Name: Minahil Mobin

Roll No: 21L-XXXX.

Section: BDS-XXX

Subject: Data Mining Lab

LAB#01:

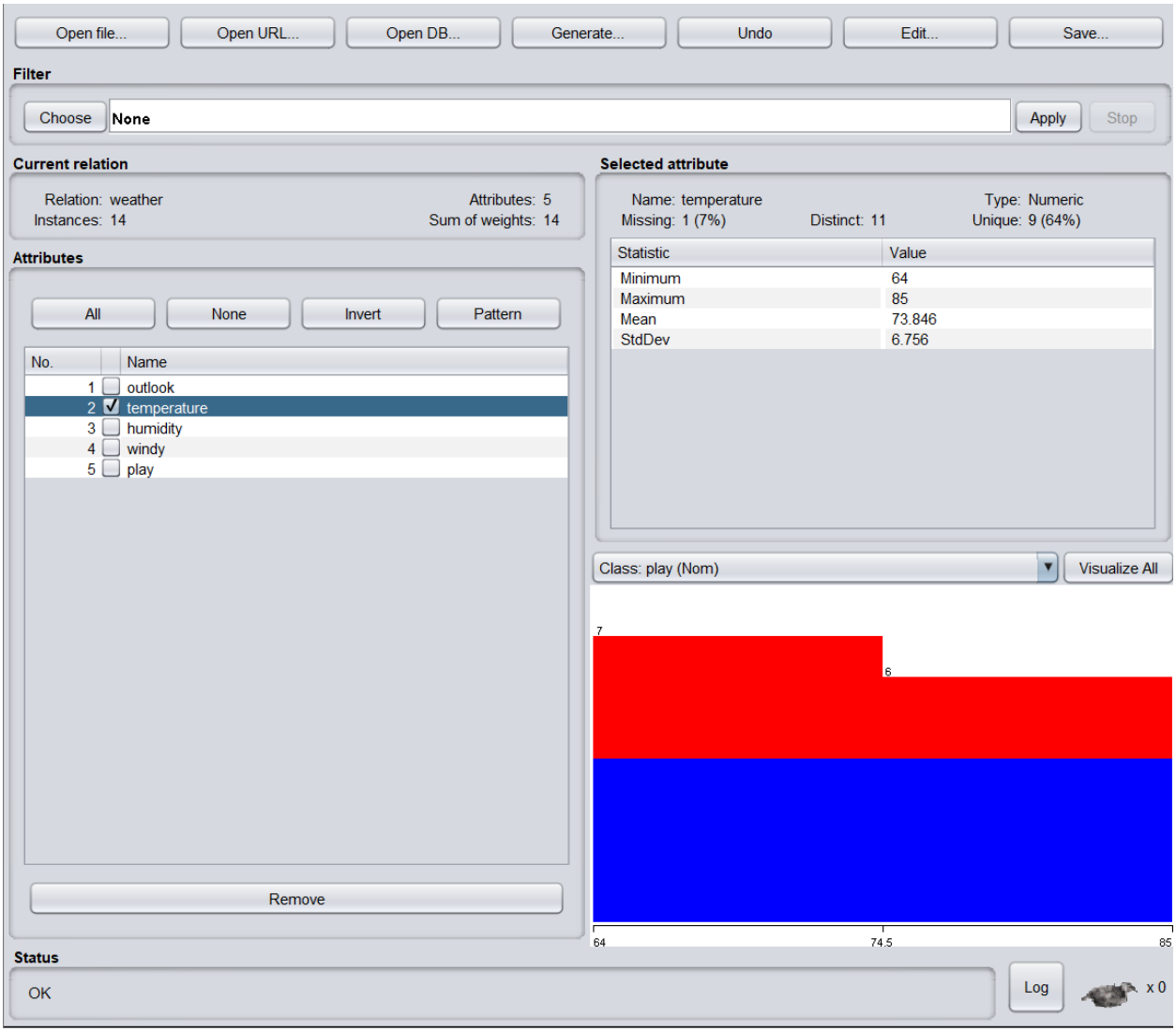
VIEW OF NOMINAL FILE:

Viewer

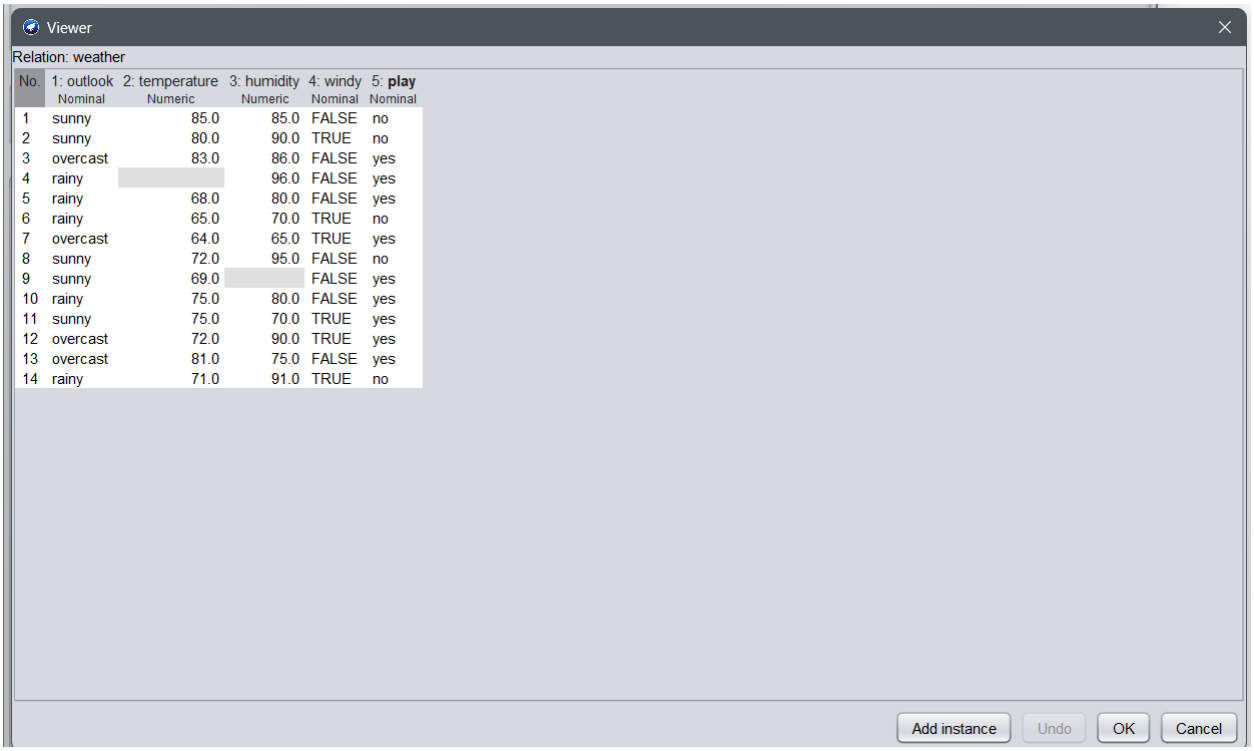
Relation: weather.symbolic

No.	1: outlook Nominal	2: temperature Nominal	3: humidity Nominal	4: windy Nominal	5: **play** Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

