**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.**

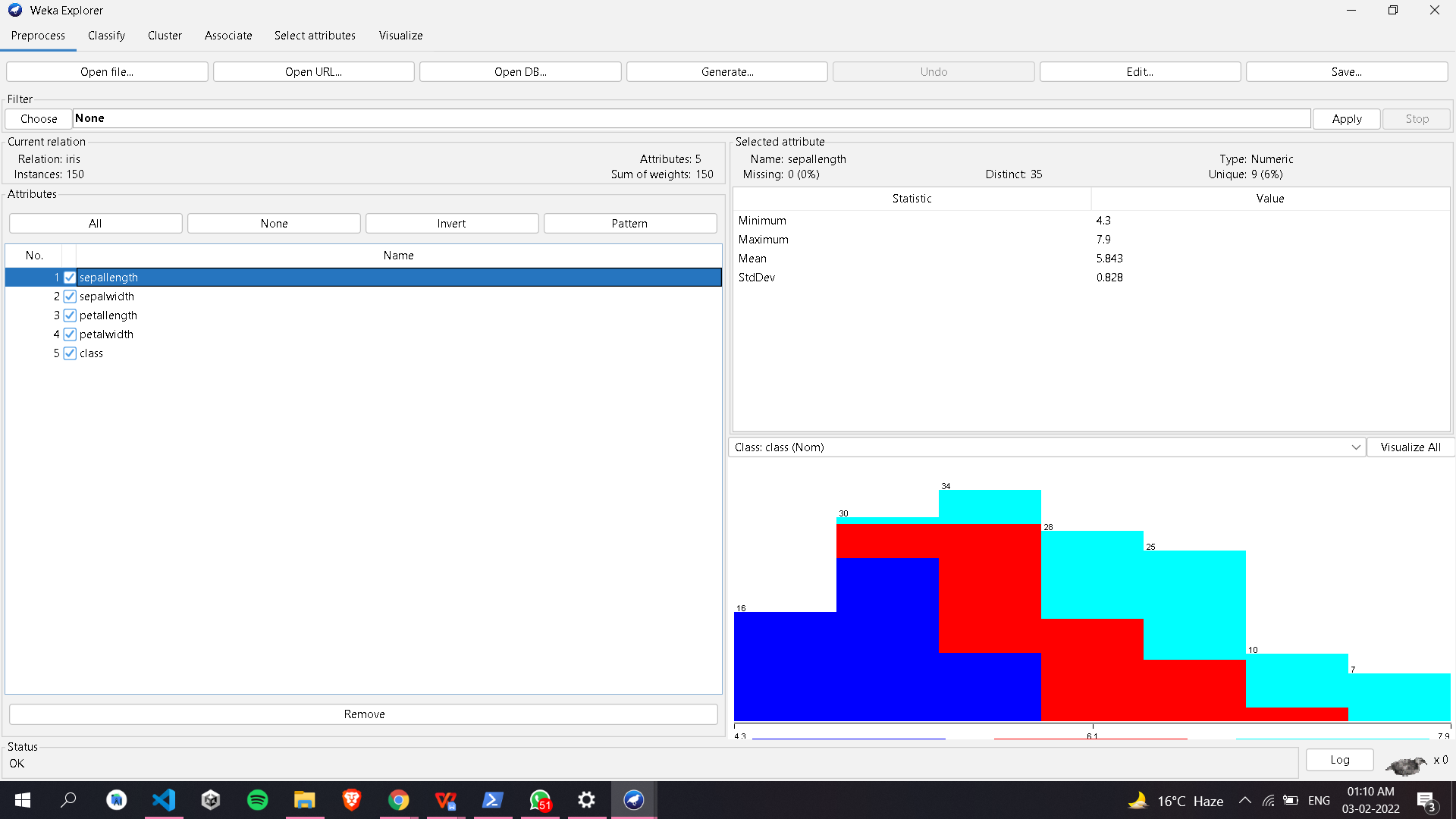
**Use the Iris dataset given**

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

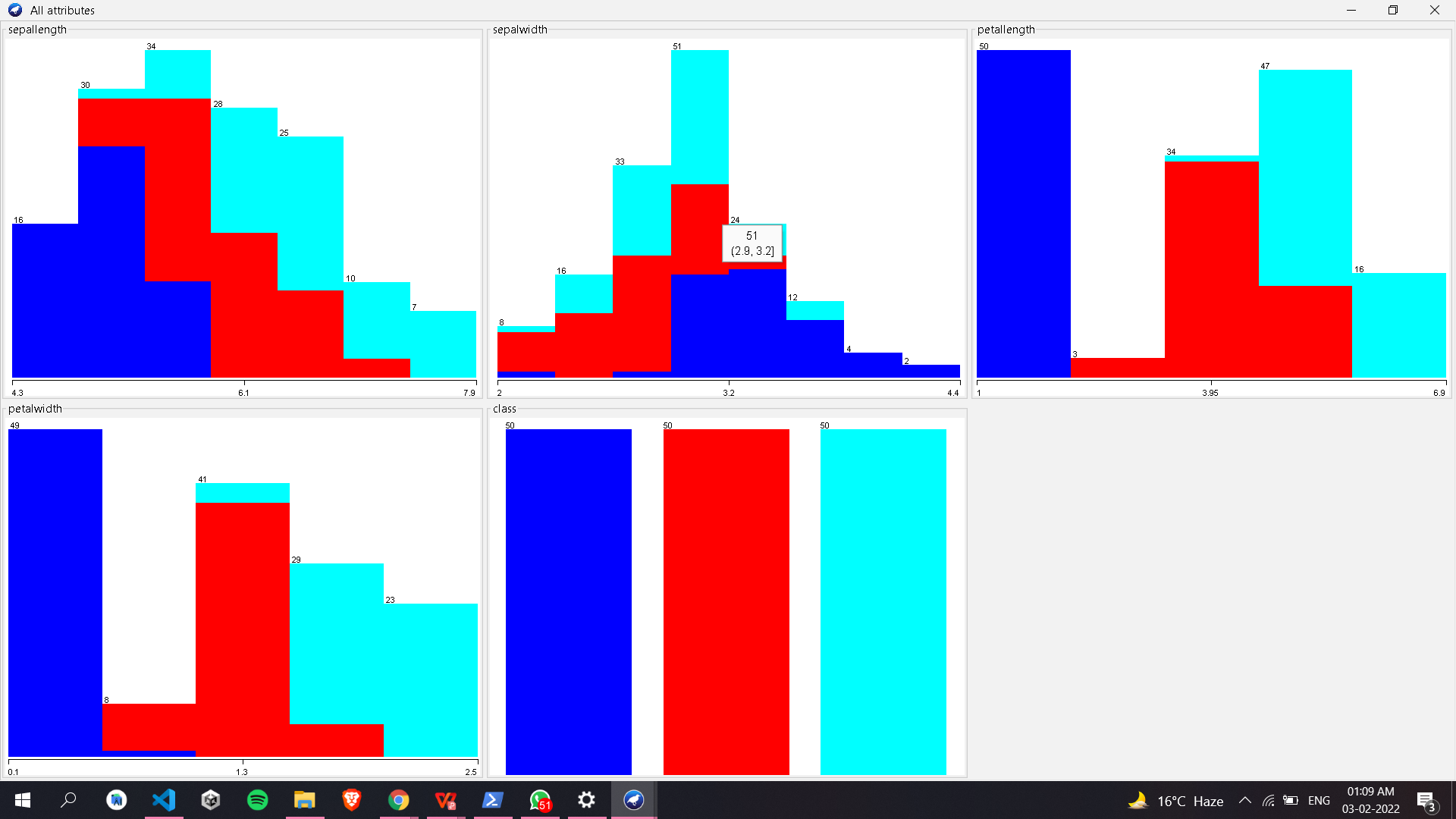
**Note-provide screen shots for every task**

