

PREPROCESSING:

It is a process of identifying the unwanted data (data cleaning) before loading the data from the data base.

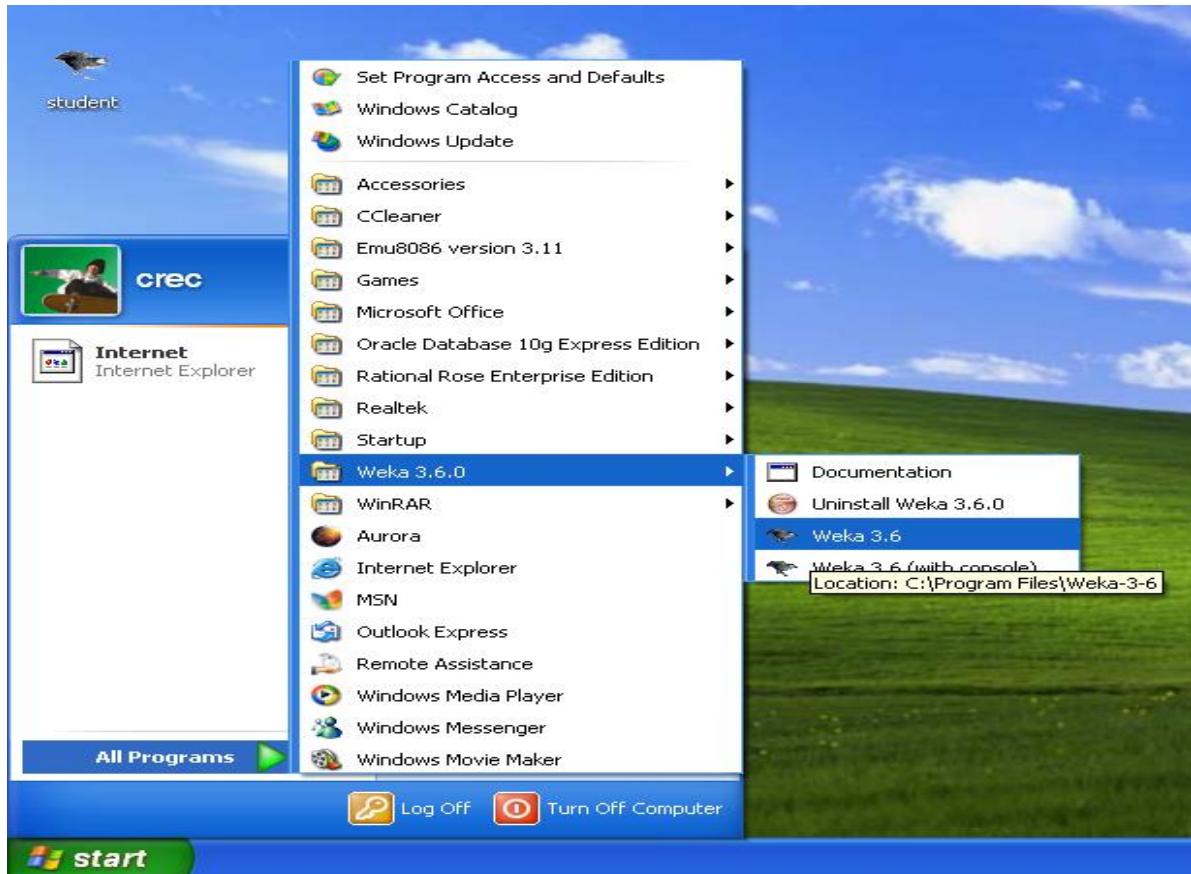


Figure-1

1. Open the WEKA application as shown in the above figure-1

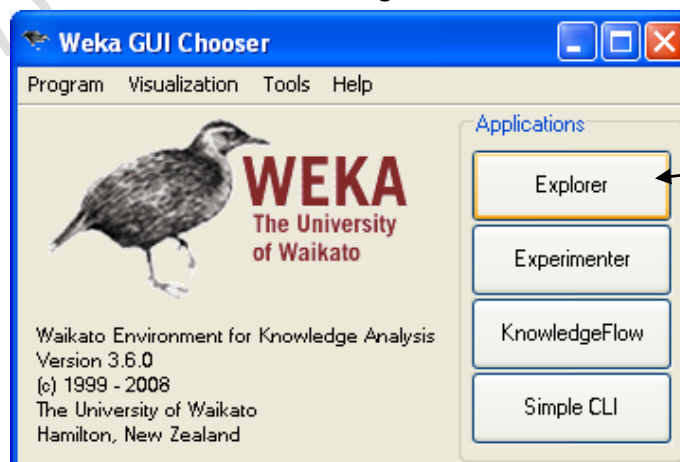


Figure-2

2. Now Click on Explorer as shown in the above figure-2.

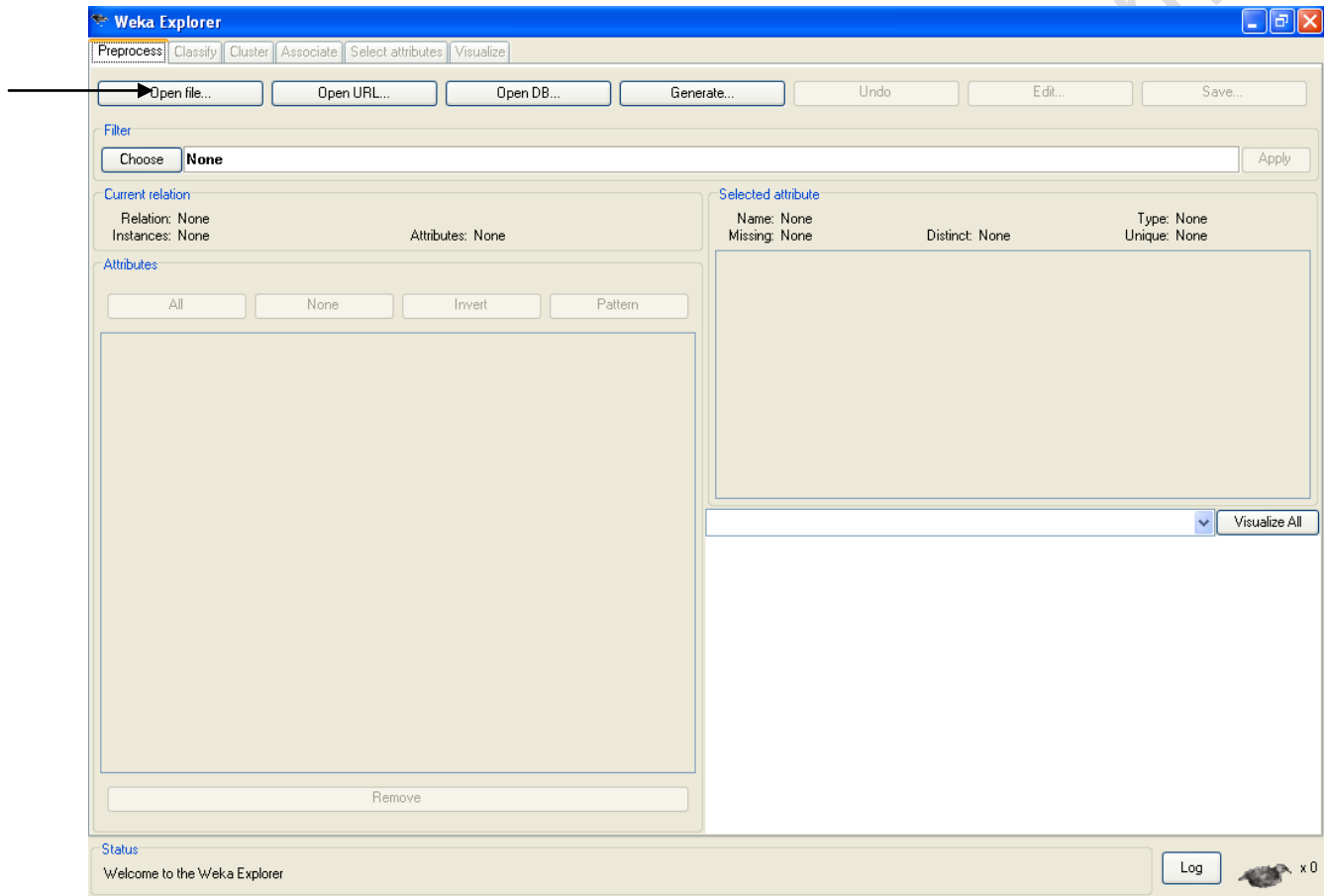


