



fundamental of data mining

Fall 2021



Date of submission: ­­­­­­­­20 – Nov – 2021 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**TEACHER PROFORMA**

Name: **­­­­­Ms. Ambreen**

Father’s Name: **Shamsuddin Hamal Khan**

**Date of Birth:**

Date: **04** Month: **August** Year: **1992**

Nationality: **Pakistani**

Qualification: **MS Computer Science**

**Contact Details:**

Tel #:­­­­­­­­­­­­­­­­­ **+923004991229**

Residence: **Gulshan-e-Iqbal Karachi**

Email ID: **ambreen.khan@bbsul.edu.pk**

**STUDENT PROFORMA**

Name: **­­­­­Daniyal Ahmed**

Father’s Name: **Ghulam Haider**

Section: **6th semster A**

Batch: **10th**

Roll #: **661**

**Date of Birth:** Date: **18** Month: **November** Year: **1999**

Nationality: **Pakistani**

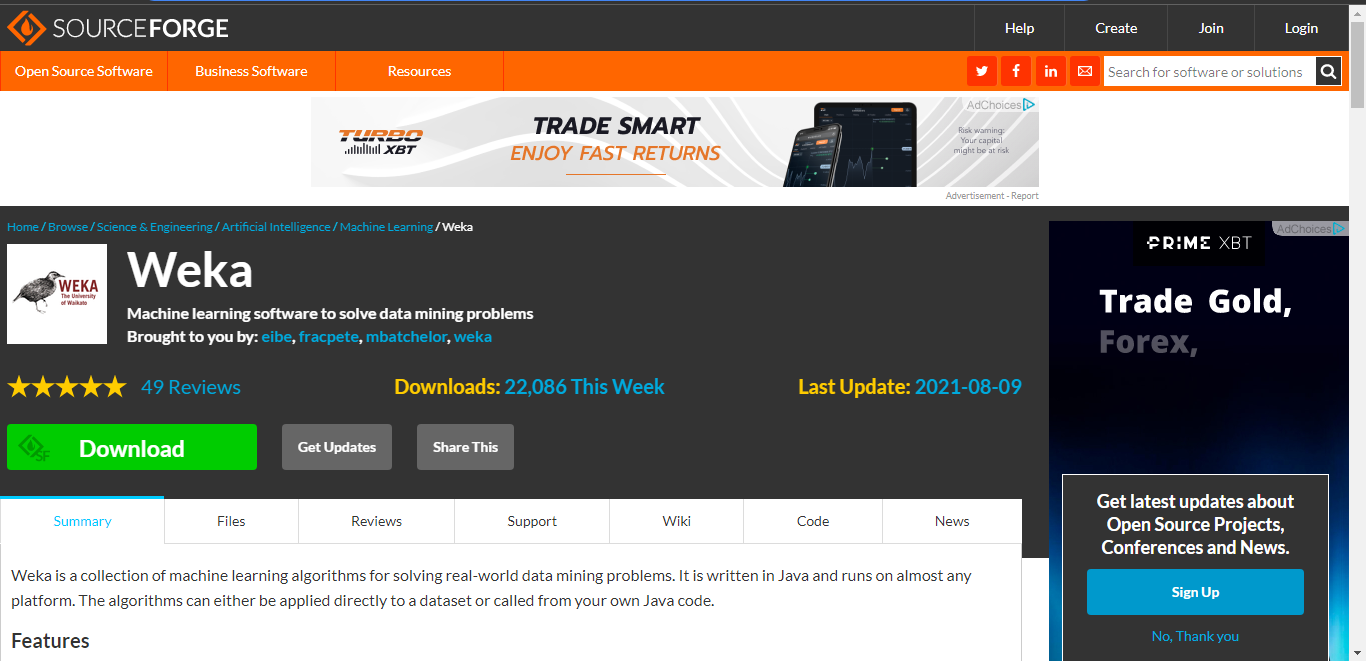
Address: **Lea market near khoja jamat khana hussain plaza 5th floor lyari karachi**

Email ID: **Daniahmedkhatri@gmail.com**

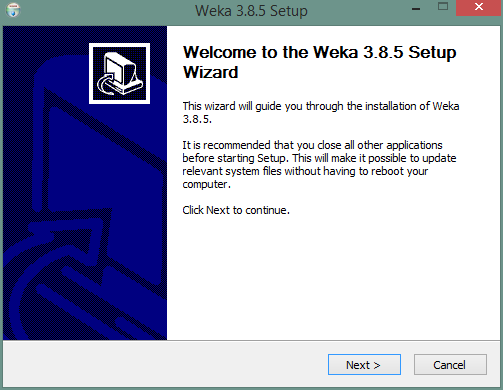
**TABLE OF CONTENT**

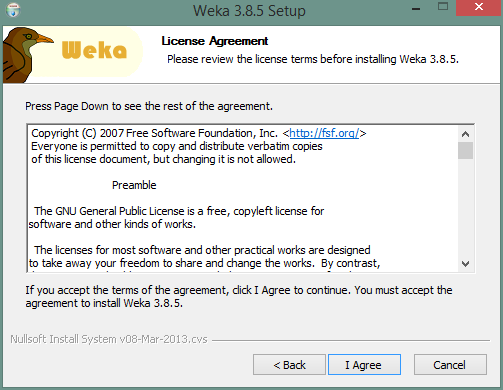
|  |  |  |
| --- | --- | --- |
| S.NO: | TASKS | PAGE # |
| 1 | Install WEKA and add screenshots of the installation steps. | 5 |
| 2 | Check each feature of WEKA and add screenshots. | 8 |
| 3 | Complete interface of WEKA with screenshots. | 11 |
| 4 | Import dummy dataset from KAGGLE and add screenshots. Also describe some features of KAGGLE briefly, and add screenshots of your work. | 14 |
| 5 | Import dummy dataset from UCI and add screenshots. Also describe some features of UCI briefly, and add screenshots of your work. | 16 |
| 6 | Import downloaded dataset from your device in WEKA and add screenshots of all steps. | 18 |
| 7 | Import any dataset in WEKA and perform preprocessing. Also add screenshots of the output. | 20 |
| 8 | Import any dataset in WEKA and perform Classification. Also add screenshots of the output. | 22 |
| 9 | Import any dataset in WEKA and perform Clustering. Also add screenshots of the output. | 24 |
| 10 | Import any dataset in WEKA and perform Decision Tree and Naïve Bayes Algorithms. Also add screenshots of the outputs. | 28 |

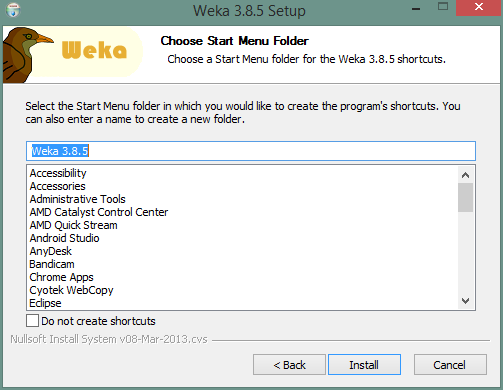
Lab #1

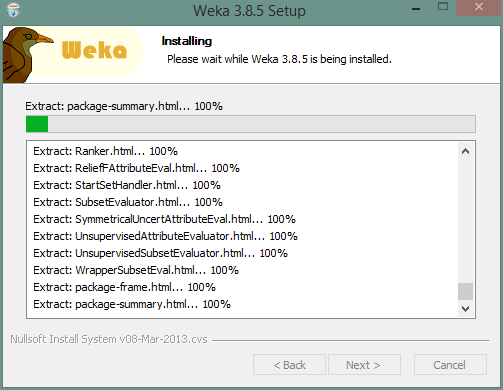
1 – Got to the [link](https://sourceforge.net/projects/weka/) and click on download:  


2 – After downlaod get completed click on the downloaded file and install it by double clicking it...then click next

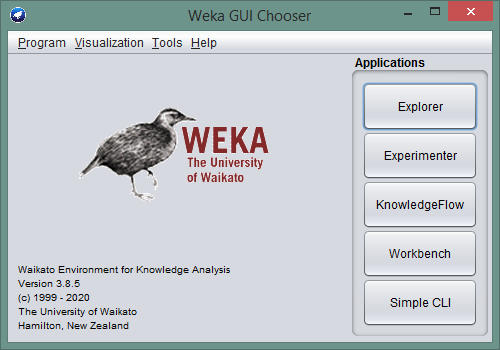


3 – After clicking next…click agree and then next  


4 – After clicking next…next  


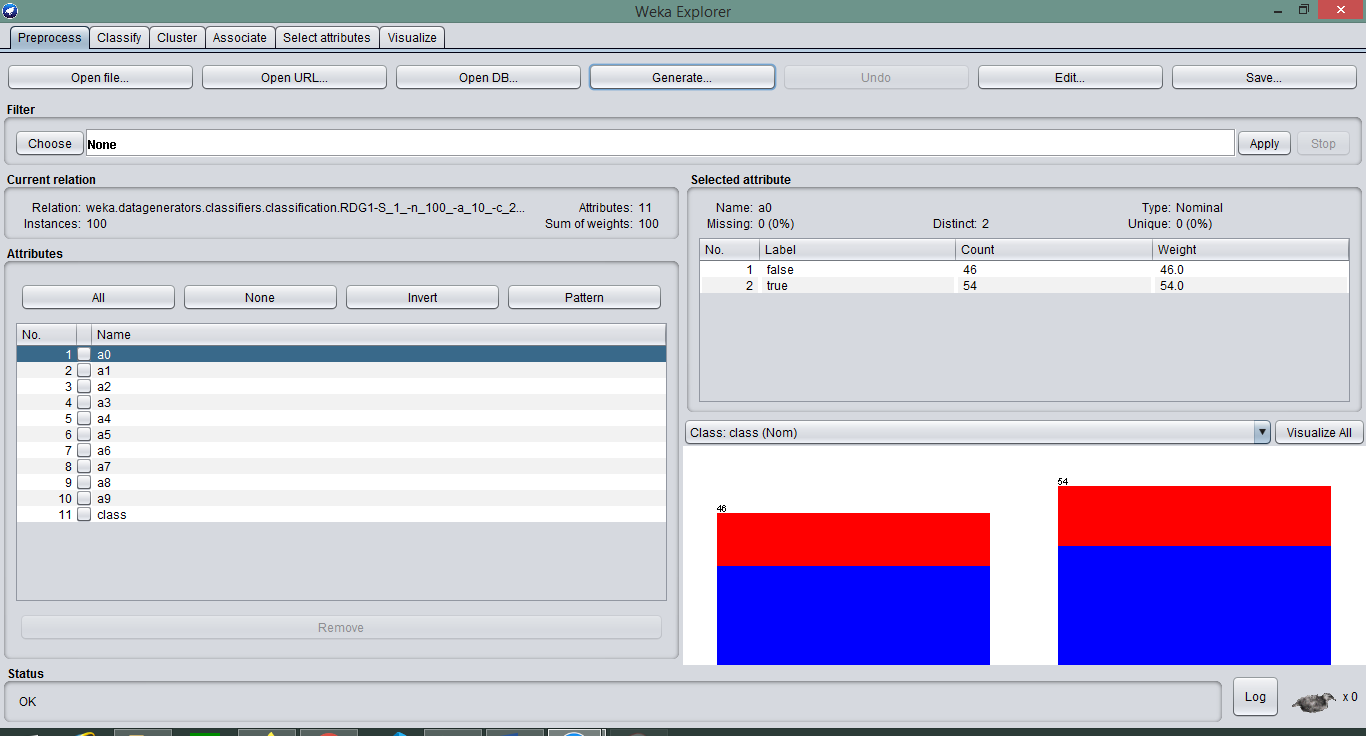
5 – After clicking next…it will start installing   


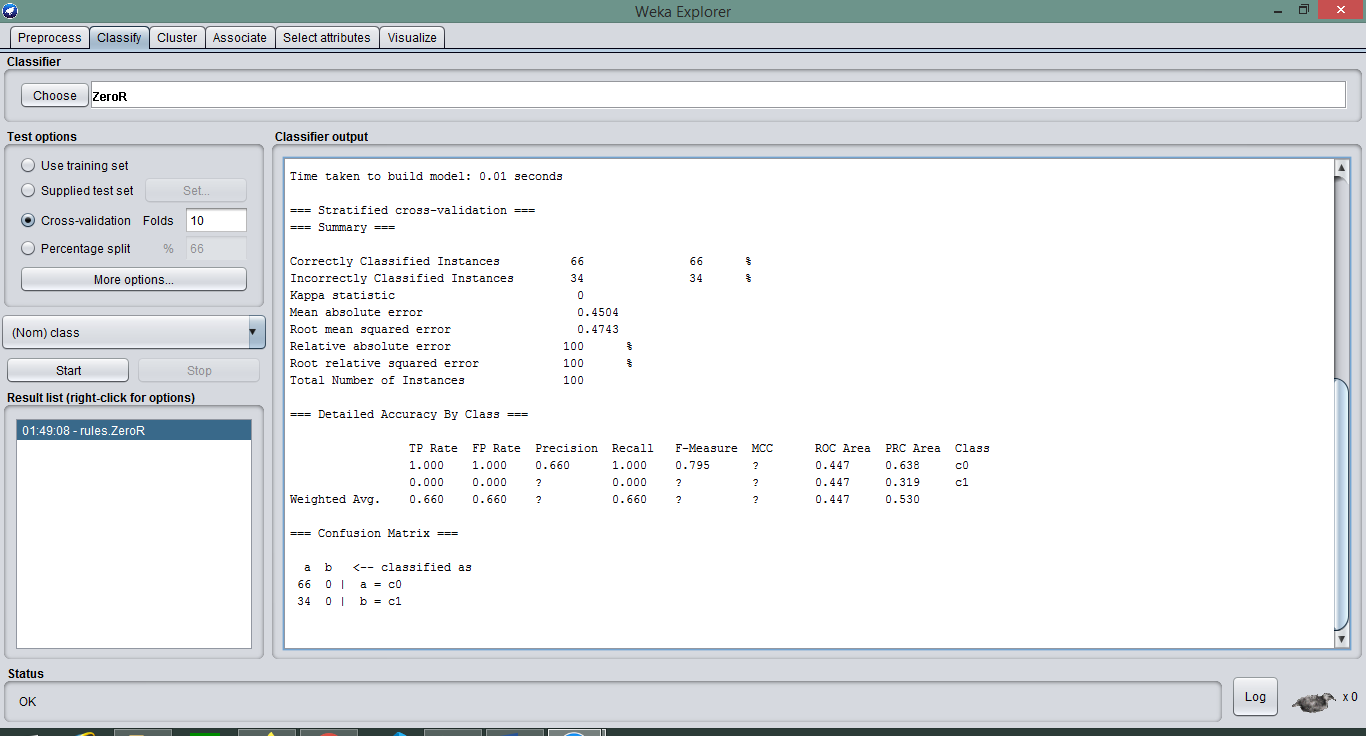
6 – After installing …it will start the **Weka Tool**

