Data Mining and Web algorithm

Lab Assignment 2:

[Feb 21-26, 2022]

Patil Amit Gurusidhappa
19104004
B11

1. Create an Employee Table with the help of Data Mining Tool WEKA.

Description:

You need to create an Employee Table with training data set which includes attributes like name,
id, salary, experience, gender, phone number.

Procedure:

Steps:

a. Open a Notepad file

b. Type the following training data set for the Employee Table.

```
@relation employee

@attribute name {x,y,z,a,b}
@attribute id numeric
@attribute salary {low,medium,high}
@attribute exp numeric
@attribute gender {male,female}
@attribute phone numeric

@data
x,101,low,2,male,250311
y,102,high,3,female,251665
z,103,medium,1,male,240238
a,104,low,5,female,2002002
b,105,high,2,male,240240
```

c. After that the file is saved with .arff file format.

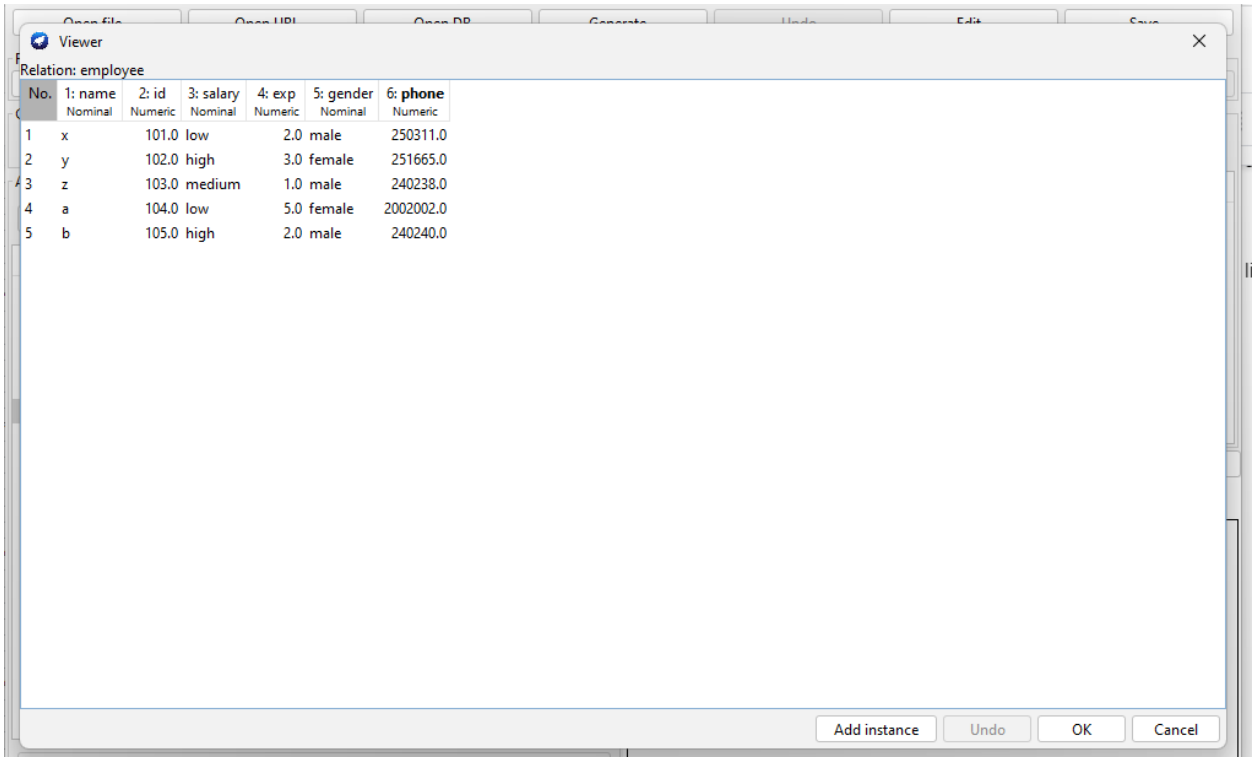
d. Minimize the arff file and then open weka.

e. The Weka dialog box is displayed on the screen.

f. In that dialog box there are four modes, click on explorer.

g. Explorer shows many options. In that click on ‘open file’ and select the arff file

h. **Click on the edit button which shows the employee table on weka**.



