EXPERIMENT NO 5

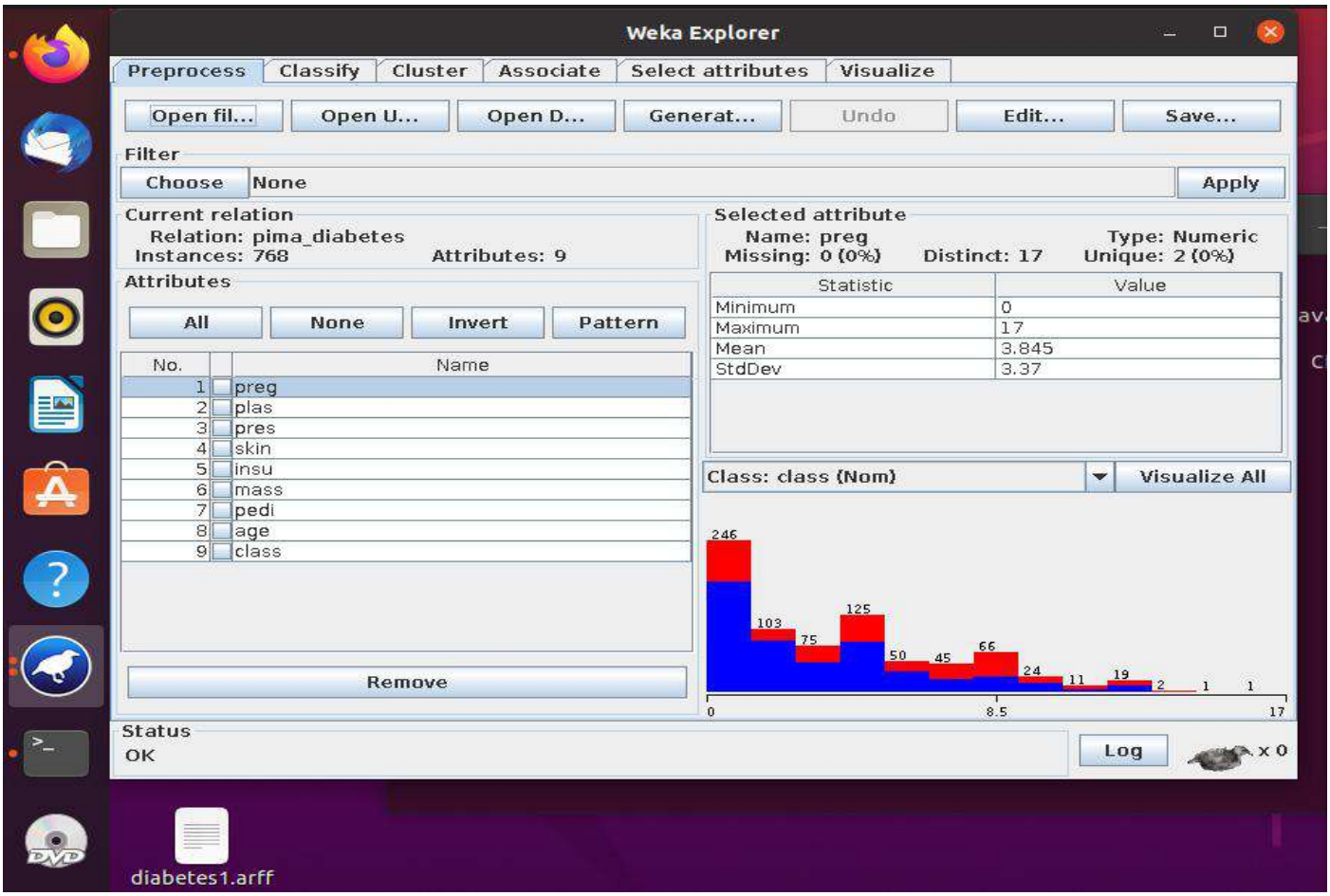
Name: Komal Mahesh Chitnis

Roll No: 26

Div: A

Moodle Id: 20102068

1. Load the diabetes.arff file into WEKA



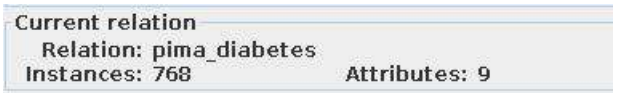
2. Study thedataset and answer following

Number of instances= ?