**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)**

****

**1. Preprocessing of data**

****

From basic visualisation of all 5 parameters, we observe that

petallength gives a relatively clear separation of classes. Hence we

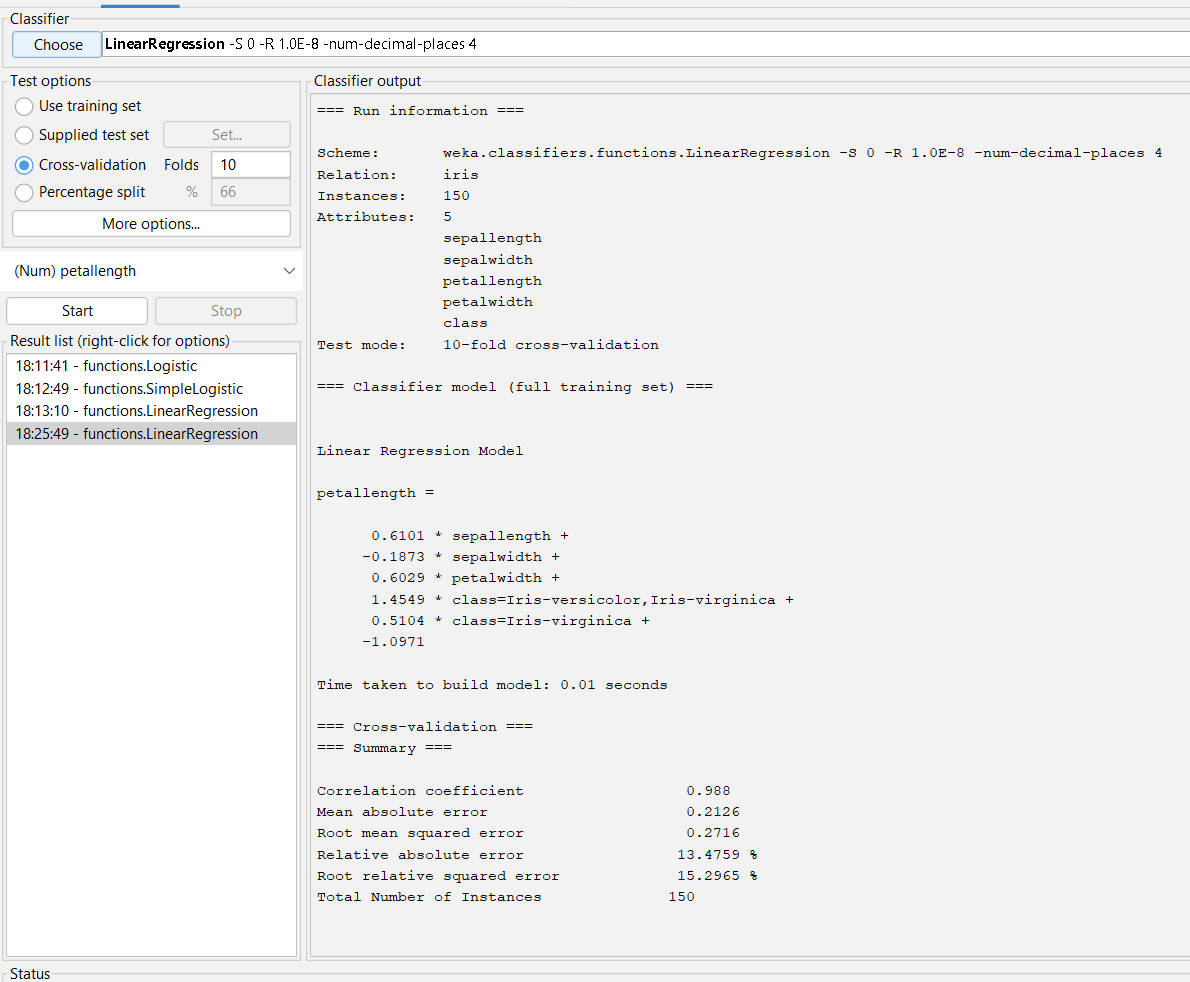
take petallength as our parameter. We will be using Linear Regression,

so we need only one parameter.