Figure-3

3. Now open file by choosing the “open file” button as shown in the above figure-3.
Relation specifies the name of the database used, instances specify the objects involved, and attributes specify the number of attributes used in the data base or relation.

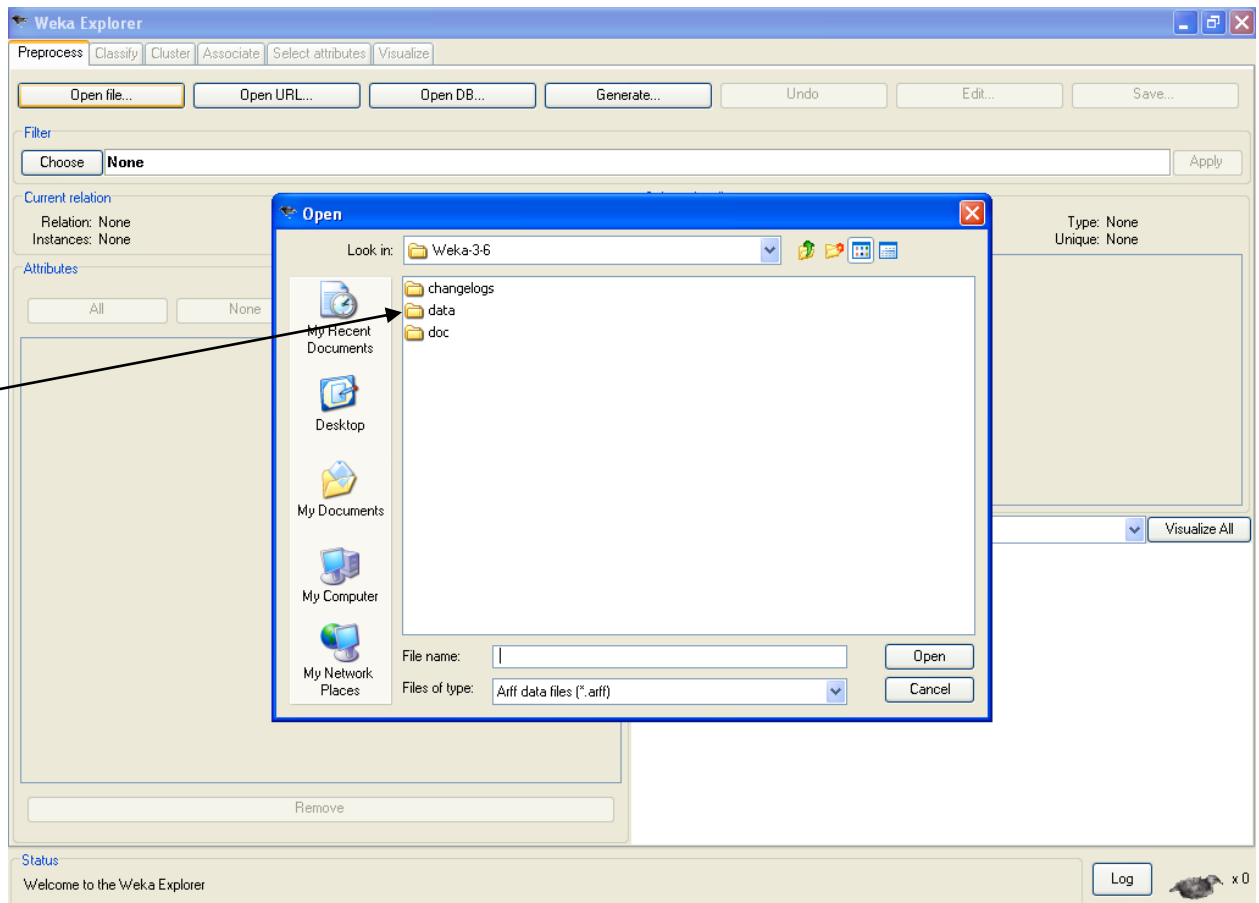


Figure-4

- Now choose the data folder in the open dialogue box as in figure-4.