Number of Attributes=?

Types of attributes other than class = ?

Type of class Attribute =?



Selected attribute
Name: skin Type: Numeric
Missing: 0 (0%) Distinct: 51 Unique: 5 (1%)

Selected attribute
Name: insu Type: Numeric
Missing: 0 (0%) Distinct: 186 Unique: 93 (12%)

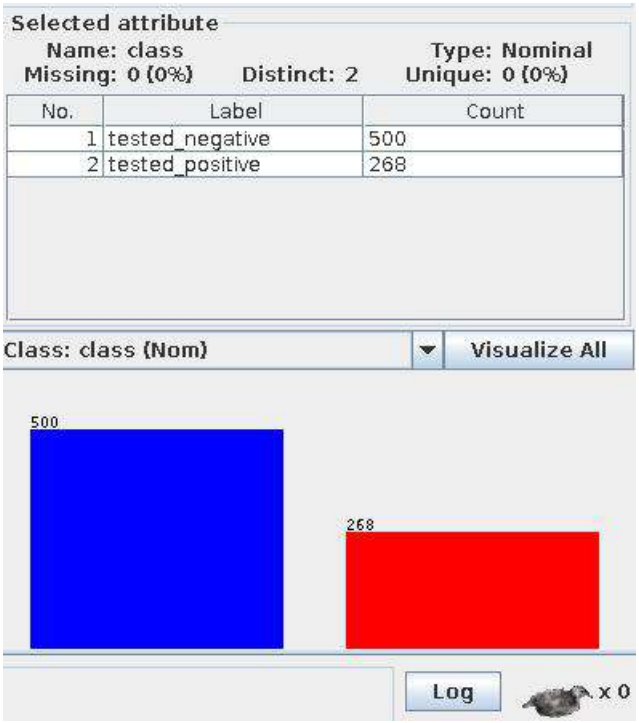
Selected attribute
Name: mass Type: Numeric
Missing: 0 (0%) Distinct: 248 Unique: 76 (10%)

Selected attribute
Name: pedi Type: Numeric
Missing: 0 (0%) Distinct: 517 Unique: 346 (45%)

Selected attribute
Name: age Type: Numeric
Missing: 0 (0%) Distinct: 52 Unique: 5 (1%)

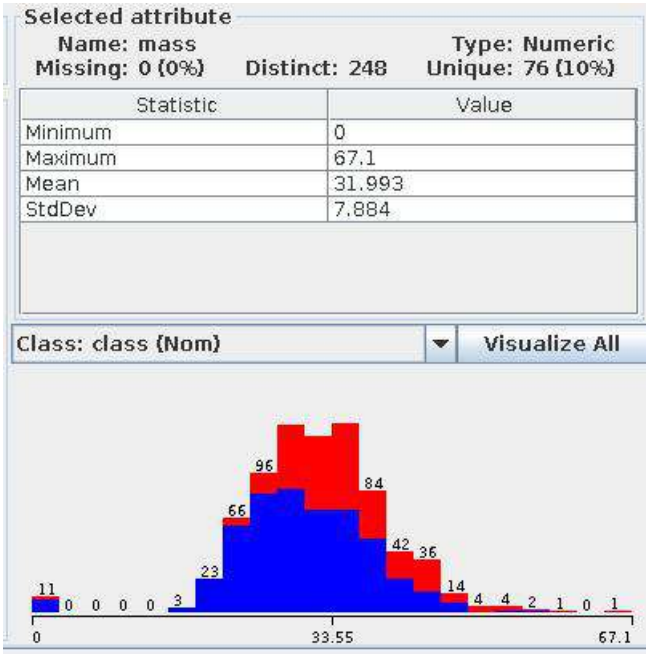
Selected attribute
Name: class Type: Nominal
Missing: 0 (0%) Distinct: 2 Unique: 0 (0%)