**2. Data Classification**

We make our model based on Linear Regression and get following

analysis of the model:

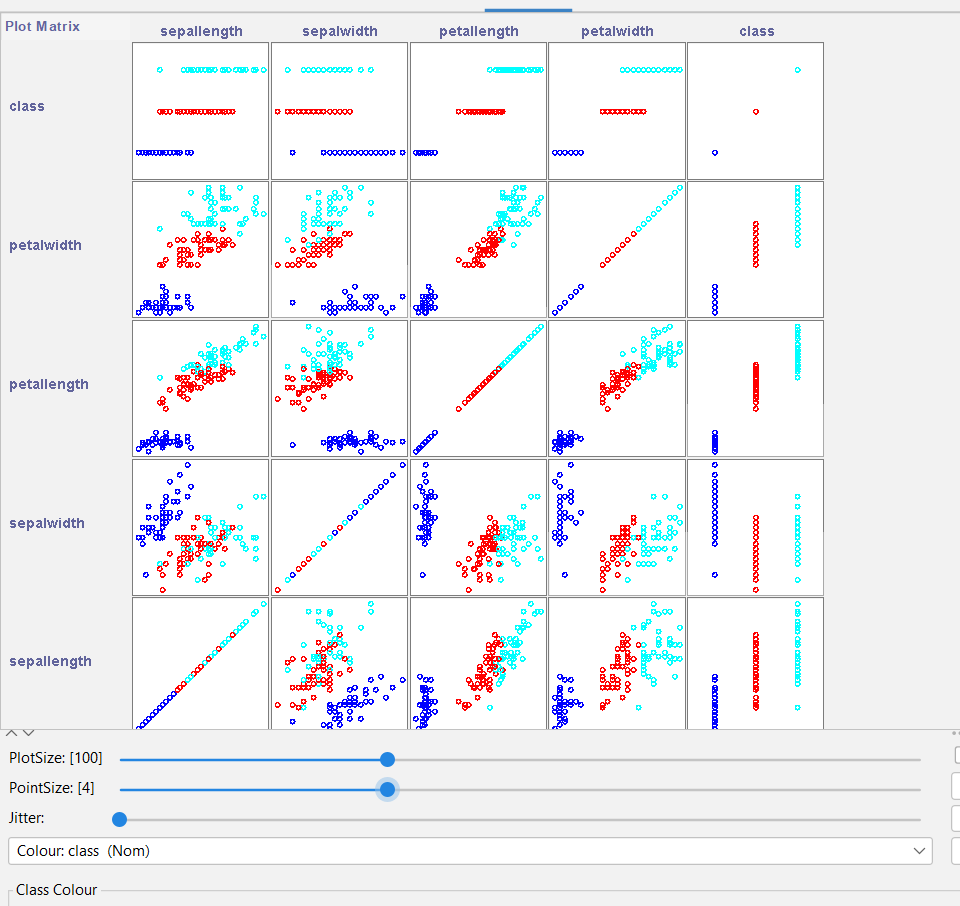
****

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. Data Visualization**

****

**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**

