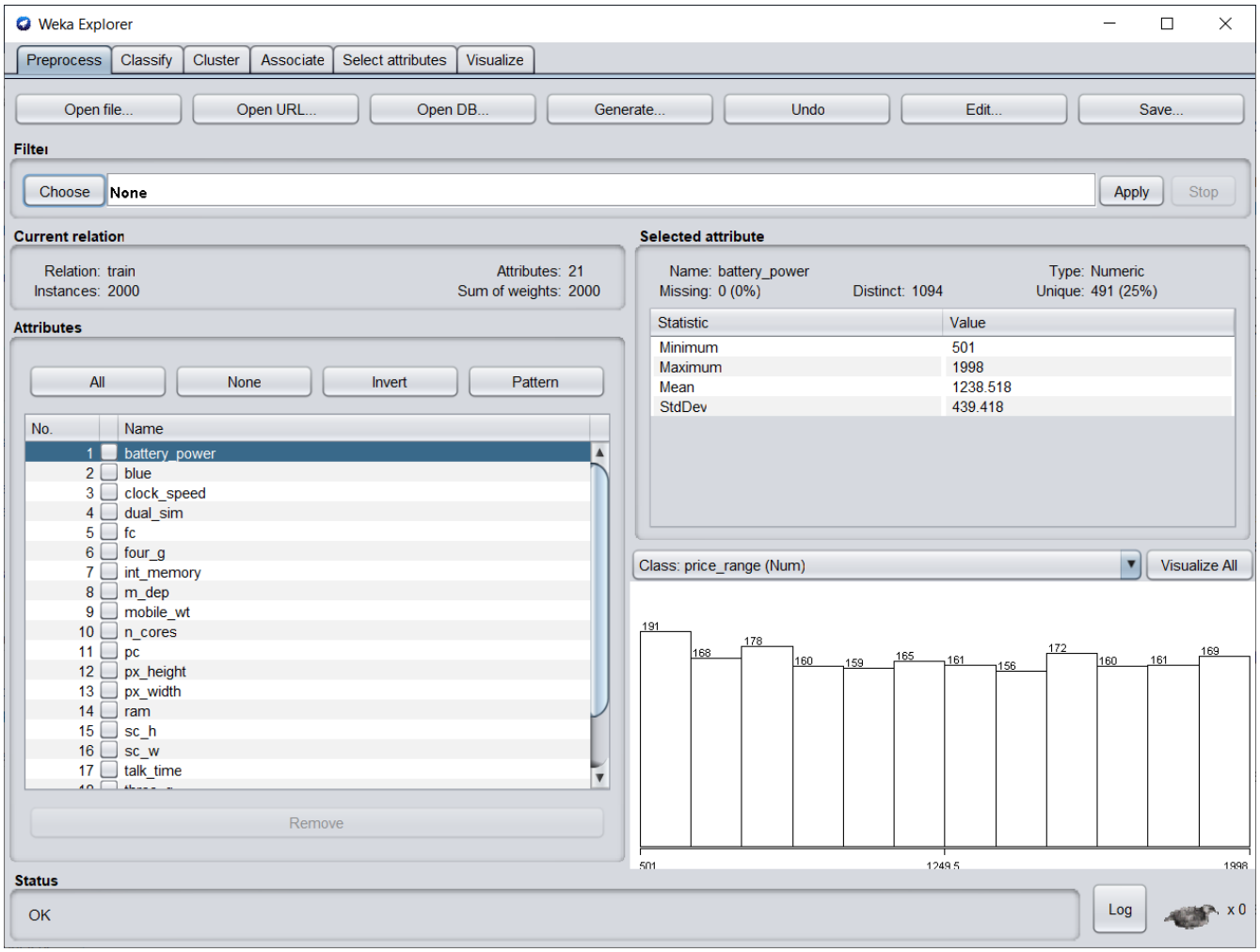
HUASCAR GABRIEL CONDORI TICONA

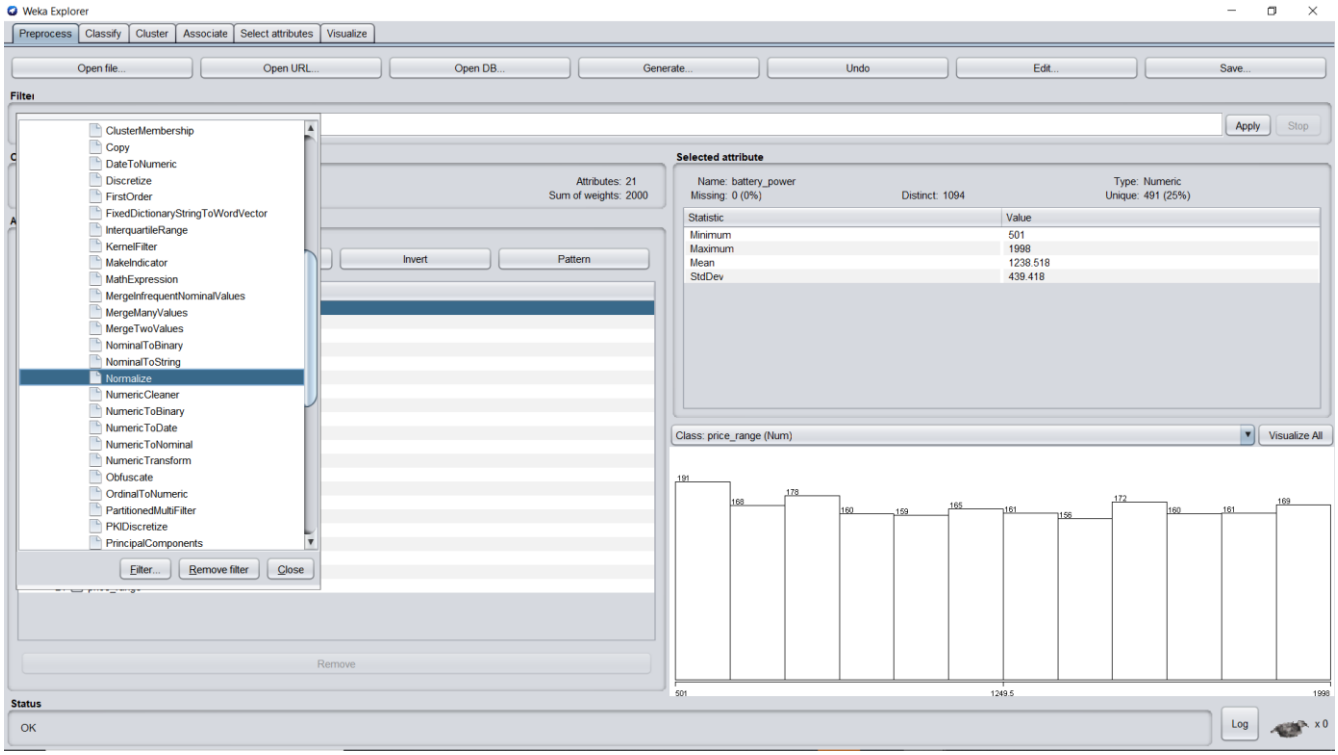
PREGUNTA 2

