



**VIT<sup>®</sup>**  
**Vellore Institute of Technology**  
(Deemed to be University under section 3 of UGC Act, 1956)

## DATA WAREHOUSING AND DATA MINING

NAME : MOTHISHWARAN C.  
REG No : 19MID0017  
COURSE CODE : CSI3010  
SLOT : L39+L40  
FACULTY : Dr. CHELLATHAMILAN T.

### Data Pre-processing

```
weather.arff - Sublime Text (UNREGISTERED)
File Edit Selection Find View Goto Tools Project Preferences Help

weather.arff x
1 @relation weather
2 @attribute outlook {sunny,rainy,overcast}
3 @attribute temperature numeric
4 @attribute humidity numeric
5 @attribute windy {true,false}
6 @attribute play {yes,no}
7 @data
8 sunny,85.0,85.0,false,no
9 overcast,80.0,90.0,true,no
10 sunny,83.0,86.0,false,yes
11 rainy,70.0,86.0,false,yes
12 rainy,68.0,80.0,false,yes
13 rainy,65.0,70.0,true,no
14 overcast,64.0,65.0,false,yes
15 sunny,72.0,95.0,true,no
16 sunny,69.0,70.0,false,yes
17 rainy,75.0,80.0,false,yes
18
```

**Weka Explorer**

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **None** Apply Stop

**Current relation**  
 Relation: weather  
 Instances: 10  
 Attributes: 5  
 Sum of weights: 10

**Attributes**  
 All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> outlook
2	<input type="checkbox"/> temperature
3	<input type="checkbox"/> humidity
4	<input type="checkbox"/> windy
5	<input type="checkbox"/> play

Remove

**Selected attribute**  
 Name: outlook  
 Missing: 0 (0%)  
 Distinct: 3  
 Type: Nominal  
 Unique: 0 (0%)

No.	Label	Count	Weight
1	sunny	4	4.0
2	rainy	4	4.0
3	overcast	2	2.0

Class: play (Nom) Visualize All

**Status**  
 OK Log x 0

**Viewer**

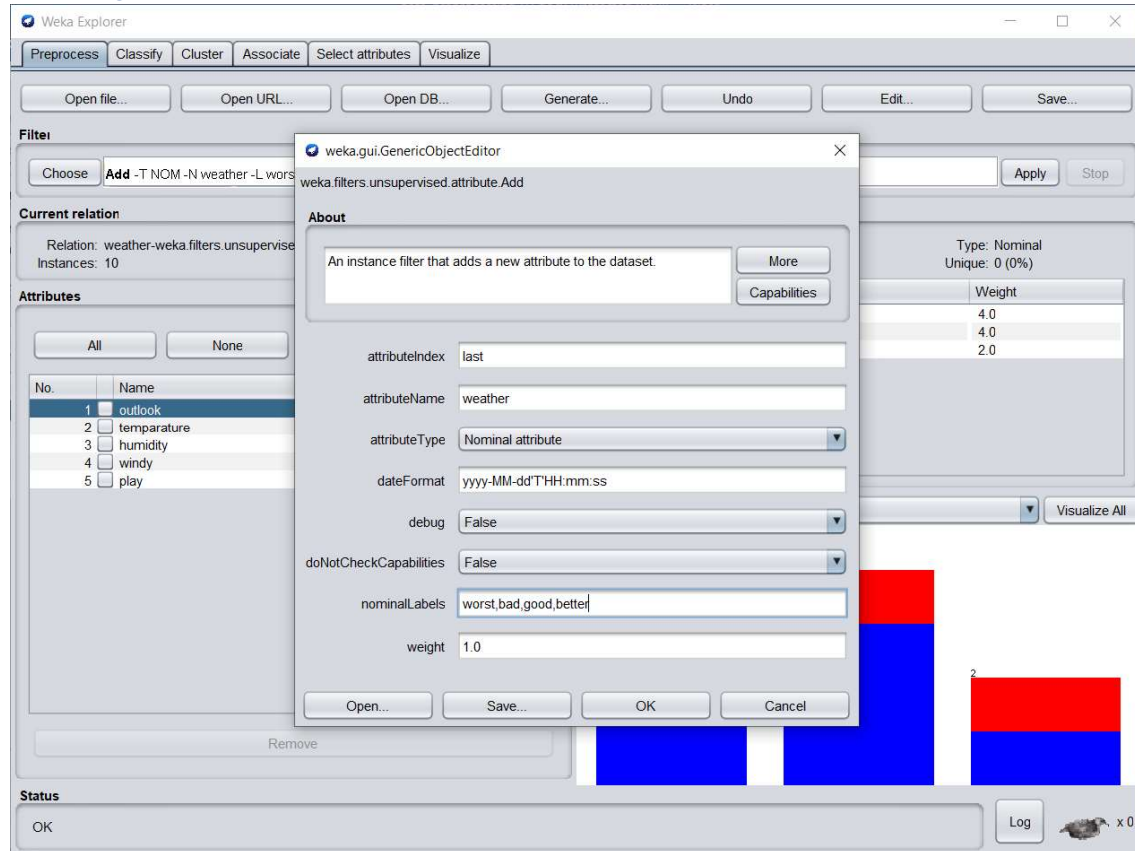
Relation: weather

No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: play Nominal
1	sunny	85.0	85.0	false	no
2	overcast	80.0	90.0	true	no
3	sunny	83.0	86.0	false	yes
4	rainy	70.0	86.0	false	yes
5	rainy	68.0	80.0	false	yes
6	rainy	65.0	70.0	true	no
7	overcast	64.0	65.0	false	yes
8	sunny	72.0	95.0	true	no
9	sunny	69.0	70.0	false	yes
10	rainy	75.0	80.0	false	yes

Add instance Undo OK Cancel

## 1) Add

### Adding new attribute(weather):



The screenshot shows the Weka Viewer interface displaying the dataset 'weather-weka.filters.unsupervised.attribute.Add-Nunamed-Clast-VW1.0-weka.filters.unsupervised.attribute.Remove-R6-weka.filters.unsupervised.attribute.Add-TNOM-Nweather-Lwo'. The dataset has 10 instances and 6 attributes: outlook, temperature, humidity, windy, play, and weather. The 'weather' attribute is highlighted in the attribute list on the right.