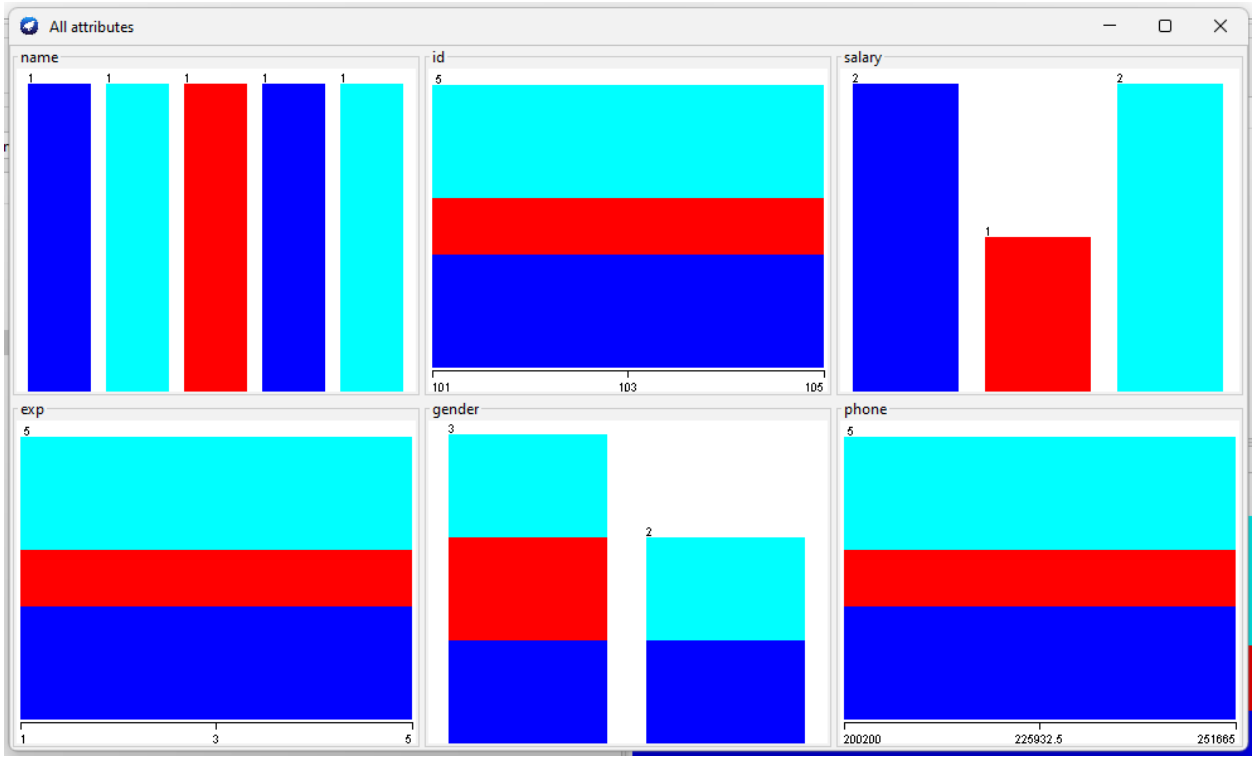
i. Identify the class attribute (if any)

Ans:

1. name
2. id
3. salary
4. Exp
5. gender
6. phone

j. Identify the missing values.