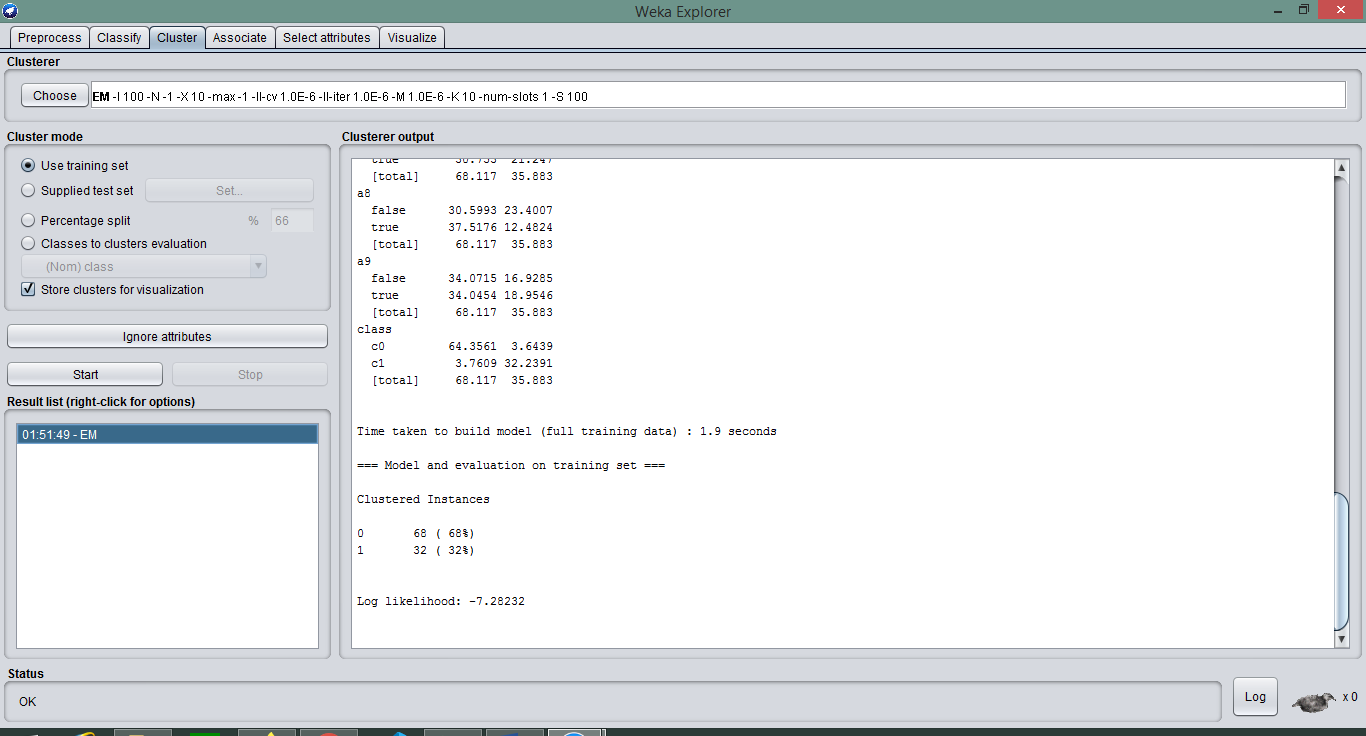


Lab #2

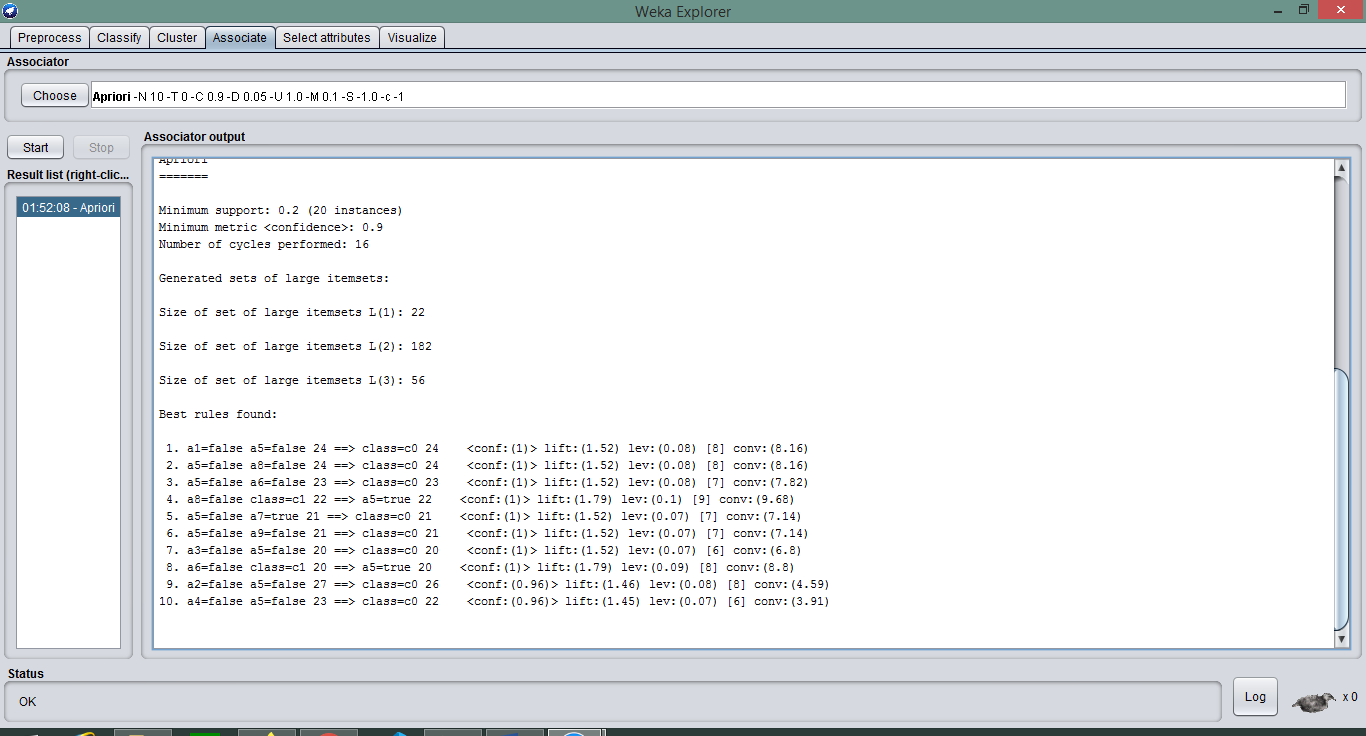
1 – In this lab we will explore all the features of **Explorer**. Like preprocessor, Classification, Association, Clustering, Select Attributes and visualization.

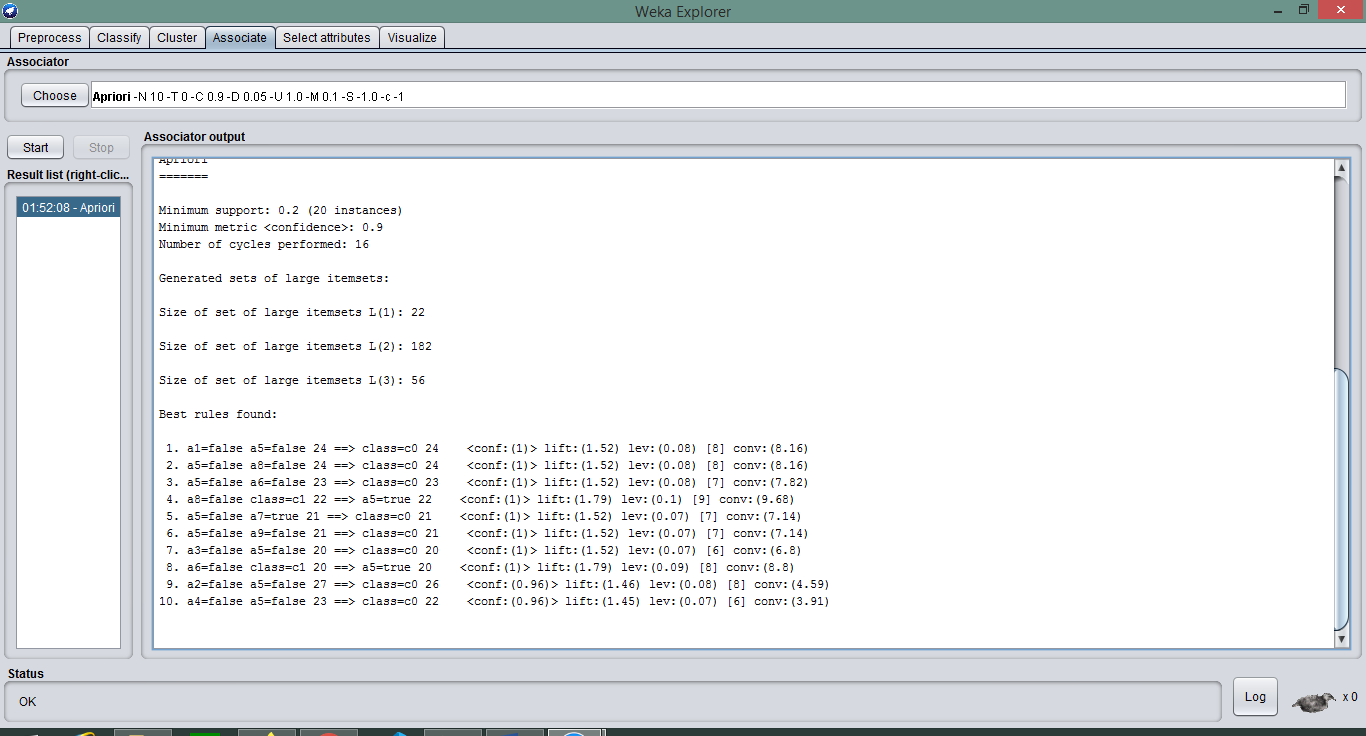
2 – First option is preprocessor… in which we will select the data from source. 

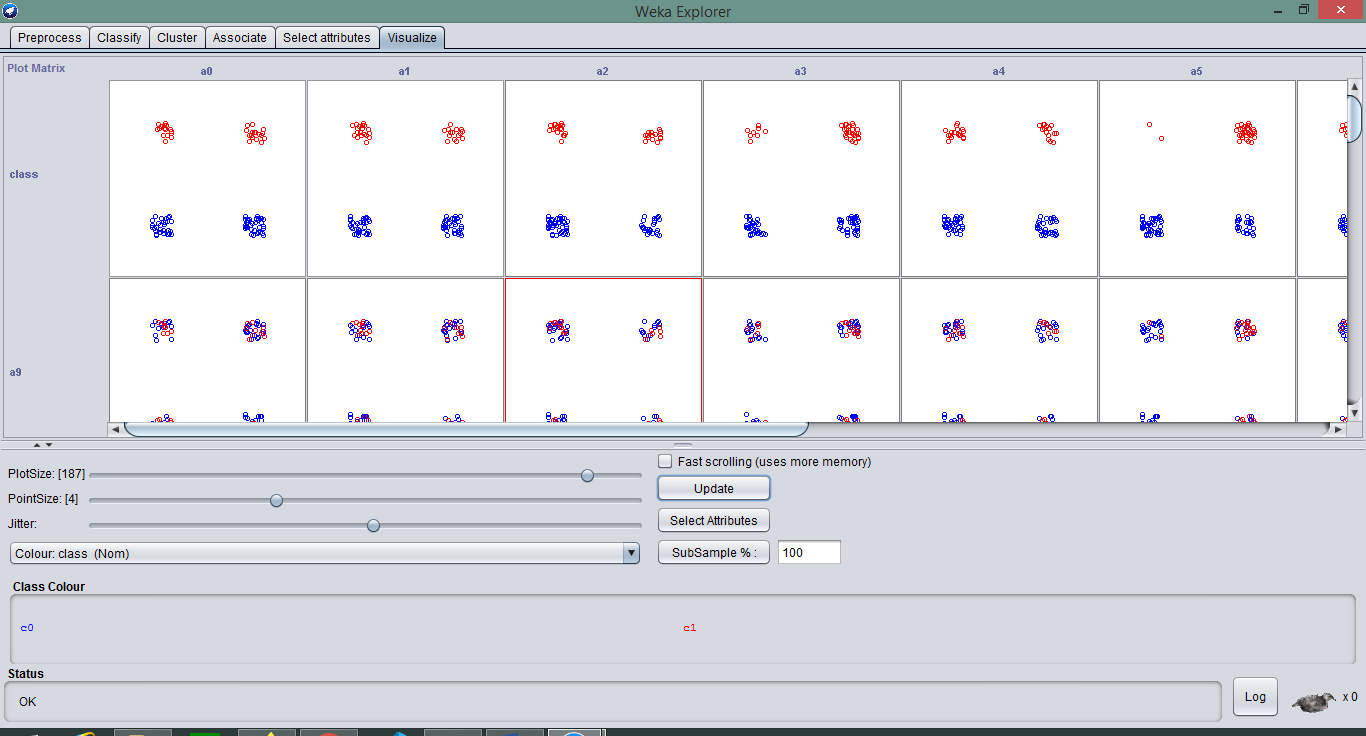
3– Second option is Classification… in which it classify the data.

3– Third option is Cluster… in which it create Cluster with the input data.  


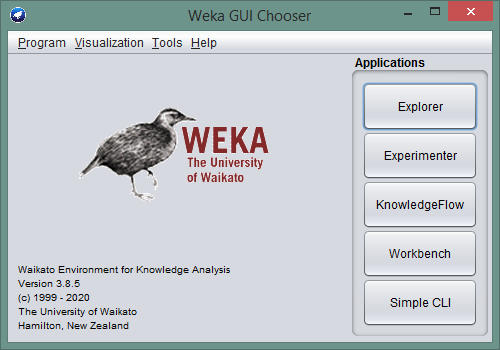
4 – Fourth option is Associate… in which it create Association with the data.