No.	1: outlook	2: temperature	3: humidity	4: windy	5: play	6: weather
1	sunny	85.0	85.0	false	no	best
2	overcast	80.0	90.0	true	no	best
3	sunny	83.0	86.0	false	yes	best
4	rainy	70.0	86.0	false	yes	best
5	rainy	68.0	80.0	false	yes	best
6	rainy	65.0	70.0	true	no	best
7	overcast	64.0	85.0	false	yes	best
8	sunny	72.0	95.0	true	no	best
9	sunny	69.0	70.0	false	yes	best
10	rainy	75.0	80.0	false	yes	best

Viewer

Relation: weather-weka.filters.unsupervised.attribute.Add-Nunamed-Clast-W1.0-weka.filters.unsupervised.attribute.Remove-R6-weka.filters.unsupervised.attribute.Add-TNOM-Nweather-Lwo

No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: play Nominal	6: weather Nominal
1	sunny	85.0	85.0	false	no	worst
2	overcast	80.0	90.0	true	no	bad
3	sunny	83.0	86.0	false	yes	better
4	rainy	70.0	86.0	false	yes	good
5	rainy	68.0	80.0	false	yes	good
6	rainy	65.0	70.0	true	no	bad
7	overcast	64.0	65.0	false	yes	good
8	sunny	72.0	95.0	true	no	better
9	sunny	69.0	70.0	false	yes	bad
10	rainy	75.0	80.0	false	yes	better

Add instance Undo OK Cancel

## Adding new instance:

Viewer

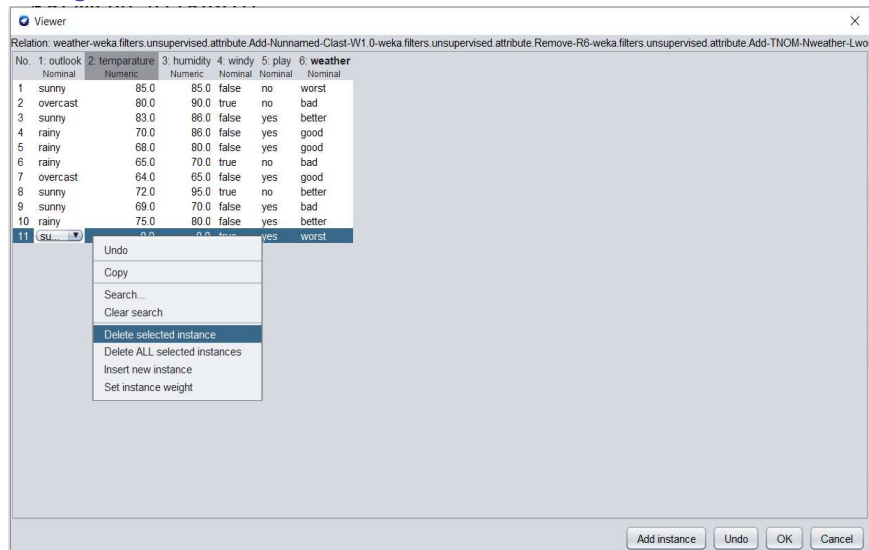
Relation: weather-weka.filters.unsupervised.attribute.Add-Nunamed-Clast-W1.0-weka.filters.unsupervised.attribute.Remove-R6-weka.filters.unsupervised.attribute.Add-TNOM-Nweather-Lwo

No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: play Nominal	6: weather Nominal
1	sunny	85.0	85.0	false	no	worst
2	overcast	80.0	90.0	true	no	bad
3	sunny	83.0	86.0	false	yes	better
4	rainy	70.0	86.0	false	yes	good
5	rainy	68.0	80.0	false	yes	good
6	rainy	65.0	70.0	true	no	bad
7	overcast	64.0	65.0	false	yes	good
8	sunny	72.0	95.0	true	no	better
9	sunny	69.0	70.0	false	yes	bad
10	rainy	75.0	80.0	false	yes	better
11	sunny	0.0	0.0	true	yes	worst