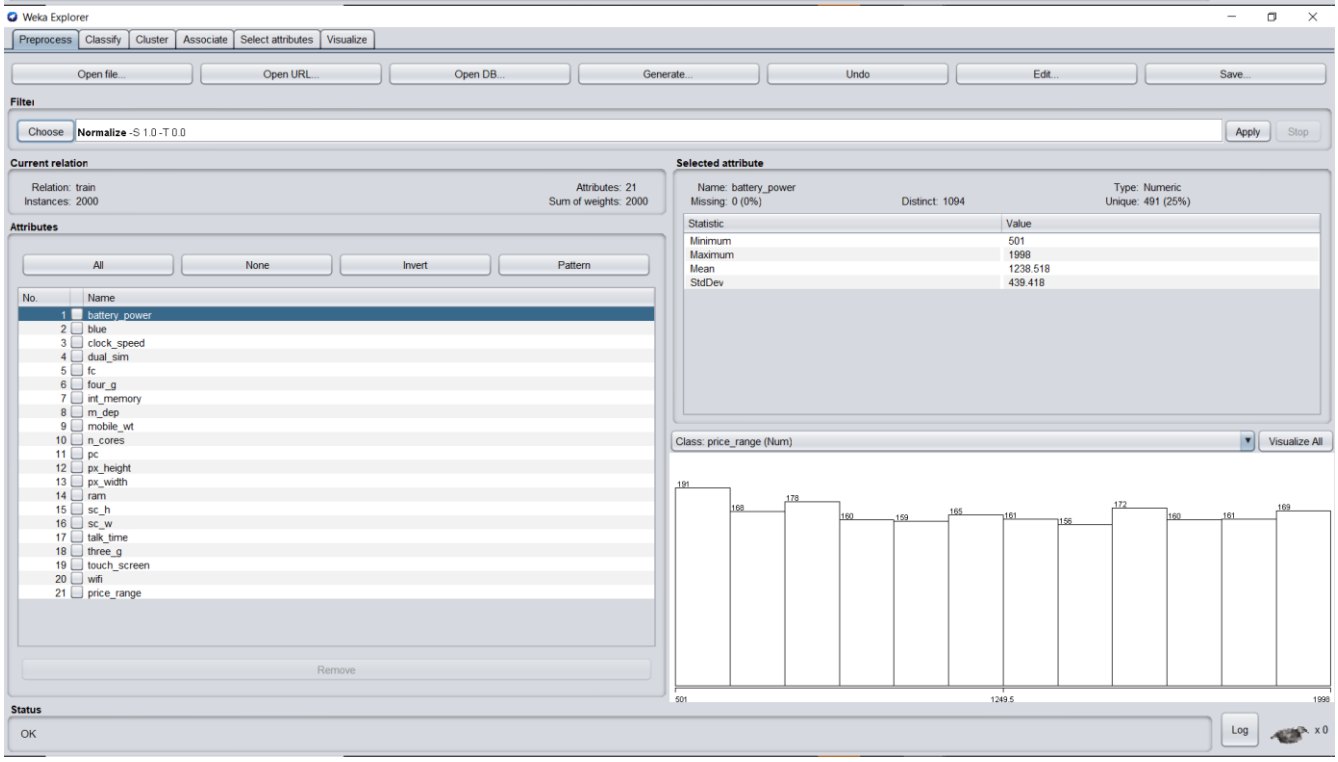
De un dataset de su tarea anterior en WEKA, realice tres algoritmos de pre procesamiento.