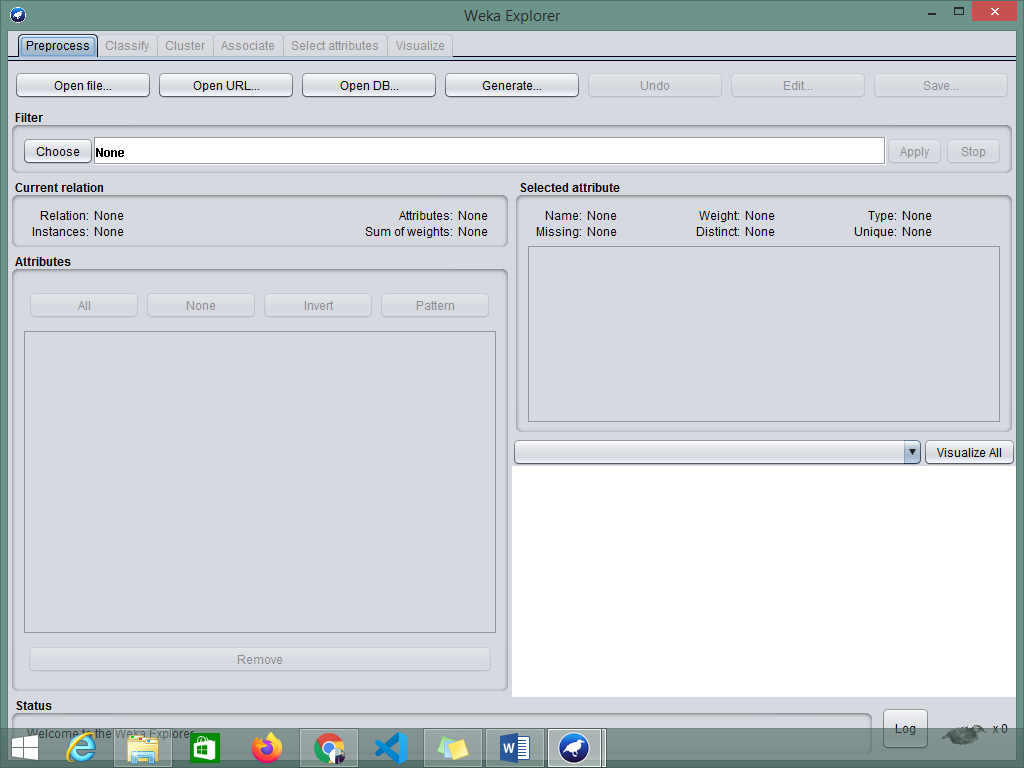


5 – Fifth option is Select Attriubute… in which it only Select specific attributes and process data.  


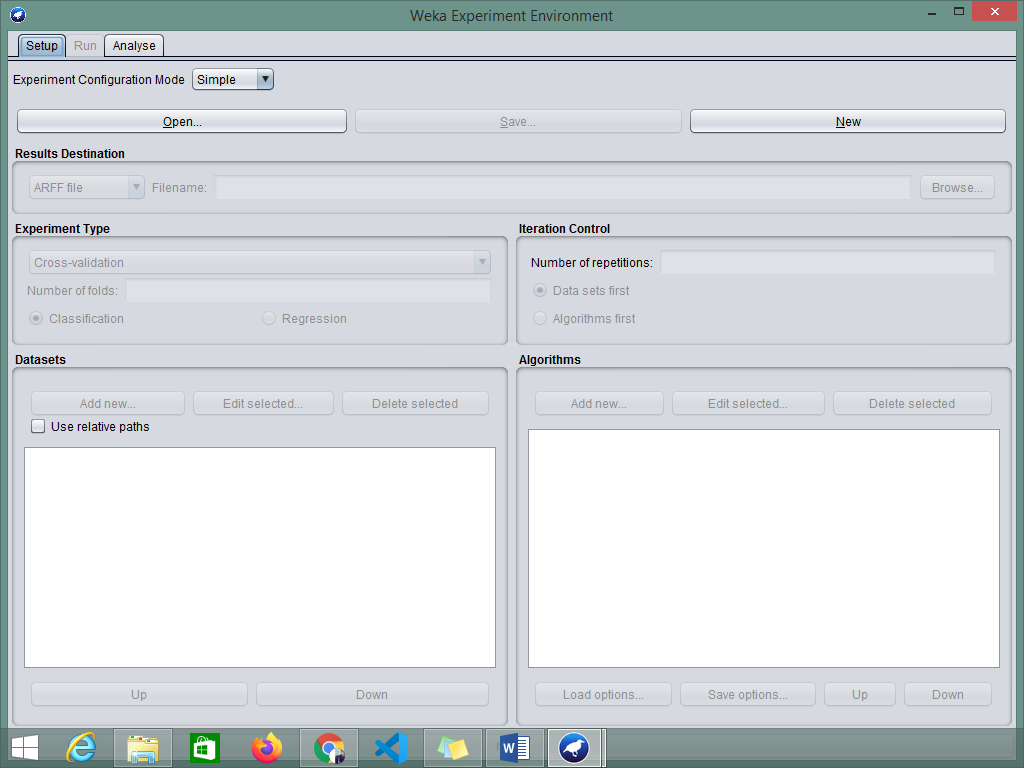
6 – Sixth option is Visualise… in which it visualize the data.  


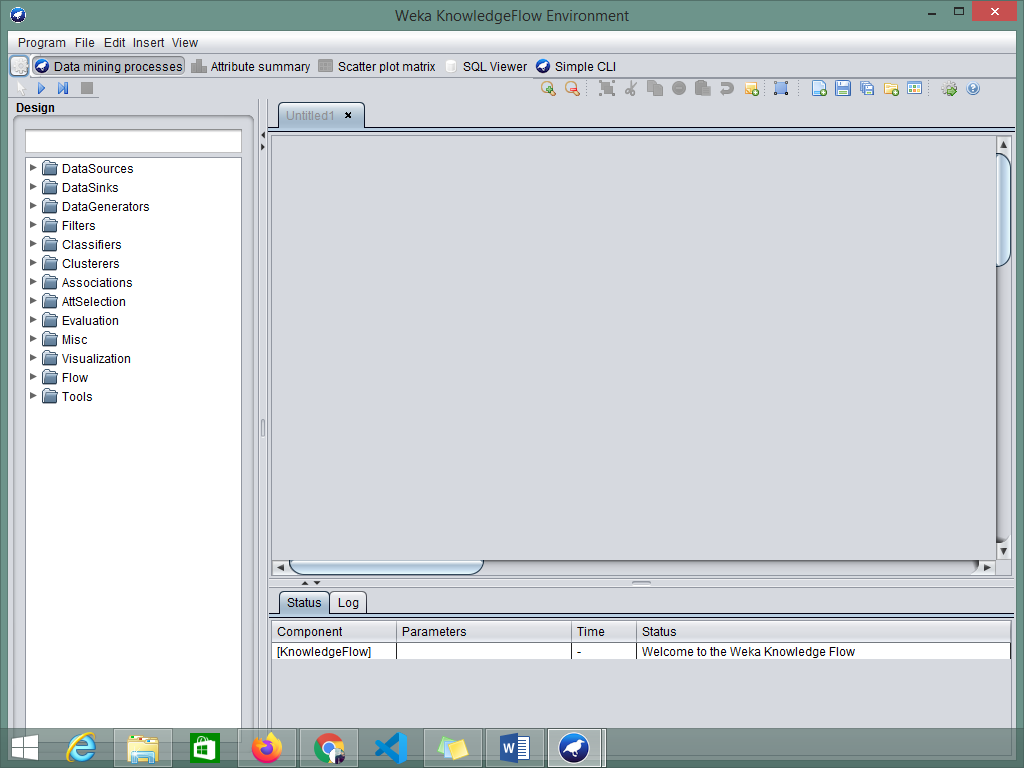
Lab #3

1 – In this lab we will explore all the main UI buttons feature  


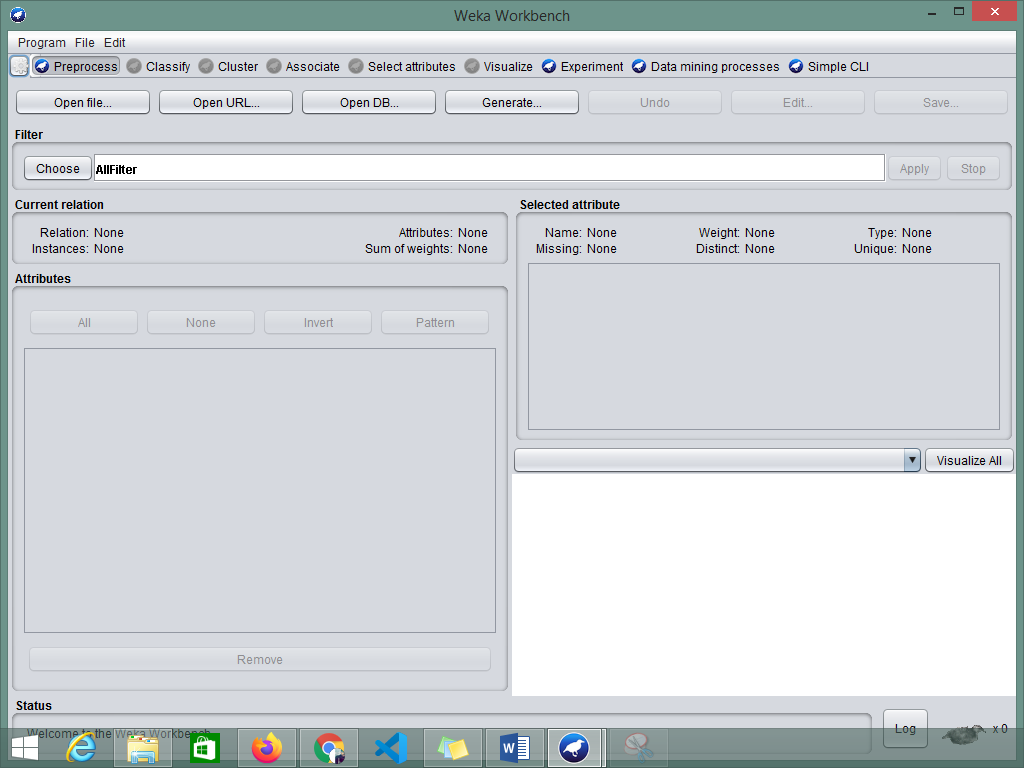
2 – First click on **Explorer**..itwill open a window which prompts for data to be input.  
This window have all the feature to evaluate the Data.

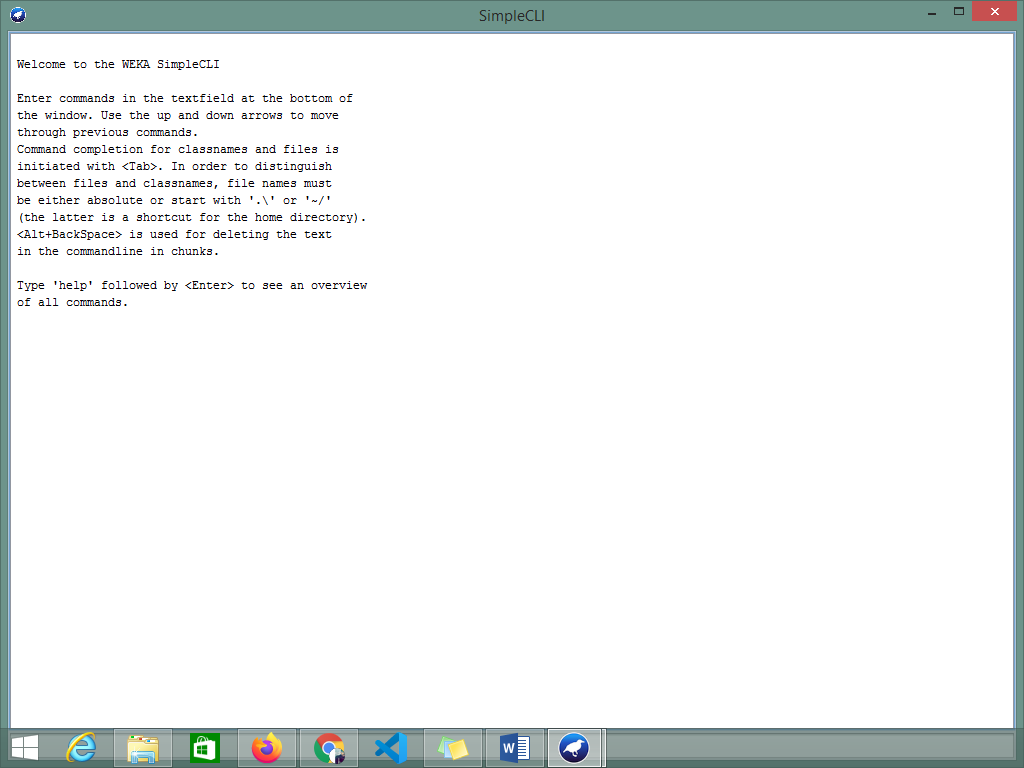
3 – Now click on **Experimenter**..it will open a window in which there is 2 tabs for setup and Analyse.



4 – Now click on **KnowledgeFlow**..it will open a window in which there is design tools for data manipulation and analysis.

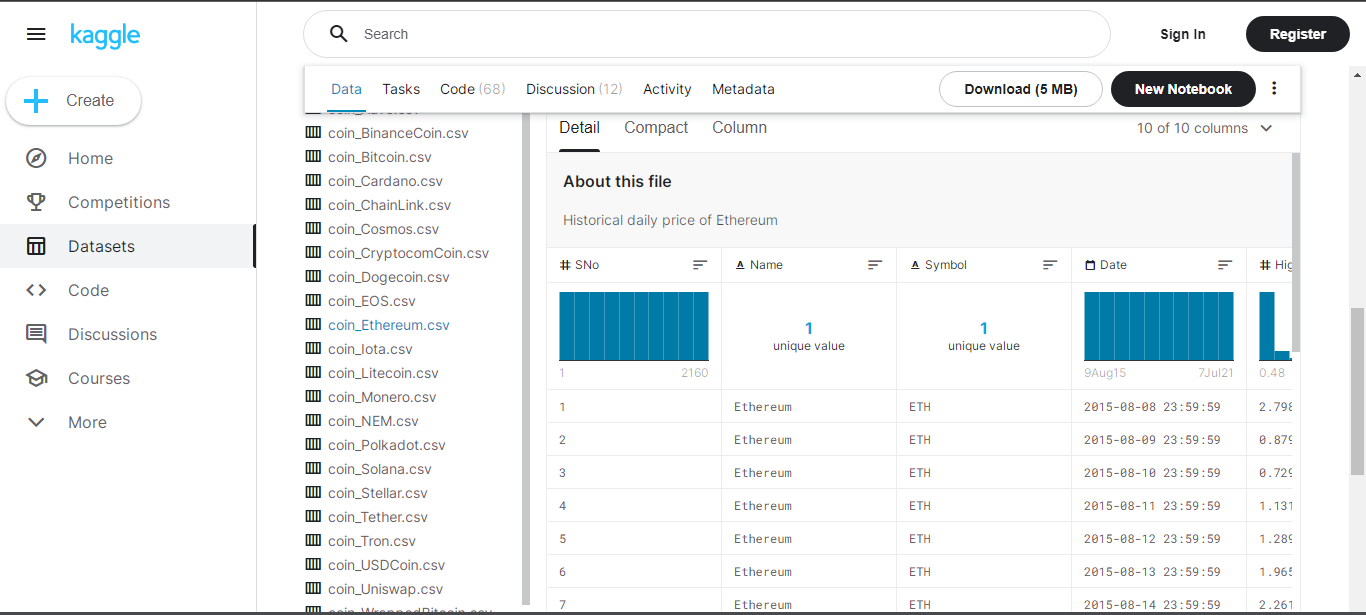
4 – Now click on **Workbench**..it will open a window in which we can mine and perform multiple task like explorer.