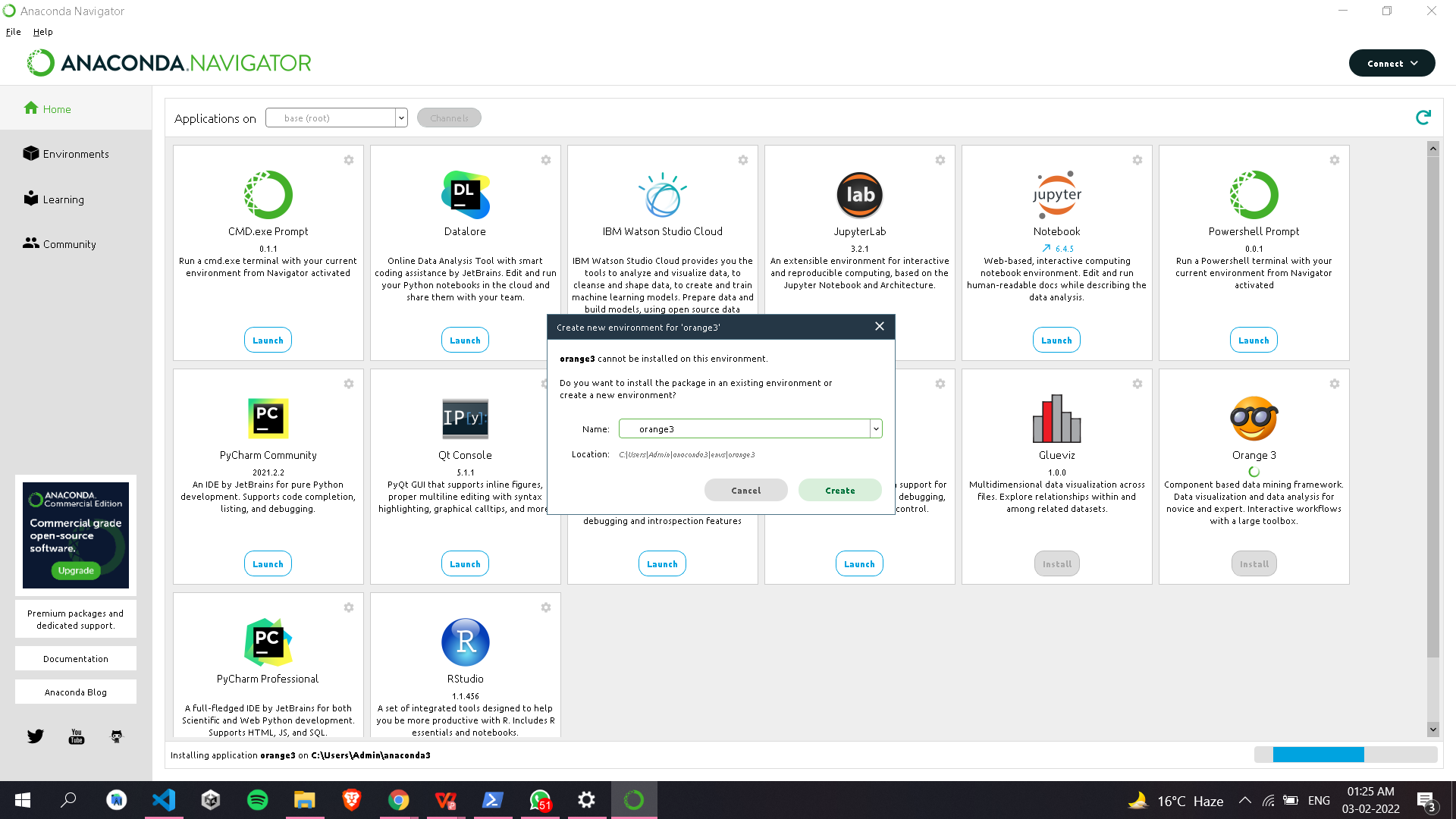
**c. Show linear projection**

**d. Show FreeViz**

**Use dataset**

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

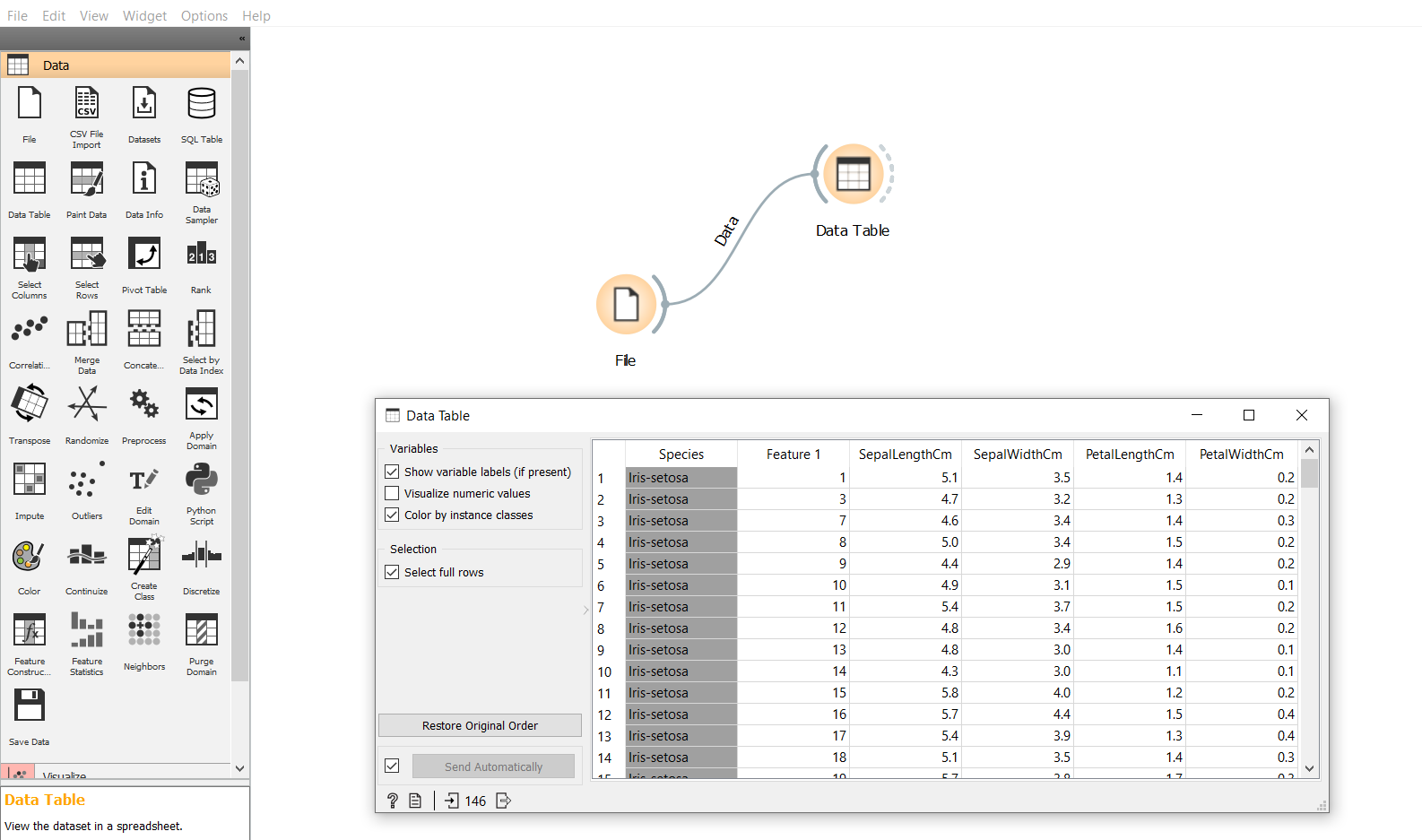
**Create a report for this task and upload screenshots for the same.**