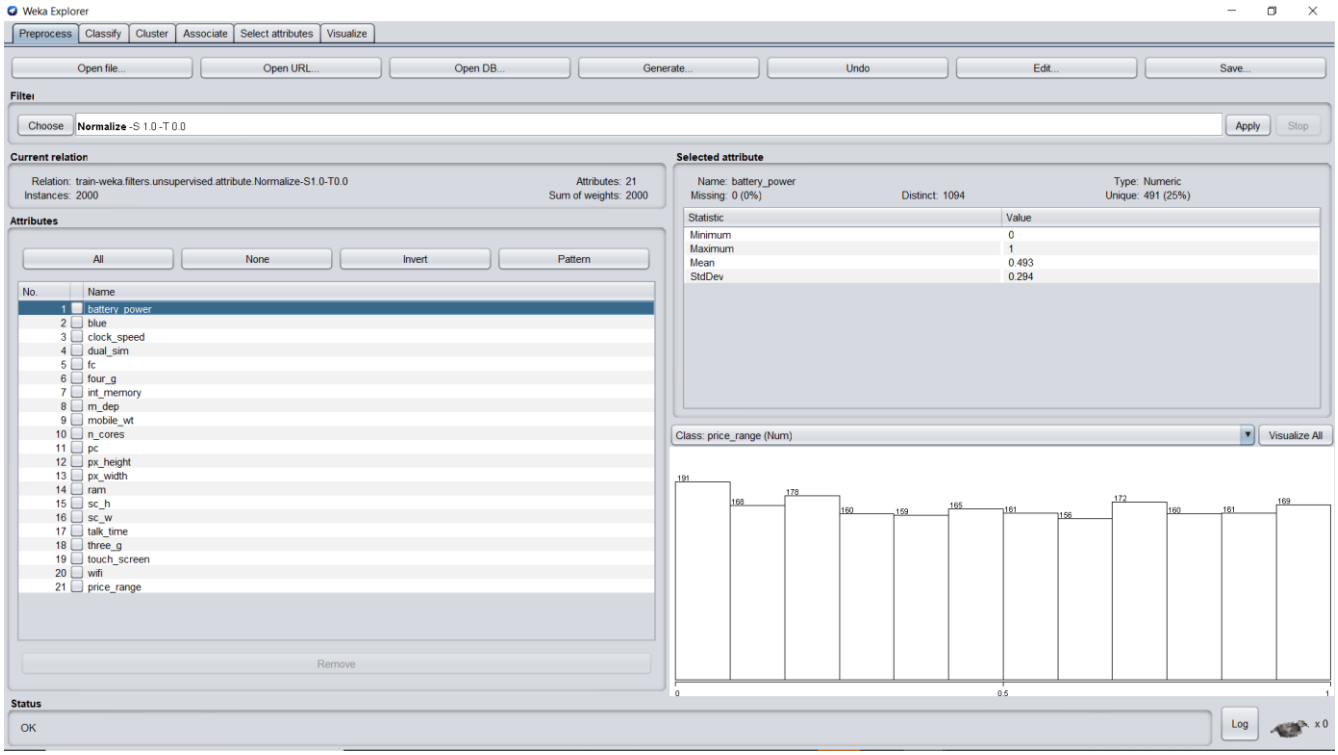
NORMALIZACION

Usa una fórmula o un algoritmo para transformar las variables medidas en diferentes escalas en una escala común para que puedan ser comparables (manzanas con manzanas) o analizadas en un modelo estadístico elegido. Un típico ejemplo es calcular el logaritmo de las variables para hacer una distribución sesgada normal (por ejemplo, desplegado en un gráfico como una curva normal). En weka realizamos la normalización llevando a una escala de 0 a 1 para todos los atributos.




Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Apply
Stop
ClusterMembership
Copy
DateToNumeric
Discretize
FirstOrder
FixedDictionaryStringToWordVector
InterquartileRange
KernelFilter
MakeIndicator
MathExpression
MergeInfrequentNominalValues
MergeManyValues
MergeTwoValues
NominalToBinary
NominalToString
Normalize
NumericCleaner
NumericToBinary
NumericToDate
NumericToNominal
NumericTransform
Obfuscate
OrdinalToNumeric
PartitionedMultiFilter
PKIDiscretize
PrincipalComponents
Filter...
Remove filter
Close
Attributes: 21
Sum of weights: 2000
Invert
Pattern
Remove
Selected attribute
Name: battery_power
Missing: 0 (0%)
Distinct: 1094
Type: Numeric
Unique: 491 (25%)
Statistic
Value
Minimum
501
Maximum
1998
Mean
1238.518
StdDev
439.418
Class: price_range (Num)
Visualize All
191
168
178
160
159
165
161
156
172
160
161
169
501
1249.5
1998
Status
OK
Log
x 0

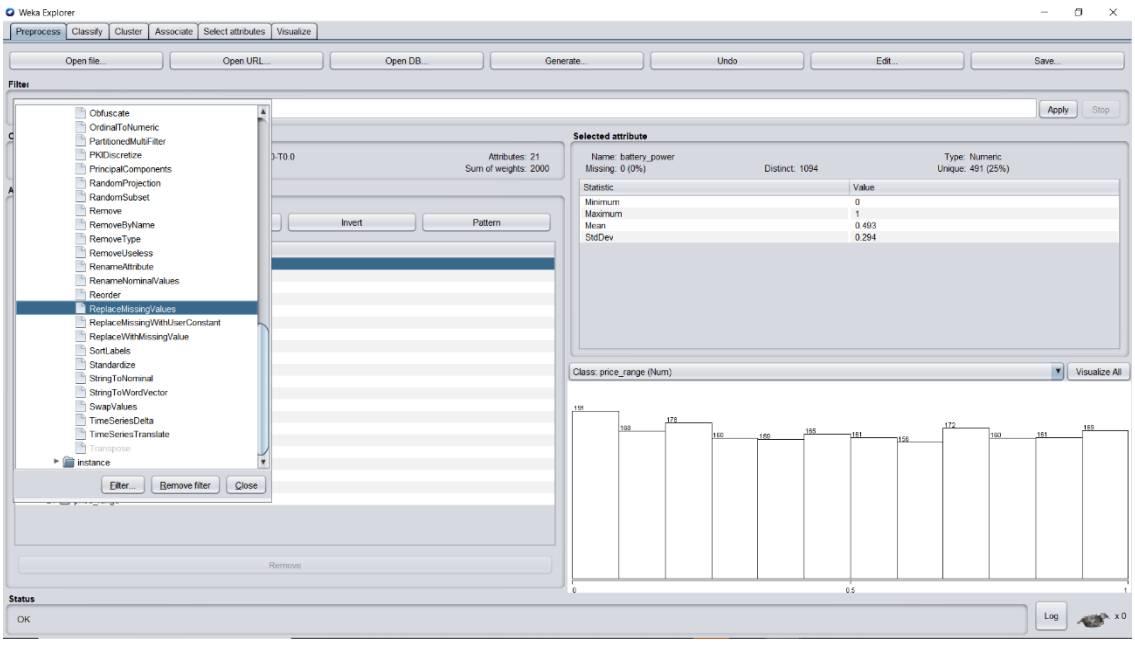

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
Normalize -S 1.0 -T 0.0
Apply
Stop
Current relation
Relation: train
Instances: 2000
Attributes: 21
Sum of weights: 2000
Attributes
All
None
Invert
Pattern
No.
Name
1 battery_power
2 blue
3 clock_speed
4 dual_sim
5 fc
6 four_g
7 int_memory
8 m_dep
9 mobile_wt
10 n_cores
11 pc
12 px_height
13 px_width
14 ram
15 sc_h
16 sc_w
17 talk_time
18 three_g
19 touch_screen
20 wifi
21 price_range
Remove
Selected attribute
Name: battery_power
Missing: 0 (0%)
Distinct: 1094
Type: Numeric
Unique: 491 (25%)
Statistic
Value
Minimum
501
Maximum
1998
Mean
1238.518
StdDev
439.418
Class: price_range (Num)
Visualize All
191
168
178
160
159
165
161
156
172
160
161
169
501
1249.5
1998
Status
OK
Log
x 0