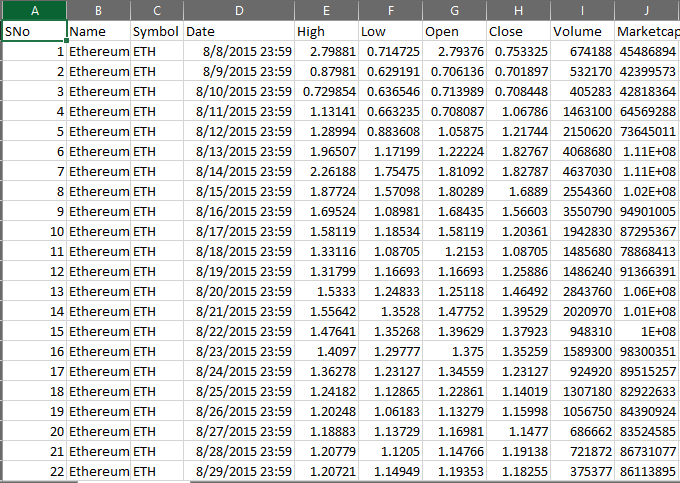


5 – Now click on **Simple CLI**..it will open a window in which we can use commands to perform task. 

Lab #4

**1 – Download the dataset from Kaggle?**Downloaded the data from Kaggle



**2 – Manage a file to describe downloaded dataset.**The data file contains the information about the Ethererium price history data which can be helped to predict the future value of ETH. This data has 10 Attributes and 2161 instances.

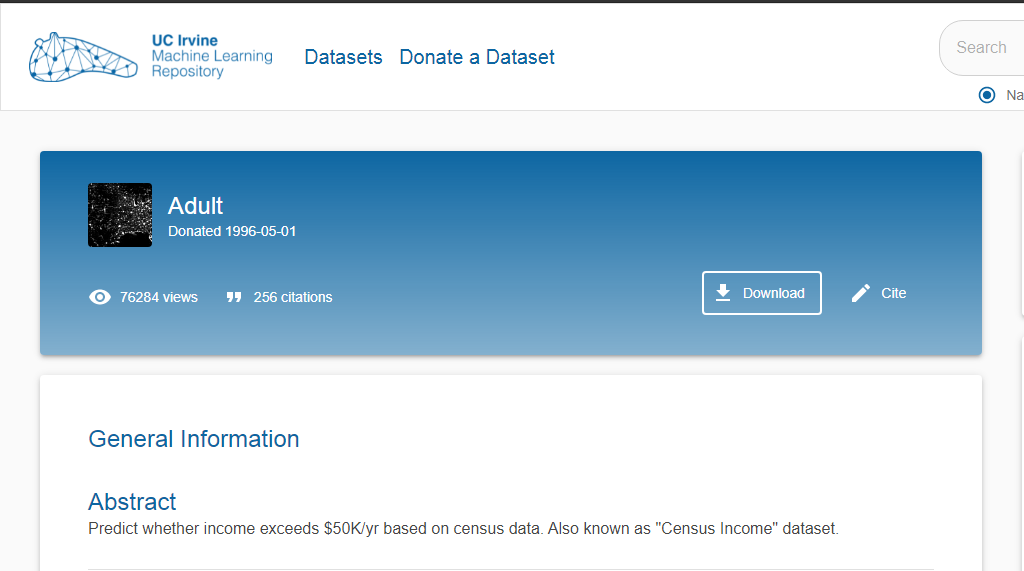
**Instances:** 2161

**Values:**

1. S.no – To count no of values
2. Name – Coin name
3. Date – price on that date
4. High – The highest price on that date
5. Low - The Lowest price on that date
6. Open – The price on which market open
7. Close - The price on which market closes
8. Volume – The number of coins traded on that date
9. Market cap – The number of coins present in market

Lab #5

**1 – Download the dataset from UCI?**Downloaded the data from UCI.



**2 – Manage a file to describe downloaded dataset.**The data file contains the information about the income and census data of adults which can be helped to predict whether income exceeds $50K/yr based on census data. Also known as "Census Income" dataset which has 15 Attributes and 48842 instances.  
  
**3 – Must check in detail like (attribute, instance, values, domain etc.)**

**Attributes:**

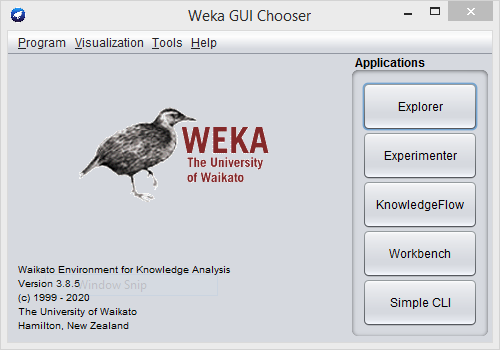
1. Age
2. Work class
3. Fnlwgt
4. Education
5. Education-num
6. marital-status:
7. occupation
8. relationship
9. race
10. sex
11. capital-gain
12. capital-loss
13. hours-per-week
14. native-country

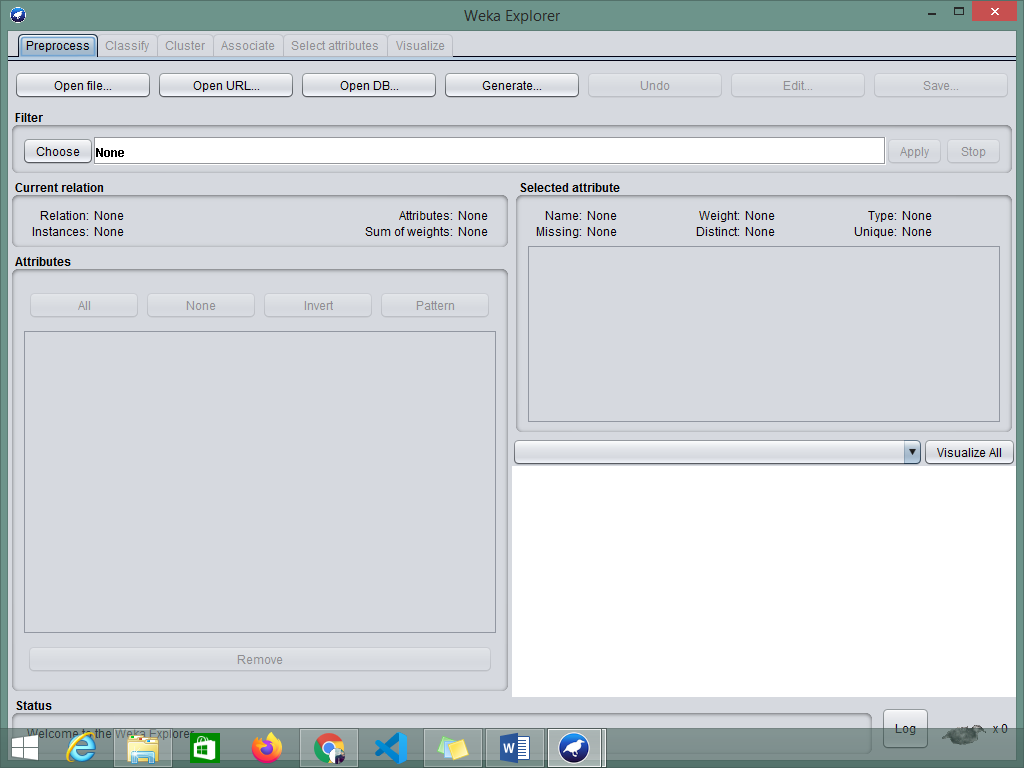
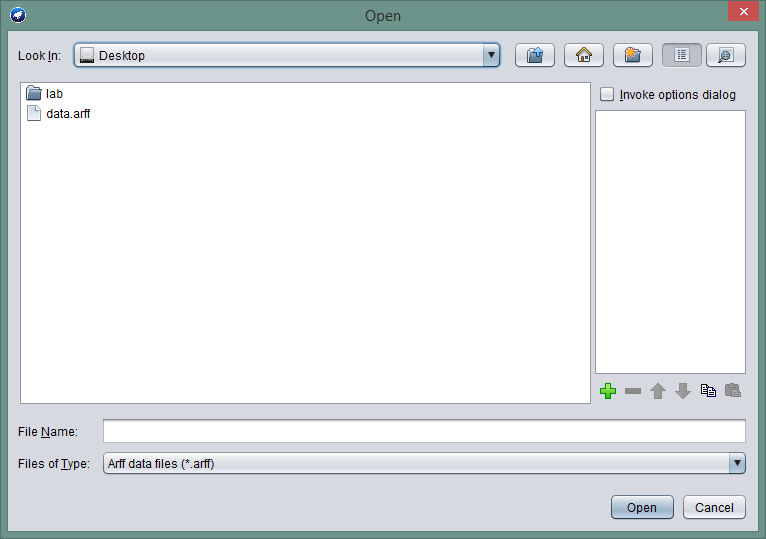
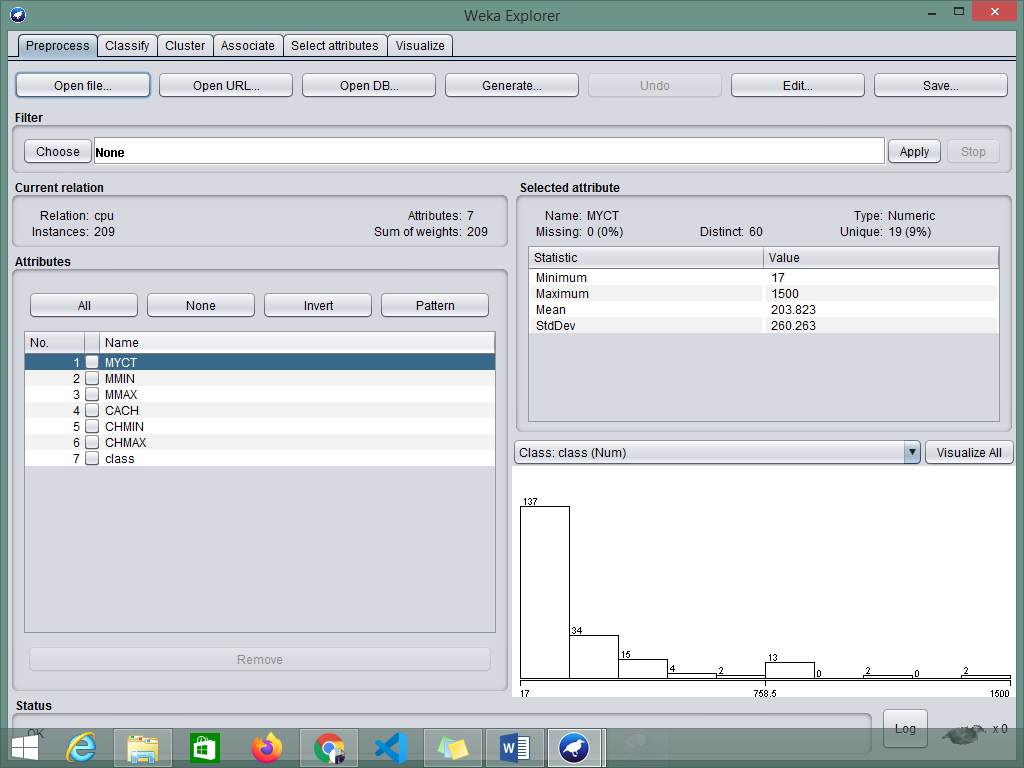
**Instances:** 48842

**Values:**

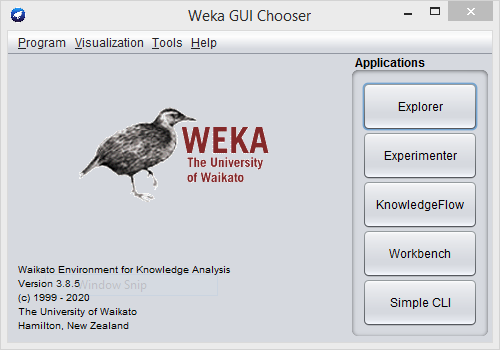
1. continuous
2. Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.
3. Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.
4. Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.
5. Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.
6. Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.
7. White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.
8. Female, Male.
9. United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc.), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad & Tobago, Peru, Hong, Holland-Netherlands.

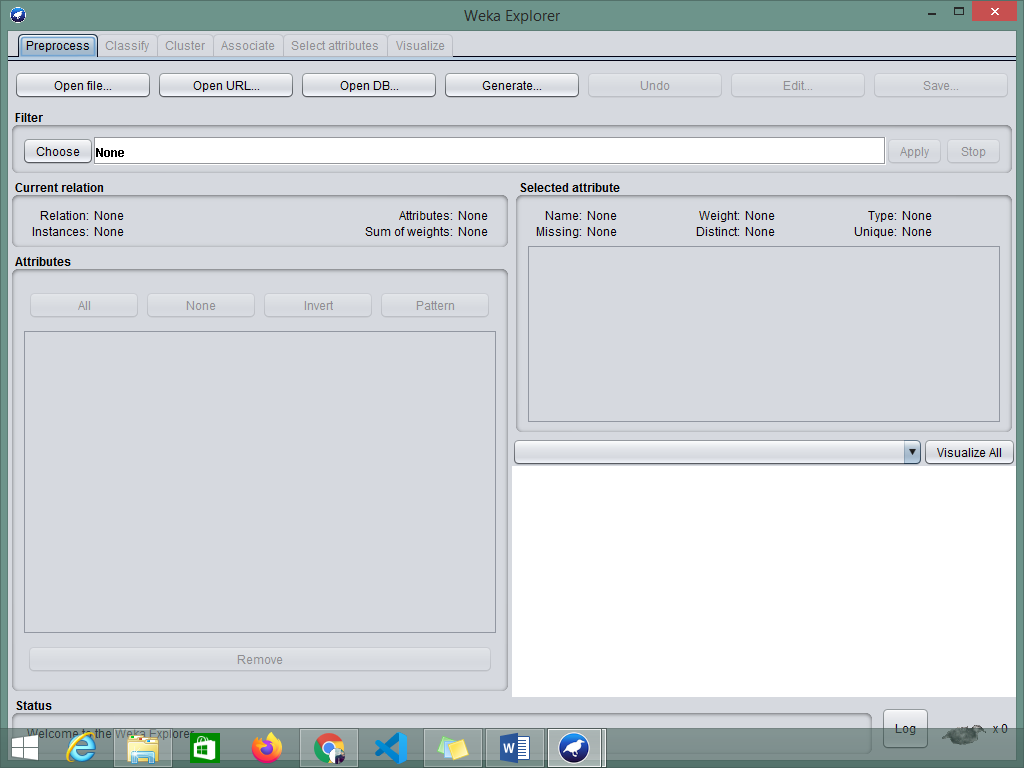
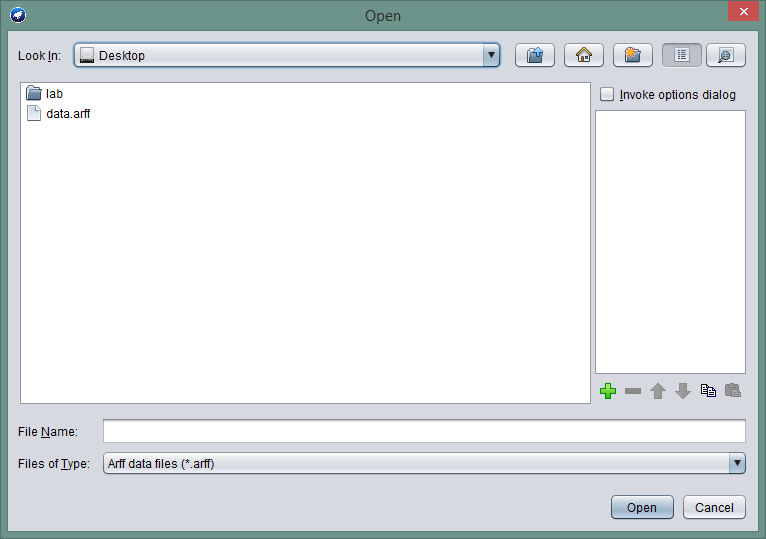
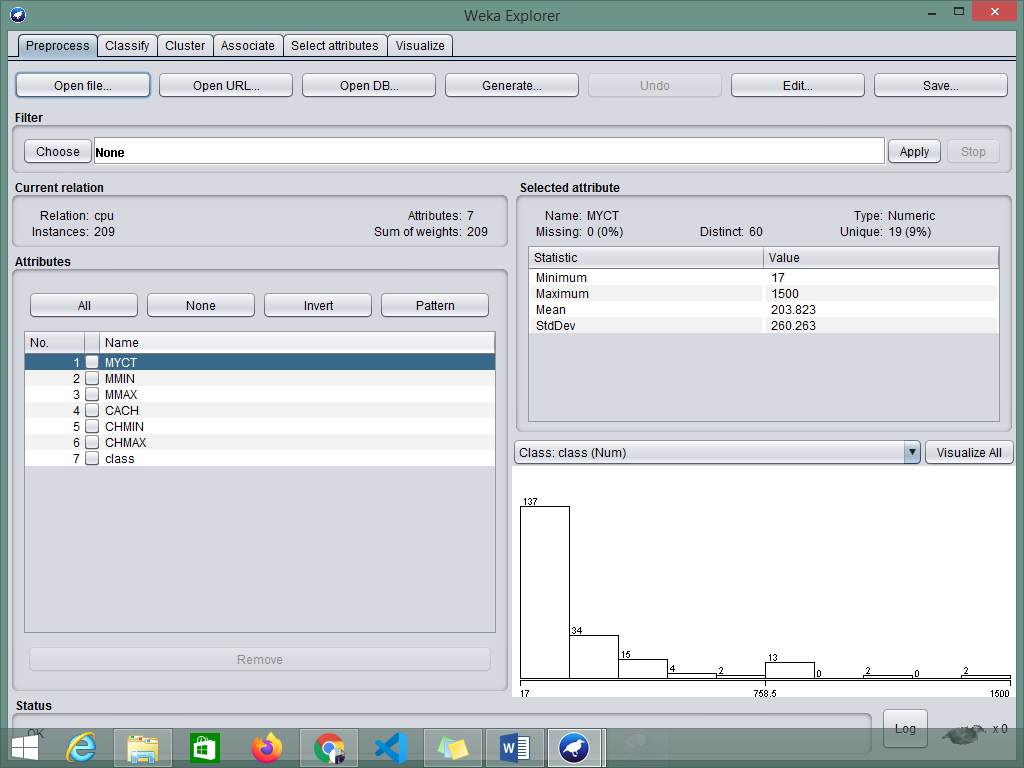
Lab #6