k. Plot Histogram

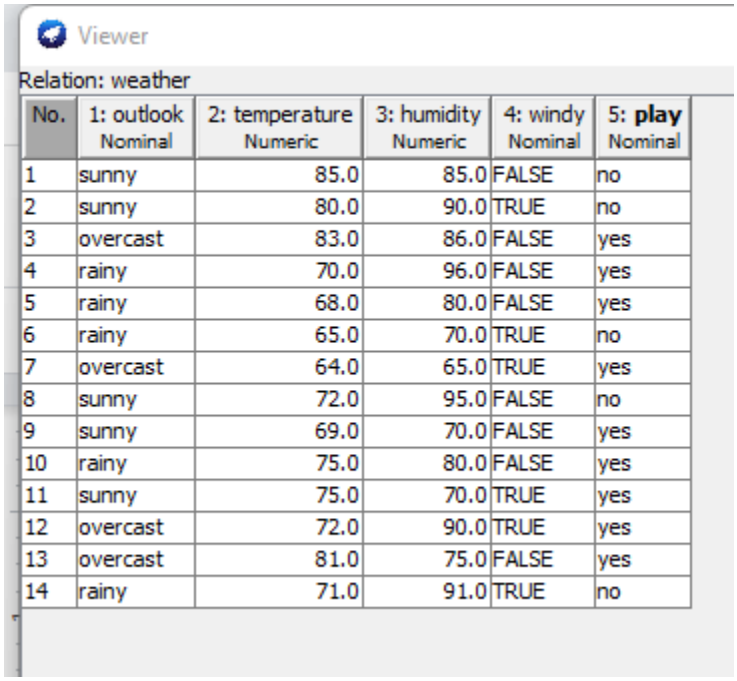


2. Create a Weather Table with the help of Data Mining Tool WEKA.
Description:
We need to create a Weather table with training data set which includes attributes like outlook, temperature, humidity, windy, play.
Weather table created in Weka should be as follows:
@relation weather

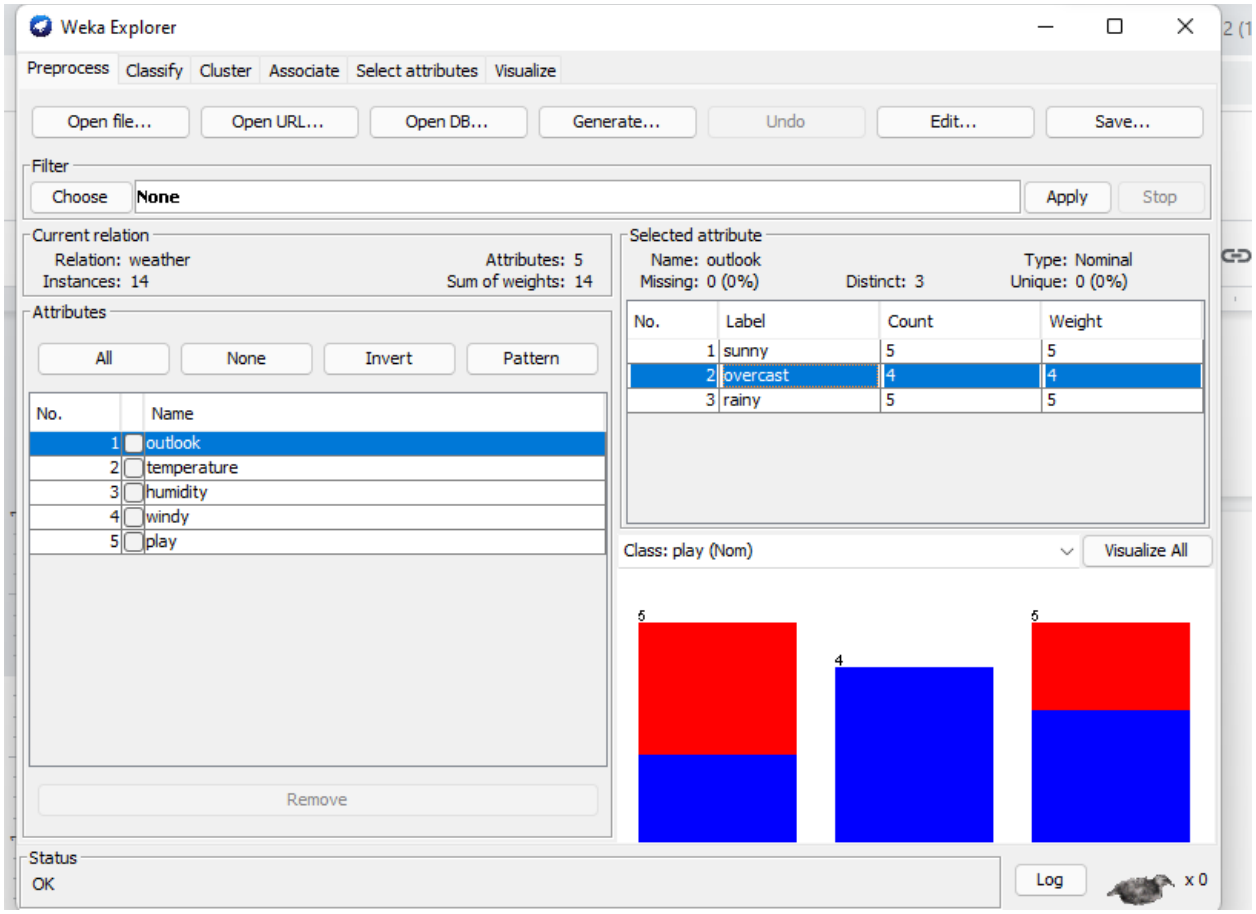
@attribute outlook {sunny, overcast, rainy}
@attribute temperature numeric
@attribute humidity numeric
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,85,85,FALSE,no
sunny,80,90,TRUE,no
overcast,83,86,FALSE,yes
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
overcast,64,65,TRUE,yes
sunny,72,95,FALSE,no
sunny,69,70,FALSE,yes