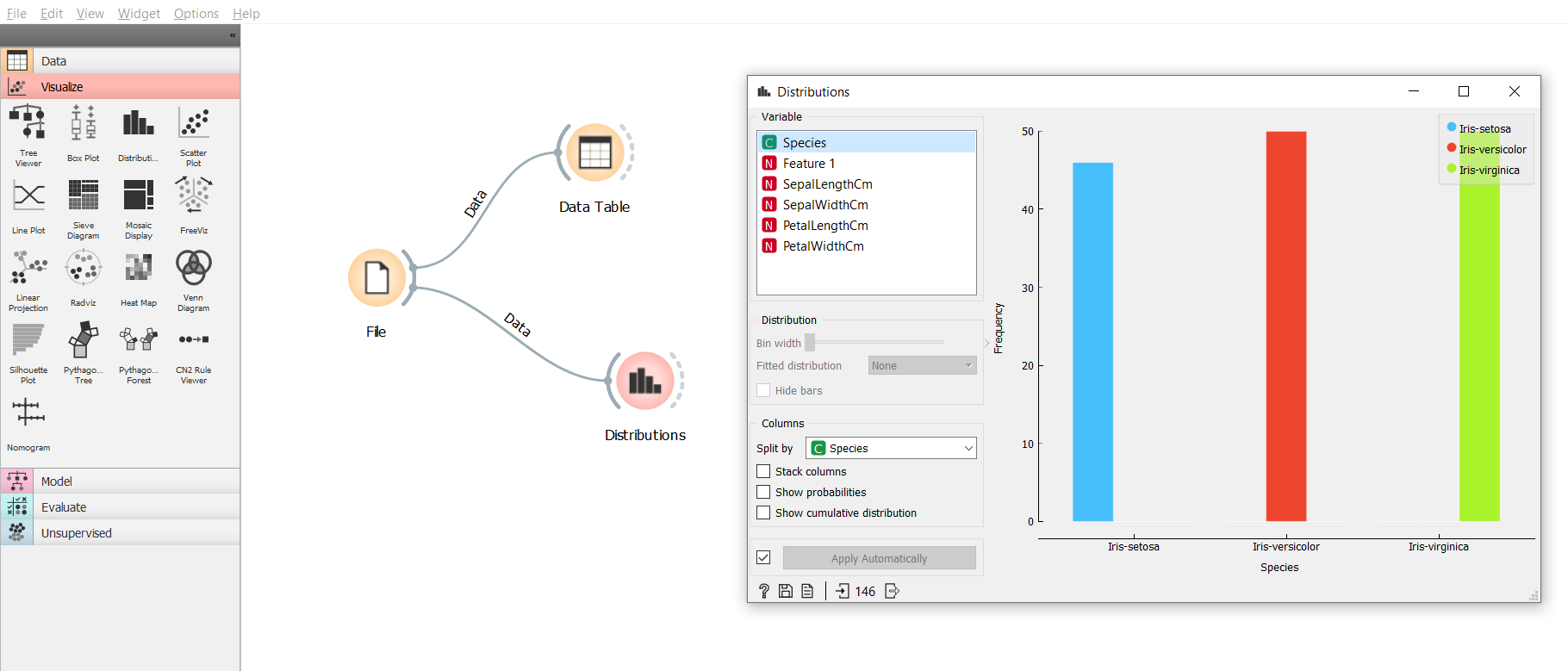
****

**1. Data distribution**

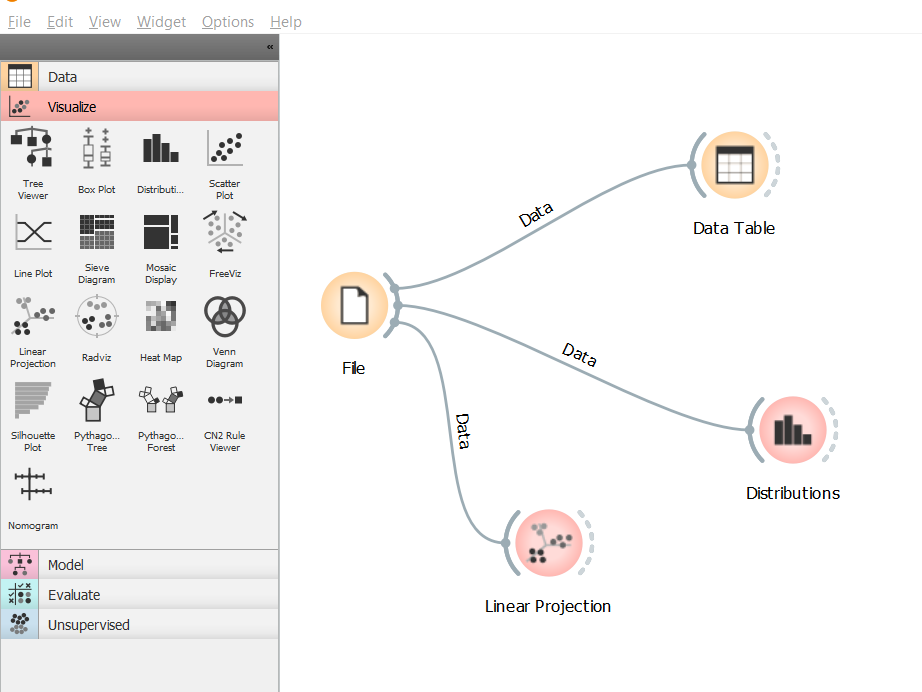
We used the File widget to include our .xlsx file and then reviewed it using the

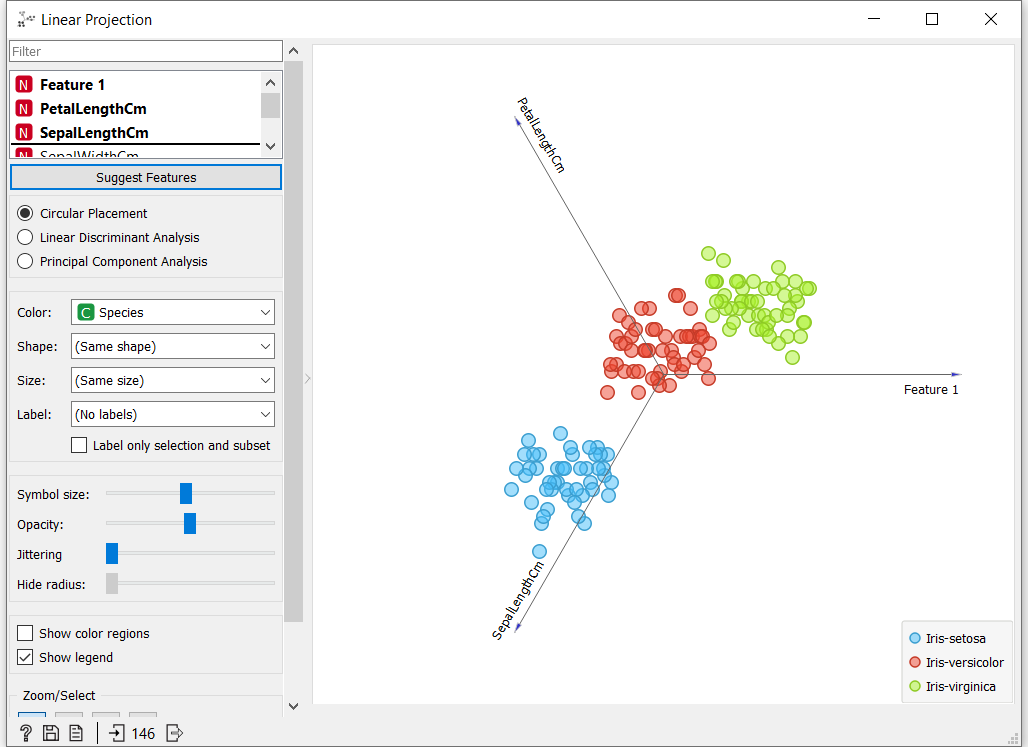
Data table widget. We get the following distribution.