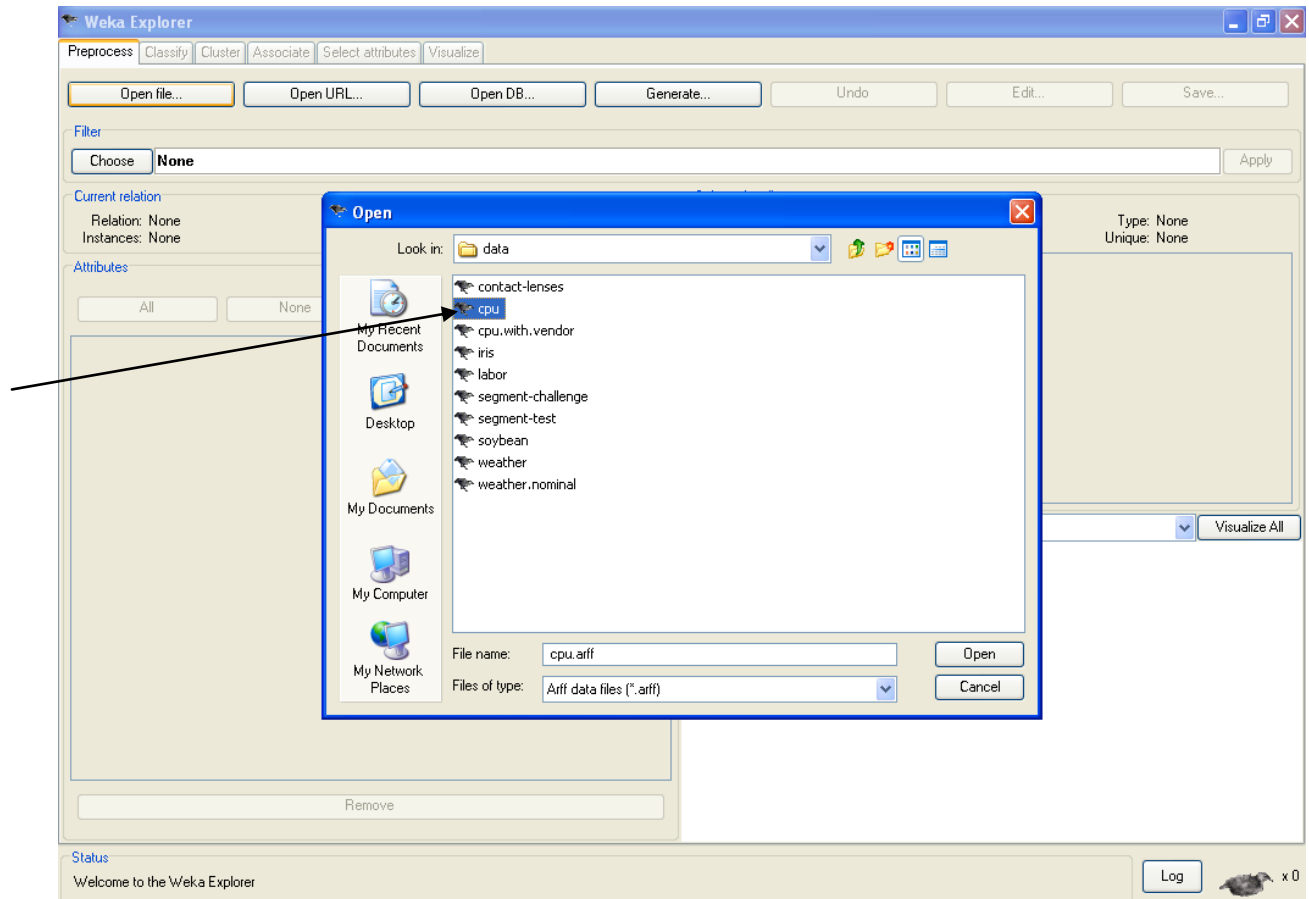


Figure-5

- Now choose the "cpu.arff" file in the above figure-5.

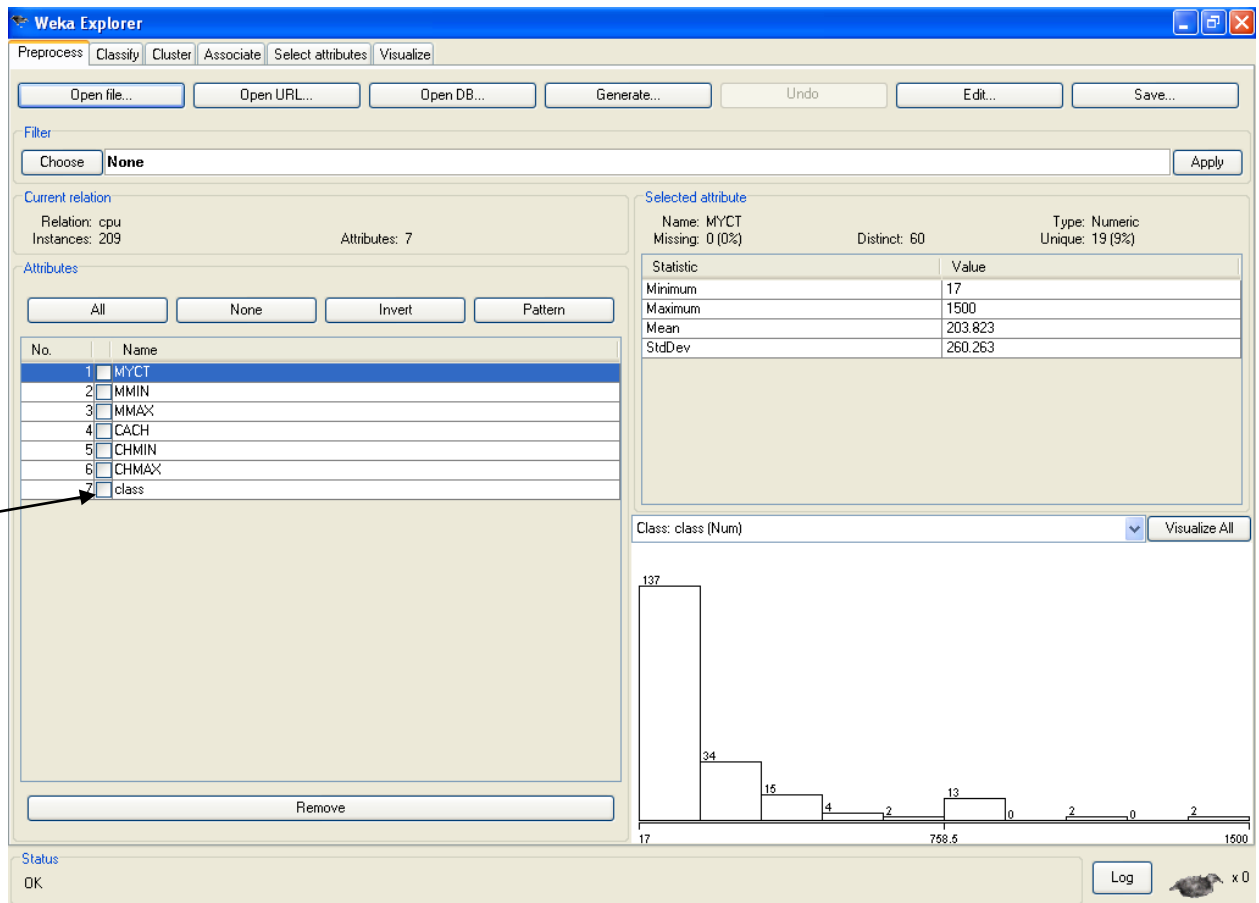


Figure-6

6. Select any attribute in the attributes section (for e.g. class as shown in Figure-6) and click on remove button as shown in the figure-7 below.