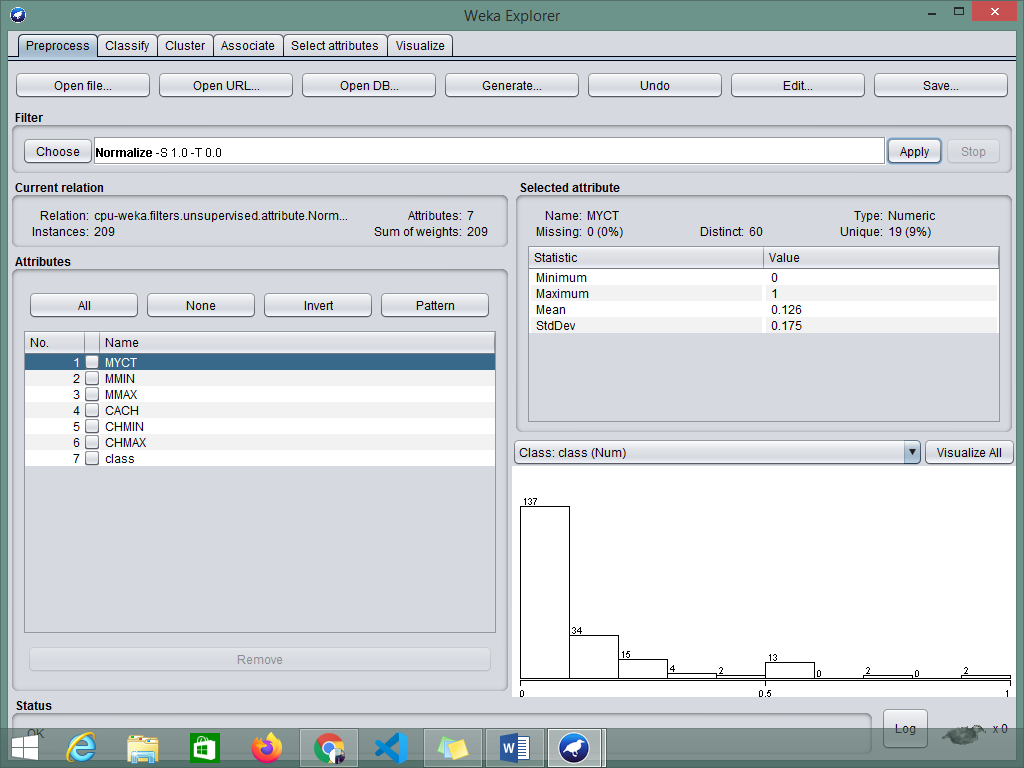
1 – In this lab we will import the downloaded data to weka first open the weka tool and press explorer.  


2 – After clicking on explorer click on open file.  
3 – After clicking on open file select the desire file for data you want to be mined.  
  
4 – After clicking on file the desired file will be opened.  


Lab #7

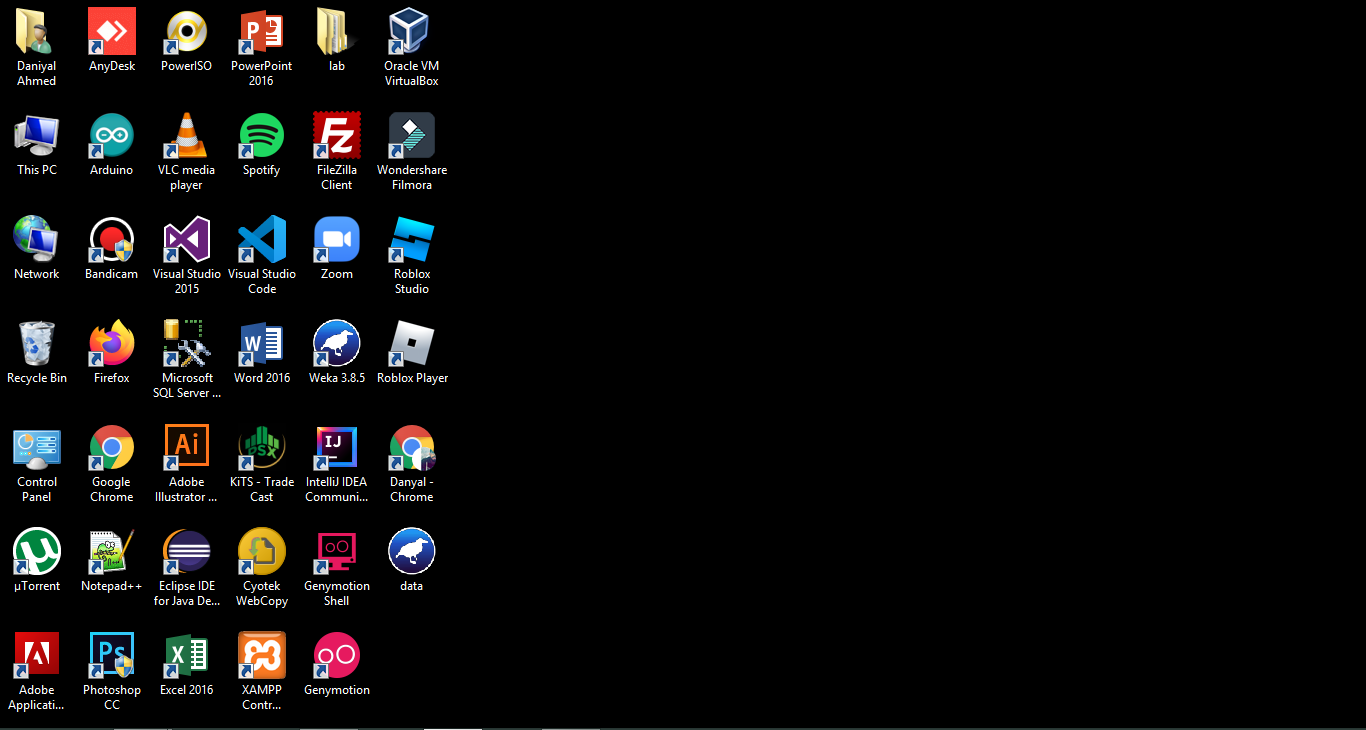
1 – In this lab we will import the downloaded data to weka first open the weka tool and press explorer.  


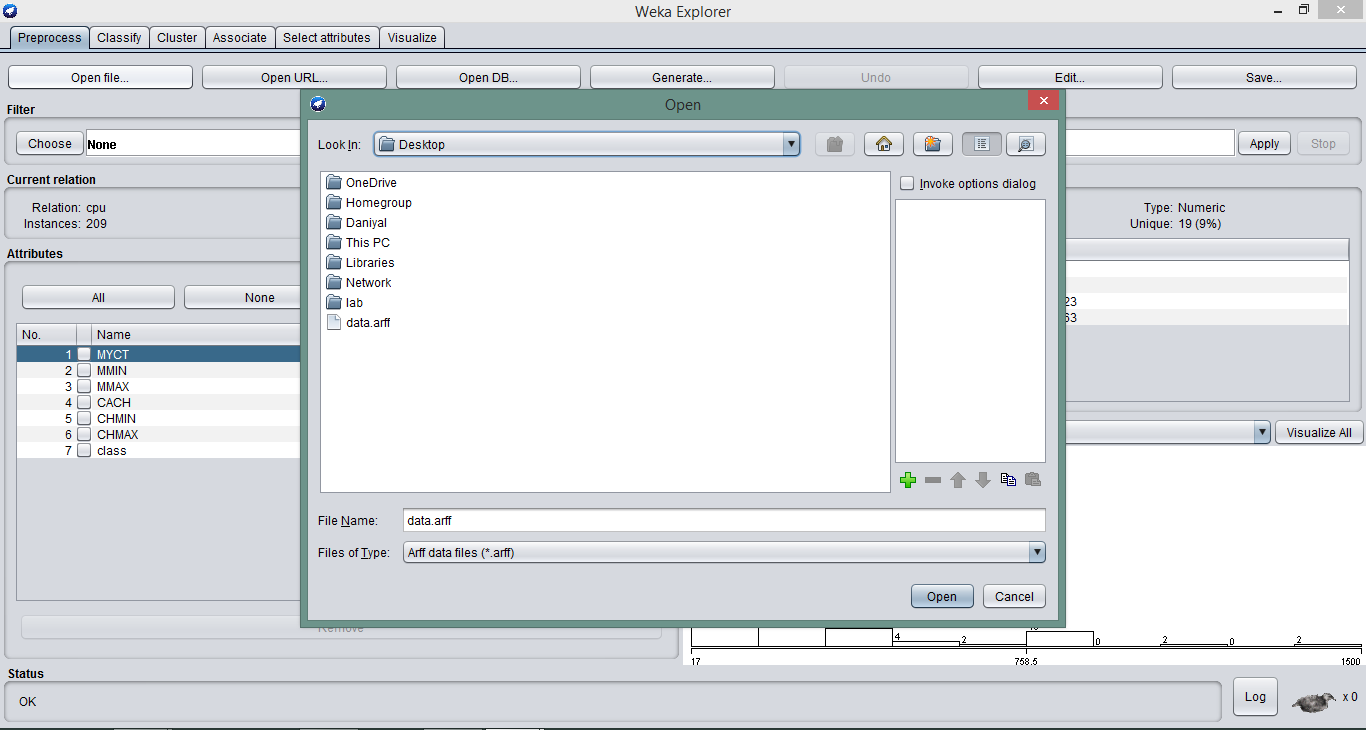
2 – After clicking on explorer click on open file.  
3 – After clicking on open file select the desire file for data you want to be mined.  
  
4 – After clicking on file the desired file will be opened and apply preprocessing.  


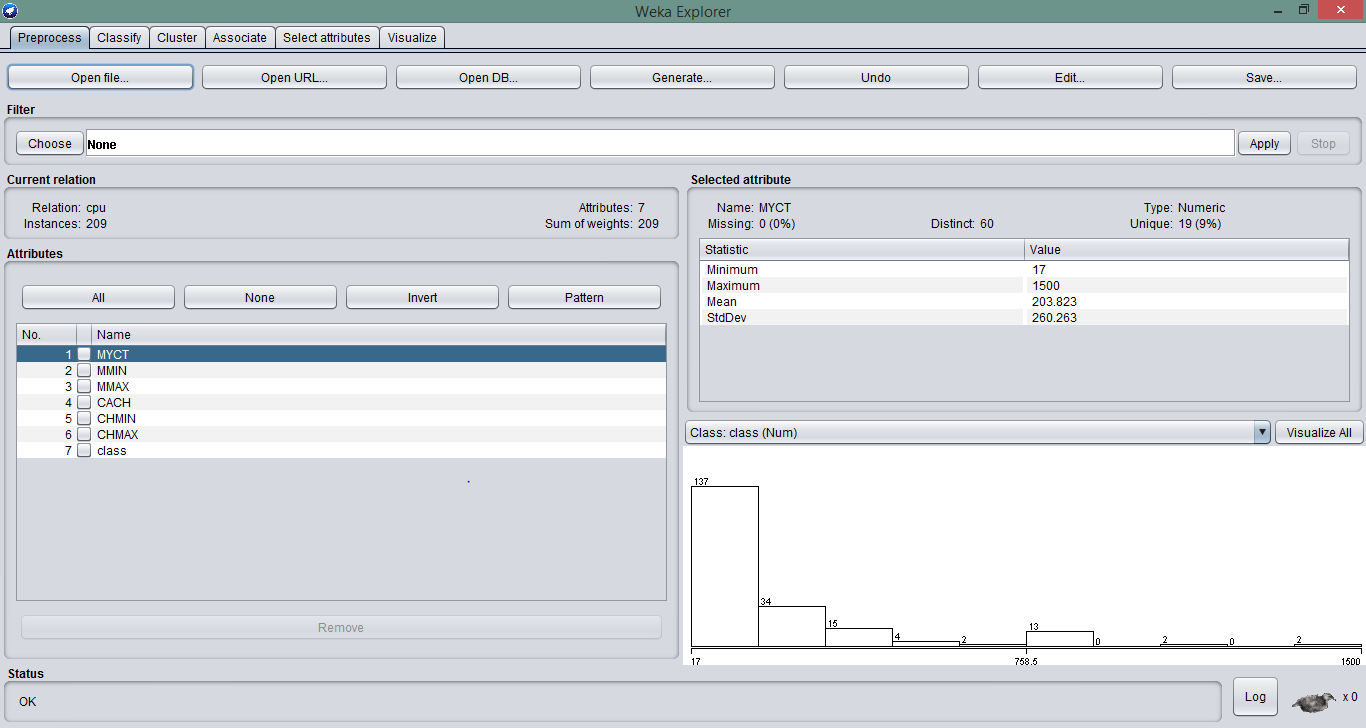
5 – click on choose and select any algorithm then press apply it will done the preprocessing.  


