REPORT

Team Names and ID:

Sai Rohith Pasala; 1001873156

Sri Naga Venkata Pavan Kalyan Sirigibattula; 1001886149

Rutvik Naga Sai Dondapati; 1001879177

Exploratory Analysis of Datasets using Statewise Testing Details Dataset using WEKA

Weka is Waikato Environment for knowledge analysis and it is a free software which is used for pre-processing of data, visualization of data and classification and clustering of data where we could implement several Machine learning algorithms and techniques.

This how the Graphical user interface of Weka looks as soon as we launch it, various applications of weka are displayed such as explorer,workbench and so on.

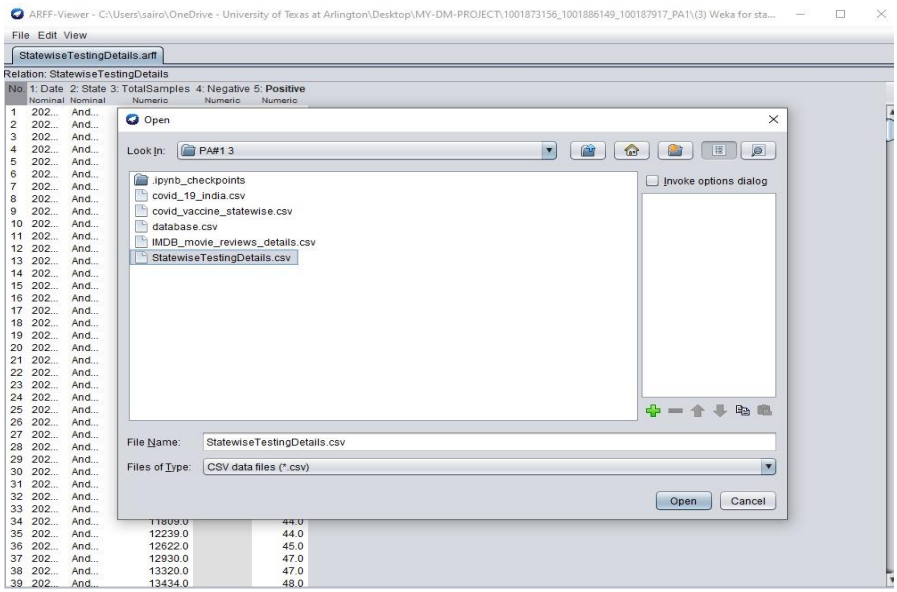
Statewise Testing Details Dataset:

The dataset that we have selected for implementing Weka is Statewise Testing details which contains the information regarding the Total samples collected in each and every state in India for every date with either positive or negative values for the number of cases. Our dataset contains a total of five columns which are:

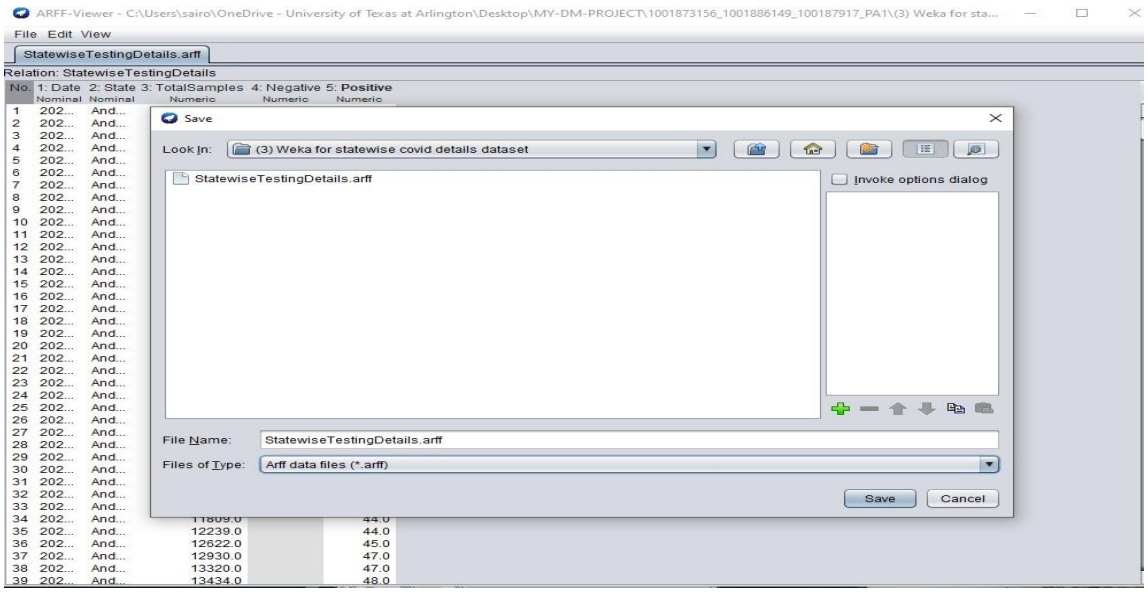
1. Date.
2. State.