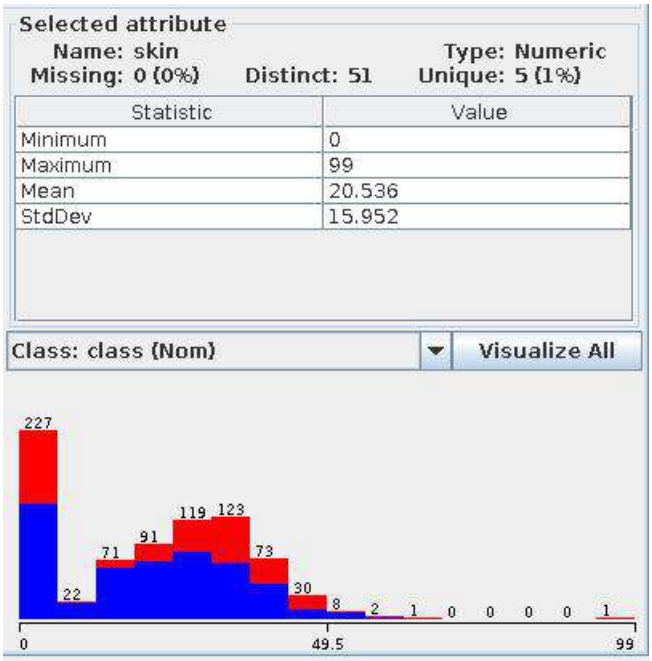
3. Study the Class labels. Do you think if there is any class imbalance.What are the class labels?Which class is dominant?Paste the screen shot - depicting the number of classes

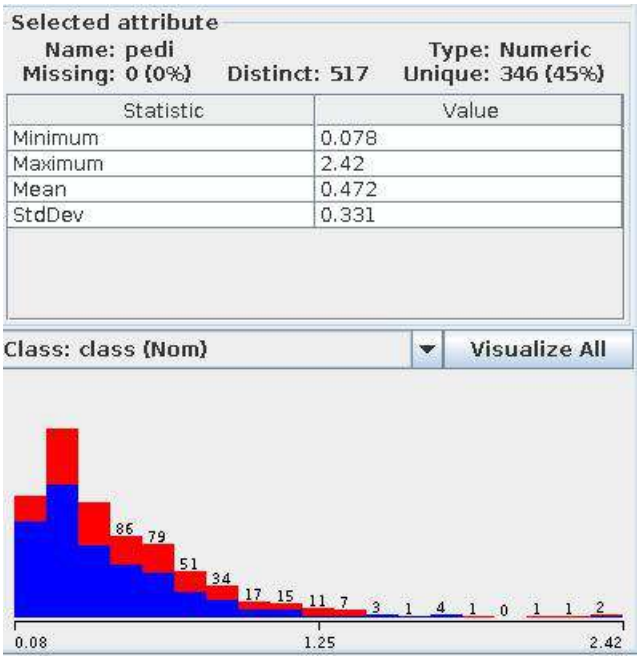


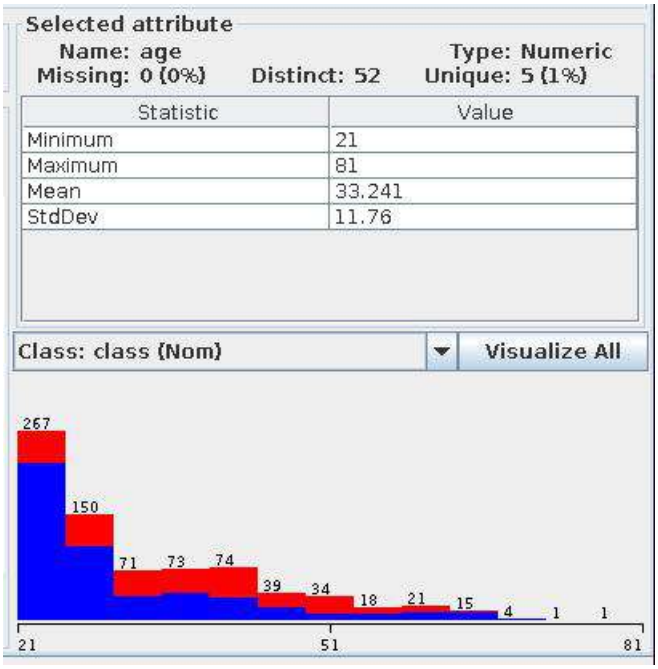
tested_negative class is dominant.