On right Side of the window we have the selected attribute field which specifies the following options like:

Name: - It specifies the name of the attribute.

Missing: - It specifies the missing values in the attribute.

Type: - It specifies the type of attribute used in the database like nominal or numeric or string etc...

Unique: - It specifies unique attributes available in the relation.

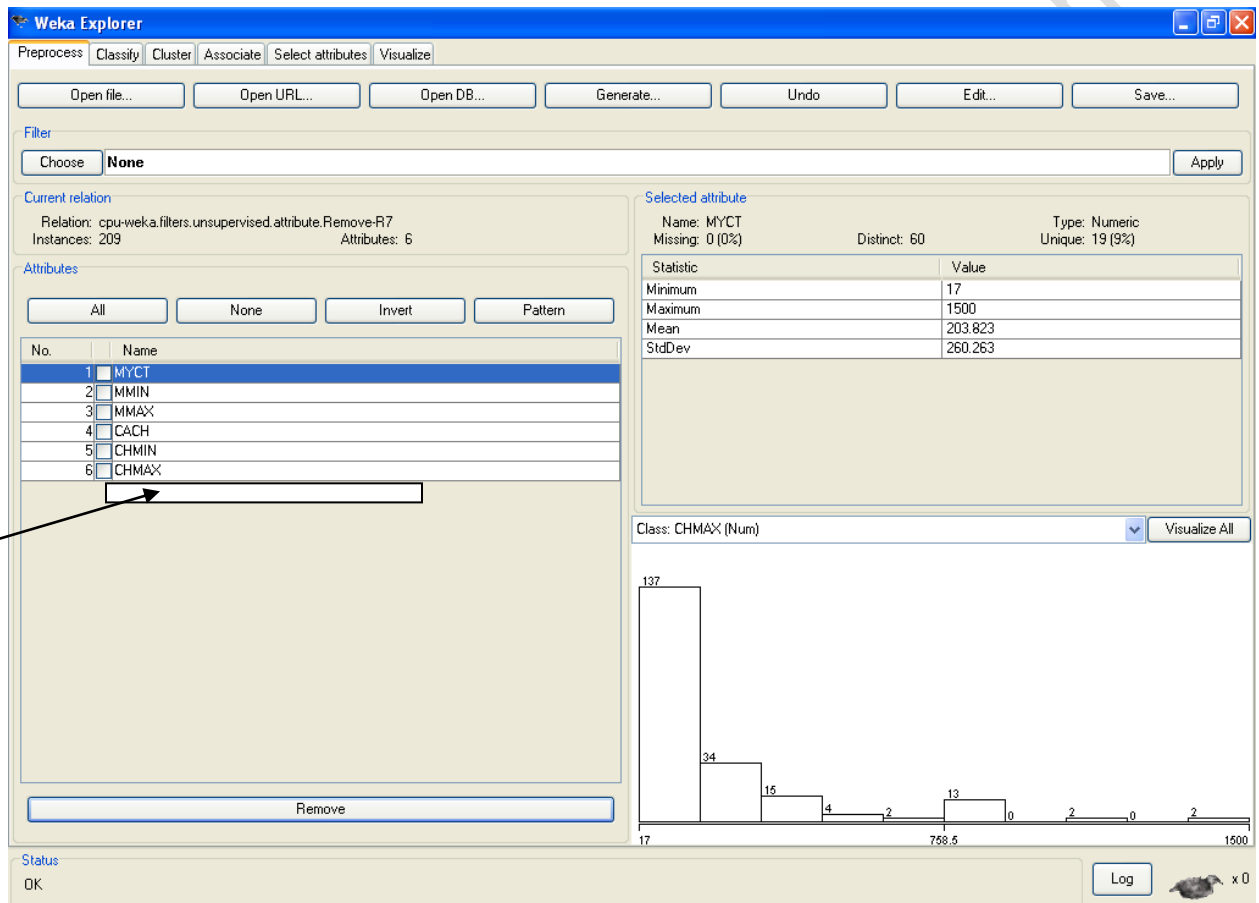


Figure-7

7. The above diagram shows the cpu.arff after removing the attribute "class".