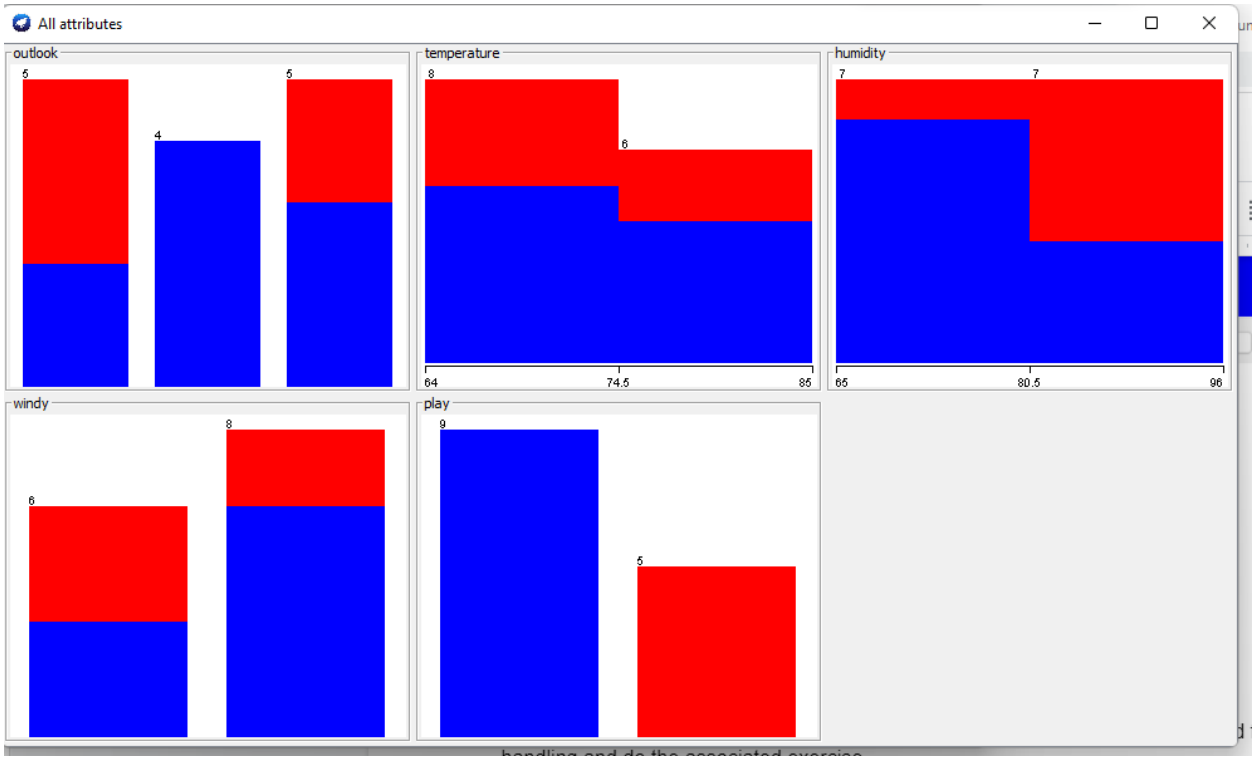
rainy,75,80,FALSE,yes
sunny,75,70,TRUE,yes
overcast,72,90,TRUE,yes
overcast,81,75,FALSE,yes
rainy,71,91,TRUE,no