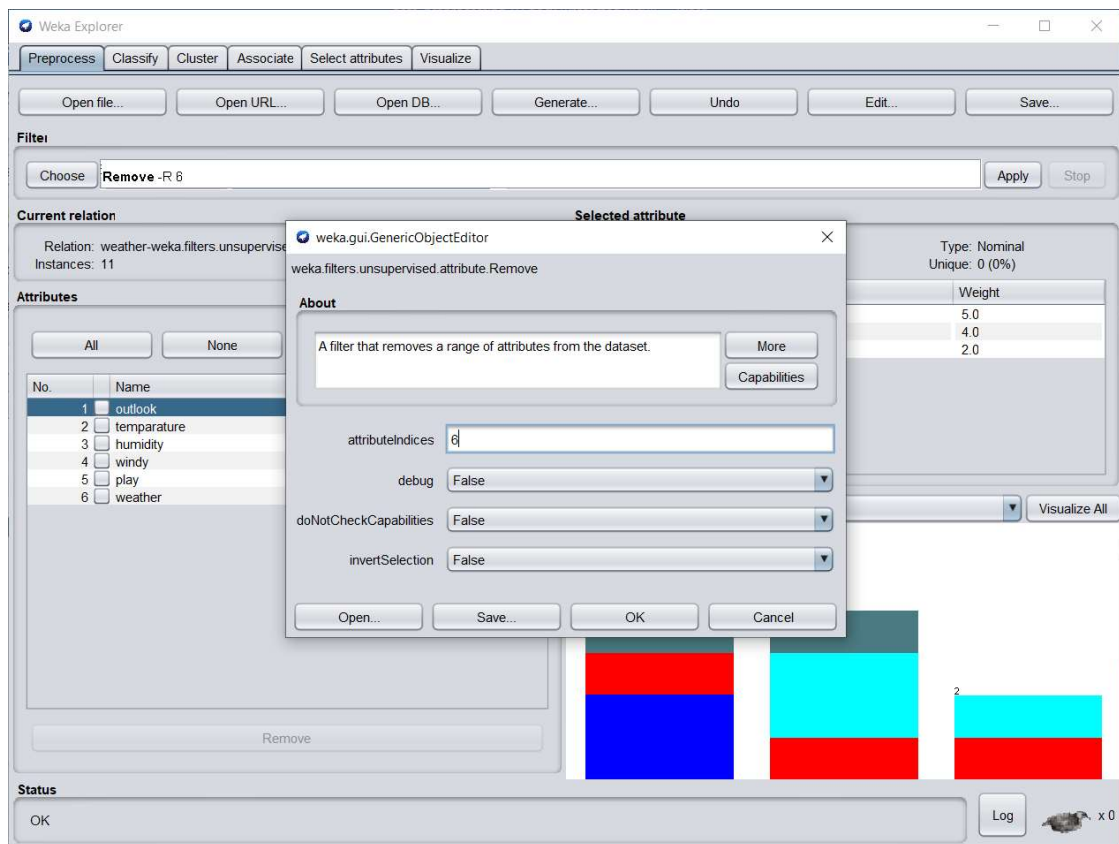
Add instance Undo OK Cancel

## 2) Remove

### Removing instances:



### Removing attributes:



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose Remove -R 6 Apply Stop

Current relation: Relation: weather-weka.filters.unsupervised.attribute.Add-Nunn... Attributes: 5 Instances: 11 Sum of weights: 11

Selected attribute: Name: outlook Missing: 0 (0%) Distinct: 3 Type: Nominal Unique: 0 (0%)

No.	Label	Count	Weight
1	sunny	5	5.0
2	rainy	4	4.0
3	overcast	2	2.0

Class: play (Nom) Visualize All

Attributes: All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> outlook
2	<input type="checkbox"/> temperature
3	<input type="checkbox"/> humidity
4	<input type="checkbox"/> windy
5	<input type="checkbox"/> play

Remove

Status: OK Log x 0

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Viewer

Relation: weather-weka.filters.unsupervised.attribute.Add-Nunnamed-Clast-W1.0-weka.filters.unsupervised.attribute.Remove-R6-weka.filters.unsupervised.attribute.Add-TNOM-Nweather-Lwo

No.	1: outlook	2: temperature	3: humidity	4: windy	5: play
	Nominal	Numeric	Numeric	Nominal	Nominal
1	sunny	85.0	85.0	false	no
2	overcast	80.0	90.0	true	no
3	sunny	83.0	86.0	false	yes
4	rainy	70.0	86.0	false	yes
5	rainy	68.0	80.0	false	yes
6	rainy	65.0	70.0	true	no
7	overcast	64.0	65.0	false	yes
8	sunny	72.0	95.0	true	no
9	sunny	69.0	70.0	false	yes
10	rainy	75.0	80.0	false	yes
11	sunny	0.0	0.0	true	yes

Add instance Undo OK Cancel

Status: OK Log x 0

### 3) Normalization:

The screenshot shows the Weka Explorer interface with the 'Preprocess' tab selected. A 'Filter' dialog is open, showing the 'Normalize' filter selected. The 'Current relation' is 'weather-weka.filters.unsupervised.attribute.Normalize' with 11 instances. The 'Attributes' list shows 'outlook', 'temperature', 'humidity', 'windy', and 'play'. The 'Normalize' dialog is open, showing the 'About' tab with the description: 'Normalizes all numeric values in the given dataset (apart from the class attribute, if set)'. The dialog also has fields for 'debug' (False), 'doNotCheckCapabilities' (False), 'ignoreClass' (False), 'scale' (1.0), and 'translation' (0.0). The 'Visualize All' button is visible. The status bar at the bottom shows 'OK' and a 'Log' button.