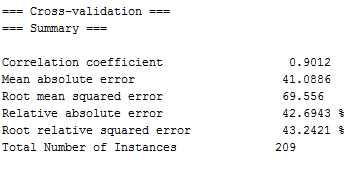
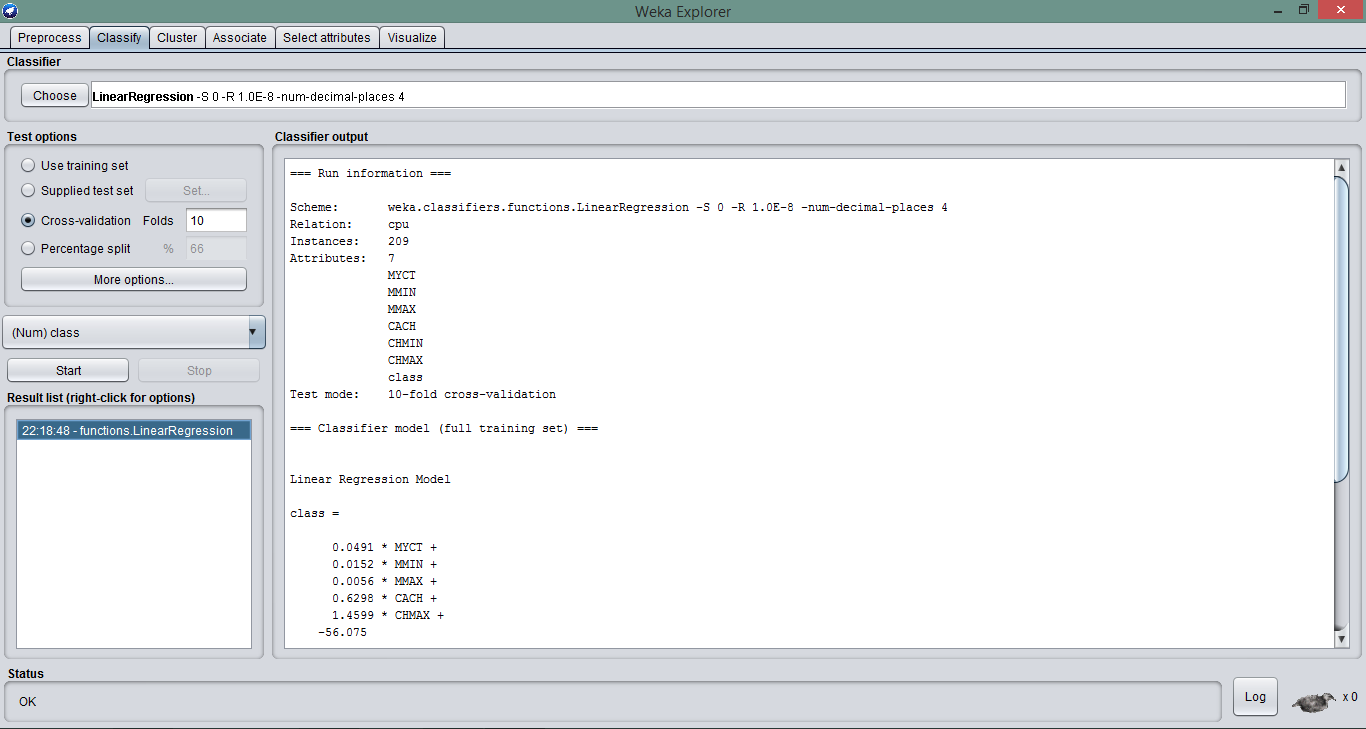
Lab #8

1 – First download the dataset from UCI/Kaggle

  
2 – Then import this data into weka using explorer….



3 – Then it will show this screen after importing… 

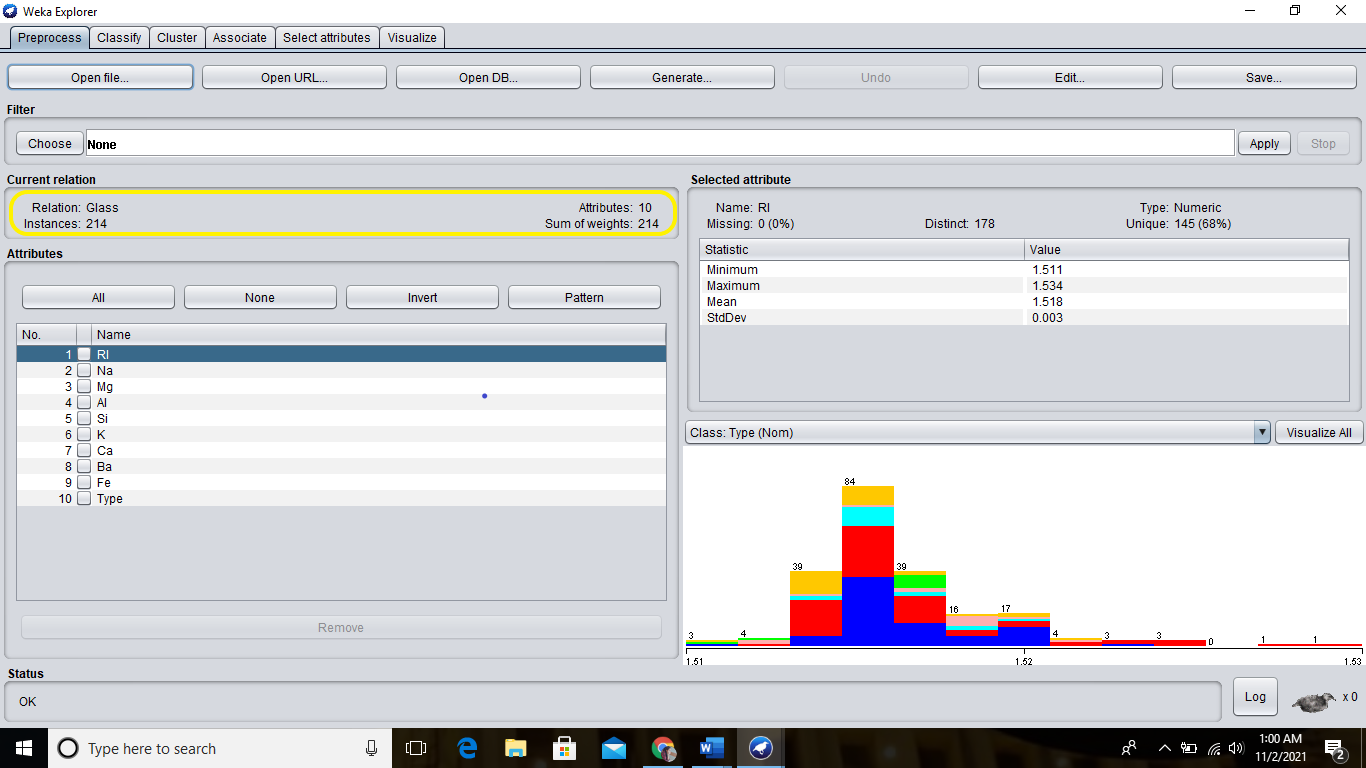
4 – Go to Classification Tab then apply the LinerRegression algorithm:  
  
**Result**

Lab #9

**1- Import a file to apply clustering algorithms?**

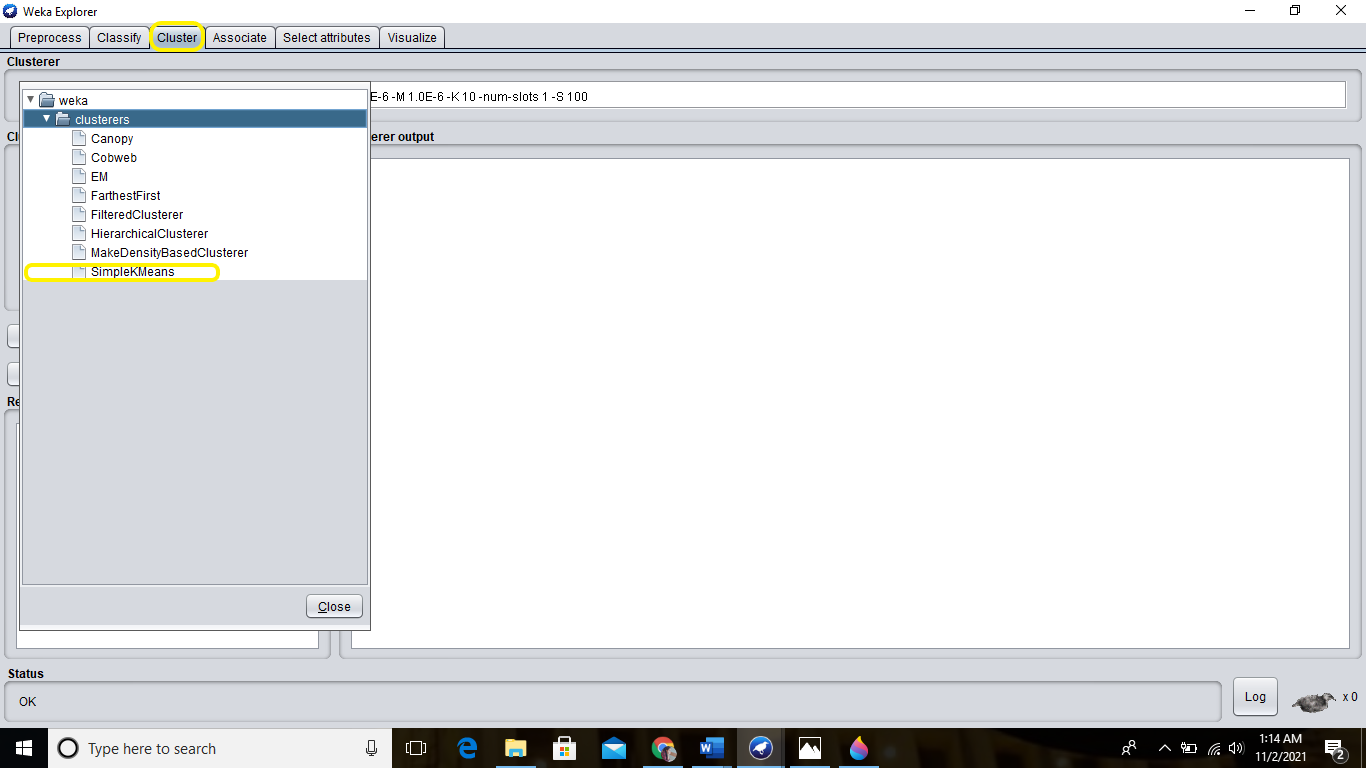
**Imported a dataset**

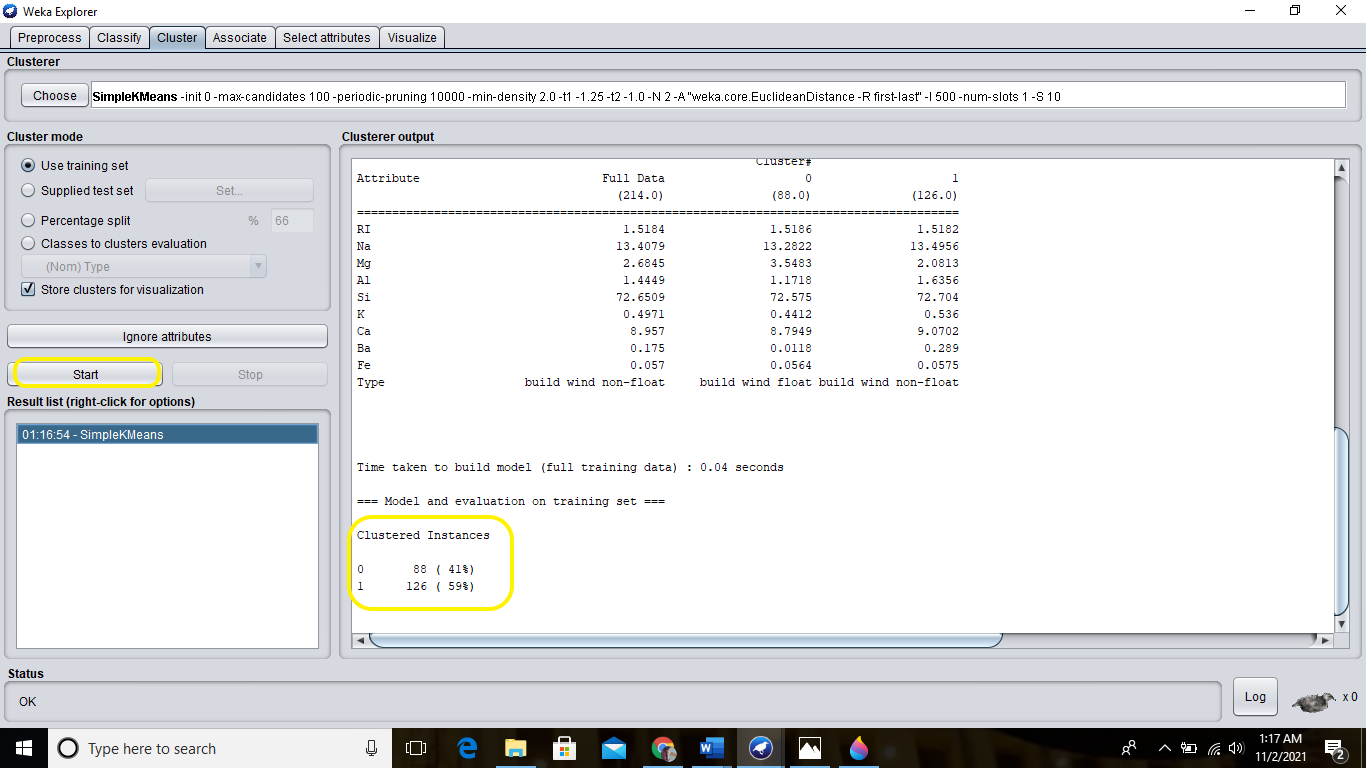
Firstly we import any dummy data set according to our choice