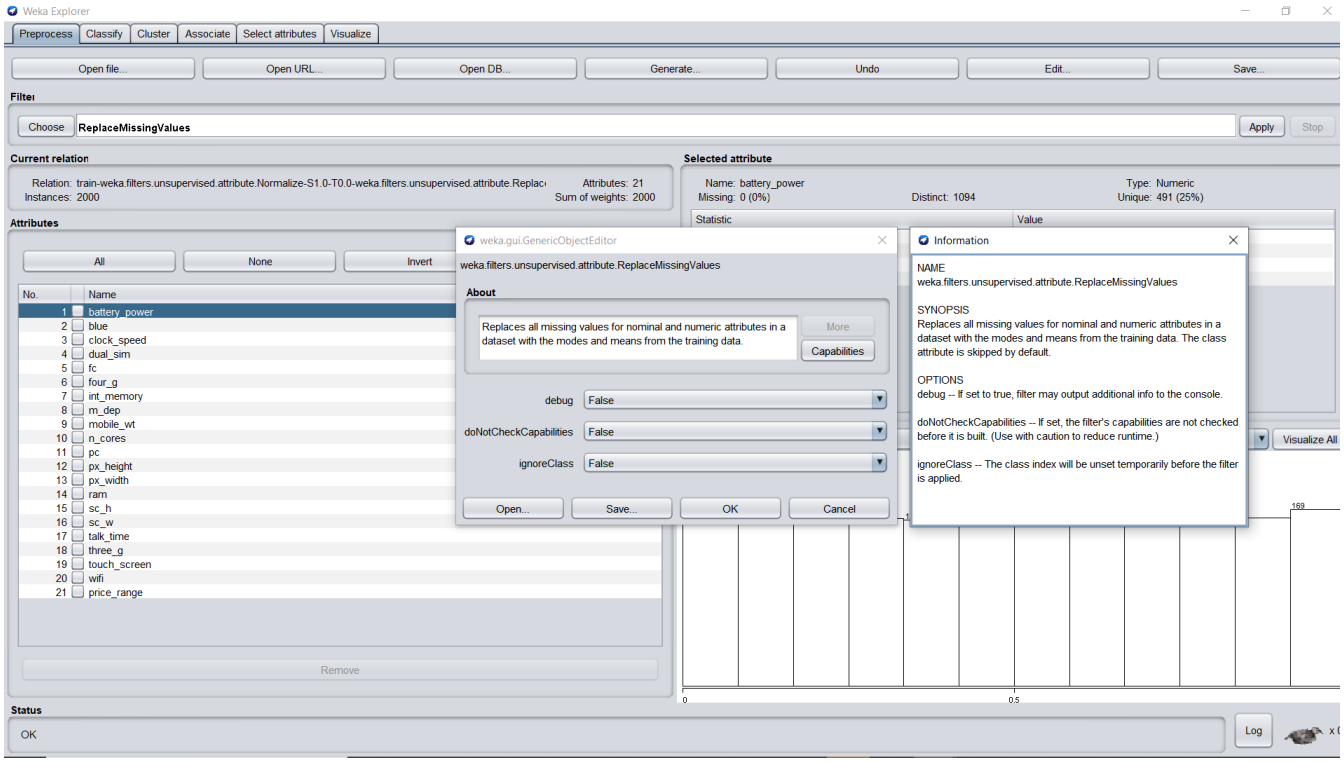


IMPUTACION (REEMPLAZAR DATOS FALTANTES)

La **imputación** es la sustitución de valores no informados en una observación por otros.

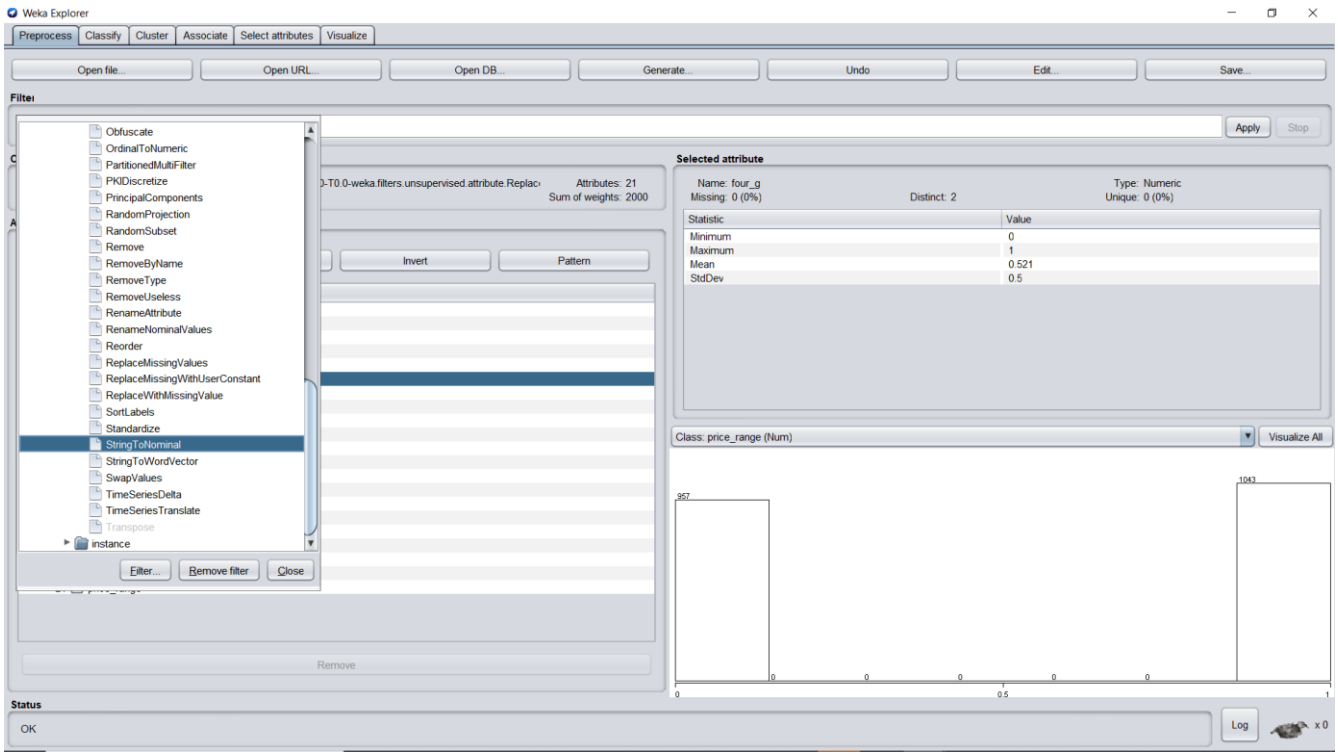
A veces es un paso necesario para poder tratar los datos con determinadas técnicas estadísticas de análisis. Idealmente, este análisis debería tener en cuenta el hecho de que algunos de los datos no son observados, sino que han sido imputados. Para este proceso podemos reemplazar los datos faltantes por la media, moda u otro estadístico.