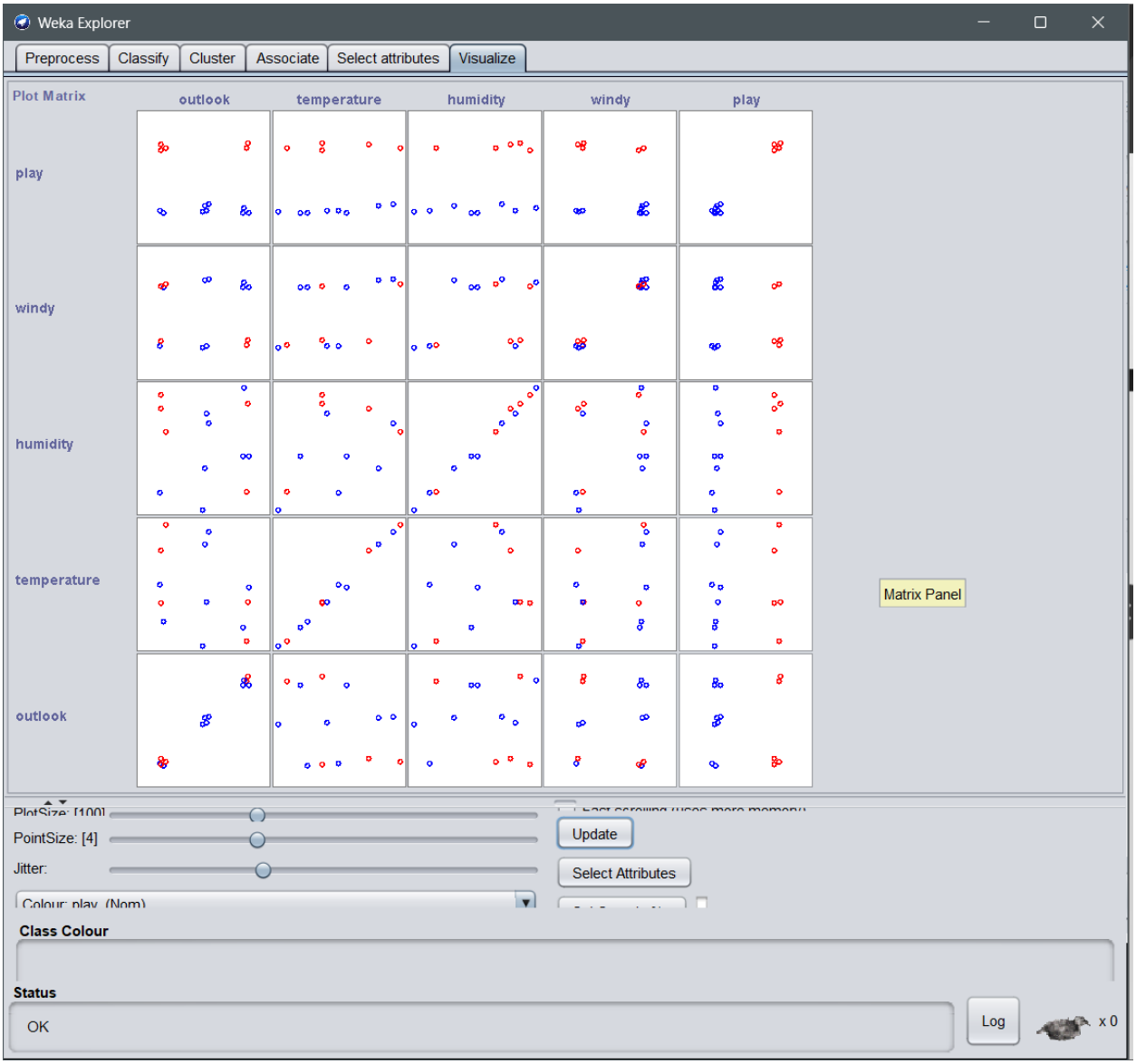
GENERAL VIEW OF NUMERICAL FILE:



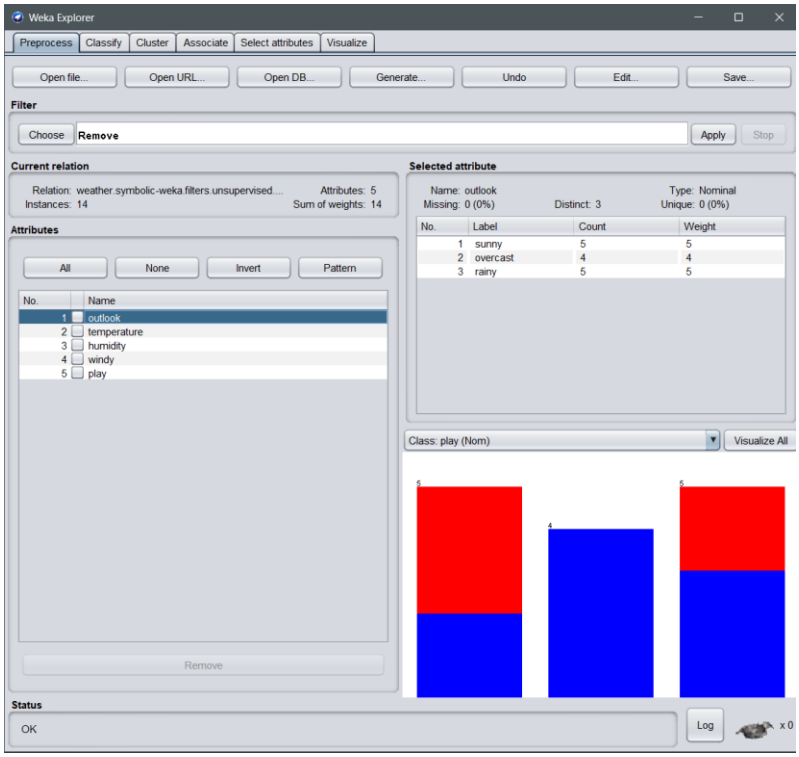
VIEW OF NUMERIC FILE:

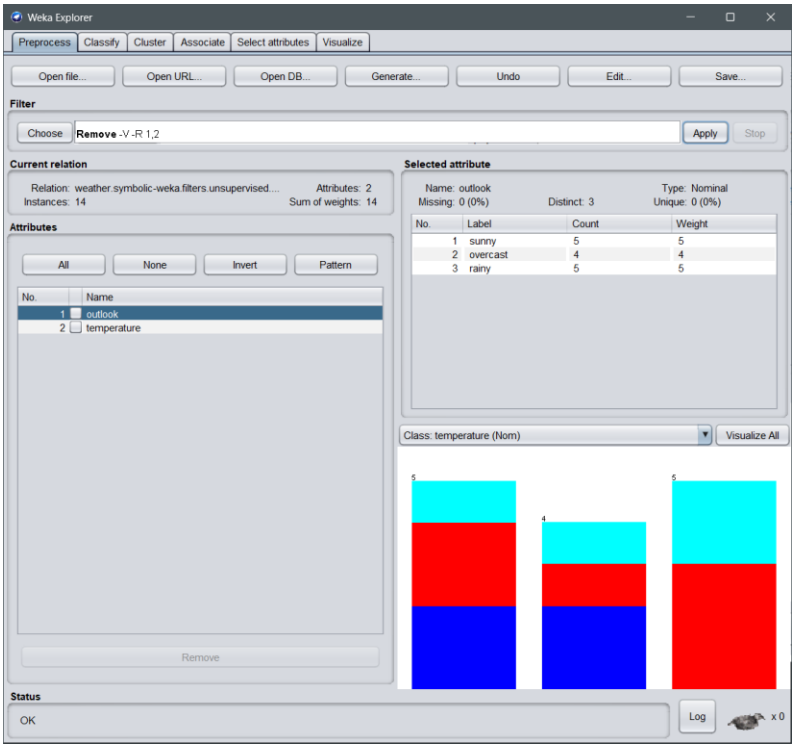


2D VIEW OF NUMERICAL FILE:

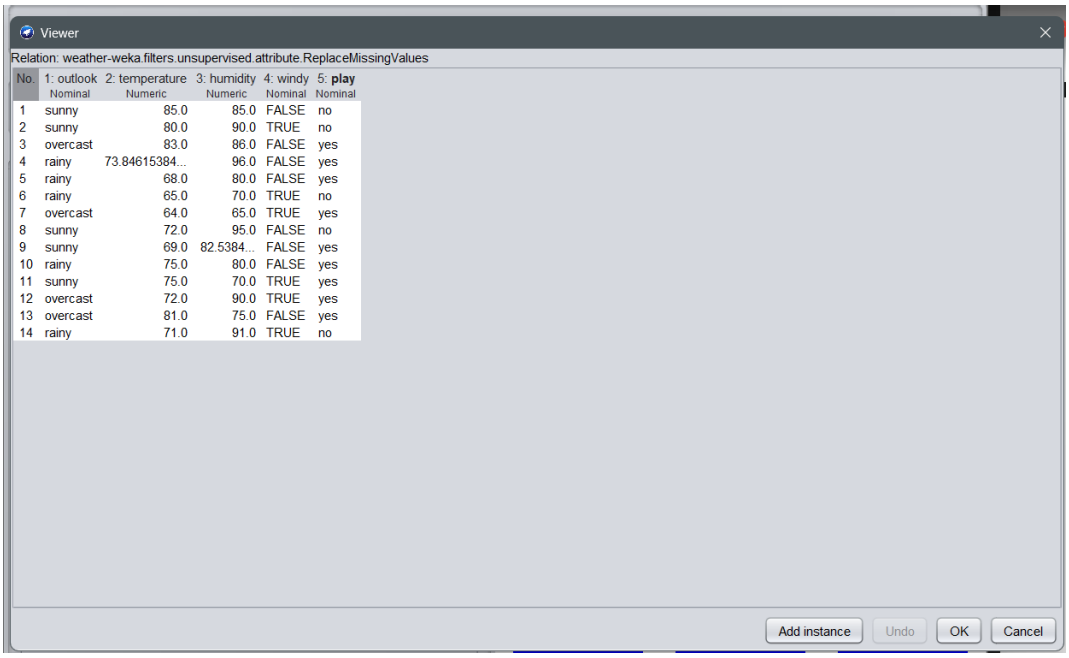


REMOVE IN NUMERIC FILE:

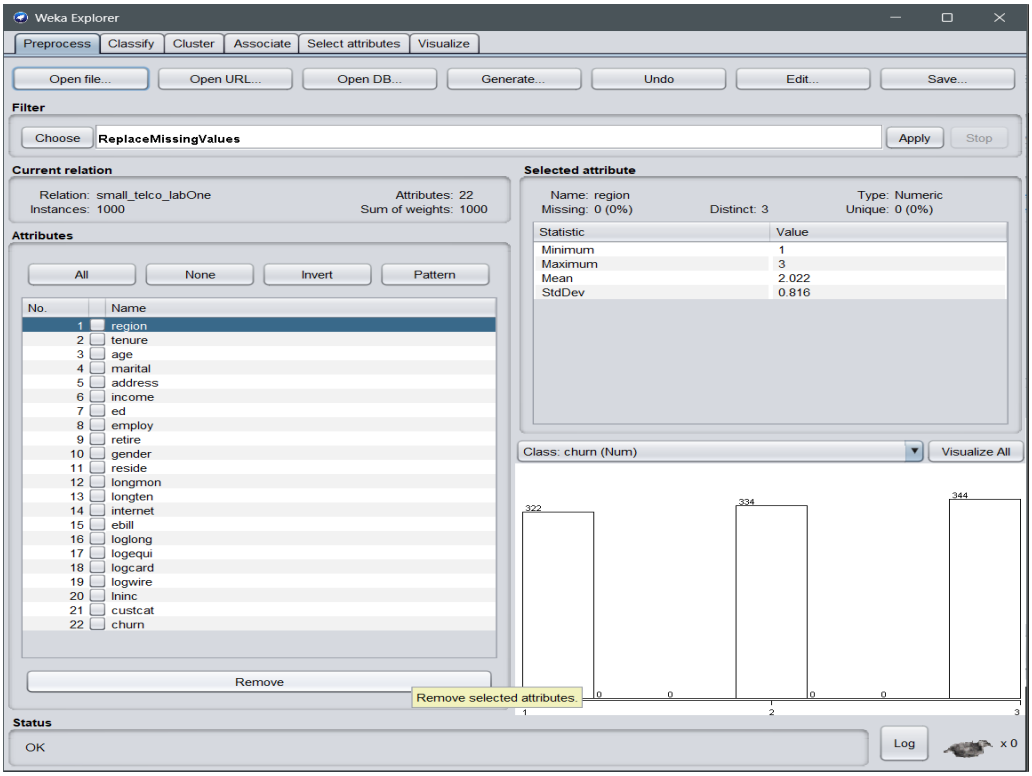




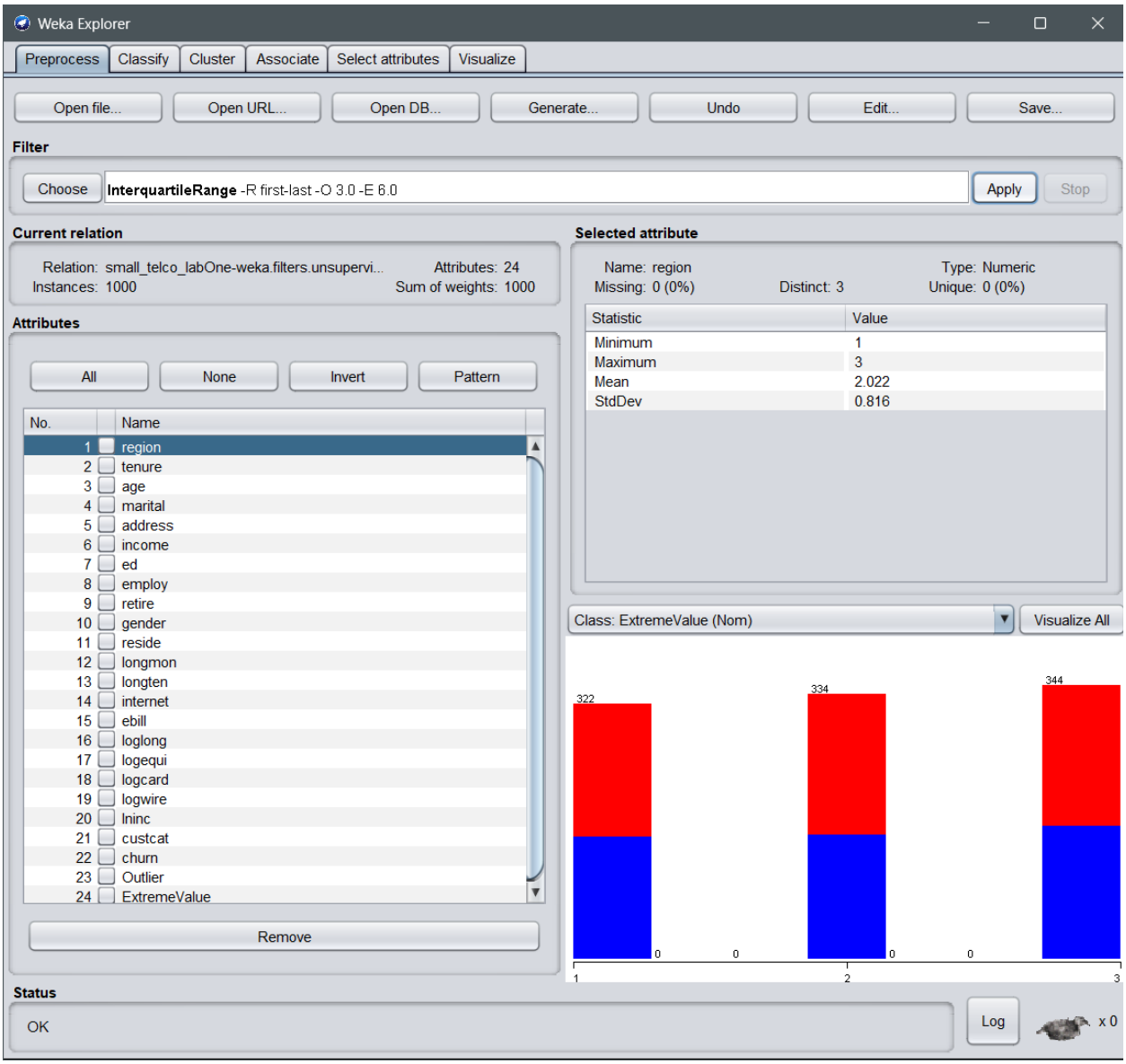
FILLING VALUES IN NUMERIC FILE:

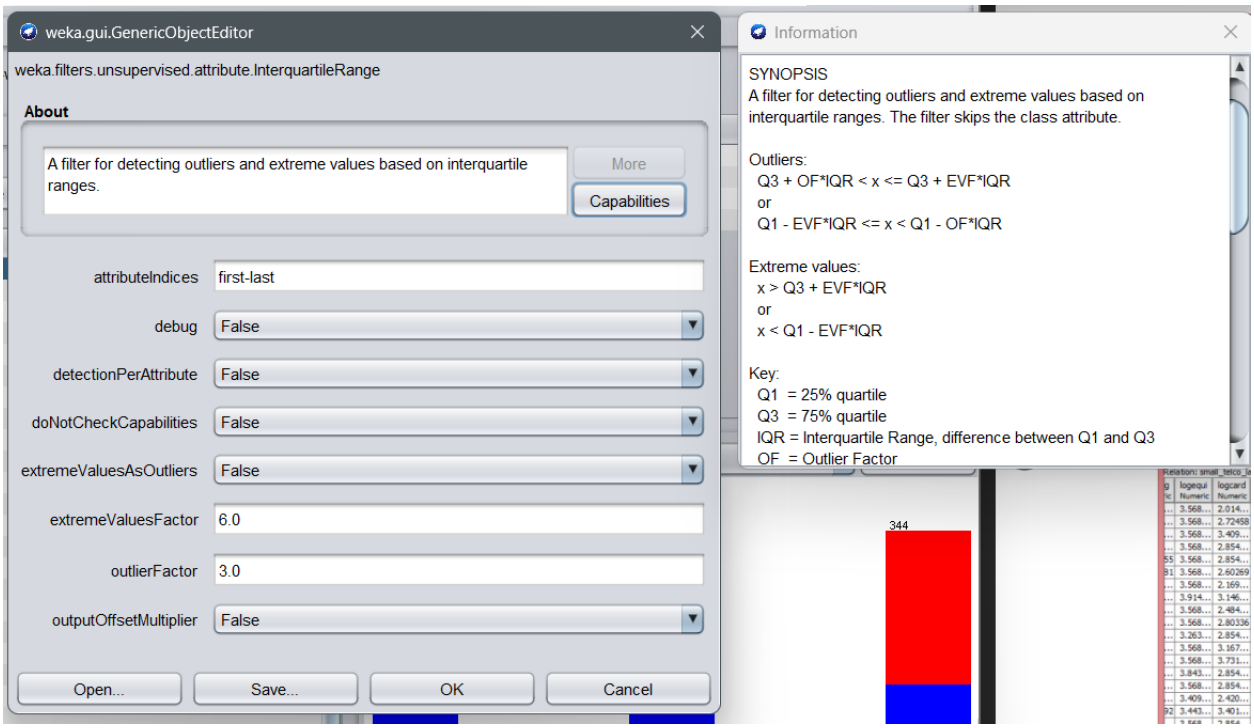


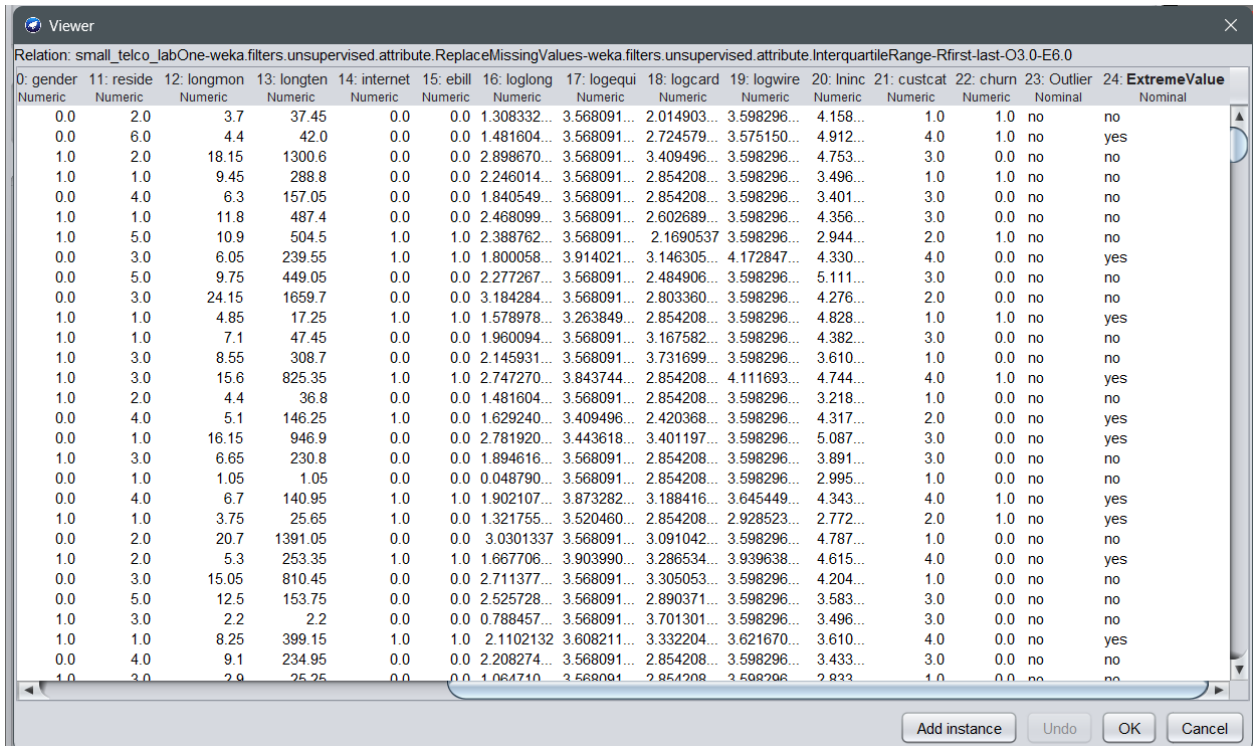
REPLACING MISSING VALUES IN CSV FILE:

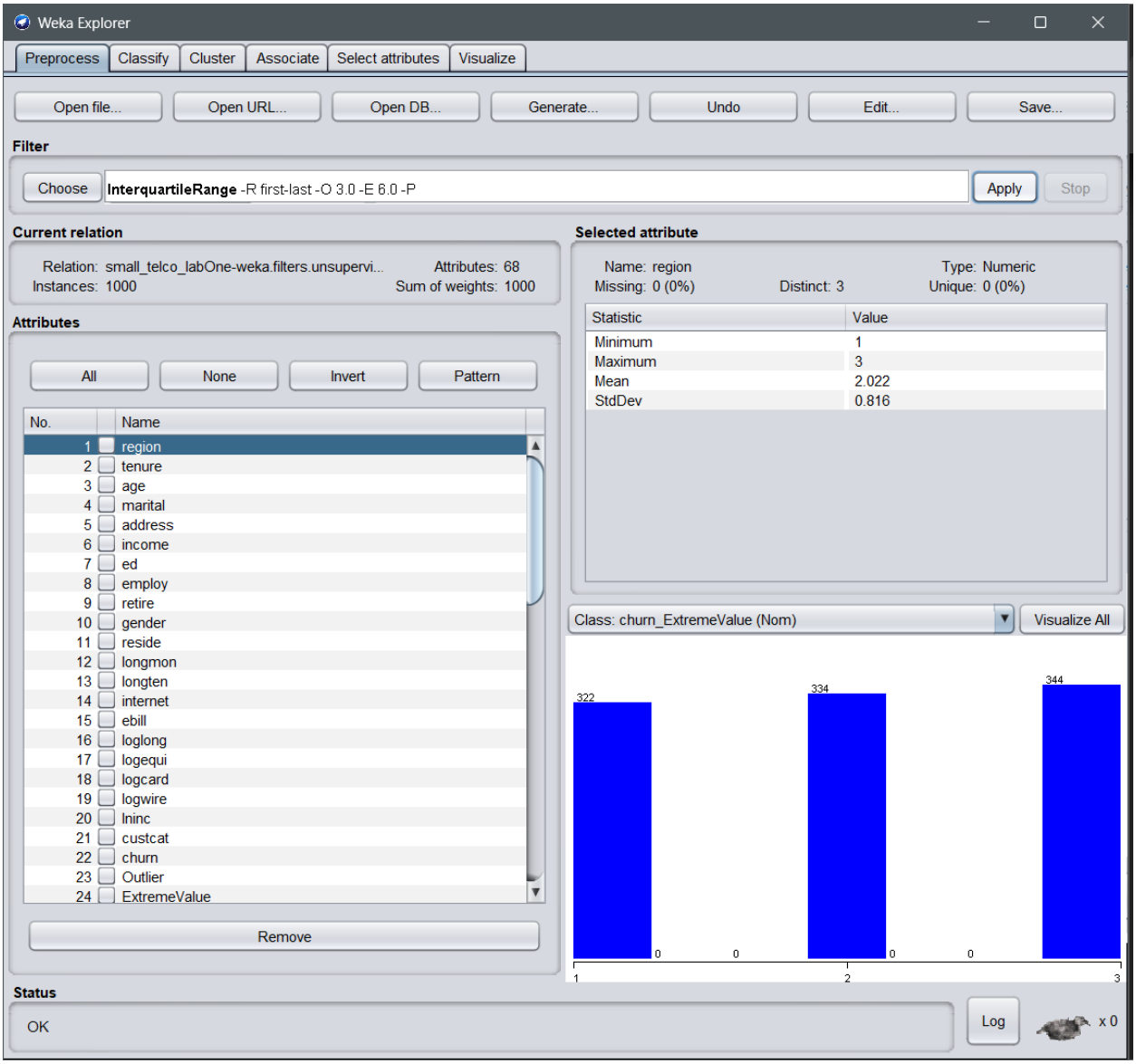


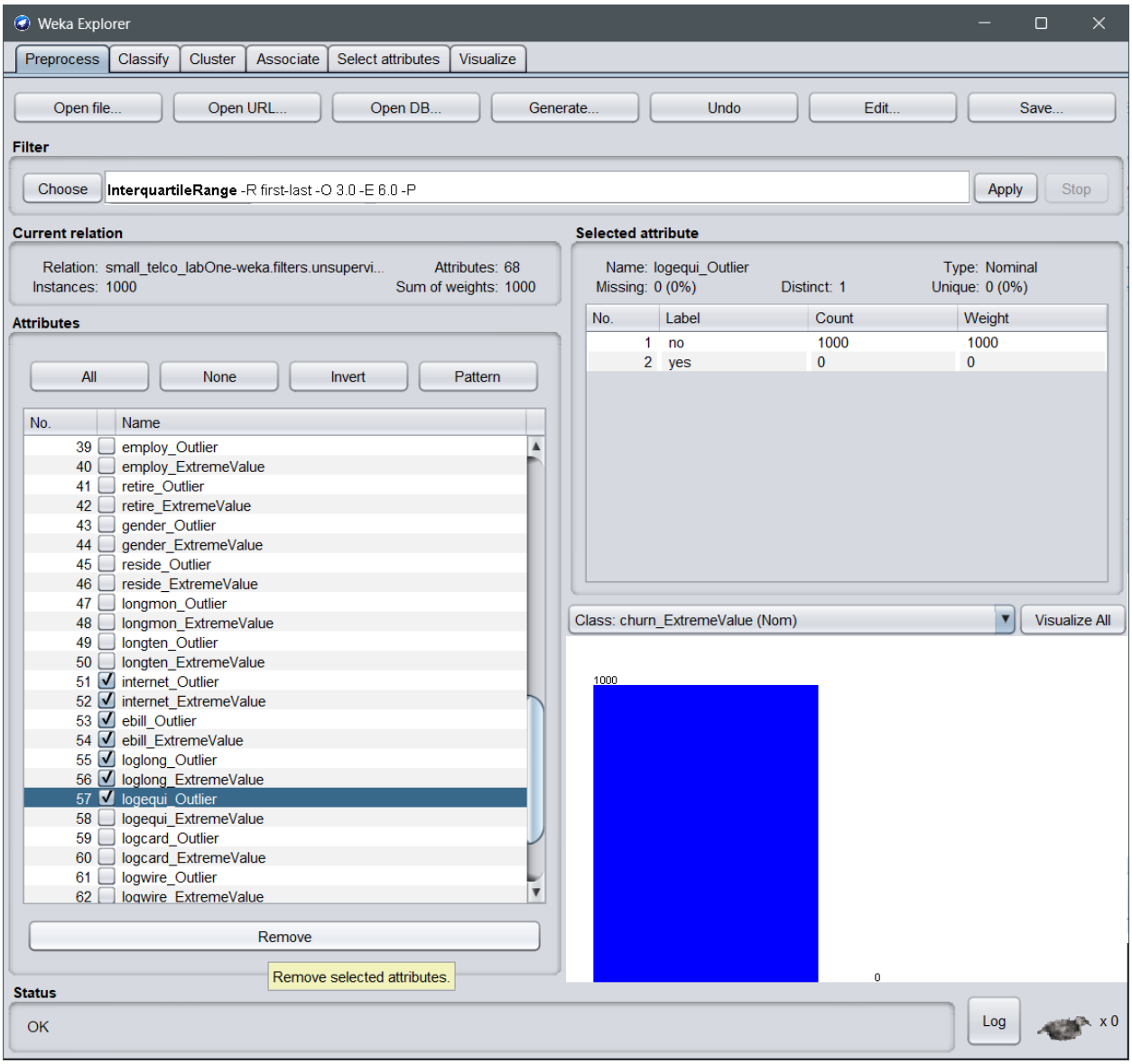
IQR ON CSV FILE:



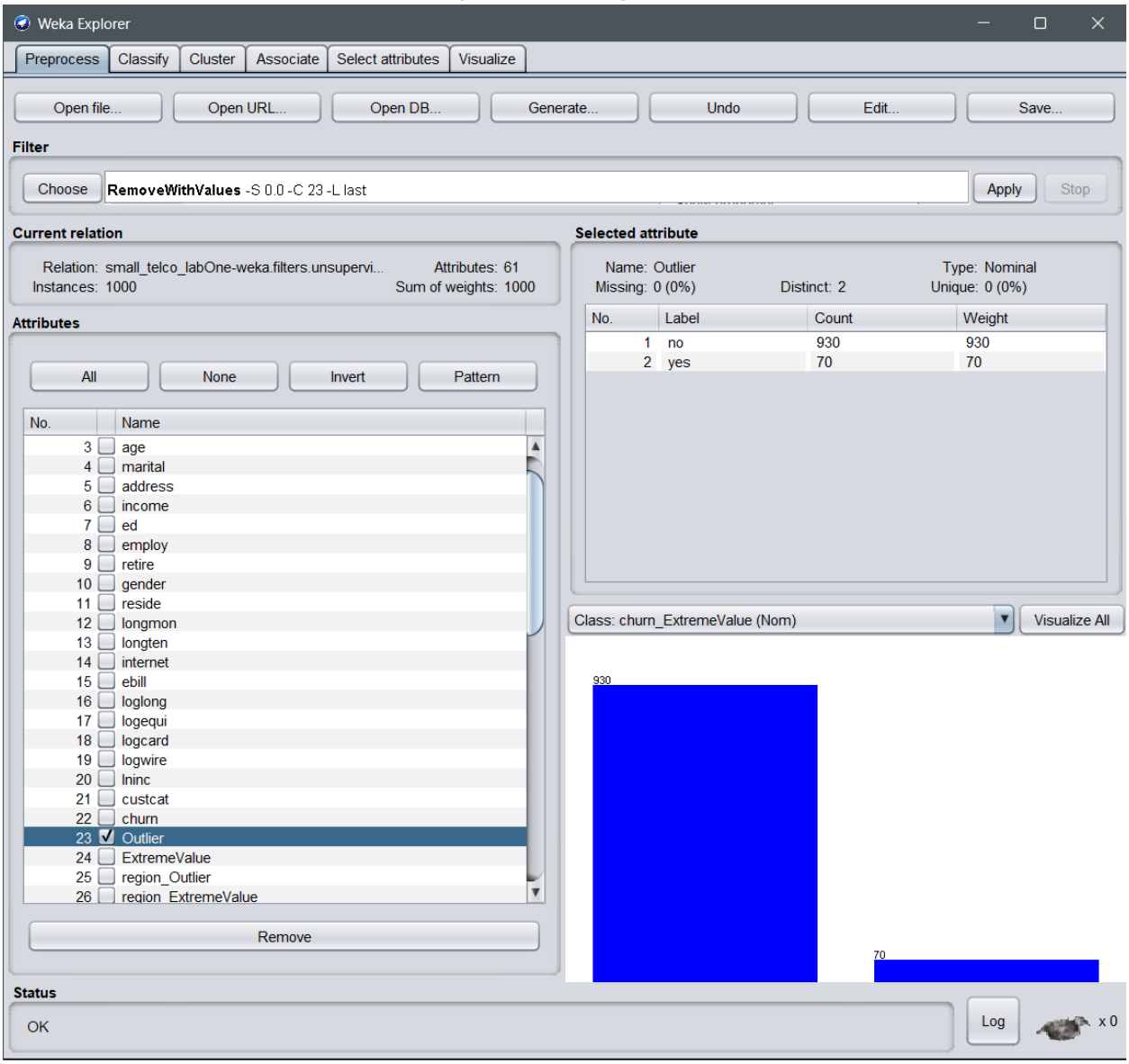

weka.gui.GenericObjectEditor
weka.filters.unsupervised.attribute.InterquartileRange
About
A filter for detecting outliers and extreme values based on interquartile ranges.
More
Capabilities
attributeIndices
first-last
debug
False
detectionPerAttribute
False
doNotCheckCapabilities
False
extremeValuesAsOutliers
False
extremeValuesFactor
6.0
outlierFactor
3.0
outputOffsetMultiplier
False
Open...
Save...
OK
Cancel
Information
SYNOPSIS
A filter for detecting outliers and extreme values based on interquartile ranges. The filter skips the class attribute.
Outliers:
Q3 + OF*IQR < x <= Q3 + EVF*IQR
or
Q1 - EVF*IQR <= x < Q1 - OF*IQR
Extreme values:
x > Q3 + EVF*IQR
or
x < Q1 - EVF*IQR
Key:
Q1 = 25% quartile
Q3 = 75% quartile
IQR = Interquartile Range, difference between Q1 and Q3
OF = Outlier Factor
344