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: weather-weka.filters.unsupervised.attribute.Normalize  
Instances: 11

Attributes:

No.	Name
1	<input type="checkbox"/> outlook
2	<input checked="" type="checkbox"/> temperature
3	<input checked="" type="checkbox"/> humidity
4	<input type="checkbox"/> windy
5	<input type="checkbox"/> play

Remove

Status: OK Log x 0

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.Normalize

About

Normalizes all numeric values in the given dataset (apart from the class attribute, if set).

debug: False

doNotCheckCapabilities: False

ignoreClass: False

scale: 1.0

translation: 0.0

Open... Save... OK Cancel

Type: Numeric  
Unique: 5 (45%)

Value
0
95
73.364
26.005

Visualize All

10

0 47.5 95

**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: weather-weka.filters.unsupervised.attribute.Add-Nunn... Attributes: 5  
 Instances: 11 Sum of weights: 11

**Attributes**  
 All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> outlook
2	<input checked="" type="checkbox"/> temperature
3	<input checked="" type="checkbox"/> humidity
4	<input checked="" type="checkbox"/> windy
5	<input checked="" type="checkbox"/> play

Remove

**Selected attribute**  
 Name: outlook  
 Missing: 0 (0%) Distinct: 3 Type: Nominal  
 Unique: 0 (0%)

No.	Label	Count	Weight
1	sunny	5	5.0
2	rainy	4	4.0
3	overcast	2	2.0

Class: play (Nom) Visualize All

**Status**  
 OK Log x 0

**Weka Explorer**

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

**Viewer**

Relation: weather-weka.filters.unsupervised.attribute.Add-Nunnamed-Class-W1.0-weka.filters.unsupervised.attribute.Remove-R6-weka.filters.unsupervised.attribute.Add-TNOM-Nweather-Lwo

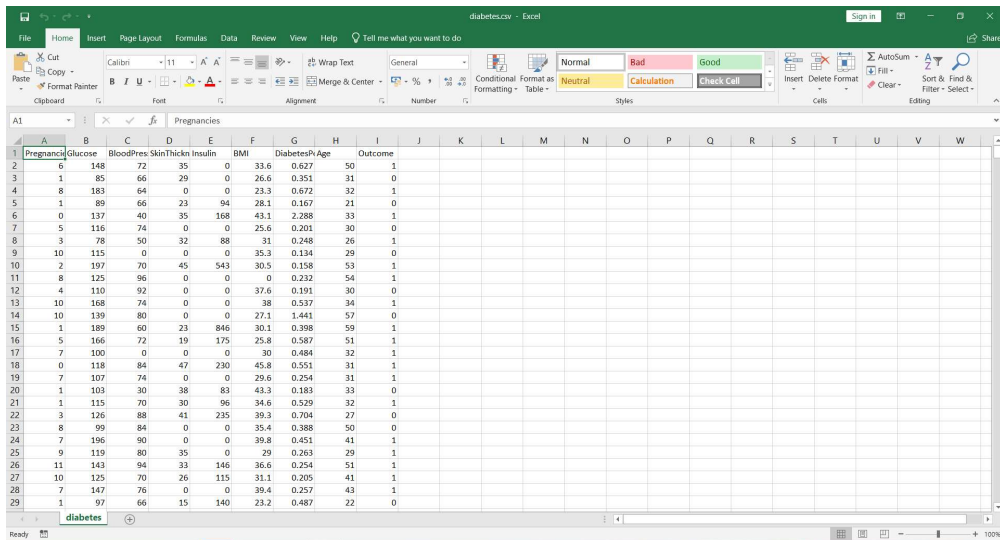
No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: play Nominal
1	sunny	1.0	0.89473...	false	no
2	overcast	0.941176470...	0.94736...	true	no
3	sunny	0.976470588...	0.90526...	false	yes
4	rainy	0.823529411...	0.90526...	false	yes
5	rainy	0.8	0.84210...	false	yes
6	rainy	0.764705882...	0.73684...	true	no
7	overcast	0.752941176...	0.68421...	false	yes
8	sunny	0.847058823...	1.0	true	no
9	sunny	0.811764705...	0.73684...	false	yes
10	rainy	0.882352941...	0.84210...	false	yes
11	sunny	0.0	0.0	true	yes