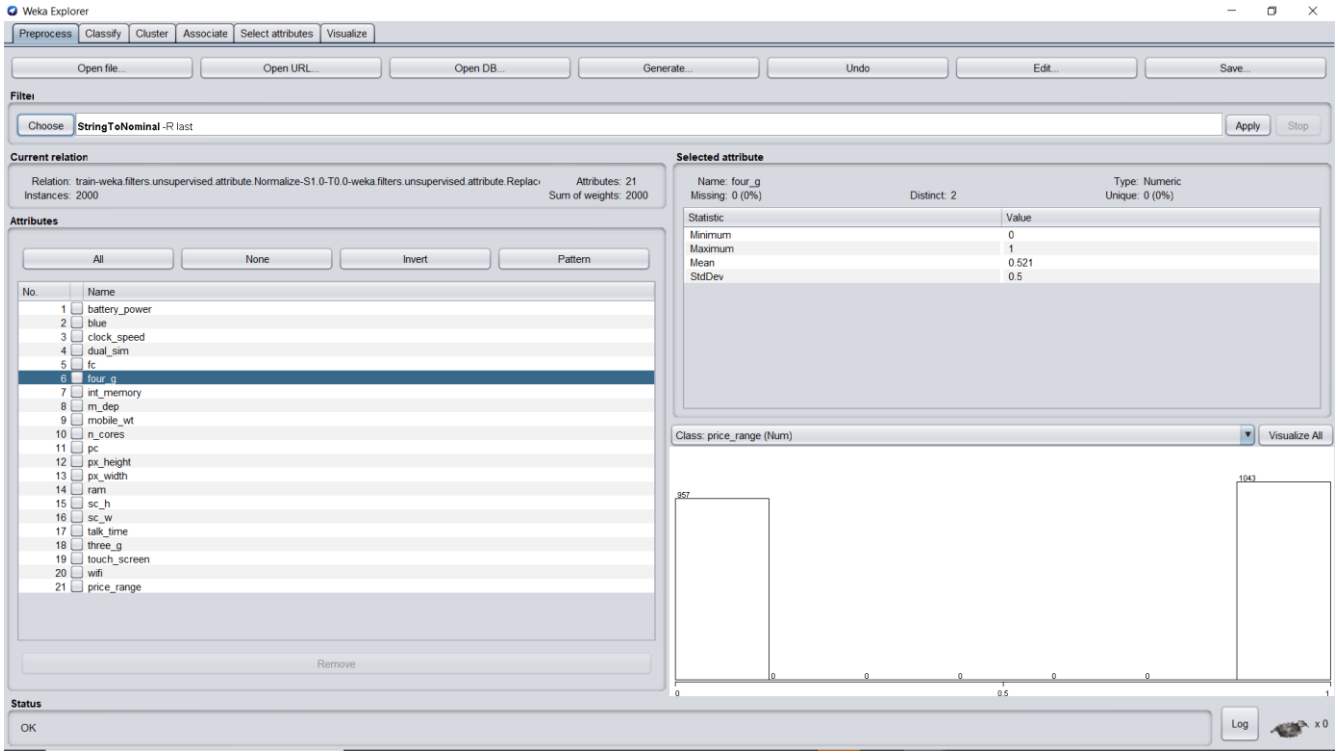


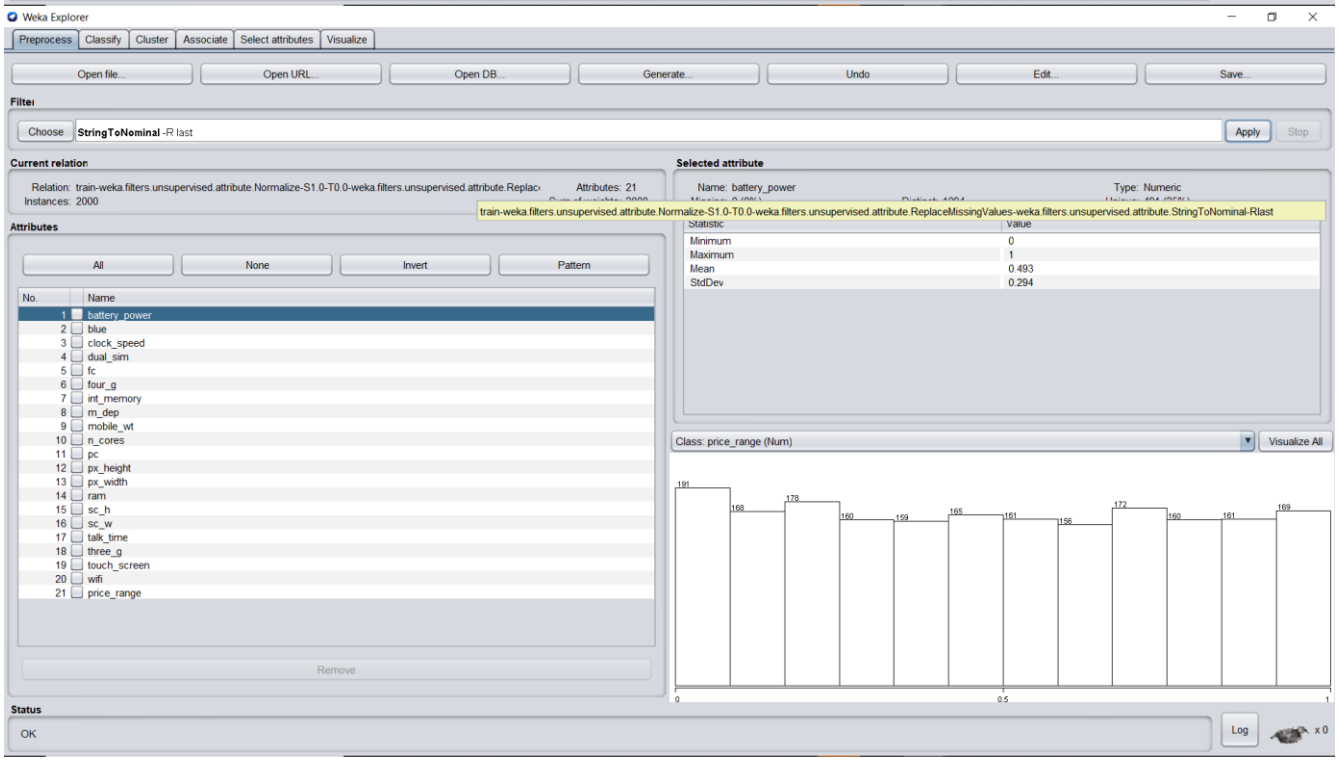