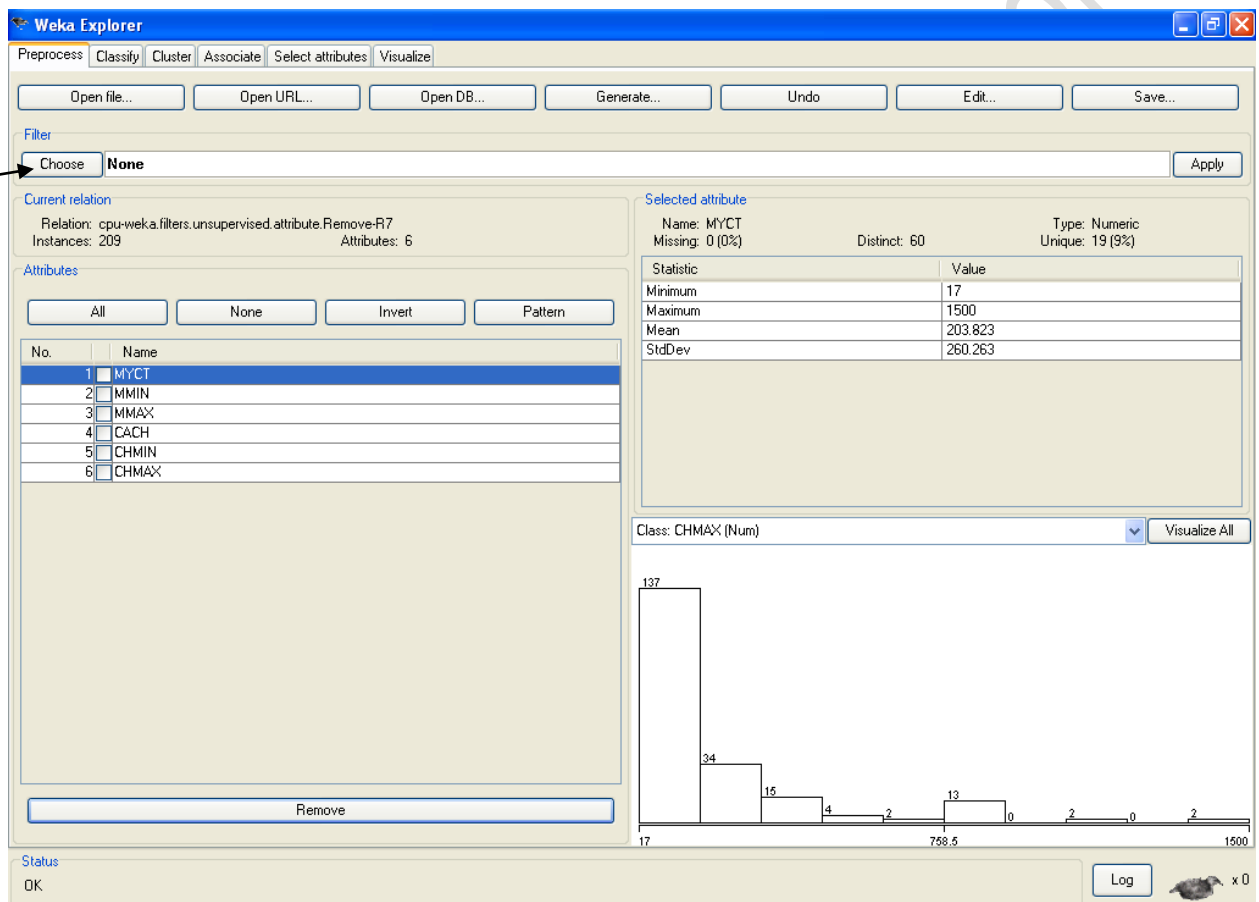


Figure-8

- Now click on the "choose" button from filter as in figure-8 and expand the "unsupervised" option and select the "Discretize" option, which is as shown in figure-9.

Discretize is used to discretize or convert numeric attributes into nominal ones, based on the class information.

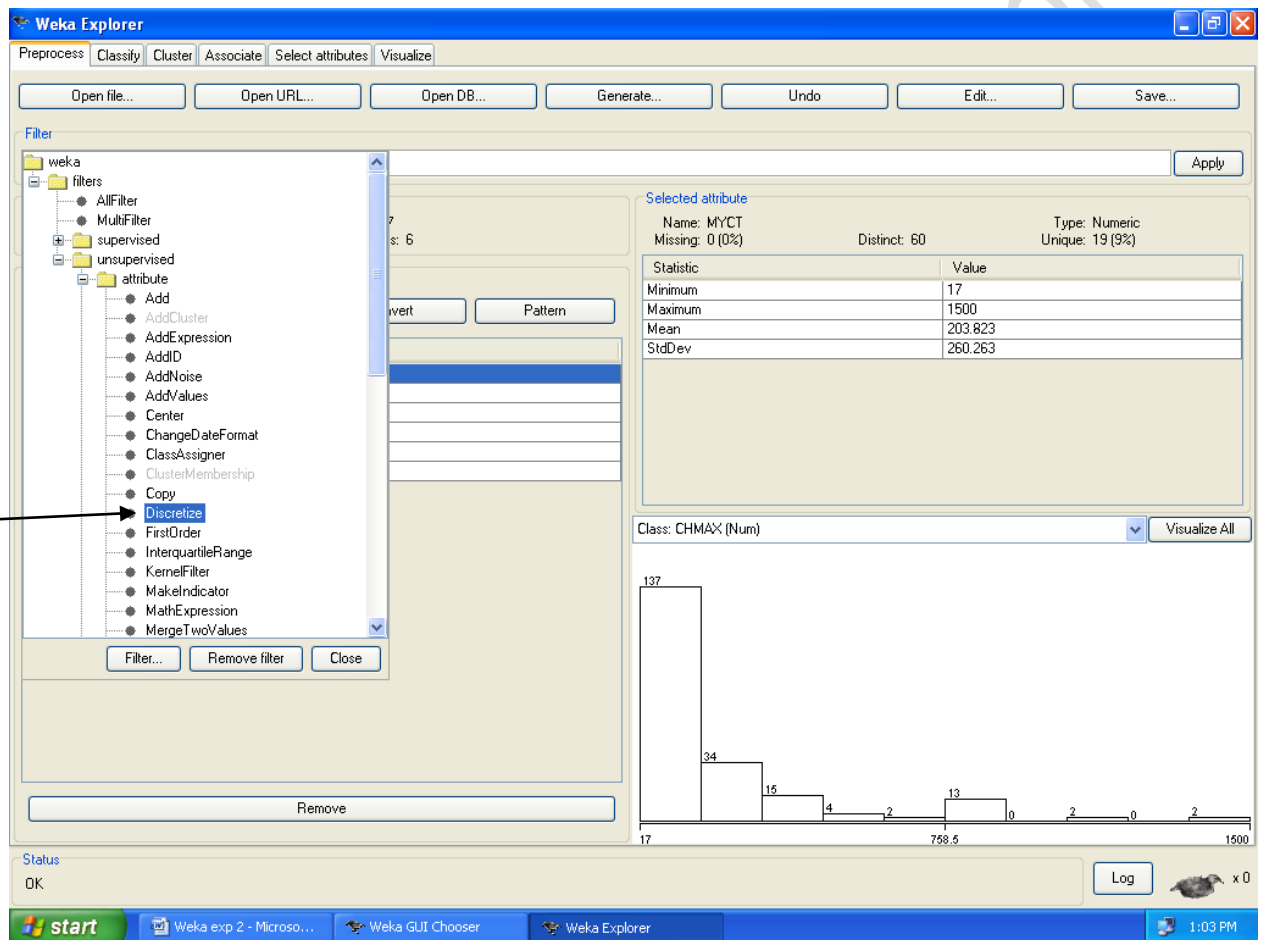


Figure-9

9. Now left click the object to edit the properties or right click and select show properties to edit the properties which is as shown in the figure-10 as follows.