****

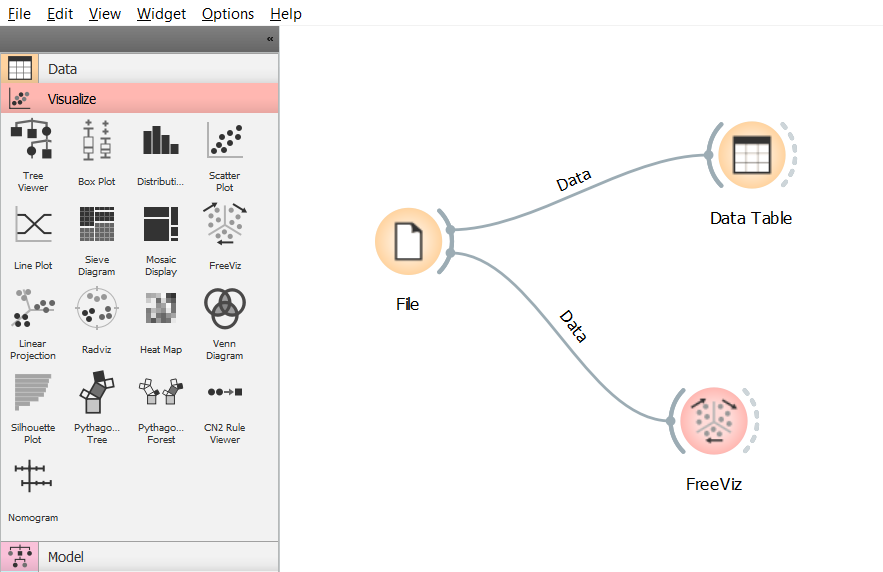
****

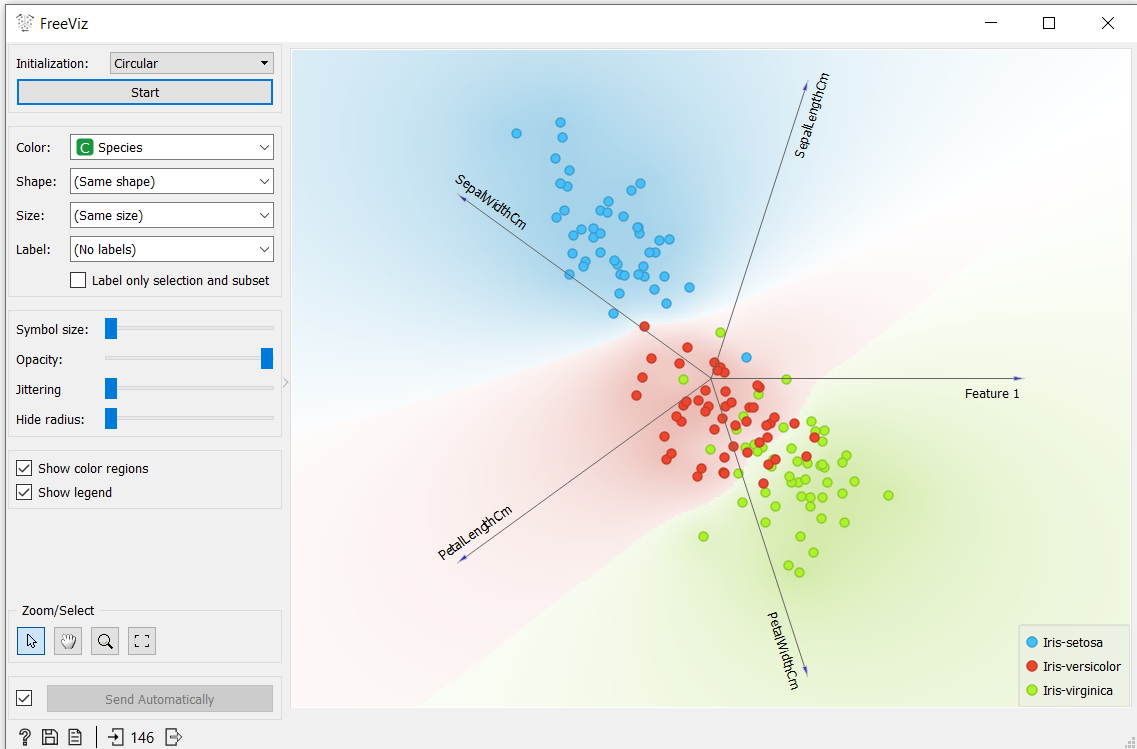
1. **Linear Projection**

****

****

**3. 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 like GNU (GPL) which allows you to edit source code and distribute it. Read these licenses and you will realize that these licenses are created to help us.

