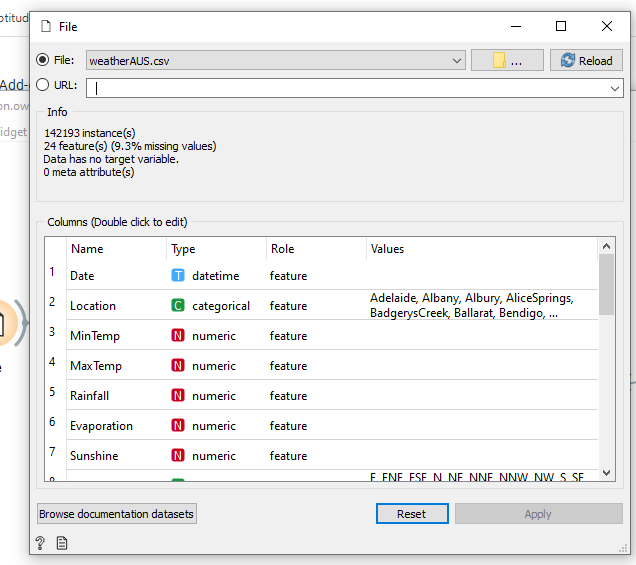
#### **Aim:** Predict whether or not it will rain tomorrow by training a prediction model on target variable RainTomorrow

#### **Description**: This dataset contains daily weather observations from numerous Australian weather stations.

The target variable RainTomorrow means: Did it rain the next day? Yes or No.

#### **Data Selection**:

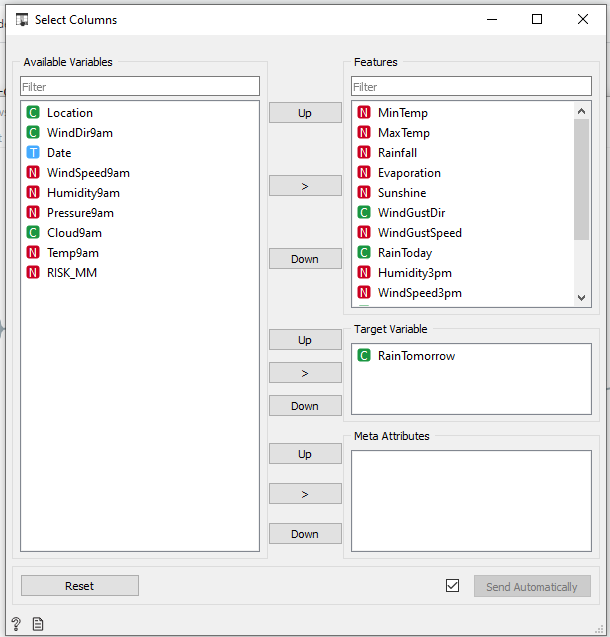
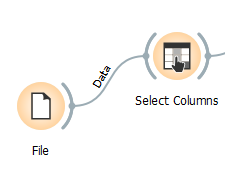
We have downloaded the dataset ( [weather dataset](https://www.kaggle.com/jsphyg/weather-dataset-rattle-package) ) of Australian weather stations which contains about 10 years of daily weather observations from numerous Australian weather stations.



#### **Data Cleaning**:

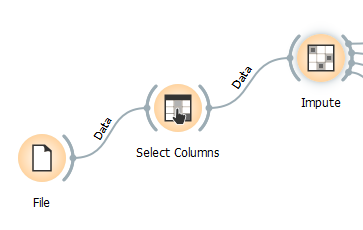
In data cleaning, we will select the required column and remove the rows with missing values as shown below:

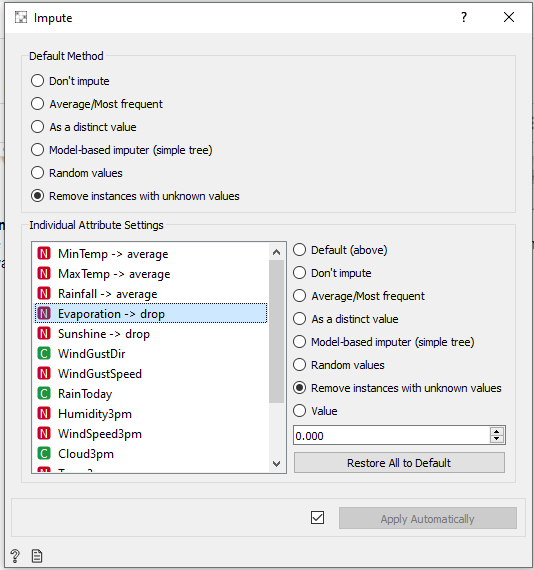
**Column Selection**



**Removing rows with missing values**

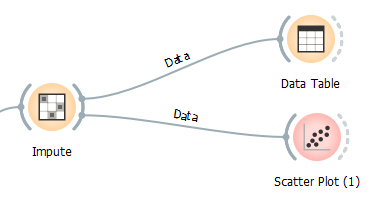
We will use the impute function to remove rows with missing values for columns “Sunshine” and “Evaporation”

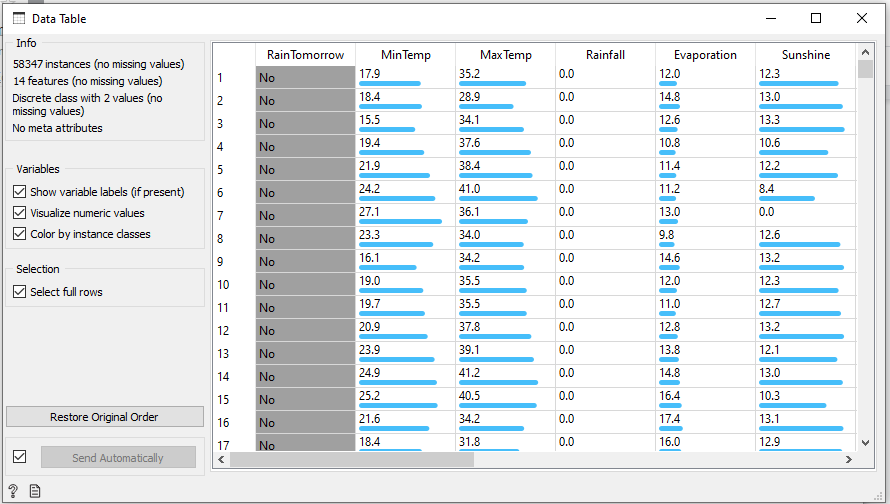
****

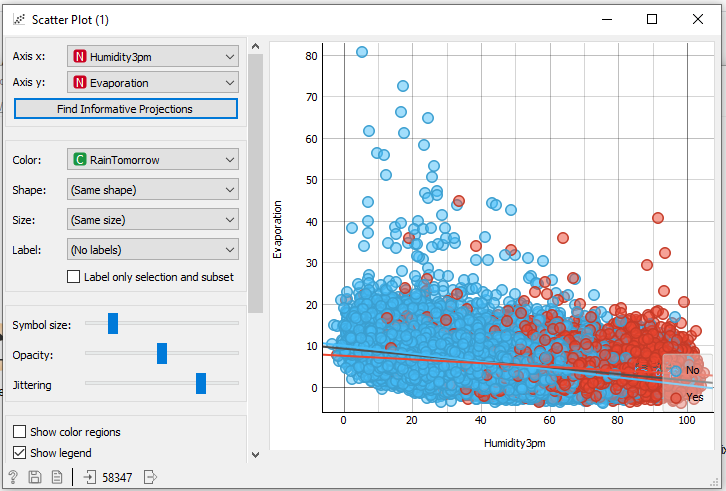
****

#### **Data Visualization**

In this step, we see the reflected data selection in data table. Then, we use scatter plot to visualize the data for better understanding of our dataset.

****

****

****

In the above scatter plot we see, as the humidity level increases and evaporation rates are low the chances of rainfall increases (red dots)

#### **Data Modeling**

In this step, we use **Logistic regression** to model our data. Then, we use “test and score” function to evaluate the performance of our model. Finally, we use “confusion matrix” to understand the prediction.

