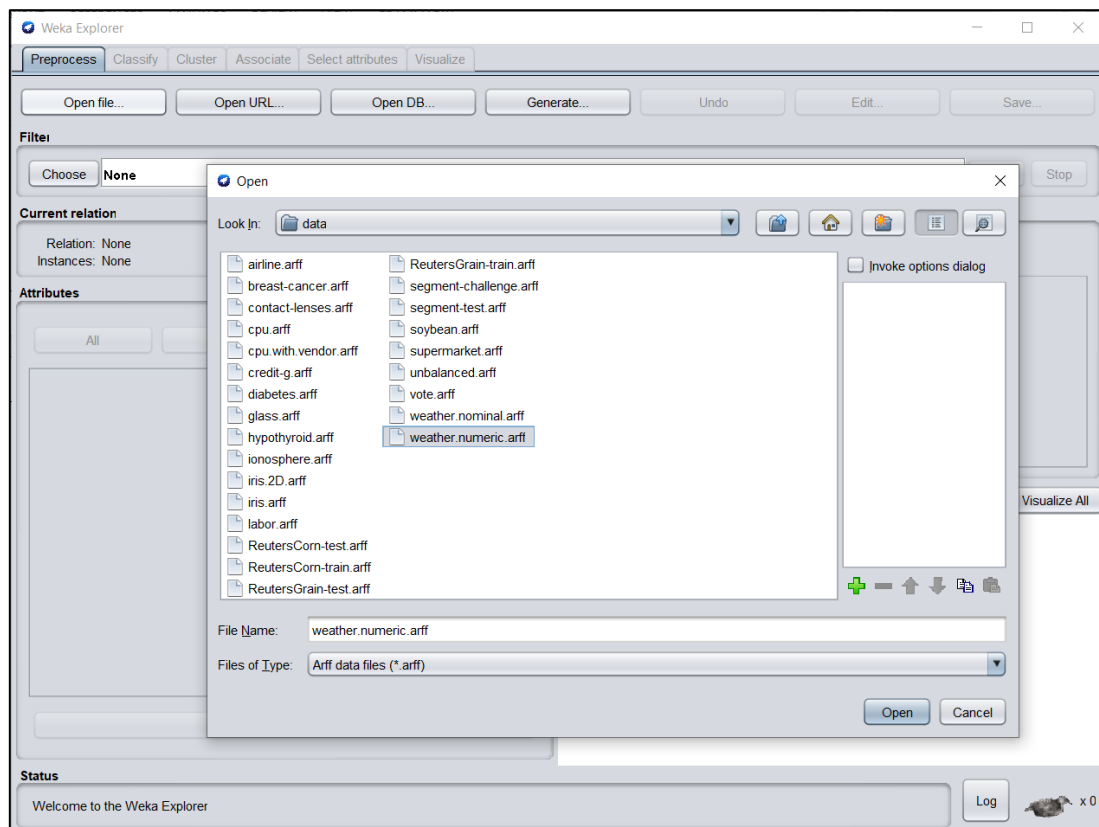
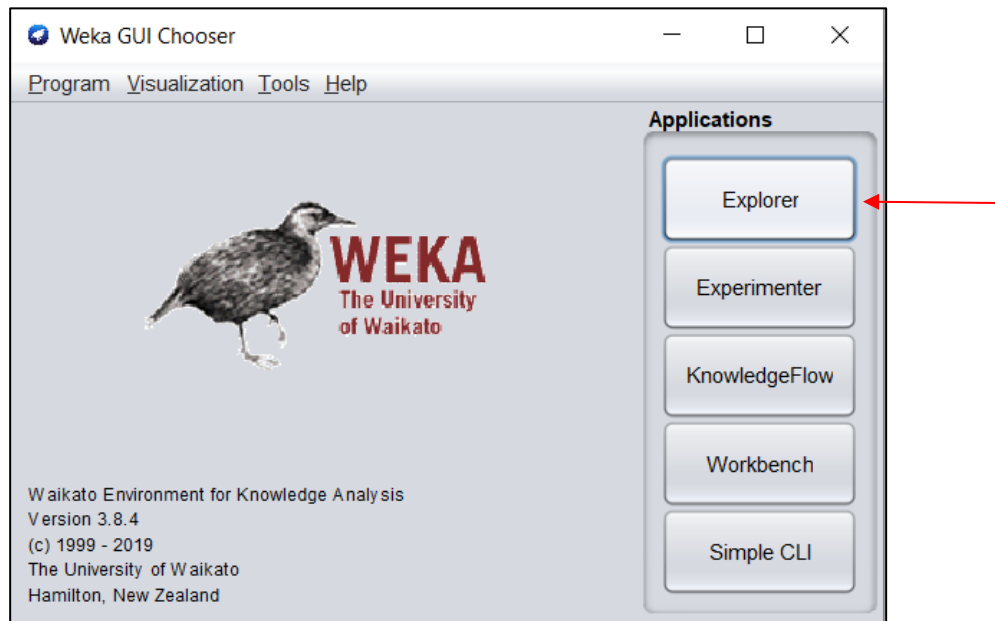
