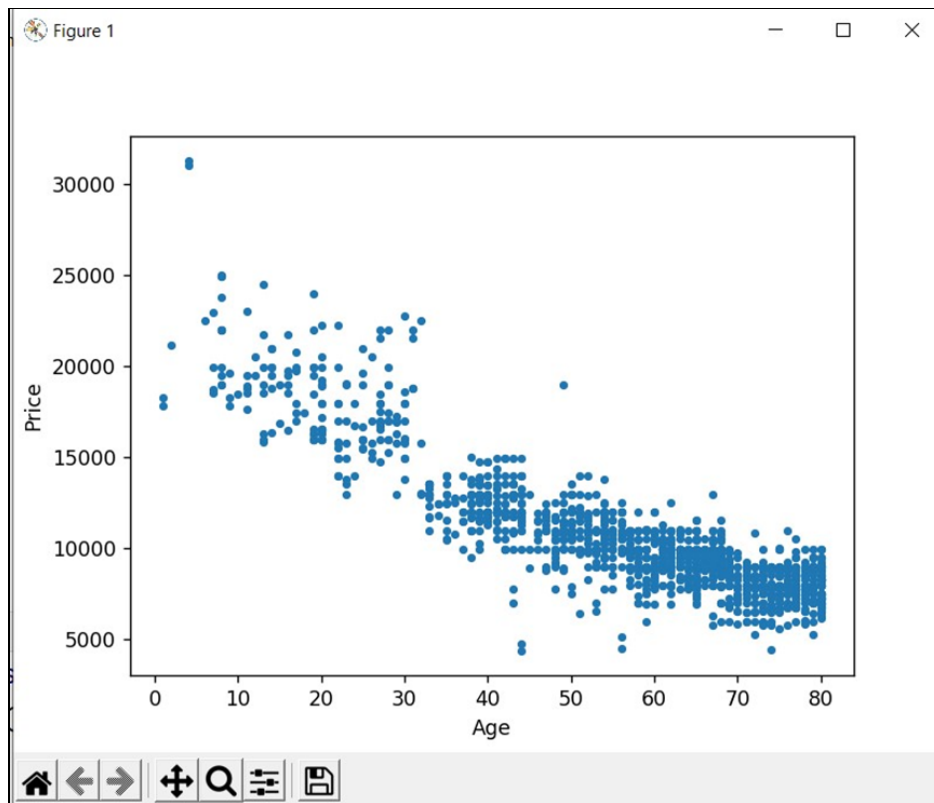
Dataset- https://drive.google.com/file/d/1i11BZFe8Xj9kNq7eeE9KOa_lz1KhEdXJ/view

- Scatter plot- Scatter plot of Price Vs Age
- Histogram- for Kilometer and CC
- Bar plot- Bar plot for different fuel types

Ans :

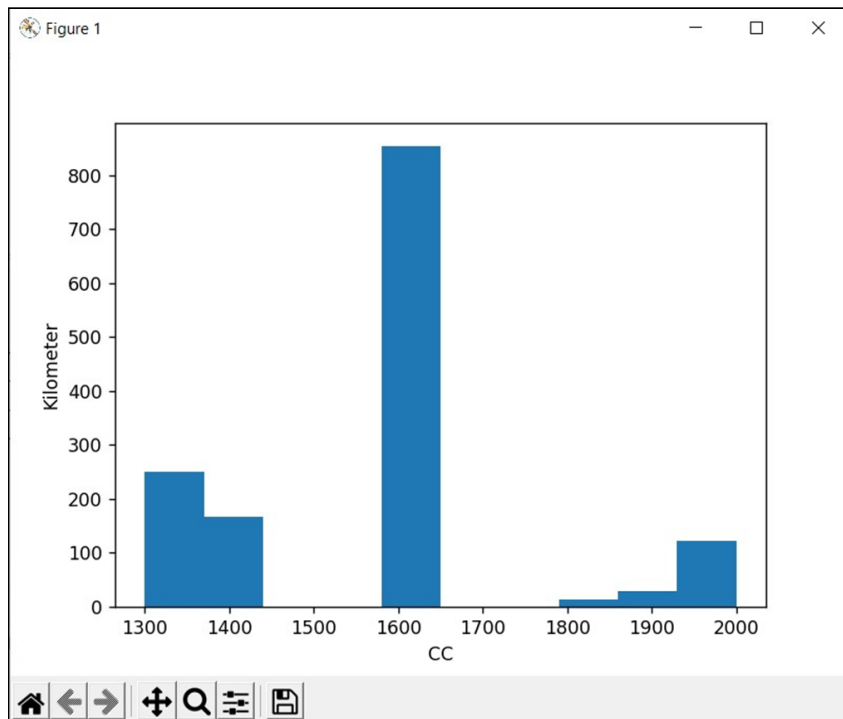
- Scatter plot - Price vs Age

```
#FOR SCATTER PLOT OF PRICE VS AGE
y = toyota_data['Price']
x = toyota_data['Age']
plt.ylabel('Price')
plt.xlabel('Age')
plt.scatter(x,y,s=10)
```