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



a. Identify the class attribute (if any)
Play (yes or no)
b. Plot Histogram



3. Consider any data set and check for following using Weka Tool:

A. Check for duplicate values

With Duplicates

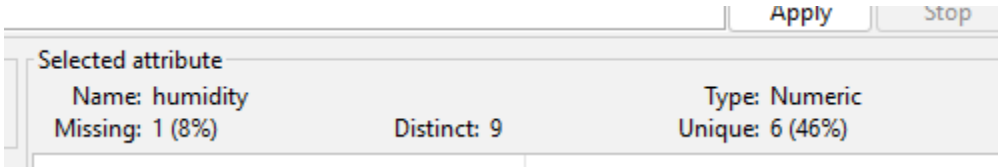
5	rainy	68.0	80.0	FALSE	yes
6	rainy	68.0	80.0	FALSE	yes

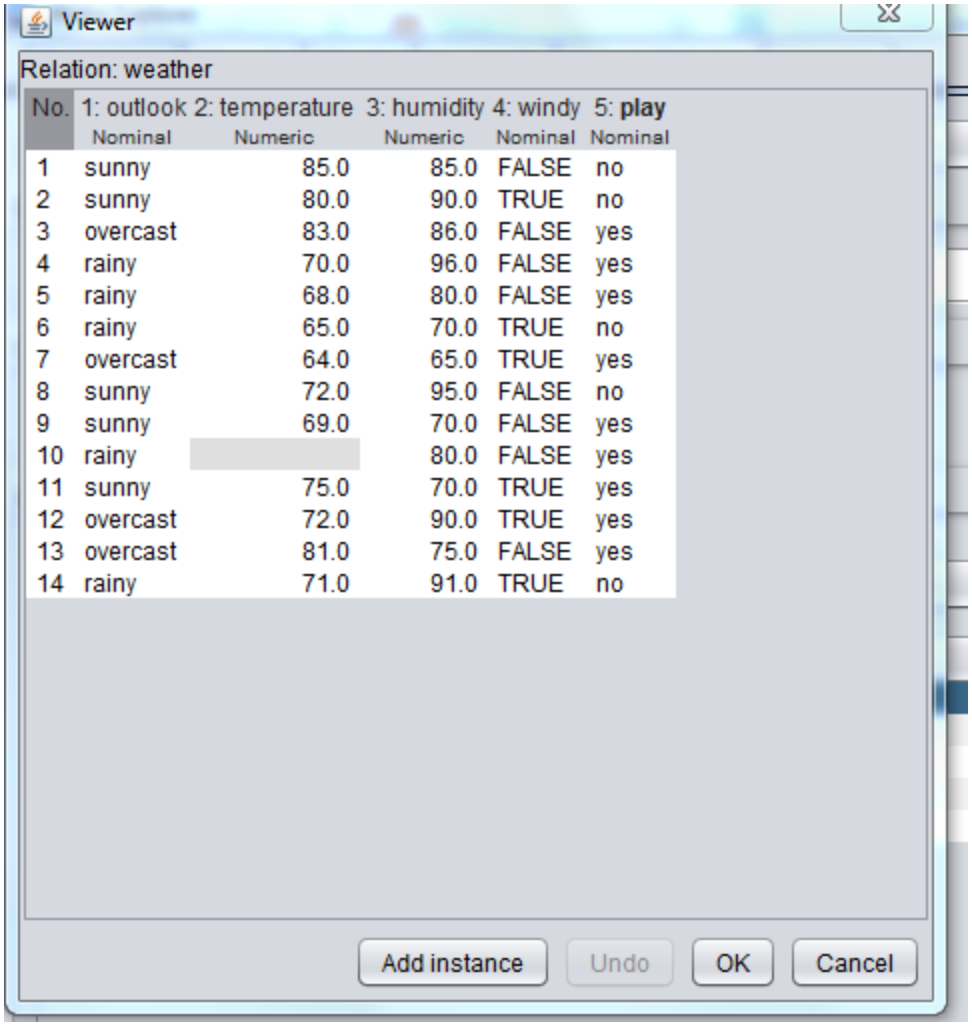
Without Duplicates

4	rainy		96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	overcast	64.0	65.0	TRUE	yes