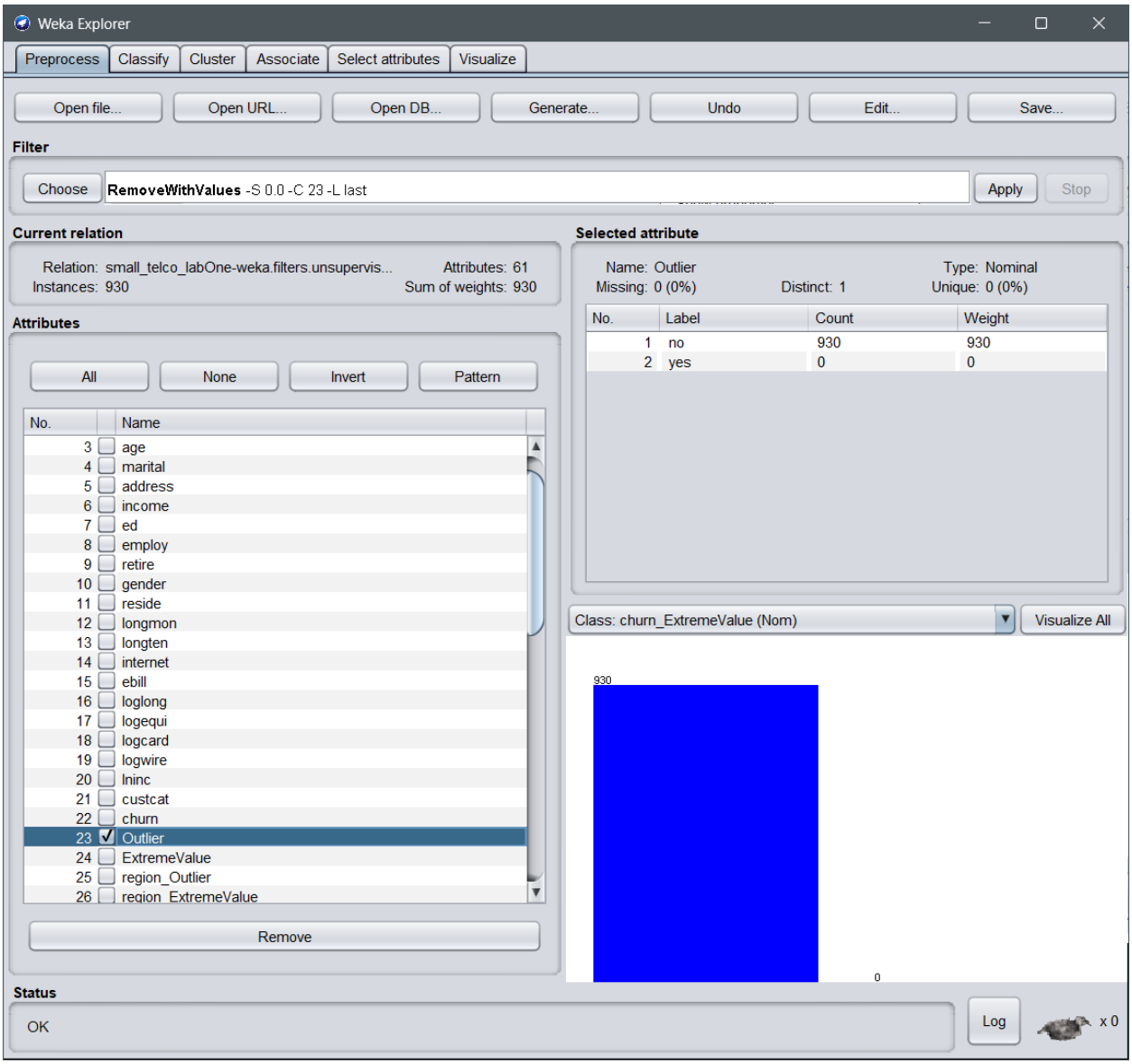

Viewer
Relation: small_telco_labOne-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.unsupervised.attribute.InterquartileRange-Rfirst-last-O3.0-E6.0
0: gender Numeric
11: reside Numeric
12: longmon Numeric
13: longten Numeric
14: internet Numeric
15: ebill Numeric
16: loglong Numeric
17: logequi Numeric
18: logcard Numeric
19: logwire Numeric
20: lninc Numeric
21: custcat Numeric
22: churn Numeric
23: Outlier Nominal
24: ExtremeValue Nominal
0.0 2.0 3.7 37.45 0.0 0.0 1.308332... 3.568091... 2.014903... 3.598296... 4.158... 1.0 1.0 no no
0.0 6.0 4.4 42.0 0.0 0.0 1.481604... 3.568091... 2.724579... 3.575150... 4.912... 4.0 1.0 no yes
1.0 2.0 18.15 1300.6 0.0 0.0 2.898670... 3.568091... 3.409496... 3.598296... 4.753... 3.0 0.0 no no
1.0 1.0 9.45 288.8 0.0 0.0 2.246014... 3.568091... 2.854208... 3.598296... 3.496... 1.0 1.0 no no
0.0 4.0 6.3 157.05 0.0 0.0 1.840549... 3.568091... 2.854208... 3.598296... 3.401... 3.0 0.0 no no
1.0 1.0 11.8 487.4 0.0 0.0 2.468099... 3.568091... 2.602689... 3.598296... 4.356... 3.0 0.0 no no
1.0 5.0 10.9 504.5 1.0 1.0 2.388762... 3.568091... 2.1690537 3.598296... 2.944... 2.0 1.0 no no
0.0 3.0 6.05 239.55 1.0 1.0 1.800058... 3.914021... 3.146305... 4.172847... 4.330... 4.0 0.0 no yes
0.0 5.0 9.75 449.05 0.0 0.0 2.277267... 3.568091... 2.484906... 3.598296... 5.111... 3.0 0.0 no no
0.0 3.0 24.15 1659.7 0.0 0.0 3.184284... 3.568091... 2.803360... 3.598296... 4.276... 2.0 0.0 no no
1.0 1.0 4.85 17.25 1.0 1.0 1.578978... 3.263849... 2.854208... 3.598296... 4.828... 1.0 1.0 no yes
1.0 1.0 7.1 47.45 0.0 0.0 1.960094... 3.568091... 3.167582... 3.598296... 4.382... 3.0 0.0 no no
1.0 3.0 8.55 308.7 0.0 0.0 2.145931... 3.568091... 3.731699... 3.598296... 3.610... 1.0 0.0 no no
1.0 3.0 15.6 825.35 1.0 1.0 2.747270... 3.843744... 2.854208... 4.111693... 4.744... 4.0 1.0 no yes
1.0 2.0 4.4 36.8 0.0 0.0 1.481604... 3.568091... 2.854208... 3.598296... 3.218... 1.0 0.0 no no
0.0 4.0 5.1 146.25 1.0 0.0 1.629240... 3.409496... 2.420368... 3.598296... 4.317... 2.0 0.0 no yes
0.0 1.0 16.15 946.9 0.0 0.0 2.781920... 3.443618... 3.401197... 3.598296... 5.087... 3.0 0.0 no yes
1.0 3.0 6.65 230.8 0.0 0.0 1.894616... 3.568091... 2.854208... 3.598296... 3.891... 3.0 0.0 no no
0.0 1.0 1.05 1.05 0.0 0.0 0.048790... 3.568091... 2.854208... 3.598296... 2.995... 1.0 0.0 no no
0.0 4.0 6.7 140.95 1.0 1.0 1.902107... 3.873282... 3.188416... 3.645449... 4.343... 4.0 1.0 no yes
1.0 1.0 3.75 25.65 1.0 0.0 1.321755... 3.520460... 2.854208... 2.928523... 2.772... 2.0 1.0 no yes
0.0 2.0 20.7 1391.05 0.0 0.0 3.0301337 3.568091... 3.091042... 3.598296... 4.787... 1.0 0.0 no no
1.0 2.0 5.3 253.35 1.0 1.0 1.667706... 3.903990... 3.286534... 3.939638... 4.615... 4.0 0.0 no yes
0.0 3.0 15.05 810.45 0.0 0.0 2.711377... 3.568091... 3.305053... 3.598296... 4.204... 1.0 0.0 no no
0.0 5.0 12.5 153.75 0.0 0.0 2.525728... 3.568091... 2.890371... 3.598296... 3.583... 3.0 0.0 no no
1.0 3.0 2.2 2.2 0.0 0.0 0.788457... 3.568091... 3.701301... 3.598296... 3.496... 3.0 0.0 no no
1.0 1.0 8.25 399.15 1.0 1.0 2.1102132 3.608211... 3.332204... 3.621670... 3.610... 4.0 0.0 no yes
0.0 4.0 9.1 234.95 0.0 0.0 2.208274... 3.568091... 2.854208... 3.598296... 3.433... 3.0 0.0 no no
Add instance
Undo
OK
Cancel