Add instance Undo OK Cancel

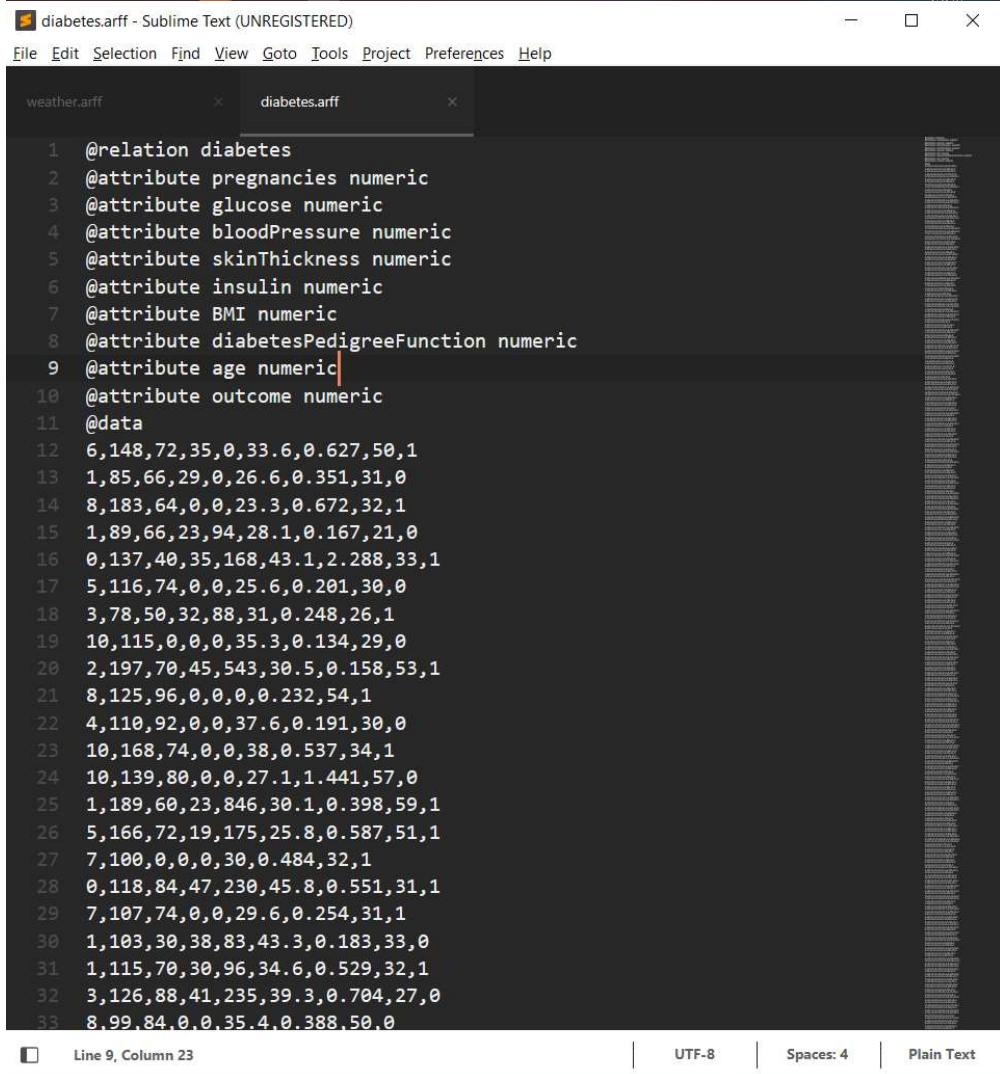
**Status**  
 OK Log x 0



## 4) Handling of missing values:



Pregnancies	Glucose	BloodPres	SkinThickn	Insulin	BMI	DiabetesPv	Age	Outcome	
6	148	72	35	0	33.6	0.627	50	1	
1	85	66	29	0	26.6	0.351	31	0	
8	183	64	0	0	23.3	0.672	32	1	
1	89	66	23	94	28.1	0.167	21	0	
0	137	40	35	168	43.1	2.288	33	1	
5	116	74	0	0	25.6	0.201	30	0	
3	78	50	32	88	31	0.248	26	1	
10	115	0	0	0	35.3	0.134	29	0	
2	197	70	45	543	30.5	0.158	53	1	
8	125	96	0	0	0	0.232	54	1	
4	110	92	0	0	37.6	0.191	30	0	
10	168	74	0	0	38	0.537	34	1	
10	139	80	0	0	27.1	1.441	57	0	
1	189	60	23	846	30.1	0.398	59	1	
5	166	72	19	175	25.8	0.587	51	1	
7	100	0	0	0	30	0.484	32	1	
0	118	84	47	230	45.8	0.551	31	1	
7	107	74	0	0	29.6	0.254	31	1	
1	103	30	38	83	43.3	0.183	33	0	
1	115	70	30	96	34.6	0.529	32	1	
3	126	88	41	235	39.3	0.704	27	0	
8	99	84	0	0	35.4	0.388	50	0	
7	196	90	0	0	38.8	0.451	43	1	
9	119	80	35	0	29	0.263	29	1	
11	143	94	33	146	36.6	0.254	51	1	
10	125	70	26	115	31.1	0.205	41	1	
7	147	76	0	0	39.4	0.237	43	1	
1	97	66	15	140	23.2	0.487	22	0	



```
@relation diabetes
@attribute pregnancies numeric
@attribute glucose numeric
@attribute bloodPressure numeric
@attribute skinThickness numeric
@attribute insulin numeric
@attribute BMI numeric
@attribute diabetesPedigreeFunction numeric
@attribute age numeric
@attribute outcome numeric
@data
6,148,72,35,0,33.6,0.627,50,1
1,85,66,29,0,26.6,0.351,31,0
8,183,64,0,0,23.3,0.672,32,1
1,89,66,23,94,28.1,0.167,21,0
0,137,40,35,168,43.1,2.288,33,1
5,116,74,0,0,25.6,0.201,30,0
3,78,50,32,88,31,0.248,26,1
10,115,0,0,0,35.3,0.134,29,0
2,197,70,45,543,30.5,0.158,53,1
8,125,96,0,0,0,0.232,54,1
4,110,92,0,0,37.6,0.191,30,0
10,168,74,0,0,38,0.537,34,1
10,139,80,0,0,27.1,1.441,57,0
1,189,60,23,846,30.1,0.398,59,1
5,166,72,19,175,25.8,0.587,51,1
7,100,0,0,0,30,0.484,32,1
0,118,84,47,230,45.8,0.551,31,1
7,107,74,0,0,29.6,0.254,31,1
1,103,30,38,83,43.3,0.183,33,0
1,115,70,30,96,34.6,0.529,32,1
3,126,88,41,235,39.3,0.704,27,0
8,99,84,0,0,35.4,0.388,50,0
```

## - Mark Missing Values

**Weka Explorer**

Preprocess Classify Cluster Associate

Open file... Open URL...