4. Study the spread of at-least three numeric attributes and provide statistical measures

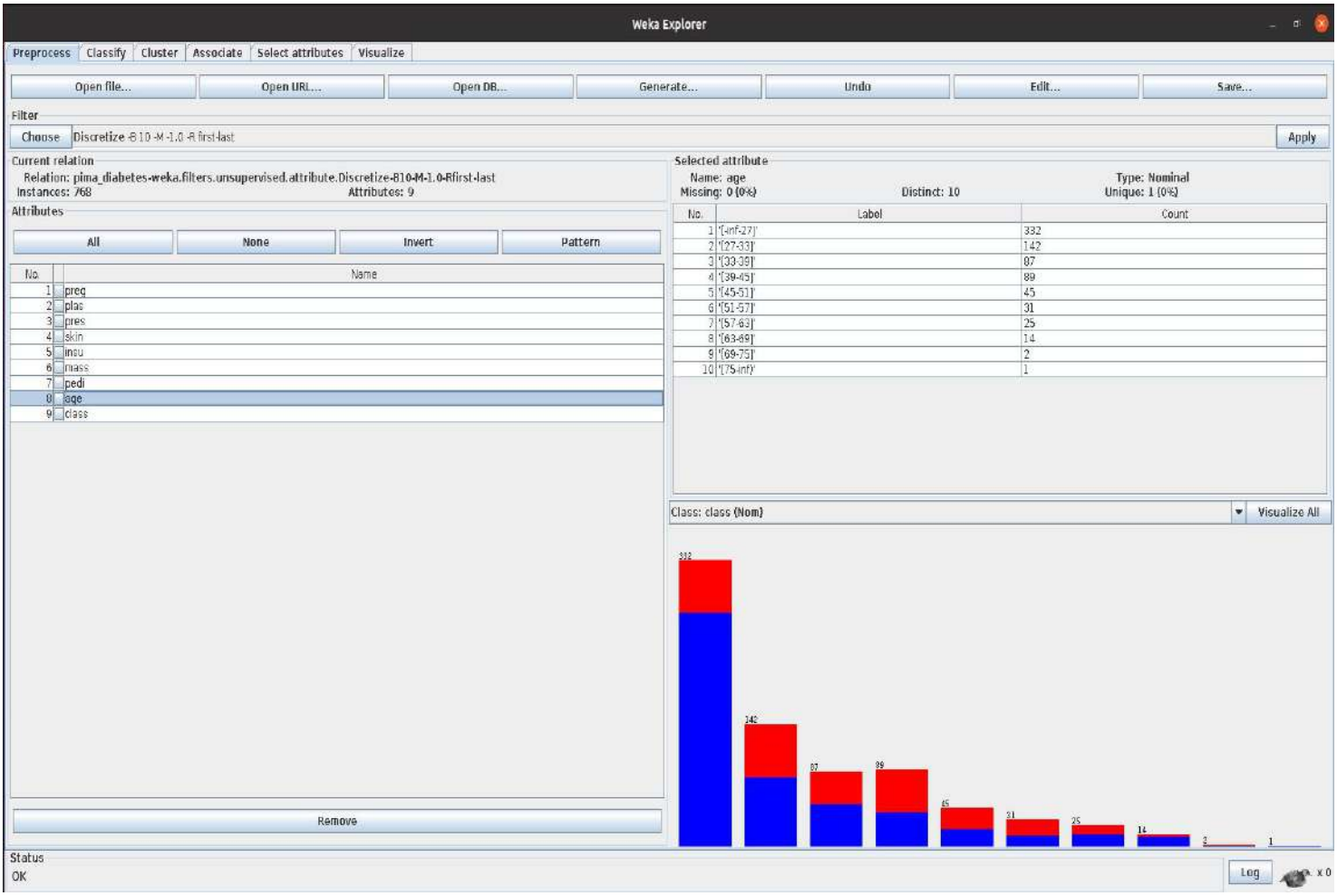




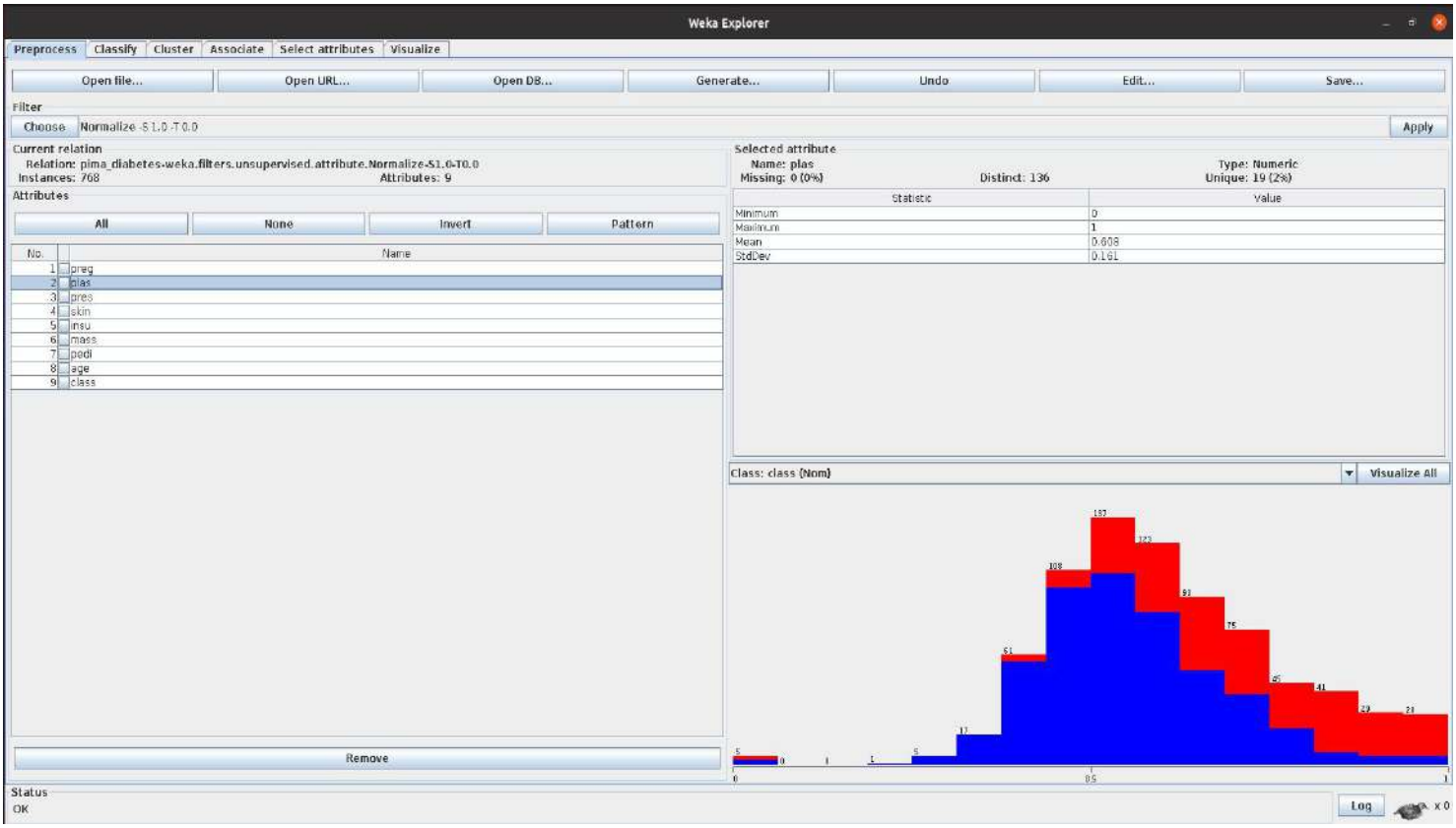




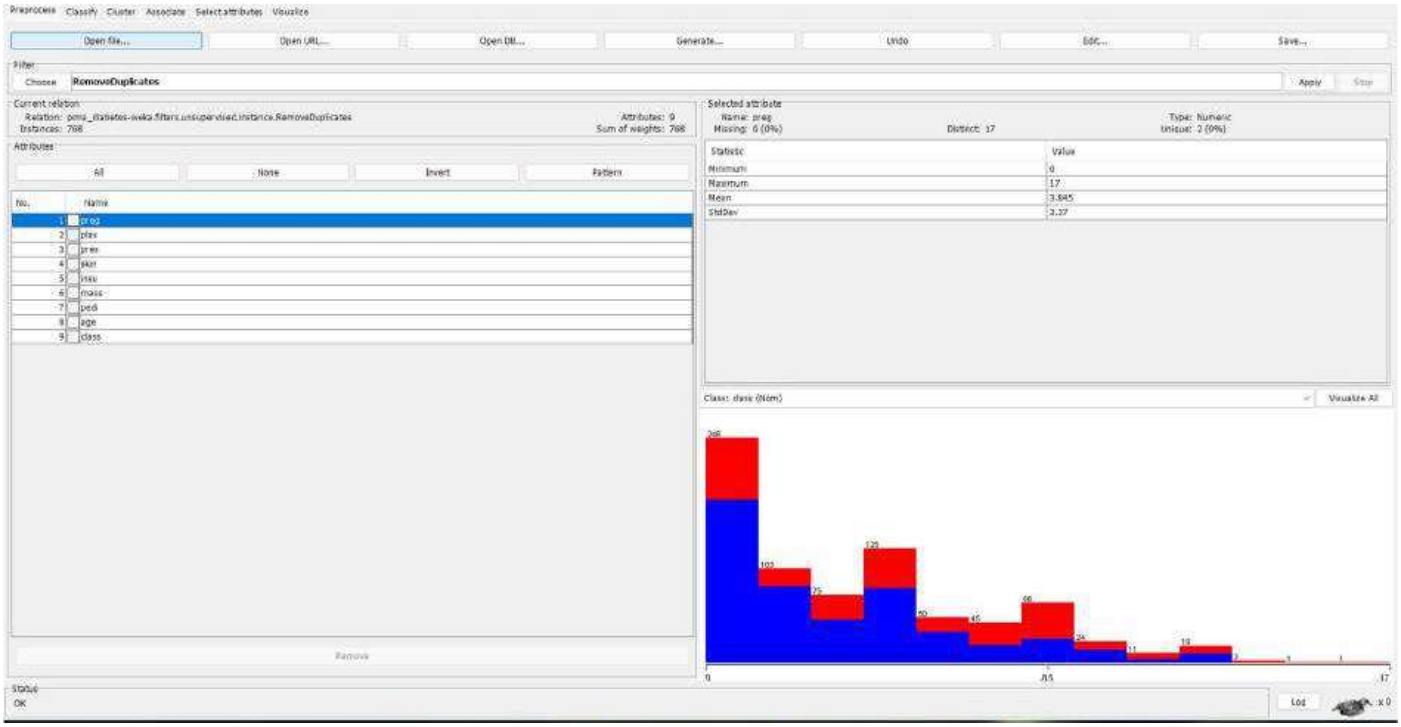
5. Discretize Age attribute –choose appropriate number of bins for discretization