3. Total Samples.
4. Positive.
5. Negative.

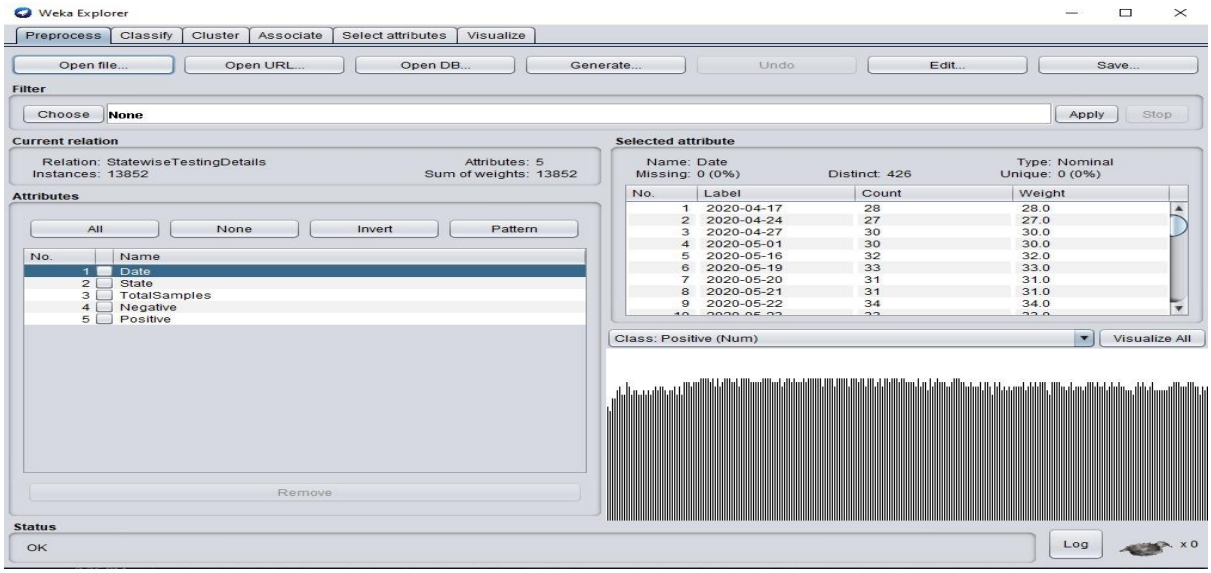
These are all the attributes that we have in the five columns of our dataset, so now first we have to import the dataset into Weka to implement the visualization on the given dataset.



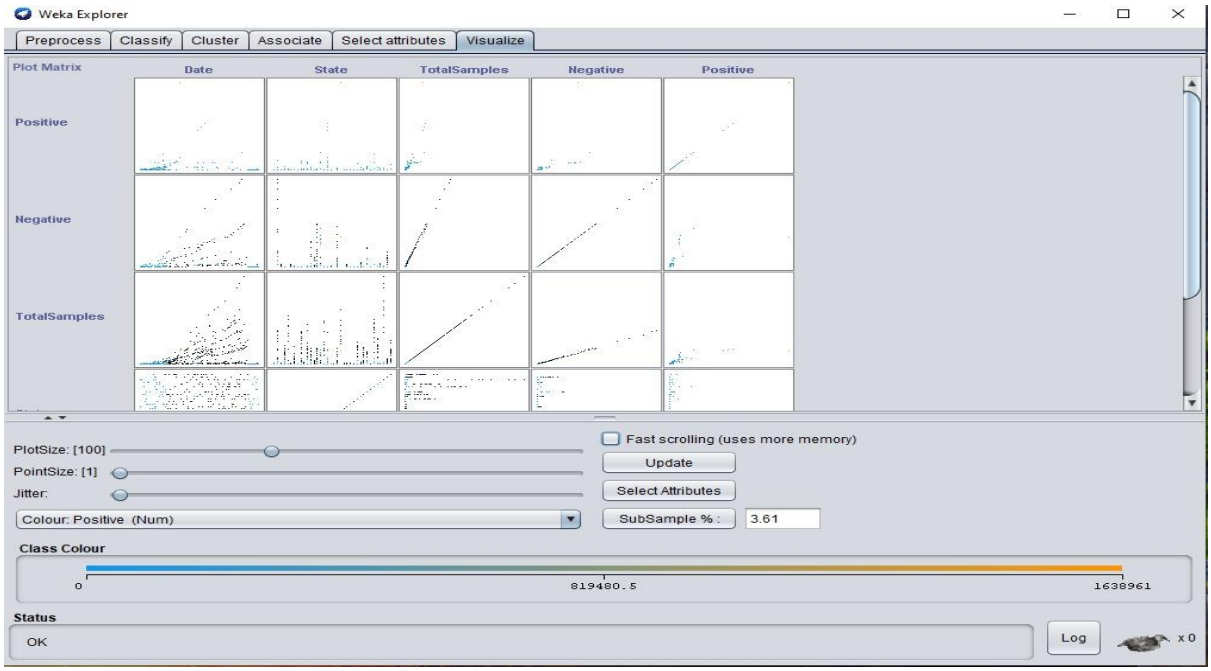
We have to import our dataset as .csv file from the current working directory, after we import the required dataset go to the explorer and select the required visulaizer and then click on the visualization we want to see.