Filter: Choose **NumericCleaner** -min 1.0E-9-max 1.7976931348623157E308-closeto 0.0-closeto-default 0.0 Apply Stop

Current relation: Relation: diabetes Instances: 768

Attributes: All None

No.	Name
1	pregnancies
2	glucose
3	bloodPressure
4	skinThickness
5	insulin
6	BMI
7	diabetesPedigreeFunction
8	age
9	outcome

Status: OK

**weka.gui.GenericObjectEditor**

weka.filters.unsupervised.attribute.NumericCleaner

About: A filter that 'cleanses' the numeric data from values that are too small, too big or very close to a certain value, and sets these values to a pre-defined default.

attributesIndices: 6

closeTo: 0.0

closeToDefault: 0.0

closeToTolerance: 1.0E-6

debug: False

decimals: -1

doNotCheckCapabilities: False

includeClass: False

invertSelection: False

maxDefault: 1.7976931348623157E308

maxThreshold: 1.7976931348623157E308

minDefault: NaN

minThreshold: 0.1E-8

Open... Save... OK Cancel

Type: Numeric Unique: 2 (0%)

Value
0
17
3.845
3.37

Visualize All

**Weka Explorer**

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **NumericCleaner** -min 1.0E-9-max 1.7976931348623157E308-closeto 0.0-closeto-default 0.0 Apply Stop

Current relation: Relation: diabetes-weka.filters.unsupervised.attribute.Numeric... Instances: 768 Attributes: 9 Sum of weights: 768

Left-click to edit properties for this object, right-click/Alt+Shift+left-click for menu

Name: BMI Type: Numeric Missing: 11 (1%) Distinct: 247 Unique: 76 (10%)

Statistic	Value
Minimum	18.2
Maximum	67.1
Mean	32.457
StdDev	6.925

Class: outcome (Num) Visualize All

Status: OK

maxThreshold: 1.7976931348623157E308

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Viewer

Relation: diabetes-weka.filters.unsupervised.attribute.NumericCleaner-min1.0E-9-min-default NaN-max1.7976931348623157E308-max-default1.7976931348623157E308-closeto0.0-closeto-default NaN

No.	1: pregnancies	2: glucose	3: bloodPressure	4: skinThickness	5: insulin	6: BMI	7: diabetesPedigreeFunction	8: age	9: outcome
43	7.0	106.0	92.0	18.0	0.0	22.7	0.235	48.0	0.0
44	9.0	171.0	110.0	24.0	240.0	45.4	0.721	54.0	1.0
45	7.0	159.0	64.0	0.0	0.0	27.4	0.294	40.0	0.0
46	0.0	180.0	66.0	39.0	0.0	42.0	1.893	25.0	1.0
47	1.0	146.0	56.0	0.0	0.0	29.7	0.564	29.0	0.0
48	2.0	71.0	70.0	27.0	0.0	28.0	0.586	22.0	0.0
49	7.0	103.0	66.0	32.0	0.0	39.1	0.344	31.0	1.0
50	7.0	105.0	0.0	0.0	0.0		0.305	24.0	0.0
51	1.0	103.0	80.0	11.0	82.0	19.4	0.491	22.0	0.0
52	1.0	101.0	50.0	15.0	36.0	missing	0.526	26.0	0.0
53	5.0	88.0	66.0	21.0	23.0		0.342	30.0	0.0
54	8.0	176.0	90.0	34.0	300.0	33.7	0.467	58.0	1.0
55	7.0	150.0	66.0	42.0	342.0	34.7	0.718	42.0	0.0
56	1.0	73.0	50.0	10.0	0.0	23.0	0.248	21.0	0.0
57	7.0	187.0	68.0	39.0	304.0	37.7	0.254	41.0	1.0
58	0.0	100.0	88.0	60.0	110.0	46.8	0.962	31.0	0.0
59	0.0	146.0	82.0	0.0	0.0	40.5	1.781	44.0	0.0
60	0.0	105.0	64.0	41.0	142.0	41.5	0.173	22.0	0.0
61	2.0	84.0	0.0	0.0	0.0		0.304	21.0	0.0
62	8.0	133.0	72.0	0.0	0.0	32.9	0.27	39.0	1.0
63	5.0	44.0	62.0	0.0	0.0	25.0	0.587	36.0	0.0
64	2.0	141.0	58.0	34.0	128.0	25.4	0.699	24.0	0.0
65	7.0	114.0	66.0	0.0	0.0	32.8	0.258	42.0	1.0
66	5.0	99.0	74.0	27.0	0.0	29.0	0.203	32.0	0.0
67	0.0	109.0	88.0	30.0	0.0	32.5	0.855	38.0	1.0
68	2.0	109.0	92.0	0.0	0.0	42.7	0.845	54.0	0.0
69	1.0	95.0	66.0	13.0	38.0	19.6	0.334	25.0	0.0
70	4.0	146.0	85.0	27.0	100.0	28.9	0.189	27.0	0.0
71	2.0	100.0	66.0	20.0	90.0	32.9	0.867	28.0	1.0
72	5.0	130.0	84.0	25.0	140.0	28.8	0.414	28.0	0.0

Add instance Undo OK Cancel

Status OK Log x 0

## - Remove instances with Missing Data

Weka Explorer

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 Apply Stop

Current relation  
Relation: diabetes-weka.filters.unsupervised.attribute.NumericCleaner-min1.0E-9-min-default NaN-max1.7976931348623157E308-max-default1.7976931348623157E308-closeto0.0-closeto-default NaN  
Instances: 768 Attributes: 9 Sum of weights: 768

Attributes  
All None Invert Pattern

No.	Name
1	<input type="checkbox"/> pregnancies
2	<input type="checkbox"/> glucose
3	<input type="checkbox"/> bloodPressure
4	<input type="checkbox"/> skinThickness
5	<input type="checkbox"/> insulin
6	<input checked="" type="checkbox"/> BMI
7	<input type="checkbox"/> diabetesPedigreeFunction
8	<input type="checkbox"/> age
9	<input type="checkbox"/> outcome

Remove

Selected attribute  
Name: BMI  
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: outcome (Num) Visualize All

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 **RemoveWithValues** -S 0.0 -C