****

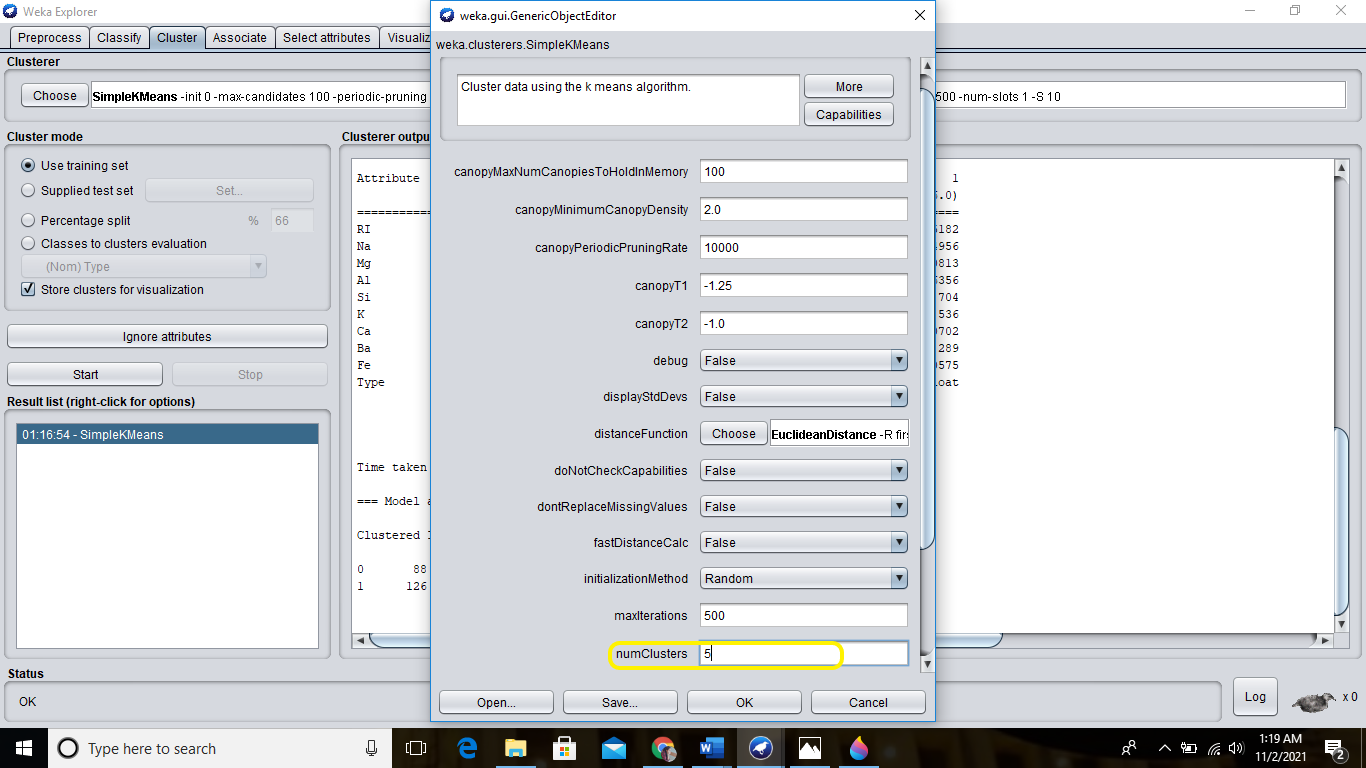
**choose simple k-mean**

Go to cluster and choose simple k-mean

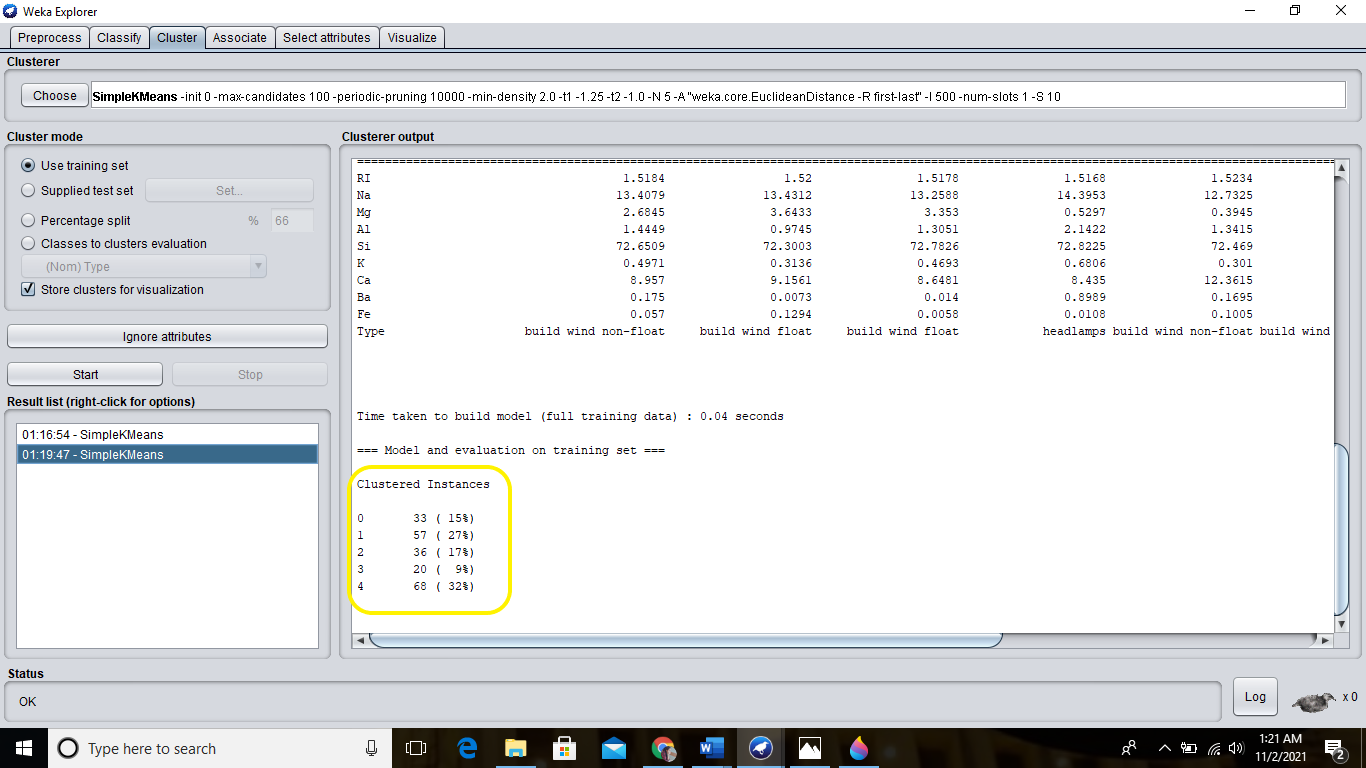
****

After Start there is two cluster in dataset****

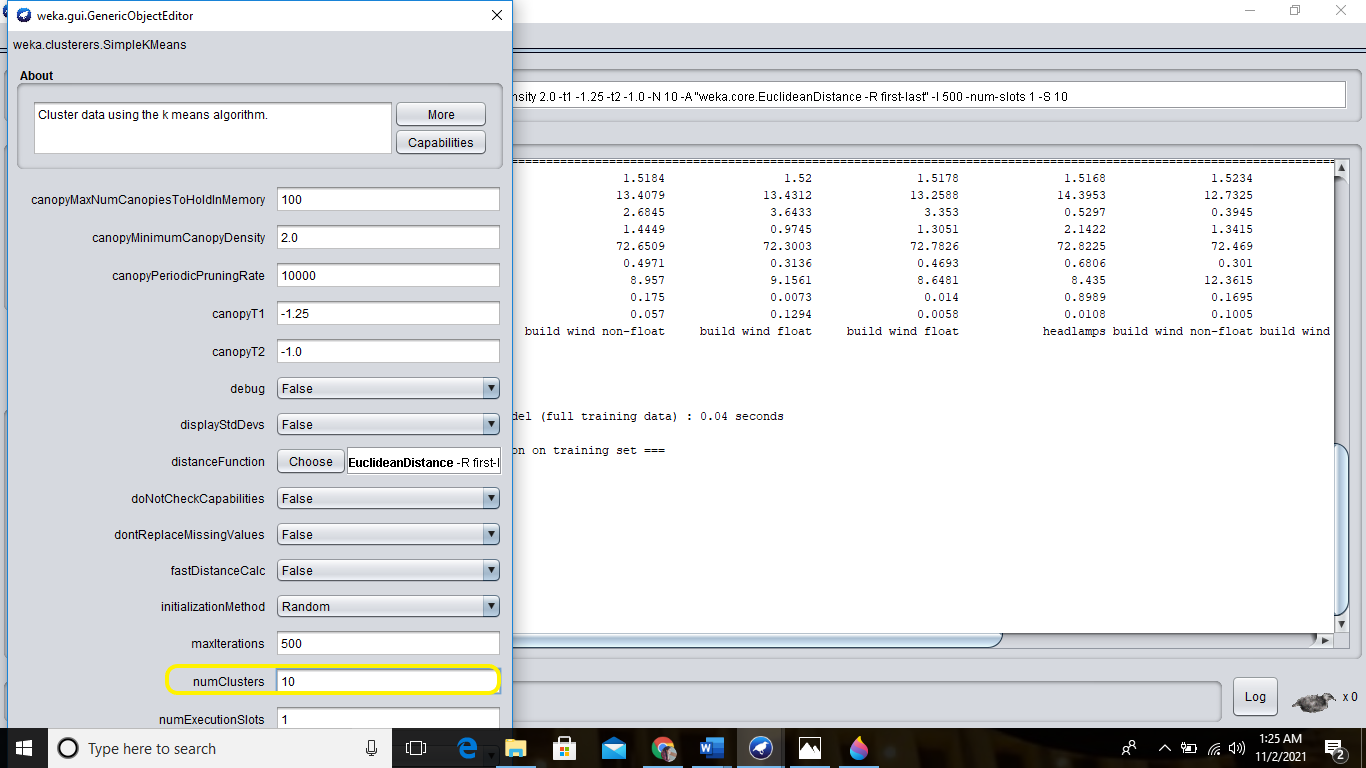
**Assign number of clusters:**

Here we assign number of cluster num=5

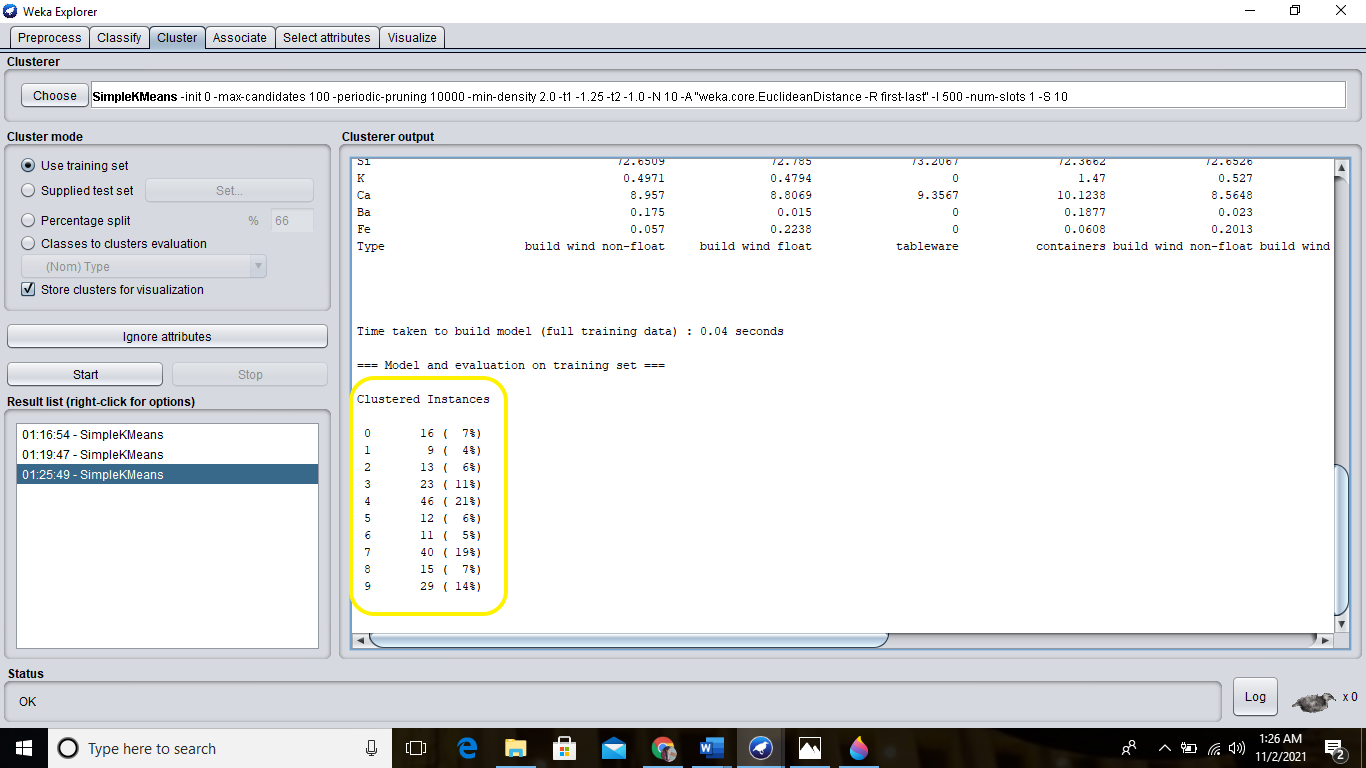
**Result:**

shows five number of clusters (0,1,2,3,4).