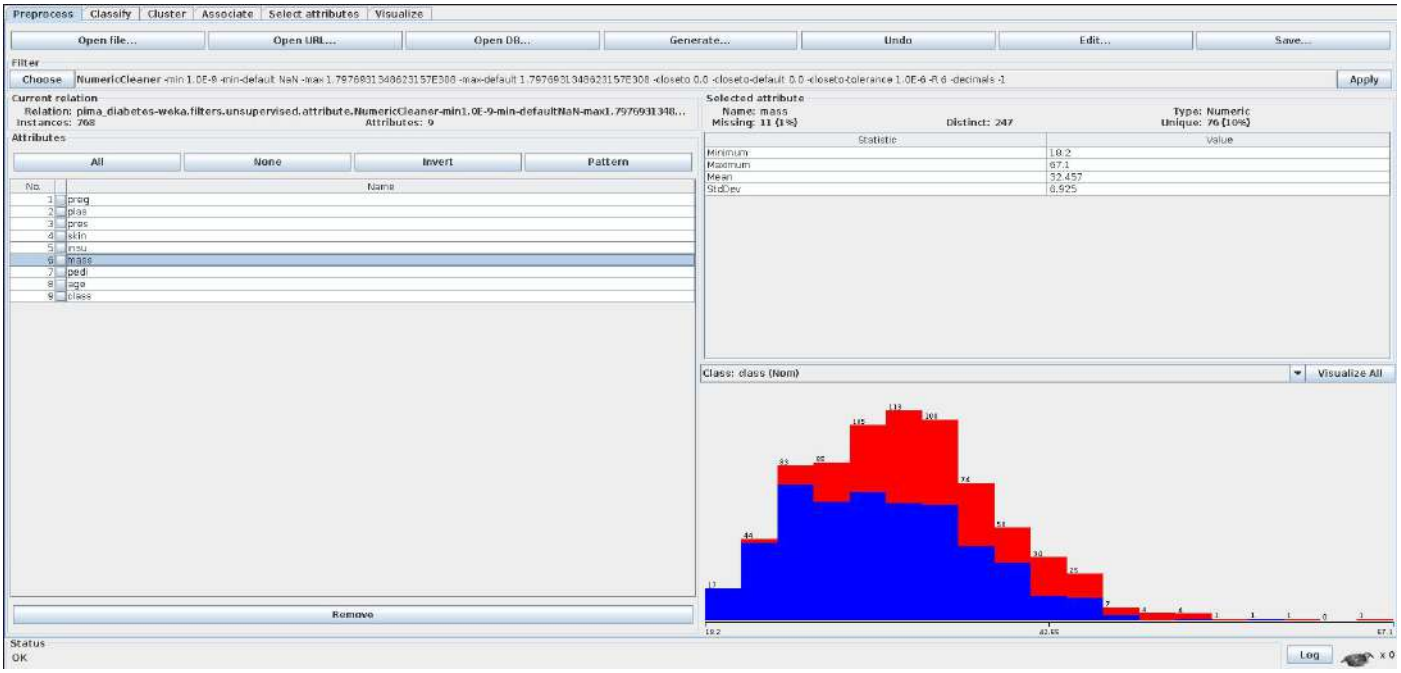
6.Normalize plas attribute using min-max normalization–use normalization range as (0 ,1)