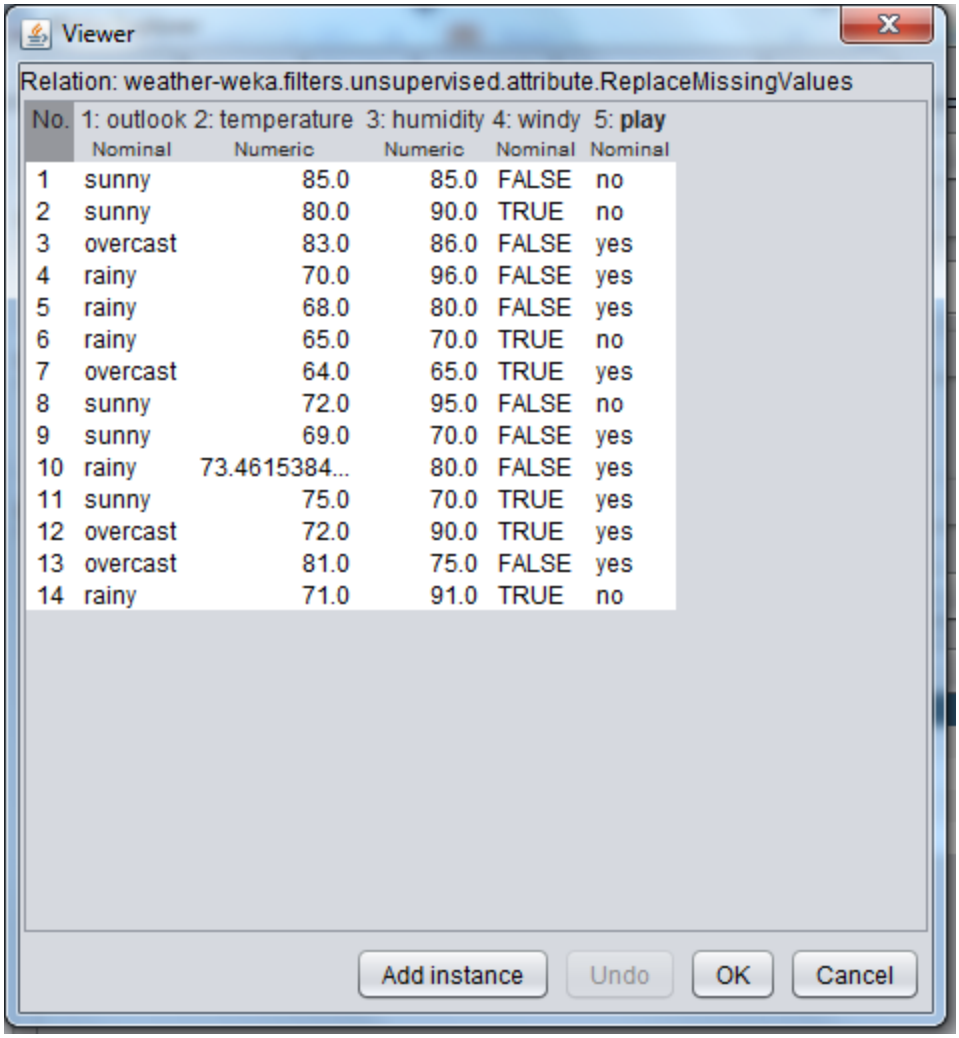
B. Check for null values

C. Check for missing values and replace them with suitable values.





After replacing values



D. Check for data values out of range (like invalid dates, invalid amounts)
E. Identify different data type that may be expected for given information
F. Discretization and Transformation (Floor, ABS etc.) of the data
G. Explore feature selection methods: filter, wrapper and dimensionality reduction.

4. Get familiarize yourself with python Numpy library (placed in Numpy folder) used for array handling and doing the associated exercise.