1. Coined by the development environments around software produced by open collaboration of software developers on the internet.
2. Later specified by the Open Source Initiative (OSI).
3. It does not explicitly state ethical values, besides those directly associated to software development

**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. But this license put some restrictions on users also like.

* Number of installations of this software into computers
* Restrictions on sharing of software illegally
* Time period up to which software will operate
* Number of features allowed to use

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

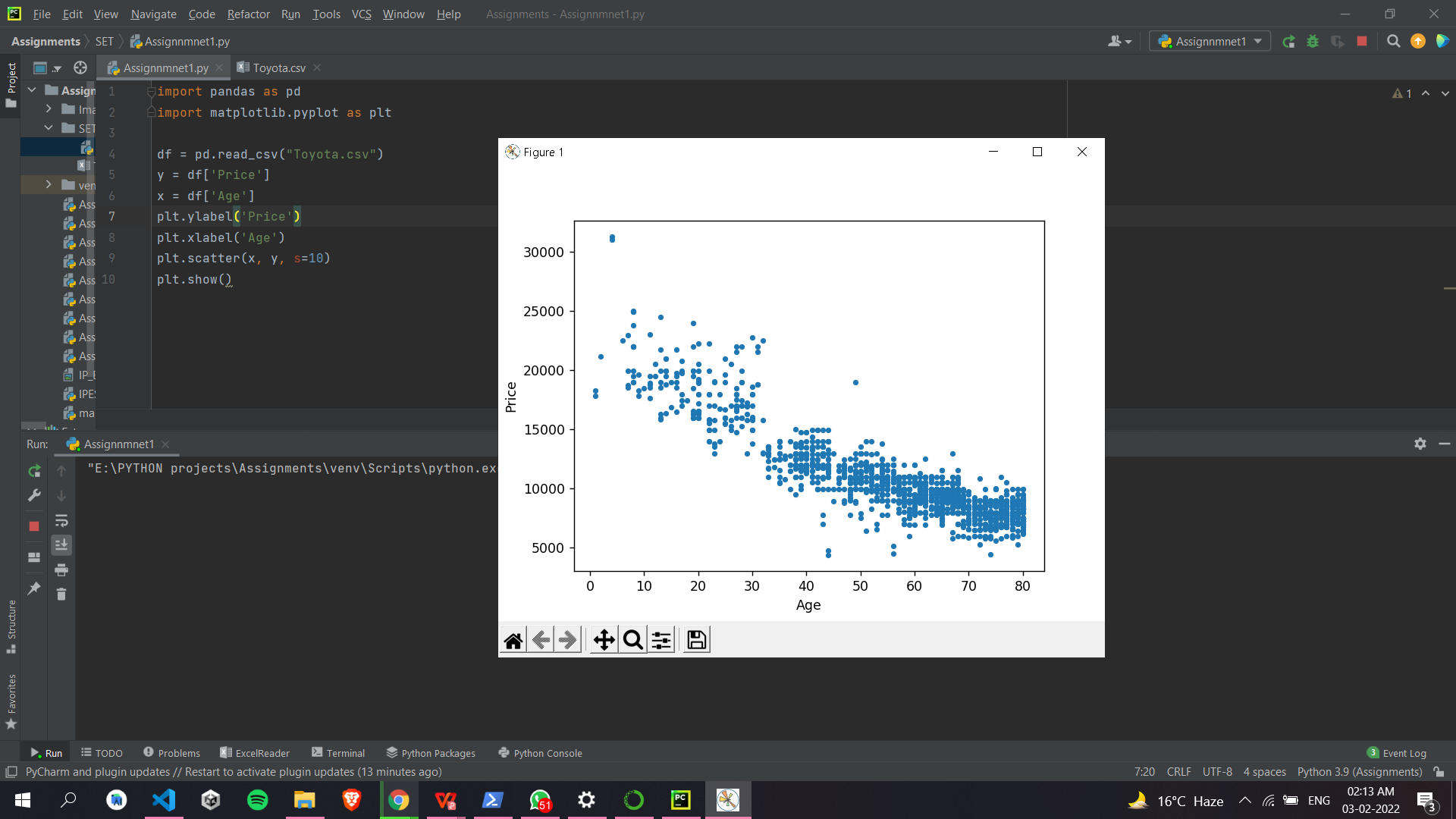
**given below.**

**Dataset- https://drive.google.com/file/d/1i11BZFe8Xj9kNq7eeE9KOa\_Iz1KhEdXJ/view**

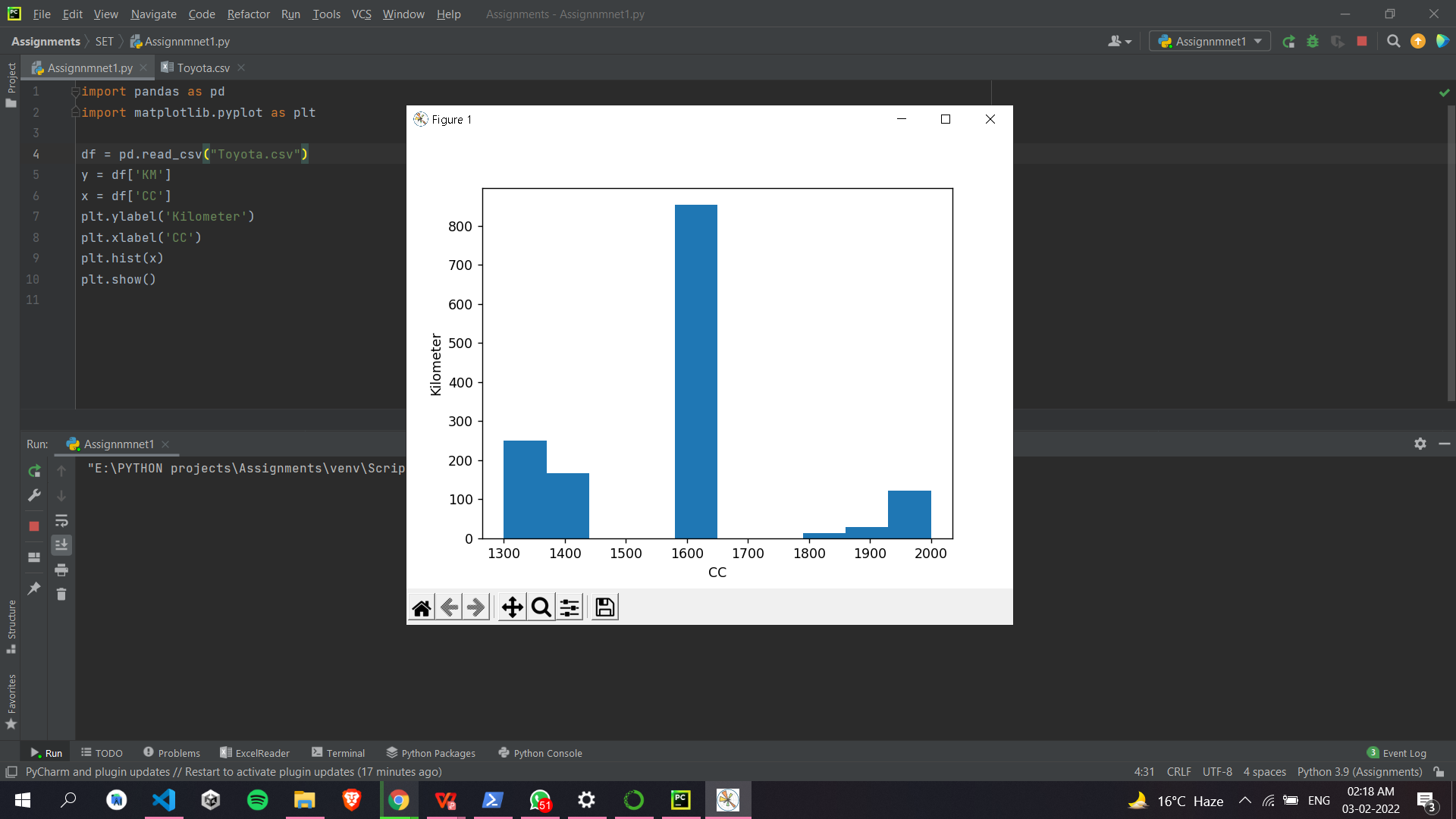
**a. Scatter plot- Scatter plot of Price Vs Age**

**b. Histogram- for Kilometer and CC**

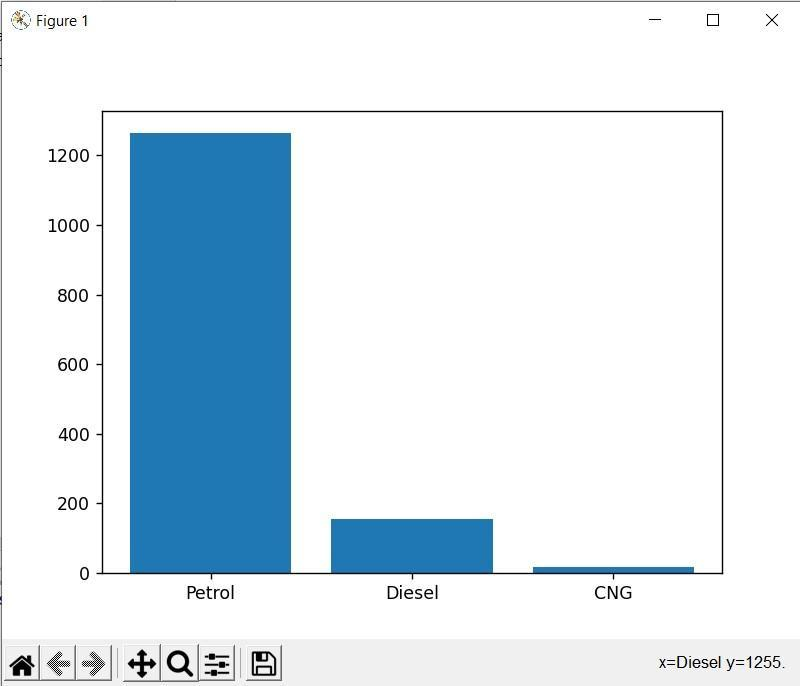
**c. Bar plot- Bar plot for different fuel types5. Enlist some examples along with its purpose and properties (at least 10) of FOSS and proprietary software with respect to database.**

****

1. **Histogram- for Kilometer and CC**

****

**3.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.**

1. FOSS

Examples : MySQL, SQLite, MongoDB

a) MySQL

● Relational Database Management System (RDBMS)

MySQL is a relational database management system. This database language is

based on the SQL queries to access and manage the records of the table.

● Easy to use

MySQL is easy to use. We have to get only the basic knowledge of SQL. We can

build and interact with MySQL by using only a few simple SQL statements.

● It is secure