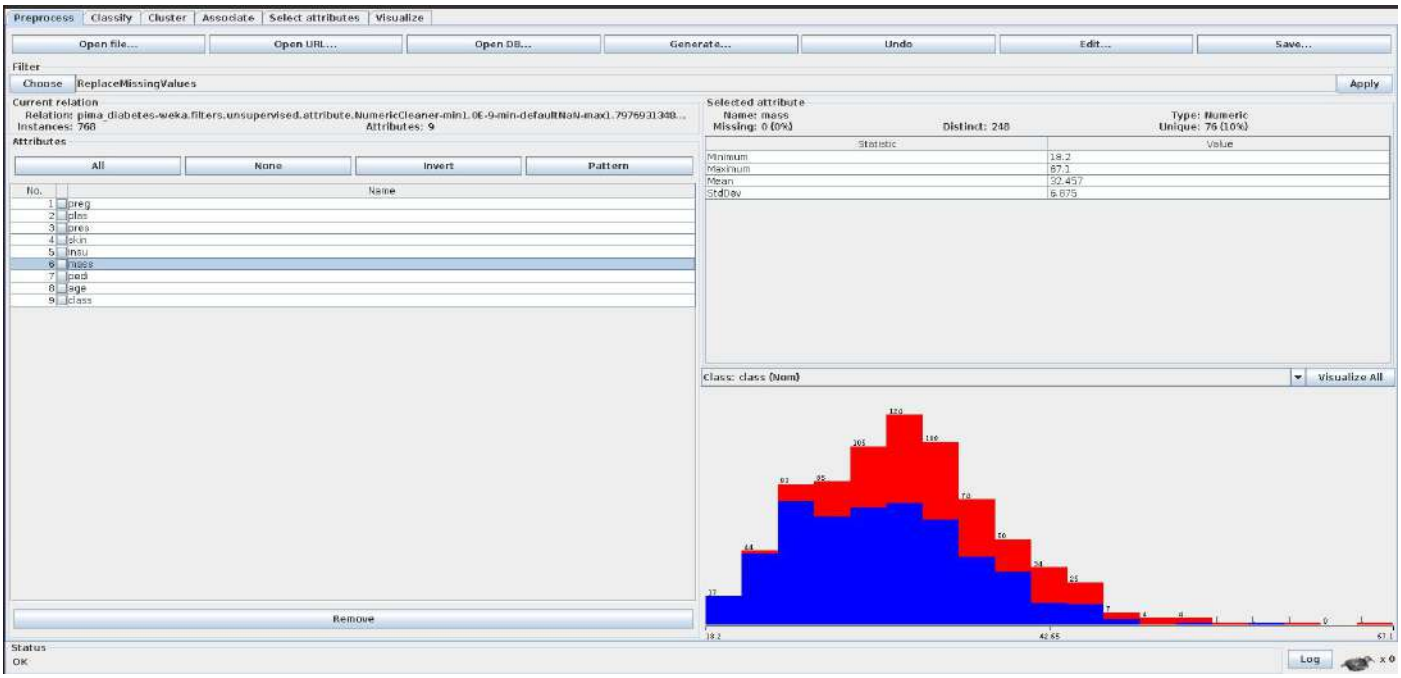


7.Remove Duplicate instances if any



8. Impute Missing values using mean value imputation.


Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
NumericCleaner -min 1.0E-9 -min-default NaN -max 1.7976931348623157E308 -max-default 1.7976931348623157E308 -closeto 0.0 -closeto-default 0.0 -closeto-tolerance 1.0E-6 -R 6 -decimals -1
Apply
Current relation
Relation: pima_diabetes-weka.filters.unsupervised.attribute.NumericCleaner-min1.0E-9-min-defaultNaN-max1.7976931348...
Instances: 768
Attributes: 9
Attributes
All
None
Invert
Pattern
No.
Name
1 preg
2 plas
3 pres
4 skin
5 insu
6 mass
7 pedi
8 age
9 class
Remove
Selected attribute
Name: mass
Missing: 11 (1%)
Distinct: 247
Type: Numeric
Unique: 76 (10%)
Statistic
Value
Minimum 18.2
Maximum 67.1
Mean 32.457
StdDev 6.925
Class: class (Nom)
Visualize All
18.2
42.65
67.1
Status
OK
Log
x 0


Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
ReplaceMissingValues
Apply
Current relation
Relation: pima_diabetes-weka.filters.unsupervised.attribute.NumericCleaner-min1.0E-9-min-defaultNaN-max1.7976931348...
Instances: 768
Attributes: 9
Attributes
All
None
Invert
Pattern
No.
Name
1 preg
2 plas
3 pres
4 skin
5 insu
6 mass
7 pedi
8 age
9 class
Remove
Selected attribute
Name: mass
Missing: 0 (0%)
Distinct: 248
Type: Numeric
Unique: 76 (10%)
Statistic
Value
Minimum 18.2
Maximum 67.1
Mean 32.457
StdDev 6.875
Class: class (Nom)
Visualize All
18.2
42.65
67.1
Status
OK
Log
x 0