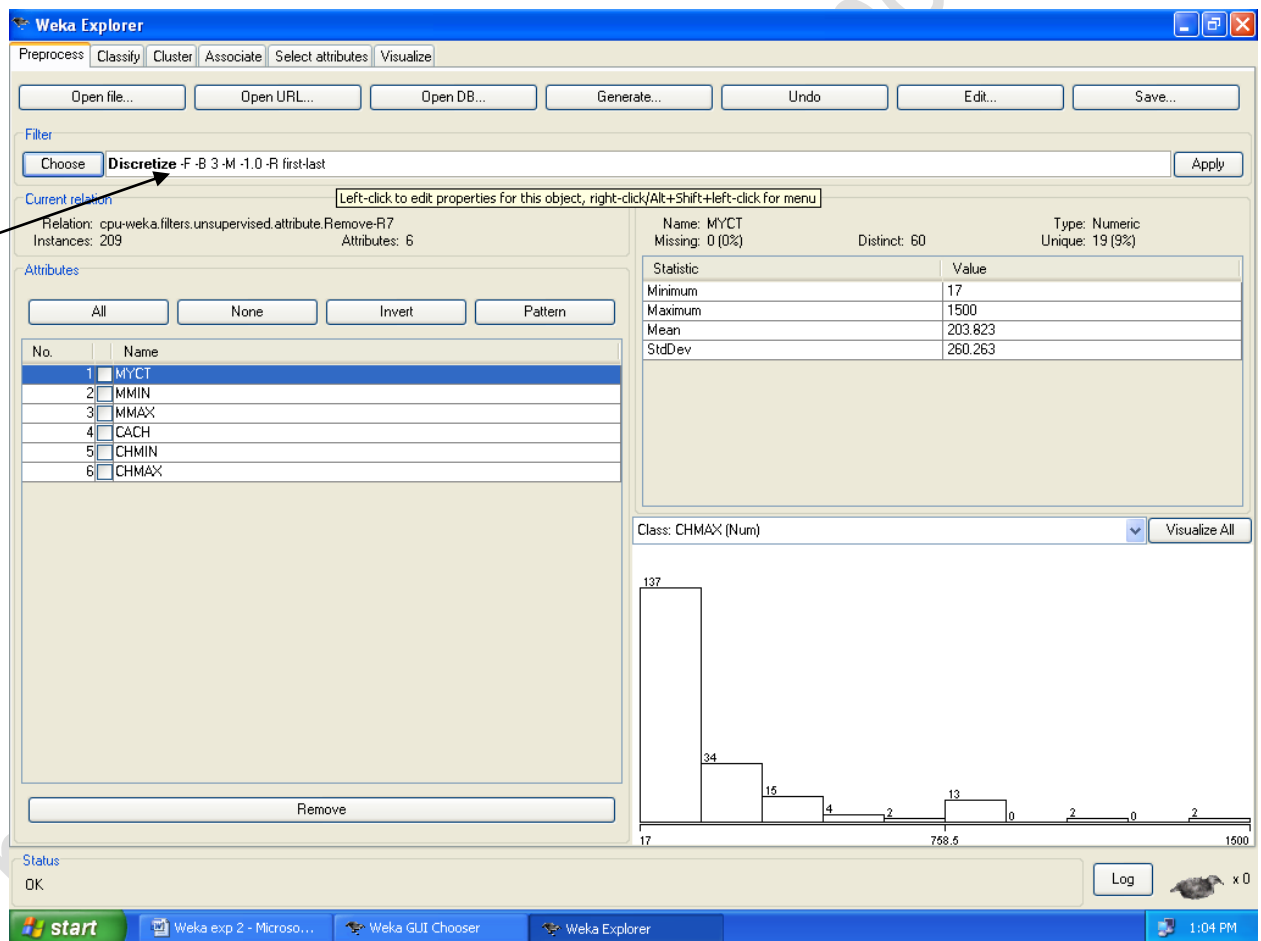


Figure-10

10. In the "GenericObjectEditor" change the bins value to either 2 or 3 or as our desire and make the "useEqualFrequency" option as "TRUE" and click on ok, which is as shown in the figure-11 as follows.

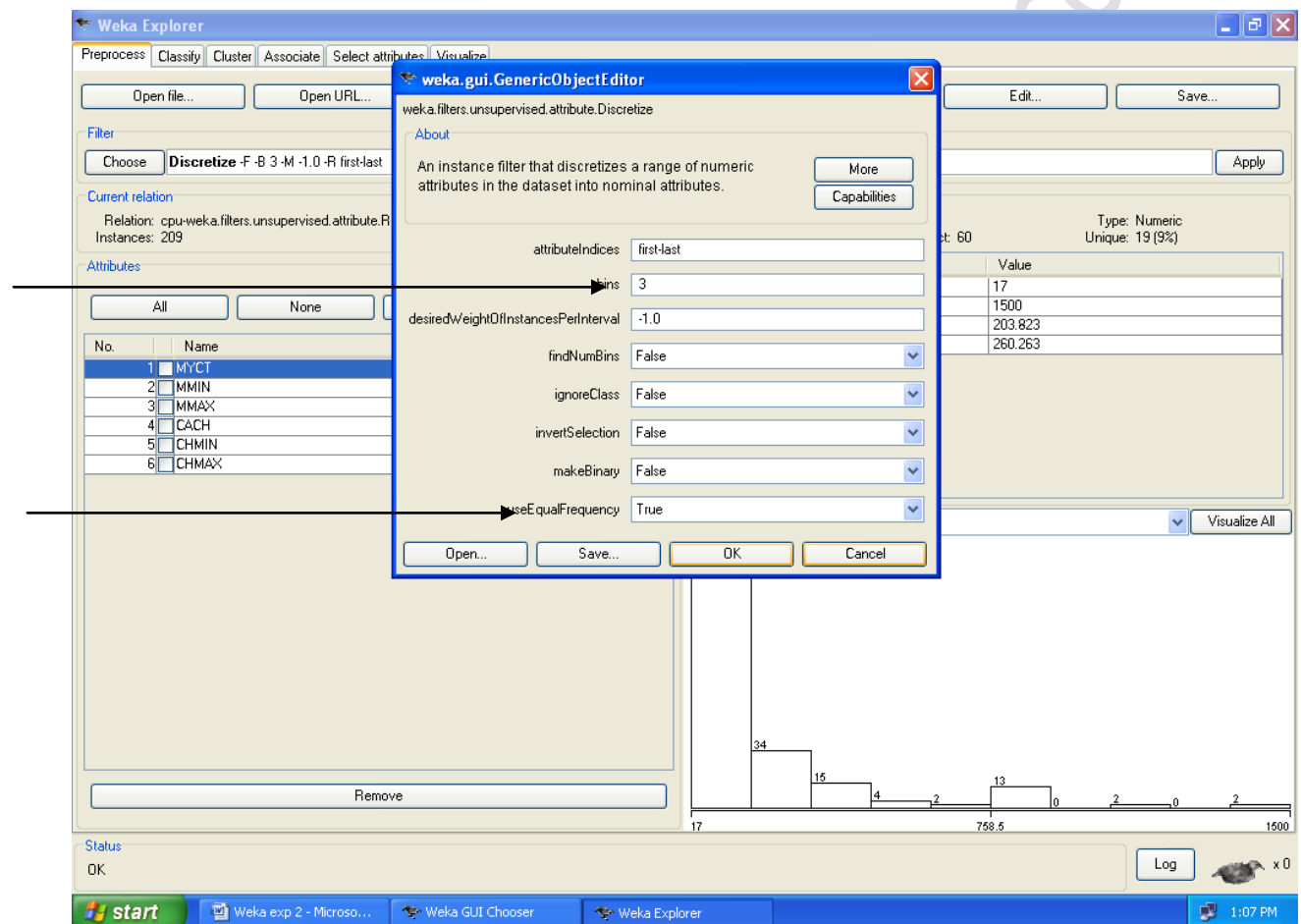


Figure-11

11. Now apply the properties by clicking on "apply" button in the filter where the Discretize object contains 3 bins and useEqualFrequency is set to TRUE. After applying the properties to the object the result obtained is as shown in the following figure-12.