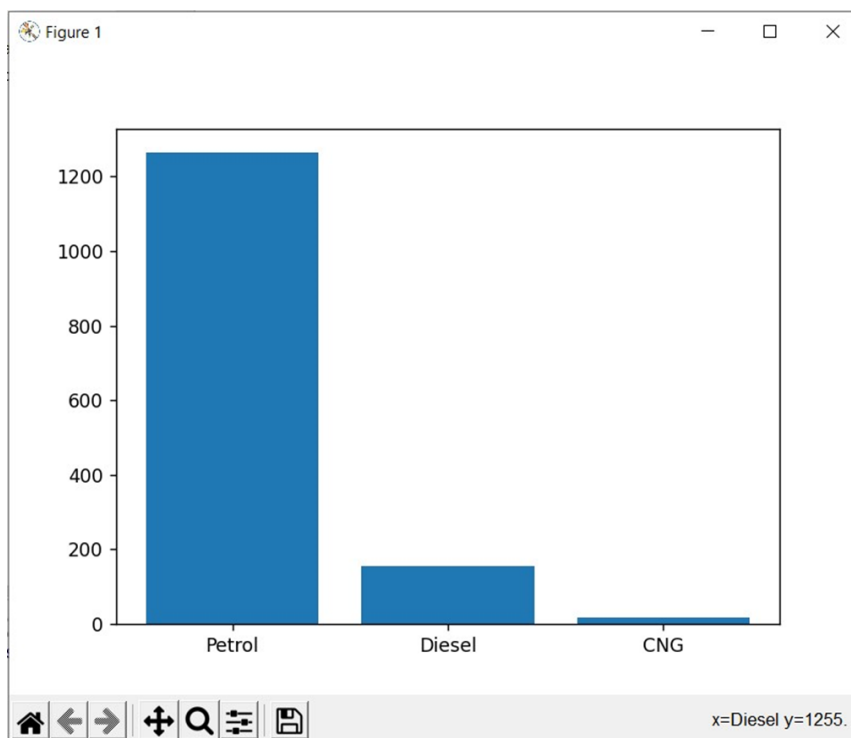


2. Histogram- for Kilometer and CC

```
#FOR HISTOGRAM OF KM & CC  
x = toyota_data['CC']  
y = toyota_data['KM']  
plt.ylabel('Kilometer')  
plt.xlabel('CC')  
plt.hist(x)
```

3 . Bar plot- Bar plot for different fuel types



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

Ans :

Example of **FOS database** :

MySQL, SQLite, MongoDB

1. **MySQL:**

Developed by Oracle Corporation and written in C, C++ language, MySQL is a leading open source and free database software system. Its stable version 8.0.13 released date is October 7, 2018.

Features:

- i. It is a relational database system like all other database systems are.
- ii. Community server and proprietary server: it supports two different database editions.
- iii. Multiple storage engines like InnoDB, CSV, NDB, etc to store data.
- iv. It supports SQL, i.e. Structured Query Language that is a standard language to query and update data along with its appropriate administration and management.
- v. MySQL executes under a variety of 20+ operating systems. Some significant ones are Apple Mac OS X, Microsoft Windows, Linux and a wide range of UNIX variants.
- vi. It is a fast database management program.
- vii. It supports ODBC interface Connector/ODBC that permits MySQL to get addressed by all the programming languages that run under Microsoft Windows, Visual Basics, Delphi, etc.
- viii. Triggers and views are used to enforce business rules, and it shows only that data which is needed by the user.

2. SQLite:

SQLite is a C programming software library that works as an excellent **relational database management software** system. The word 'lite' in the name indicates the lightweight nature of the database administration, setup, and essential source. Developed by D. Richard Hipp on August 17, 2000, the stable version release date of SQLite 3.26.0 is December 1, 2018.

Features:

- i. Before using, you don't need to 'install' SQLite and also, no 'setup' procedure is required. It is accessing just through simple download with zero-configuration requirements.
- ii. Majority of SQL database engines perform over separate server process. However, there is no different server process for installing, setting up, configuring, managing, initializing and troubleshooting the software program.
- iii. A single ordinary disk file – SQLite database storage anywhere in the hierarchy of the directory.
- iv. The SQLite software supports cross-platform file format.
- v. The software is compact. The entire library optimizes to less than 500 KiB in size.
- vi. Static typing format usable in most of SQL database engines. Whereas, this restriction is a 'no follow' in case of SQLite with involving manifest typing.
- vii. SQLite uses variable-length records. It only uses the amount of disk space required for storing the information in a row.
- viii. The SQLite source code is designed to read and access by the average level of programmer.
- ix. Compilation of the SQL statements into virtual machine code.
- x. Public domain based source code of SQLite brings no copyright claim from the company's end.
- xi. SQLite provides a significant level of enhancements to the SQL language that generally remains absent in other database engines.

11. MongoDB:

MongoDB is a 'No SQL' free and open source database management system. It is a document-oriented software program that runs on multiple operating systems i.e. Linux, Windows Vista, OS X, FreeBSD and Solaris. The software is developed by MongoDB Inc. and licensed under the Server Side Public License (SSPL). Its stable version 4.0.6 release date is February 5, 2019.

Features:

- i. MongoDB supports ad hoc queries. You can search and index any field in the document.
- ii. It possesses an exceptional document data model to present the data in the best possible way.
- iii. It supports Master Slave Replication.
- iv. It supports map reduce and aggregation tools.
- v. It is facilitated by an automatic load balancing configuration because of the data placed in the shards.
- vi. It possesses a distributed design system to put data intelligently in the sections.
- vii. Operates on multiple platforms and is a schema-less database written in C++.
- viii. It uses JavaScript instead of procedures.

Example of **proprietary** :

Skype- Azure Cosmos DB(NoSQL)

- i. They have higher scalability.
- ii. They use distributed computing
- iii. They are cost effective
- iv. They support flexible schema.
- v. They can process both unstructured and semi-structured data.
- vi. There are no complex relationships, such as the ones between tables in an RDBMS.