MySQL consists of a solid data security layer that protects sensitive data from

intruders. Also, passwords are encrypted in MySQL.

● Client/ Server Architecture

MySQL follows the working of a client/server architecture. There is a database

server (MySQL) and arbitrarily many clients (application programs), which

communicate with the server; that is, they can query data, save changes, etc.

● Free to download

MySQL is free to use so that we can download it from MySQL official website

without any cost

● It is scalable MySQL supports multi-threading that makes it easily scalable. It can handle

almost any amount of data, up to as much as 50 million rows or more. The default

file size limit is about 4 GB. However, we can increase this number to a theoretical

limit of 8 TB of data.

● Speed

MySQL is considered one of the very fast database languages, backed by a

large number of the benchmark test.

● High Flexibility

MySQL supports a large number of embedded applications, which makes

MySQL very flexible.

● Compatible on many operating systems

MySQL is compatible to run on many operating systems, like Novell NetWare,

Windows\* Linux\*, many varieties of UNIX\* (such as Sun\* Solaris\*, AIX, and

DEC\* UNIX), OS/2, FreeBSD\*, and others

● Allows roll-back

MySQL allows transactions to be rolled back, commit, and crash recovery.

● Memory efficiency

Its efficiency is high because it has a very low memory leakage problem.

b) MongoDB

1. Support ad hoc queries

In MongoDB, you can search by field, range query and it also supports regular

expression searches.

2. Indexing

You can index any field in a document.

3. Replication

MongoDB supports Master Slave replication.