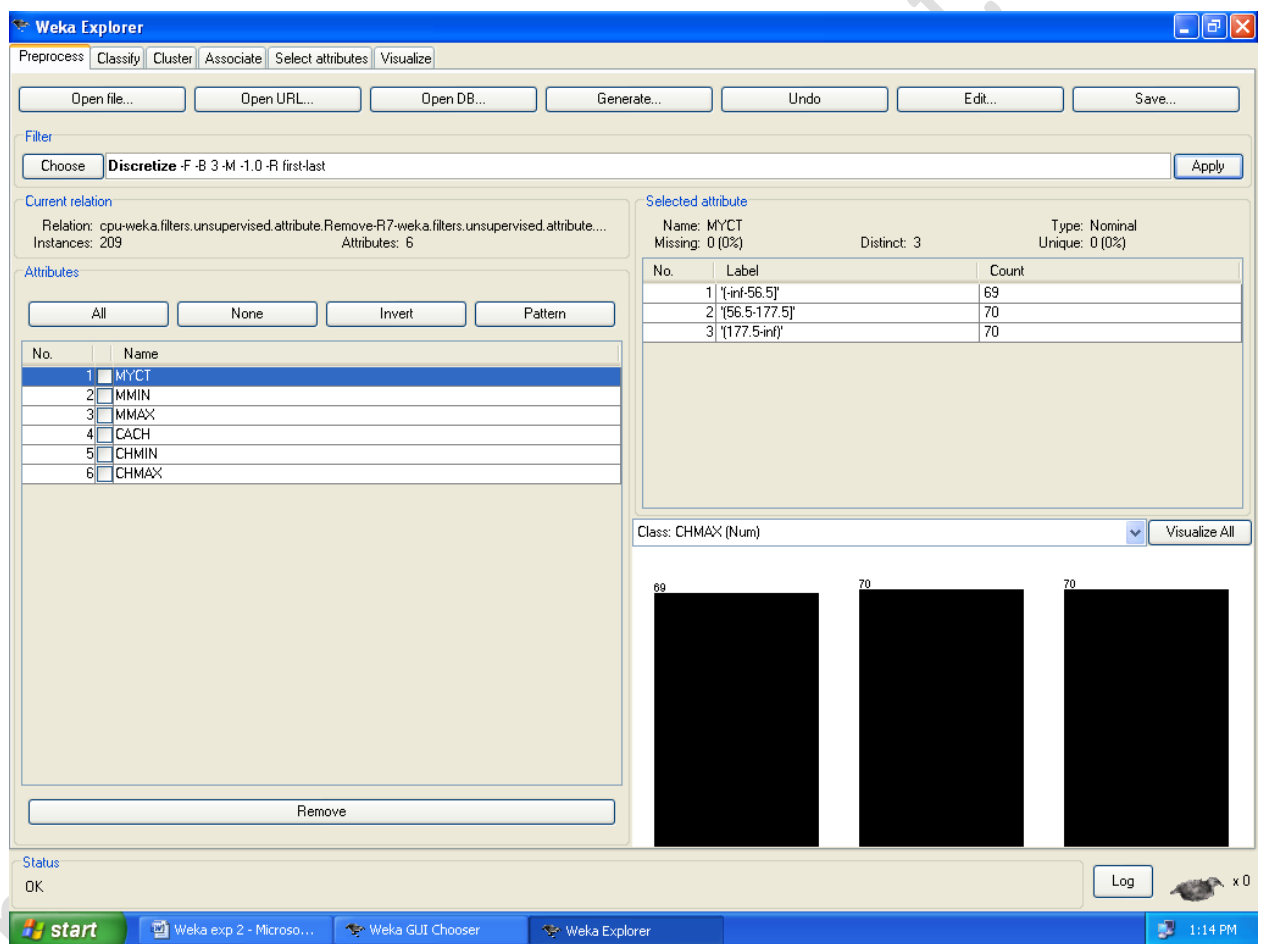
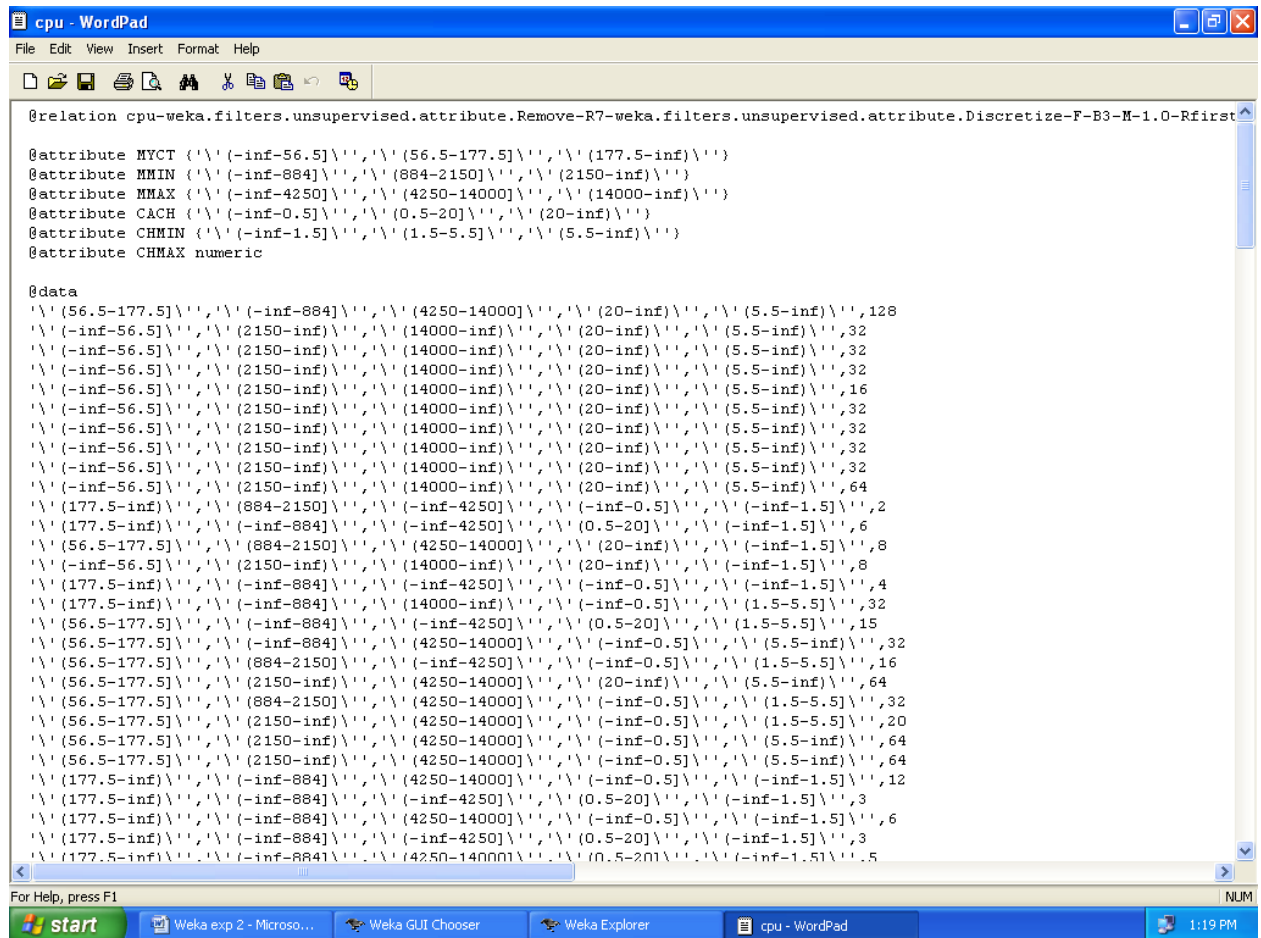