**Current relation:**  
Relation: diabetes-weka.filters.unsupervised...  
Instances: 768

**Attributes:**

No.	Name
1	<input type="checkbox"/> pregnancies
2	<input type="checkbox"/> glucose
3	<input type="checkbox"/> bloodPressure
4	<input type="checkbox"/> skinThickness
5	<input type="checkbox"/> insulin
6	<input checked="" type="checkbox"/> BMI
7	<input type="checkbox"/> diabetesPedigreeFunction
8	<input type="checkbox"/> age
9	<input type="checkbox"/> outcome

**RemoveWithValues** dialog box:

weka.filters.unsupervised.instance.RemoveWithValues

About: Filters instances according to the value of an attribute.

attributeIndex: 6

debug: False

doNotCheckCapabilities: False

dontFilterAfterFirstBatch: False

invertSelection: False

matchMissingValues: True

modifyHeader: False

nominalIndices: first-last

splitPoint: 0.0

Buttons: Open... | Save... | OK | Cancel

**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 **RemoveWithValues** -S 0.0 -C 6 -L first-last -M

**Current relation:**  
Relation: diabetes-weka.filters.unsupervised.attribute.Numeric...  
Instances: 757  
Attributes: 9  
Sum of weights: 757

**Selected attribute:**

Name: BMI  
Missing: 0 (0%)  
Distinct: 247  
Type: Numeric  
Unique: 76 (10%)

Statistic	Value
Minimum	18.2
Maximum	67.1
Mean	32.457
StdDev	6.925

**Attributes:**

No.	Name
1	<input type="checkbox"/> pregnancies
2	<input type="checkbox"/> glucose
3	<input type="checkbox"/> bloodPressure
4	<input type="checkbox"/> skinThickness
5	<input type="checkbox"/> insulin
6	<input checked="" type="checkbox"/> BMI
7	<input type="checkbox"/> diabetesPedigreeFunction
8	<input type="checkbox"/> age
9	<input type="checkbox"/> outcome

**Class: outcome (Num)**

**Visualize All**

**Status:** OK

Log x 0

## - Impute Missing Values (replace the missing values with some other value)

**Weka Explorer**

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 Apply Stop

**Current relation**

Relation: diabetes-weka.filters.unsupervised.attribute.Numeric... Attributes: 9  
Instances: 768 Sum of weights: 768

**Attributes**

All None Invert Pattern

No.	Name
1	<input type="checkbox"/> pregnancies
2	<input type="checkbox"/> glucose
3	<input type="checkbox"/> bloodPressure
4	<input type="checkbox"/> skinThickness
5	<input type="checkbox"/> insulin
6	<input checked="" type="checkbox"/> BMI
7	<input type="checkbox"/> diabetesPedigreeFunction
8	<input type="checkbox"/> age
9	<input type="checkbox"/> outcome

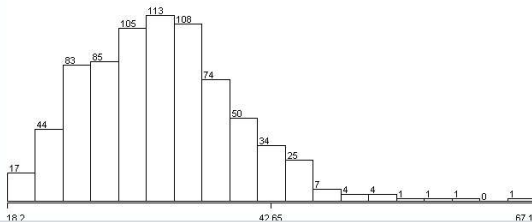
Remove

**Selected attribute**

Name: BMI 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: outcome (Num) Visualize All



18.2 47.65 67.1

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 **ReplaceMissingValues** Apply Stop

**Current relation**

Relation: diabetes-weka.filters.unsupervised.attribute.Numeric... Attributes: 9  
Instances: 768 Sum of weights: 768

**Attributes**

All None Invert Pattern

No.	Name
1	<input type="checkbox"/> pregnancies
2	<input type="checkbox"/> glucose
3	<input type="checkbox"/> bloodPressure
4	<input type="checkbox"/> skinThickness
5	<input type="checkbox"/> insulin
6	<input checked="" type="checkbox"/> BMI
7	<input type="checkbox"/> diabetesPedigreeFunction
8	<input type="checkbox"/> age
9	<input type="checkbox"/> outcome

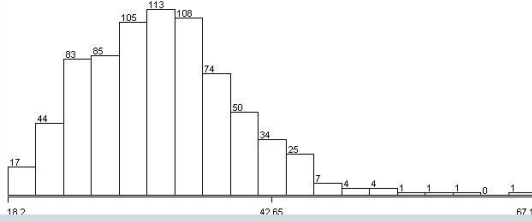
Remove

**Selected attribute**

Name: BMI 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: outcome (Num) Visualize All



18.2 47.65 67.1

OK Log x 0



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Viewer

Relation: diabetes-weka.filters.unsupervised.attribute.NumericCleaner-min1.0E-9-min-defaultNaN-max1.7976931348623157E308-max-default1.7976931348623157E308-closeto0.0-closeto-de