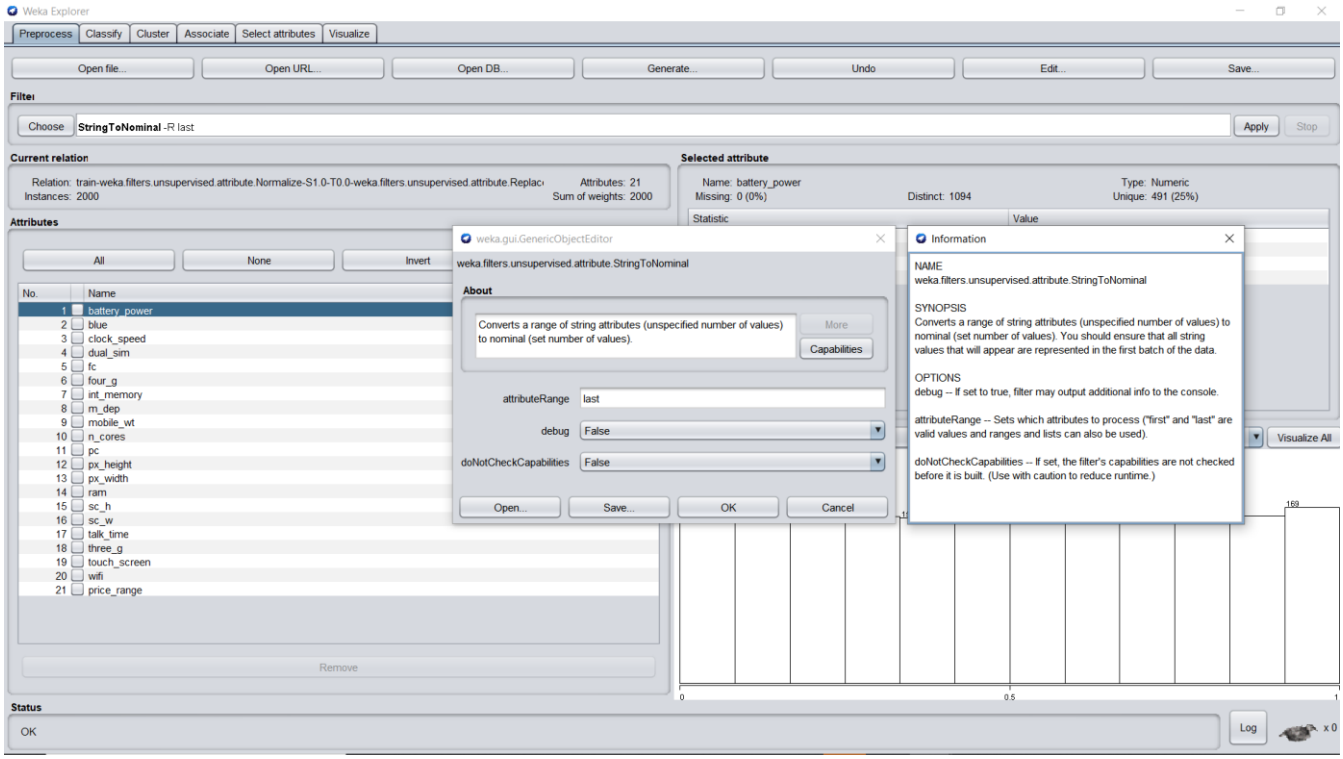
CONVERTIR STRINGS A NUMEROS (String to Nominal)