Figure-12

12. We can observe the change in the result in the visualize which is as shown in the figure-12.



```
@relation cpu-weka.filters.unsupervised.attribute.Remove-R7-weka.filters.unsupervised.attribute.Discretize-F-B3-M-1.0-Rfirst

@attribute MYCT (''(-inf-56.5)''',''(56.5-177.5)''',''(177.5-inf)''')
@attribute MMIN (''(-inf-884)''',''(884-2150)''',''(2150-inf)''')
@attribute MMAX (''(-inf-4250)''',''(4250-14000)''',''(14000-inf)''')
@attribute CACH (''(-inf-0.5)''',''(0.5-20)''',''(20-inf)''')
@attribute CHMIN (''(-inf-1.5)''',''(1.5-5.5)''',''(5.5-inf)''')
@attribute CHMAX numeric

@data
''(56.5-177.5)''',''(-inf-884)''',''(4250-14000)''',''(20-inf)''',''(5.5-inf)''',128
''(-inf-56.5)''',''(2150-inf)''',''(14000-inf)''',''(20-inf)''',''(5.5-inf)''',32
''(-inf-56.5)''',''(2150-inf)''',''(14000-inf)''',''(20-inf)''',''(5.5-inf)''',32
''(-inf-56.5)''',''(2150-inf)''',''(14000-inf)''',''(20-inf)''',''(5.5-inf)''',32
''(-inf-56.5)''',''(2150-inf)''',''(14000-inf)''',''(20-inf)''',''(5.5-inf)''',16
''(-inf-56.5)''',''(2150-inf)''',''(14000-inf)''',''(20-inf)''',''(5.5-inf)''',32
''(-inf-56.5)''',''(2150-inf)''',''(14000-inf)''',''(20-inf)''',''(5.5-inf)''',32
''(-inf-56.5)''',''(2150-inf)''',''(14000-inf)''',''(20-inf)''',''(5.5-inf)''',32
''(-inf-56.5)''',''(2150-inf)''',''(14000-inf)''',''(20-inf)''',''(5.5-inf)''',32
''(-inf-56.5)''',''(2150-inf)''',''(14000-inf)''',''(20-inf)''',''(5.5-inf)''',64
''(177.5-inf)''',''(884-2150)''',''(-inf-4250)''',''(-inf-0.5)''',''(-inf-1.5)''',2
''(177.5-inf)''',''(-inf-884)''',''(-inf-4250)''',''(0.5-20)''',''(-inf-1.5)''',6
''(56.5-177.5)''',''(884-2150)''',''(4250-14000)''',''(20-inf)''',''(-inf-1.5)''',8
''(-inf-56.5)''',''(2150-inf)''',''(14000-inf)''',''(20-inf)''',''(-inf-1.5)''',8
''(177.5-inf)''',''(-inf-884)''',''(-inf-4250)''',''(-inf-0.5)''',''(-inf-1.5)''',4
''(177.5-inf)''',''(-inf-884)''',''(14000-inf)''',''(-inf-0.5)''',''(1.5-5.5)''',32
''(56.5-177.5)''',''(-inf-884)''',''(-inf-4250)''',''(0.5-20)''',''(1.5-5.5)''',15
''(56.5-177.5)''',''(-inf-884)''',''(4250-14000)''',''(-inf-0.5)''',''(5.5-inf)''',32
''(56.5-177.5)''',''(884-2150)''',''(-inf-4250)''',''(-inf-0.5)''',''(1.5-5.5)''',16
''(56.5-177.5)''',''(2150-inf)''',''(4250-14000)''',''(20-inf)''',''(5.5-inf)''',64
''(56.5-177.5)''',''(884-2150)''',''(4250-14000)''',''(-inf-0.5)''',''(1.5-5.5)''',32
''(56.5-177.5)''',''(2150-inf)''',''(4250-14000)''',''(-inf-0.5)''',''(1.5-5.5)''',20
''(56.5-177.5)''',''(2150-inf)''',''(4250-14000)''',''(-inf-0.5)''',''(5.5-inf)''',64
''(56.5-177.5)''',''(2150-inf)''',''(4250-14000)''',''(-inf-0.5)''',''(5.5-inf)''',64
''(177.5-inf)''',''(-inf-884)''',''(4250-14000)''',''(-inf-0.5)''',''(-inf-1.5)''',12
''(177.5-inf)''',''(-inf-884)''',''(-inf-4250)''',''(0.5-20)''',''(-inf-1.5)''',3
''(177.5-inf)''',''(-inf-884)''',''(4250-14000)''',''(-inf-0.5)''',''(-inf-1.5)''',6
''(177.5-inf)''',''(-inf-884)''',''(-inf-4250)''',''(0.5-20)''',''(-inf-1.5)''',3
''(177.5-inf)''',''(-inf-884)''',''(4250-14000)''',''(0.5-20)''',''(-inf-1.5)''',5
```

Figure-13

13. Finally the result is obtained as shown in the figure-13.