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
InterquartileRange -R first-last -O 3.0 -E 6.0 -P
Apply
Stop
Current relation
Relation: small_telco_labOne-weka.filters.unsupervi...
Attributes: 68
Instances: 1000
Sum of weights: 1000
Selected attribute
Name: region
Type: Numeric
Missing: 0 (0%)
Distinct: 3
Unique: 0 (0%)
Statistic
Value
Minimum 1
Maximum 3
Mean 2.022
StdDev 0.816
Attributes
All
None
Invert
Pattern
No.
Name
1 region
2 tenure
3 age
4 marital
5 address
6 income
7 ed
8 employ
9 retire
10 gender
11 reside
12 longmon
13 longten
14 internet
15 ebill
16 loglong
17 logequi
18 logcard
19 logwire
20 lninc
21 custcat
22 churn
23 Outlier
24 ExtremeValue
Remove
Class: churn_ExtremeValue (Nom)
Visualize All
322
0
0
334
0
0
344
1
2
3
Status
OK
Log
x 0

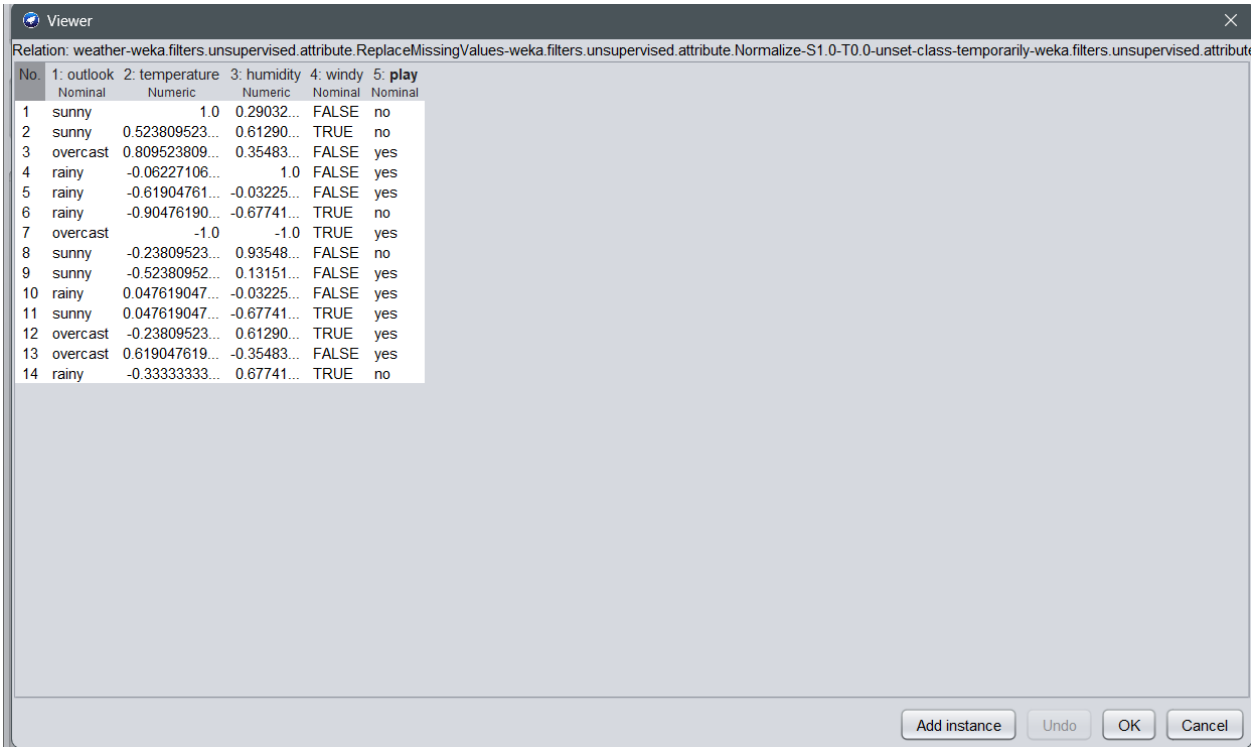

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
InterquartileRange -R first-last -O 3.0 -E 6.0 -P
Apply
Stop
Current relation
Relation: small_telco_labOne-weka.filters.unsupervi...
Attributes: 68
Instances: 1000
Sum of weights: 1000
Selected attribute
Name: logequi_Outlier
Type: Nominal
Missing: 0 (0%)
Distinct: 1
Unique: 0 (0%)
No. Label Count Weight
1 no 1000 1000
2 yes 0 0
Attributes
All
None
Invert
Pattern
No. Name
39 employ_Outlier
40 employ_ExtremeValue
41 retire_Outlier
42 retire_ExtremeValue
43 gender_Outlier
44 gender_ExtremeValue
45 reside_Outlier
46 reside_ExtremeValue
47 longmon_Outlier
48 longmon_ExtremeValue
49 longten_Outlier
50 longten_ExtremeValue
51 internet_Outlier
52 internet_ExtremeValue
53 ebill_Outlier
54 ebill_ExtremeValue
55 loglong_Outlier
56 loglong_ExtremeValue
57 logequi_Outlier
58 logequi_ExtremeValue
59 logcard_Outlier
60 logcard_ExtremeValue
61 logwire_Outlier
62 logwire_ExtremeValue
Class: churn_ExtremeValue (Nom)
Visualize All
1000
0
Remove
Remove selected attributes.
Status
OK
Log
x 0