No.	1: pregnancies	2: glucose	3: bloodPressure	4: skinThickness	5: insulin	6: BMI	7: diabetesPedigreeFunction	8: age	9: outcome
	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
1	6.0	148.0	72.0	35.0	0.0	33.6	0.627	50.0	1.0
2	1.0	85.0	66.0	29.0	0.0	26.6	0.351	31.0	0.0
3	8.0	183.0	64.0	0.0	0.0	23.3	0.672	32.0	1.0
4	1.0	89.0	66.0	23.0	94.0	28.1	0.167	21.0	0.0
5	0.0	137.0	40.0	35.0	168.0	43.1	2.288	33.0	1.0
6	5.0	116.0	74.0	0.0	0.0	25.6	0.201	30.0	0.0
7	3.0	78.0	50.0	32.0	88.0	31.0	0.248	26.0	1.0
8	10.0	115.0	0.0	0.0	0.0	35.3	0.134	29.0	0.0
9	2.0	197.0	70.0	45.0	543.0	30.5	0.158	53.0	1.0
10	8.0	125.0	96.0	0.0	0.0		0.232	54.0	1.0
11	4.0	110.0	92.0	0.0	0.0	37.6	0.191	30.0	0.0
12	10.0	168.0	74.0	0.0	0.0	38.0	0.537	34.0	1.0
13	10.0	139.0	80.0	0.0	0.0	27.1	1.441	57.0	0.0
14	1.0	189.0	60.0	23.0	846.0	30.1	0.398	59.0	1.0
15	5.0	166.0	72.0	19.0	175.0	25.8	0.587	51.0	1.0
16	7.0	100.0	0.0	0.0	0.0	30.0	0.484	32.0	1.0
17	0.0	118.0	84.0	47.0	230.0	45.8	0.551	31.0	1.0
18	7.0	107.0	74.0	0.0	0.0	29.6	0.254	31.0	1.0
19	1.0	103.0	30.0	38.0	83.0	43.3	0.183	33.0	0.0
20	1.0	115.0	70.0	30.0	96.0	34.6	0.529	32.0	1.0
21	3.0	126.0	88.0	41.0	235.0	39.3	0.704	27.0	0.0
22	8.0	99.0	84.0	0.0	0.0	35.4	0.388	50.0	0.0
23	7.0	196.0	90.0	0.0	0.0	39.8	0.451	41.0	1.0
24	9.0	119.0	80.0	35.0	0.0	29.0	0.263	29.0	1.0
25	11.0	143.0	94.0	33.0	146.0	36.6	0.254	51.0	1.0
26	10.0	125.0	70.0	26.0	115.0	31.1	0.205	41.0	1.0
27	7.0	147.0	76.0	0.0	0.0	39.4	0.257	43.0	1.0
28	1.0	97.0	66.0	15.0	140.0	23.2	0.487	22.0	0.0
29	13.0	145.0	82.0	19.0	110.0	22.2	0.245	57.0	0.0
30	5.0	117.0	92.0	0.0	0.0	24.1	0.237	38.0	0.0

Add instance Undo OK Cancel

Status

OK

Log

x 0

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Viewer

Relation: diabetes-weka.filters.unsupervised.attribute.NumericCleaner-min1.0E-9-min-defaultNaN-max1.7976931348623157E308-max-default1.7976931348623157E308-closeto0.0-closeto-de

No.	1: pregnancies	2: glucose	3: bloodPressure	4: skinThickness	5: insulin	6: BMI	7: diabetesPedigreeFunction	8: age	9: outcome
	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
1	6.0	148.0	72.0	35.0	0.0	33.6	0.627	50.0	1.0
2	1.0	85.0	66.0	29.0	0.0	26.6	0.351	31.0	0.0
3	8.0	183.0	64.0	0.0	0.0	23.3	0.672	32.0	1.0
4	1.0	89.0	66.0	23.0	94.0	28.1	0.167	21.0	0.0
5	0.0	137.0	40.0	35.0	168.0	43.1	2.288	33.0	1.0
6	5.0	116.0	74.0	0.0	0.0	25.6	0.201	30.0	0.0
7	3.0	78.0	50.0	32.0	88.0	31.0	0.248	26.0	1.0
8	10.0	115.0	0.0	0.0	0.0	35.3	0.134	29.0	0.0
9	2.0	197.0	70.0	45.0	543.0	30.5	0.158	53.0	1.0
10	8.0	125.0	96.0	0.0	0.0	32.4	0.232	54.0	1.0
11	4.0	110.0	92.0	0.0	0.0	37.6	0.191	30.0	0.0
12	10.0	168.0	74.0	0.0	0.0	38.0	0.537	34.0	1.0
13	10.0	139.0	80.0	0.0	0.0	27.1	1.441	57.0	0.0
14	1.0	189.0	60.0	23.0	846.0	30.1	0.398	59.0	1.0
15	5.0	166.0	72.0	19.0	175.0	25.8	0.587	51.0	1.0
16	7.0	100.0	0.0	0.0	0.0	30.0	0.484	32.0	1.0
17	0.0	118.0	84.0	47.0	230.0	45.8	0.551	31.0	1.0
18	7.0	107.0	74.0	0.0	0.0	29.6	0.254	31.0	1.0
19	1.0	103.0	30.0	38.0	83.0	43.3	0.183	33.0	0.0
20	1.0	115.0	70.0	30.0	96.0	34.6	0.529	32.0	1.0
21	3.0	126.0	88.0	41.0	235.0	39.3	0.704	27.0	0.0
22	8.0	99.0	84.0	0.0	0.0	35.4	0.388	50.0	0.0
23	7.0	196.0	90.0	0.0	0.0	39.8	0.451	41.0	1.0
24	9.0	119.0	80.0	35.0	0.0	29.0	0.263	29.0	1.0
25	11.0	143.0	94.0	33.0	146.0	36.6	0.254	51.0	1.0
26	10.0	125.0	70.0	26.0	115.0	31.1	0.205	41.0	1.0
27	7.0	147.0	76.0	0.0	0.0	39.4	0.257	43.0	1.0
28	1.0	97.0	66.0	15.0	140.0	23.2	0.487	22.0	0.0
29	13.0	145.0	82.0	19.0	110.0	22.2	0.245	57.0	0.0
30	5.0	117.0	92.0	0.0	0.0	24.1	0.237	38.0	0.0

Add instance Undo OK Cancel

Status

OK

Log

x 0