Convierte datos nominales a numeros.




Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
StringToNominal -R last
Apply
Stop
Current relation
Relation: train-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.filters.unsupervised.attribute.Replac
Attributes: 21
Instances: 2000
Sum of weights: 2000
Attributes
All
None
Invert
Pattern
No.
Name
1 battery_power
2 blue
3 clock_speed
4 dual_sim
5 fc
6 four_g
7 int_memory
8 m_dep
9 mobile_wt
10 n_cores
11 pc
12 px_height
13 px_width
14 ram
15 sc_h
16 sc_w
17 talk_time
18 three_g
19 touch_screen
20 wifi
21 price_range
Remove
Selected attribute
Name: four_g
Type: Numeric
Missing: 0 (0%)
Distinct: 2
Unique: 0 (0%)
Statistic
Value
Minimum 0
Maximum 1
Mean 0.521
StdDev 0.5
Class: price_range (Num)
Visualize All
957
1043
0
0.5
1
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
StringToNominal -R last
Apply
Stop
Current relation
Relation: train-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.filters.unsupervised.attribute.Replac
Attributes: 21
Instances: 2000
train-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.unsupervised.attribute.StringToNominal-Rlast
Attributes
All
None
Invert
Pattern
No.
Name
1 battery_power
2 blue
3 clock_speed
4 dual_sim
5 fc
6 four_g
7 int_memory
8 m_dep
9 mobile_wt
10 n_cores
11 pc
12 px_height
13 px_width
14 ram
15 sc_h
16 sc_w
17 talk_time
18 three_g
19 touch_screen
20 wifi
21 price_range
Remove
Selected attribute
Name: battery_power
Type: Numeric
Statistic
Value
Minimum 0
Maximum 1
Mean 0.493
StdDev 0.294
Class: price_range (Num)
Visualize All
191
168
178
160
159
165
161
156
172
160
161
169
0
0.5
1
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
StringToNominal -R last
Apply
Stop
Current relation
Relation: train-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.filters.unsupervised.attribute.Replac
Instances: 2000
Attributes: 21
Sum of weights: 2000
Selected attribute
Name: battery_power
Missing: 0 (0%)
Distinct: 1094
Type: Numeric
Unique: 491 (25%)
Statistic
Value
Attributes
All
None
Invert
No.
Name
1 battery_power
2 blue
3 clock_speed
4 dual_sim
5 fc
6 four_g
7 int_memory
8 m_dep
9 mobile_wt
10 n_cores
11 pc
12 px_height
13 px_width
14 ram
15 sc_h
16 sc_w
17 talk_time
18 three_g
19 touch_screen
20 wifi
21 price_range
Remove
Visualize All
169
0
0.5
1
weka.gui.GenericObjectEditor
weka.filters.unsupervised.attribute.StringToNominal
About
Converts a range of string attributes (unspecified number of values) to nominal (set number of values).
More
Capabilities
attributeRange
last
debug
False
doNotCheckCapabilities
False
Open...
Save...
OK
Cancel
Information
NAME
weka.filters.unsupervised.attribute.StringToNominal
SYNOPSIS
Converts a range of string attributes (unspecified number of values) to nominal (set number of values). You should ensure that all string values that will appear are represented in the first batch of the data.
OPTIONS
debug -- If set to true, filter may output additional info to the console.
attributeRange -- Sets which attributes to process ("first" and "last" are valid values and ranges and lists can also be used).
doNotCheckCapabilities -- If set, the filter's capabilities are not checked before it is built. (Use with caution to reduce runtime.)
Status
OK
Log
x 0