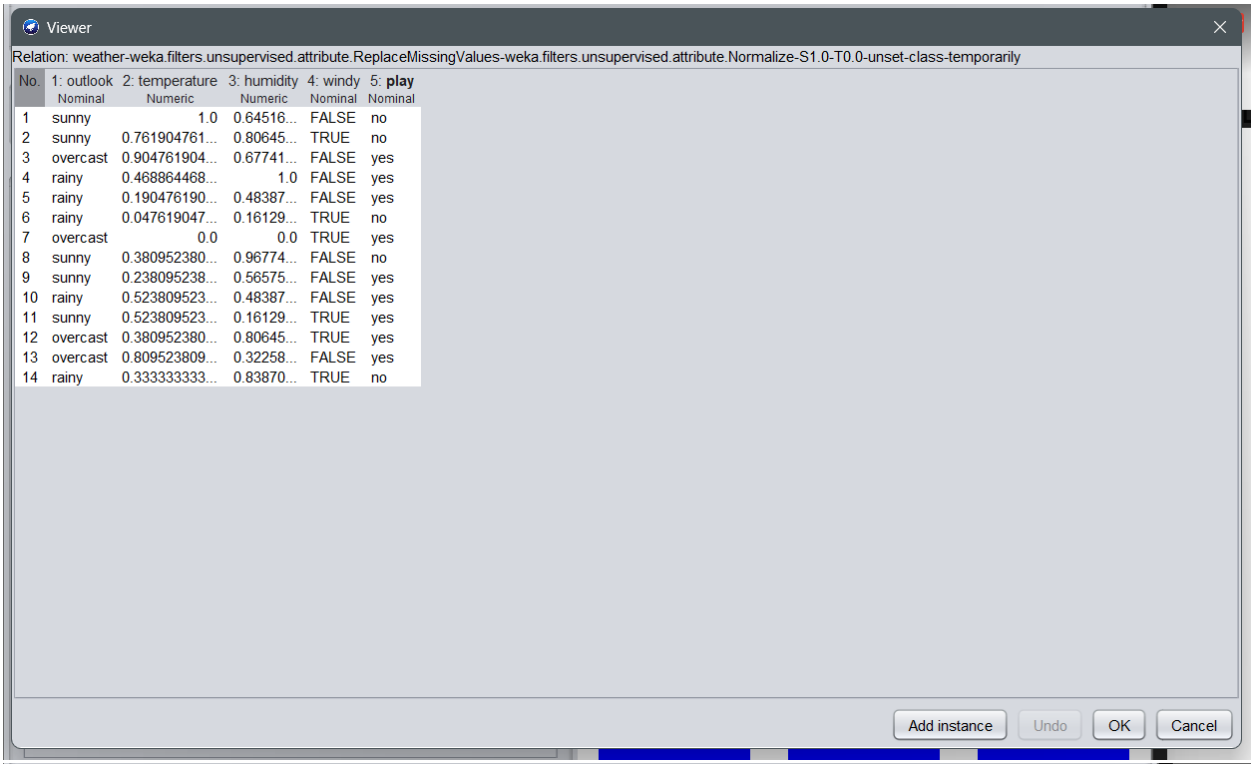
REMOVING WITH VALUES IN CSV FILE (Before & After**):**

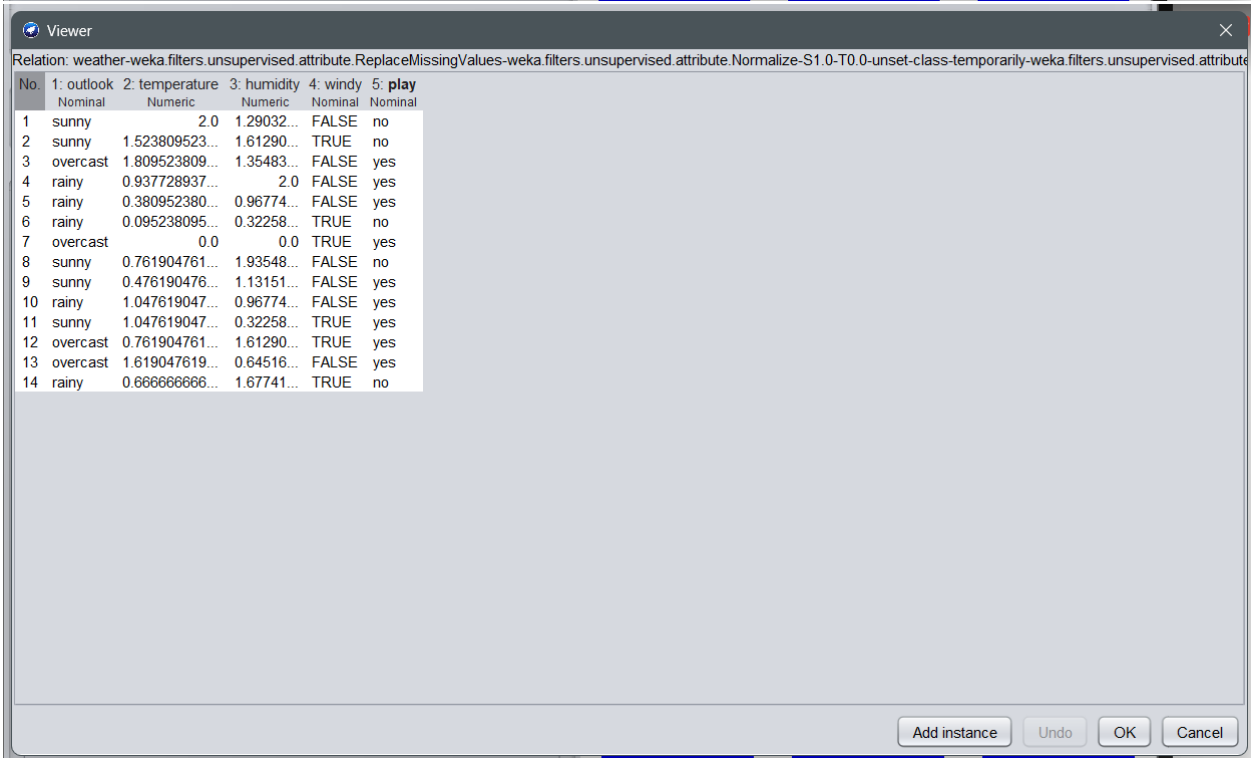



Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
RemoveWithValues -S 0.0 -C 23 -L last
Apply
Stop
Current relation
Relation: small_telco_labOne-weka.filters.unsupervis...
Attributes: 61
Instances: 930
Sum of weights: 930
Selected attribute
Name: Outlier
Type: Nominal
Missing: 0 (0%)
Distinct: 1
Unique: 0 (0%)
No. Label Count Weight
1 no 930 930
2 yes 0 0
Attributes
All
None
Invert
Pattern
No. Name
3 age
4 marital
5 address
6 income
7 ed
8 employ
9 retire
10 gender
11 reside
12 longmon
13 longten
14 internet
15 ebill
16 loglong
17 logequi
18 logcard
19 logwire
20 lninc
21 custcat
22 churn
23 Outlier
24 ExtremeValue
25 region_Outlier
26 region_ExtremeValue
Remove
Class: churn_ExtremeValue (Nom)
Visualize All
930
0
Status
OK
Log
x 0

NORMALIZATION ON NUMERIC FILE:




Viewer
Relation: weather-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-unset-class-temporarily
No. 1: outlook Nominal 2: temperature Numeric 3: humidity Numeric 4: windy Nominal 5: play Nominal
1 sunny 1.0 0.64516... FALSE no
2 sunny 0.761904761... 0.80645... TRUE no
3 overcast 0.904761904... 0.67741... FALSE yes
4 rainy 0.468864468... 1.0 FALSE yes
5 rainy 0.190476190... 0.48387... FALSE yes
6 rainy 0.047619047... 0.16129... TRUE no
7 overcast 0.0 0.0 TRUE yes
8 sunny 0.380952380... 0.96774... FALSE no
9 sunny 0.238095238... 0.56575... FALSE yes
10 rainy 0.523809523... 0.48387... FALSE yes
11 sunny 0.523809523... 0.16129... TRUE yes
12 overcast 0.380952380... 0.80645... TRUE yes
13 overcast 0.809523809... 0.32258... FALSE yes
14 rainy 0.333333333... 0.83870... TRUE no
Add instance
Undo
OK
Cancel


Viewer
Relation: weather-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-unset-class-temporarily-weka.filters.unsupervised.attribute
No. 1: outlook Nominal 2: temperature Numeric 3: humidity Numeric 4: windy Nominal 5: play Nominal
1 sunny 2.0 1.29032... FALSE no
2 sunny 1.523809523... 1.61290... TRUE no
3 overcast 1.809523809... 1.35483... FALSE yes
4 rainy 0.937728937... 2.0 FALSE yes
5 rainy 0.380952380... 0.96774... FALSE yes
6 rainy 0.095238095... 0.32258... TRUE no
7 overcast 0.0 0.0 TRUE yes
8 sunny 0.761904761... 1.93548... FALSE no
9 sunny 0.476190476... 1.13151... FALSE yes
10 rainy 1.047619047... 0.96774... FALSE yes
11 sunny 1.047619047... 0.32258... TRUE yes
12 overcast 0.761904761... 1.61290... TRUE yes
13 overcast 1.619047619... 0.64516... FALSE yes
14 rainy 0.666666666... 1.67741... TRUE no
Add instance
Undo
OK
Cancel