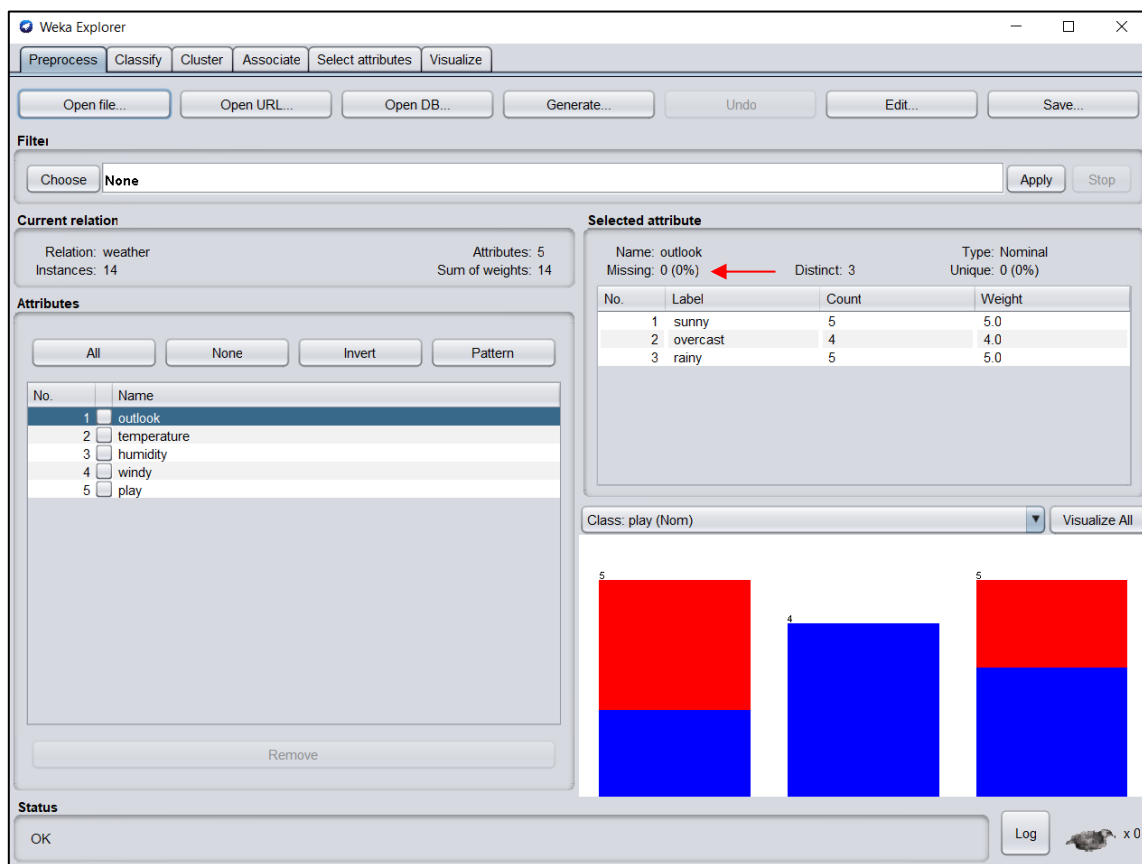