**Validate the data:**

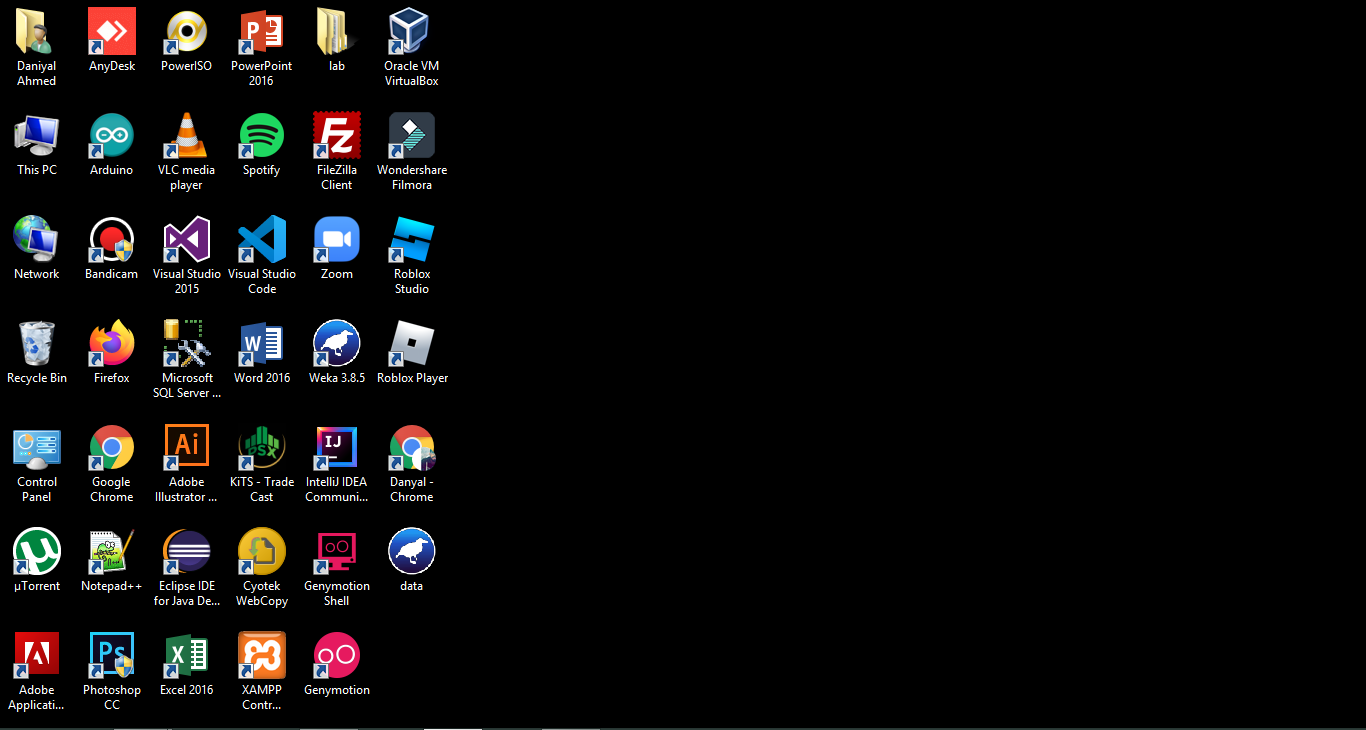
****Now here we can put value of k=10

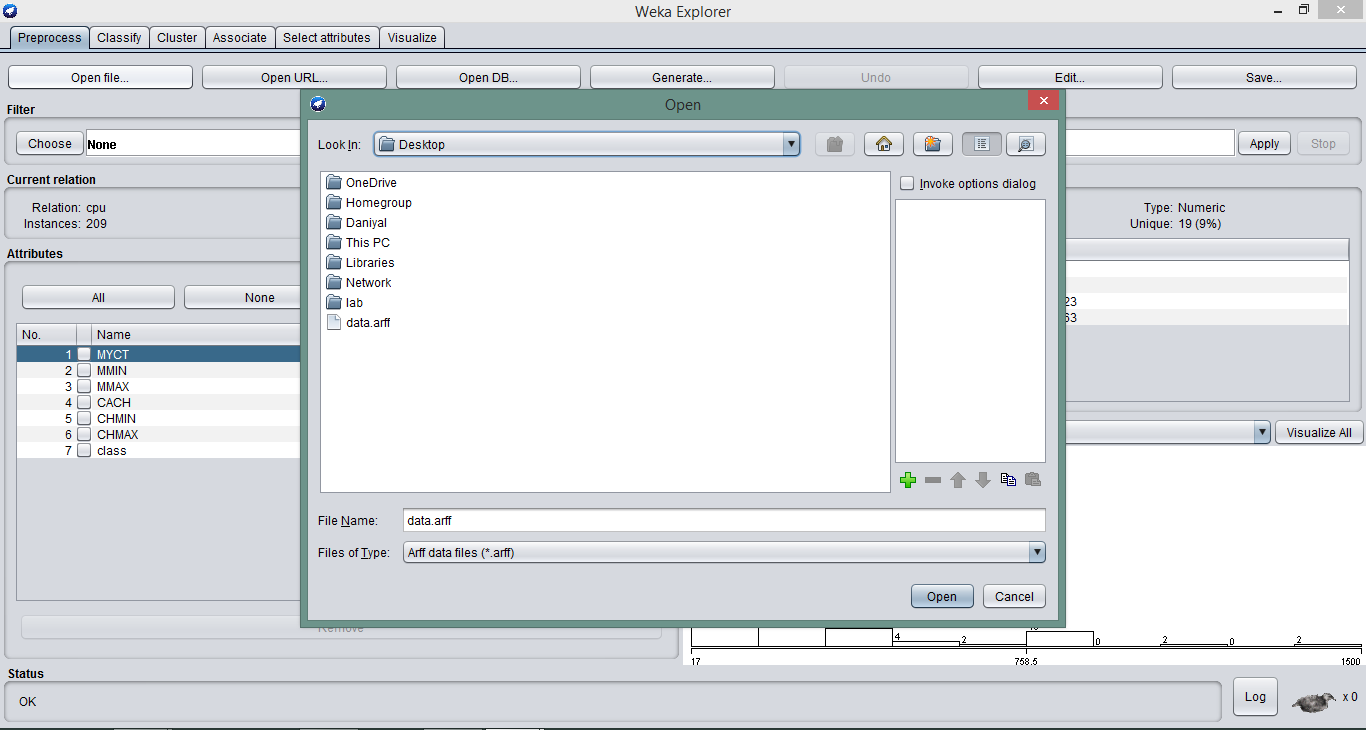
**Result:**

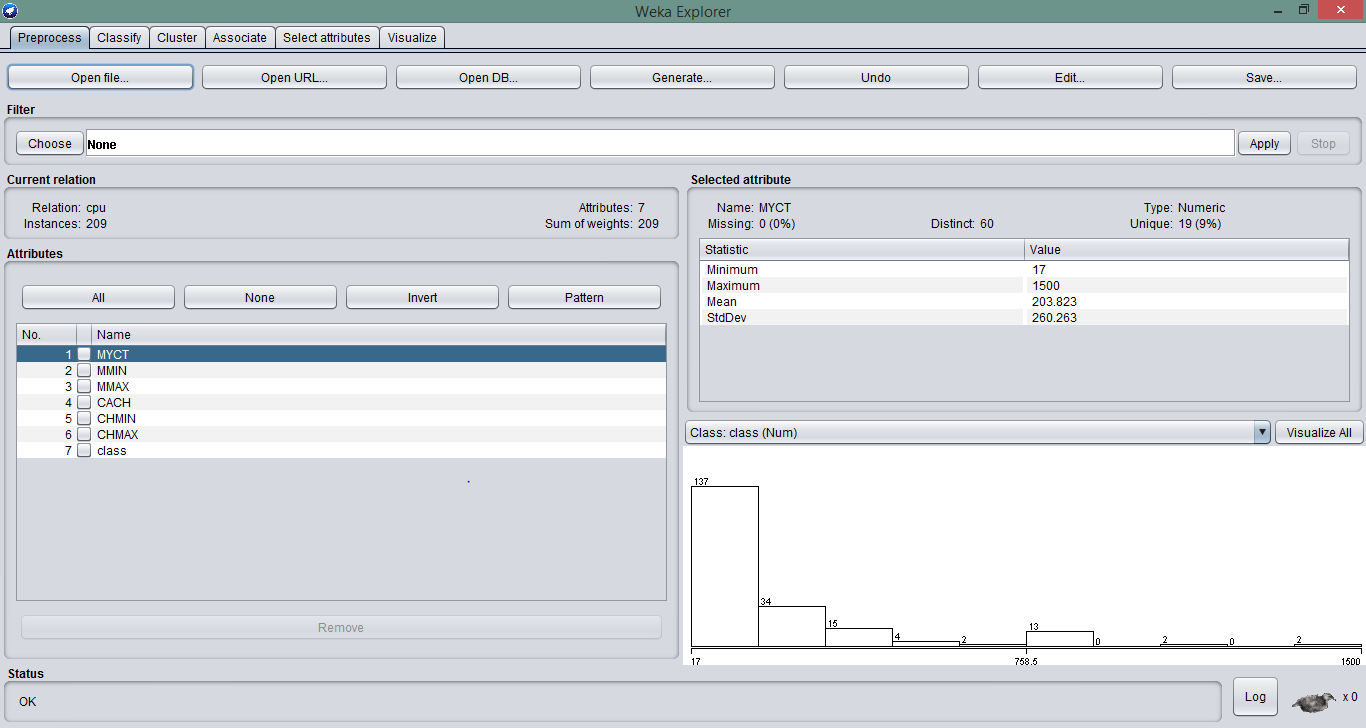
****so the data point are divided into 10 ten clusters (0,1,2,3,4,5,6,7,8,9)

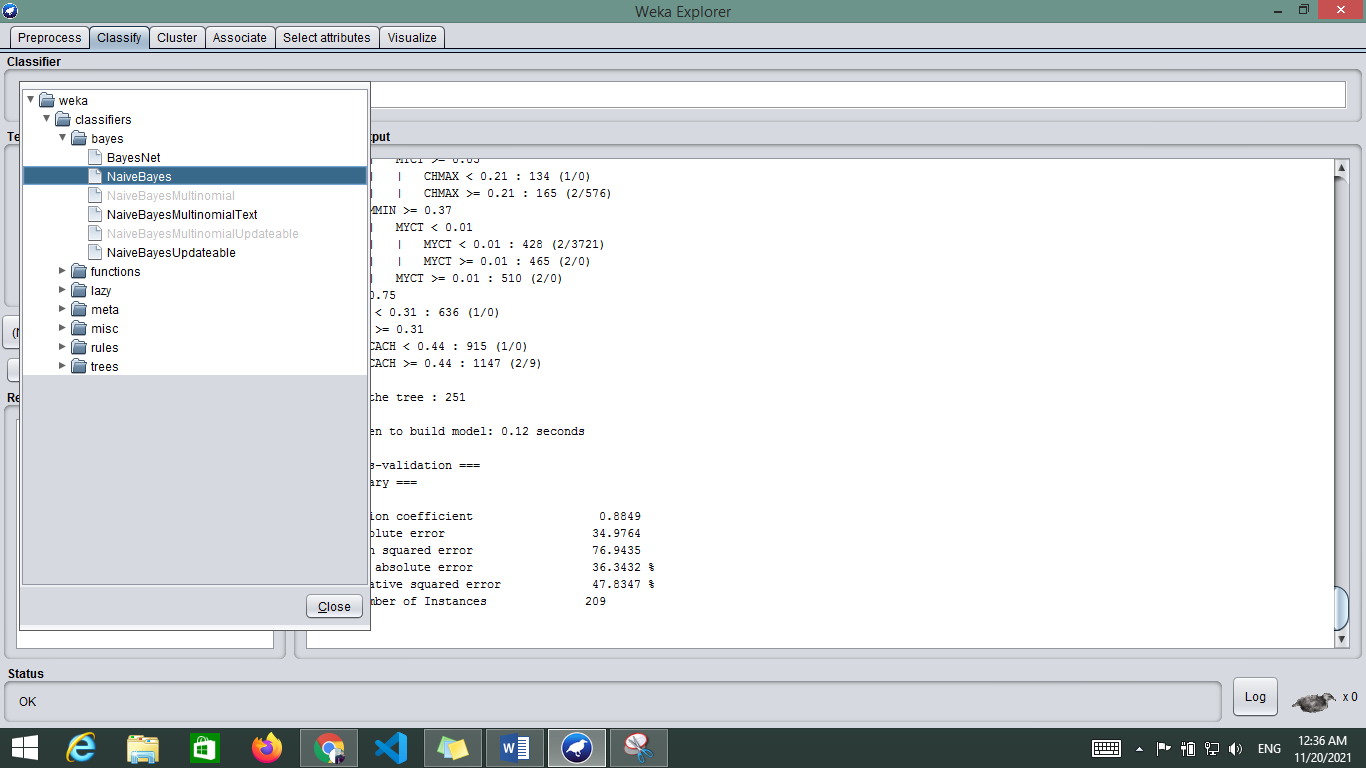
Lab #10

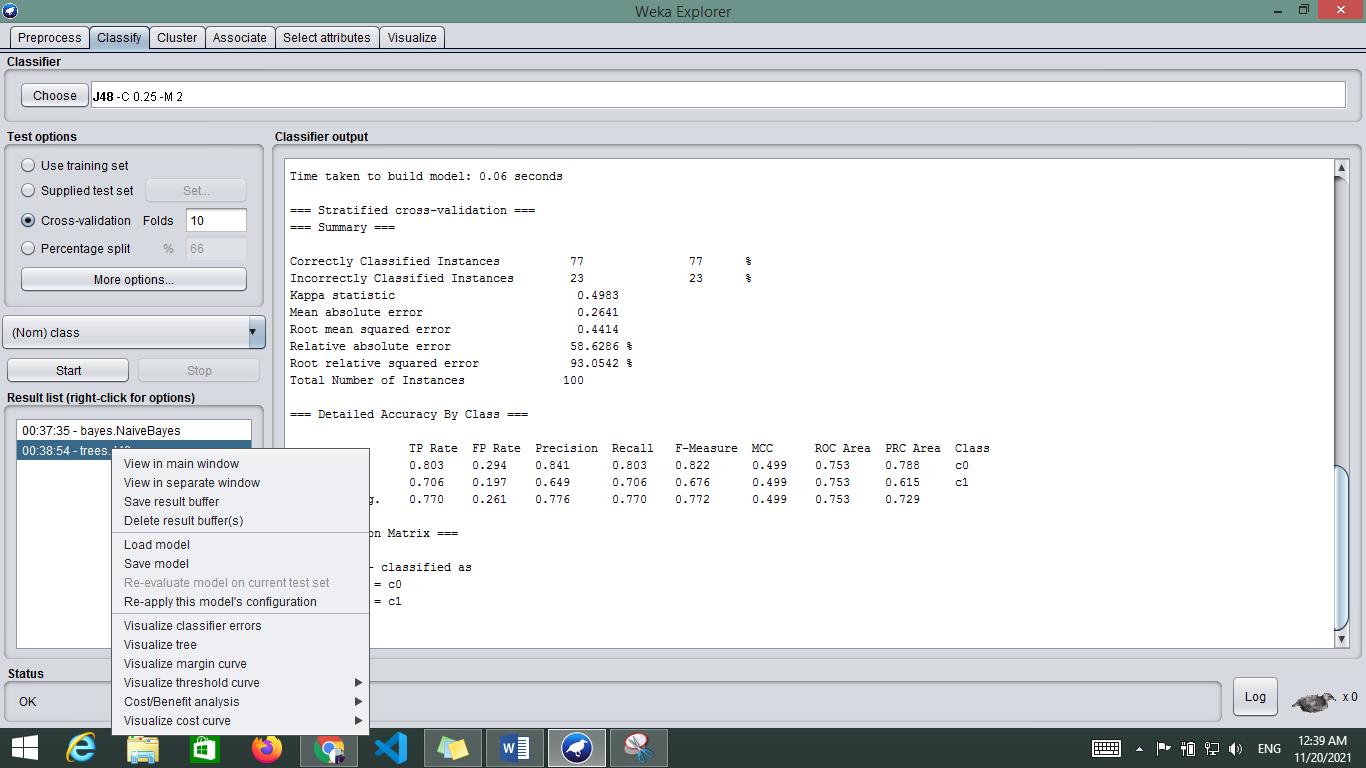
1 – First download the dataset from UCI/Kaggle

  
2 – Then import this data into weka using explorer….



3 – Then it will show this screen after importing… 

4 – Go to Classification Tab then apply the Naive Bayes algorithm:  


5 – Then go to tree and apply j48 algorithm and then click on visualize tree.  


6 – then it will show the tree for the relative dataset.  