## Apply Data Cleaning methods on a Weather data set using Weka.

**Description:** In this part, we have taken the Weather Numeric dataset which comes preloaded with Weka Tool.

**Step 1:** Open Weka and click on Explore button on top right corner that will open new windows Weka explorer in which load data by clicking open file as shown below snapshot.





**Step 2:** As you can see on above snapshot there is no missing value so we have to manually delete same value for that click on edit button on top right side which will open new window as shown below.

**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	sunny	80.0	90.0	TRUE	no
3	overcast	83.0	86.0	FALSE	yes
4	rainy	70.0	96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	rainy	65.0	70.0	TRUE	no
7	overcast	64.0	65.0	TRUE	yes
8	sunny	72.0	95.0	FALSE	no
9	sunny	69.0	70.0	FALSE	yes
10	rainy	75.0	80.0	FALSE	yes
11	sunny	75.0	70.0	TRUE	yes
12	overcast	72.0	90.0	TRUE	yes
13	overcast	81.0	75.0	FALSE	yes
14	rainy	71.0	91.0	TRUE	no

Add instance | Undo | OK | Cancel

**Step 3:** Now delete some value by directly clicking on value which you want to delete.

Viewer

Relation: weather

No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: play Nominal
1	sunny	80.0	90.0	TRUE	no
2	rainy	65.0		TRUE	no
3	overcast	64.0	65.0	TRUE	yes
4		72.0	90.0	TRUE	yes
5	rainy	71.0		TRUE	no
6		85.0	85.0	FALSE	no
7	overcast	83.0	86.0	FALSE	yes
8	rainy		96.0	FALSE	
9	rainy	68.0	80.0	FALSE	yes
10	sunny		95.0	FALSE	no
11	sunny	69.0	70.0	FALSE	
12	rainy	75.0	80.0	FALSE	yes
13	overcast	81.0	75.0	FALSE	yes
14	sunny	75.0	70.0		yes

Add instance Undo OK Cancel

**Step 4:** To apply data cleaning method click on filter option in which select unsupervised\Attribute\ ReplaceByMissingValues. In below snapshot it can be seen that there is missing value and in filter ReplaceMissingValue option.

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: weather Instances: 14 Attributes: 5 Sum of weights: 14

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

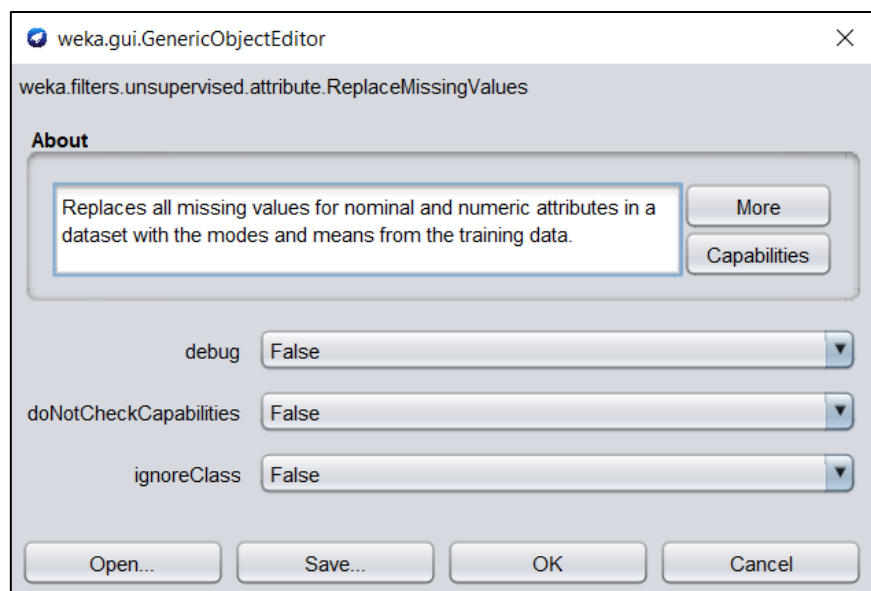
Remove

Selected attribute: Name: outlook Missing: 2 (14%) Distinct: 3 Type: Nominal Unique: 0 (0%)

No.	Label	Count	Weight
1	sunny	4	4.0
2	overcast	3	3.0
3	rainy	5	5.0

Class: play (Nom) Visualize All

Status: OK Log x 0



**Step 5:** Click on apply to replace missing value it will replace all missing value with new value as shown below.

Viewer

Relation: weather-weka.filters.unsupervised.attribute.ReplaceMissingValues

No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: play Nominal
1	overcast	64.0	65.0	TRUE	yes
2	rainy	72.0	90.0	TRUE	yes
3	overcast	83.0	86.0	FALSE	yes
4	rainy	74.0	96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	rainy	75.0	80.0	FALSE	yes
7	overcast	81.0	75.0	FALSE	yes
8	sunny	75.0	70.0	FALSE	yes
9	sunny	80.0	90.0	TRUE	no
10	rainy	65.0	81.8333...	TRUE	no
11	rainy	71.0	81.8333...	TRUE	no
12	rainy	85.0	85.0	FALSE	no
13	sunny	74.0	95.0	FALSE	no
14	sunny	69.0	70.0	FALSE	no

Add instance Undo OK Cancel

We have successfully applied the data cleaning methods on the Weather Numeric Dataset.