Convert the imported .csv file to .arff format.



All the null data and redundant data is very less and is automatically gets filtered, further wecan remove and compare variables against each other to get a clear idea of the dataset.



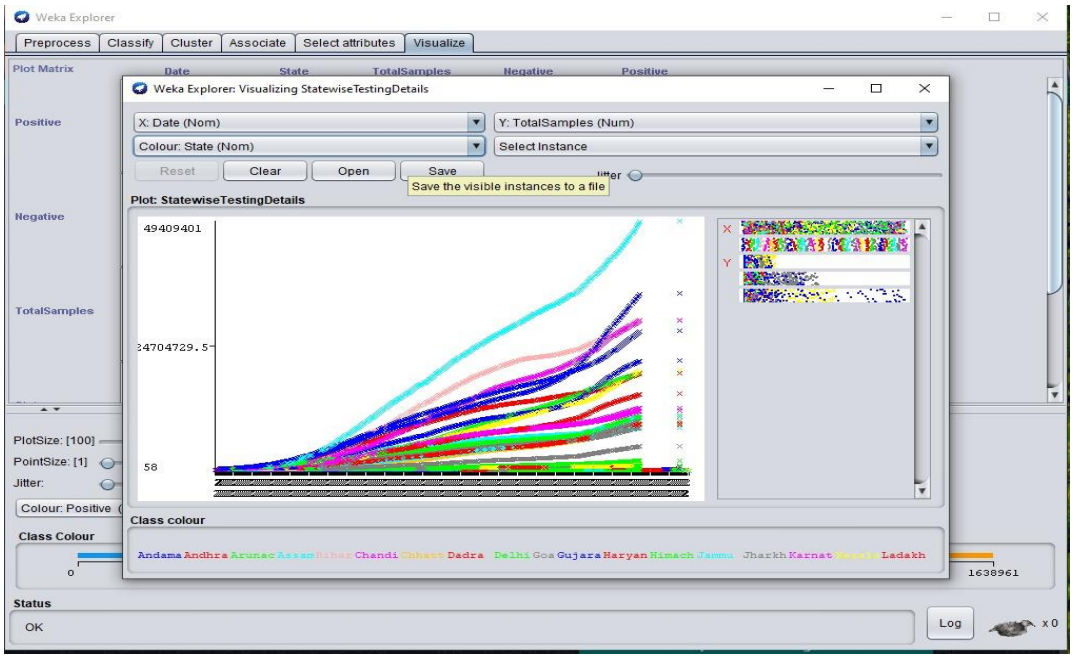
Visualising all attributes in weka visulaise tab, as seen in above picture we can see the grpahs of different varibales plotted against frequency.

The graph below demonstrates the visualization of State vs Total Samples:

X label:Date

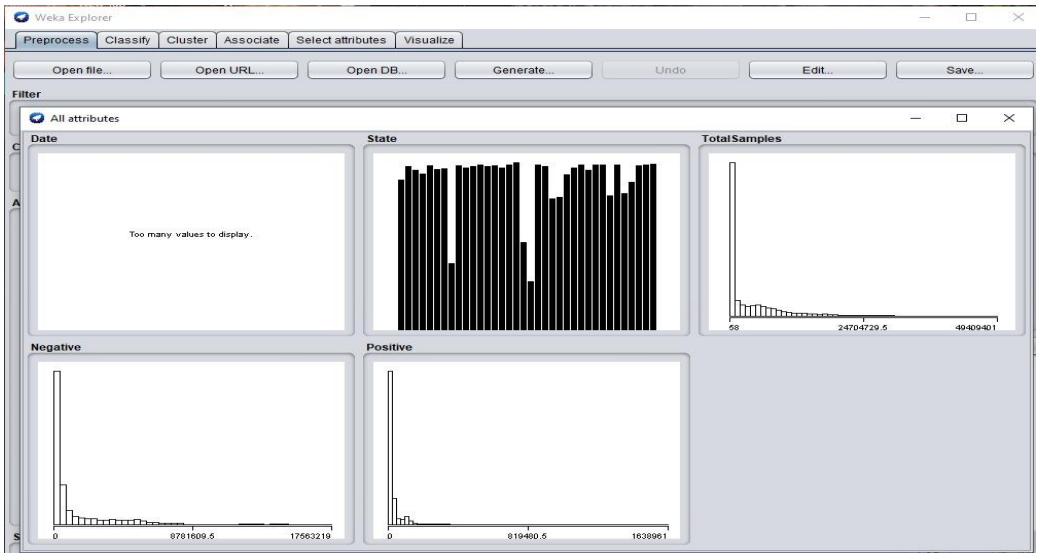
Y label: Total Samples

Color label: States



The graph below demonstrates the visualization of State vs Total Samples:

All the variables graphs plotted against frequency

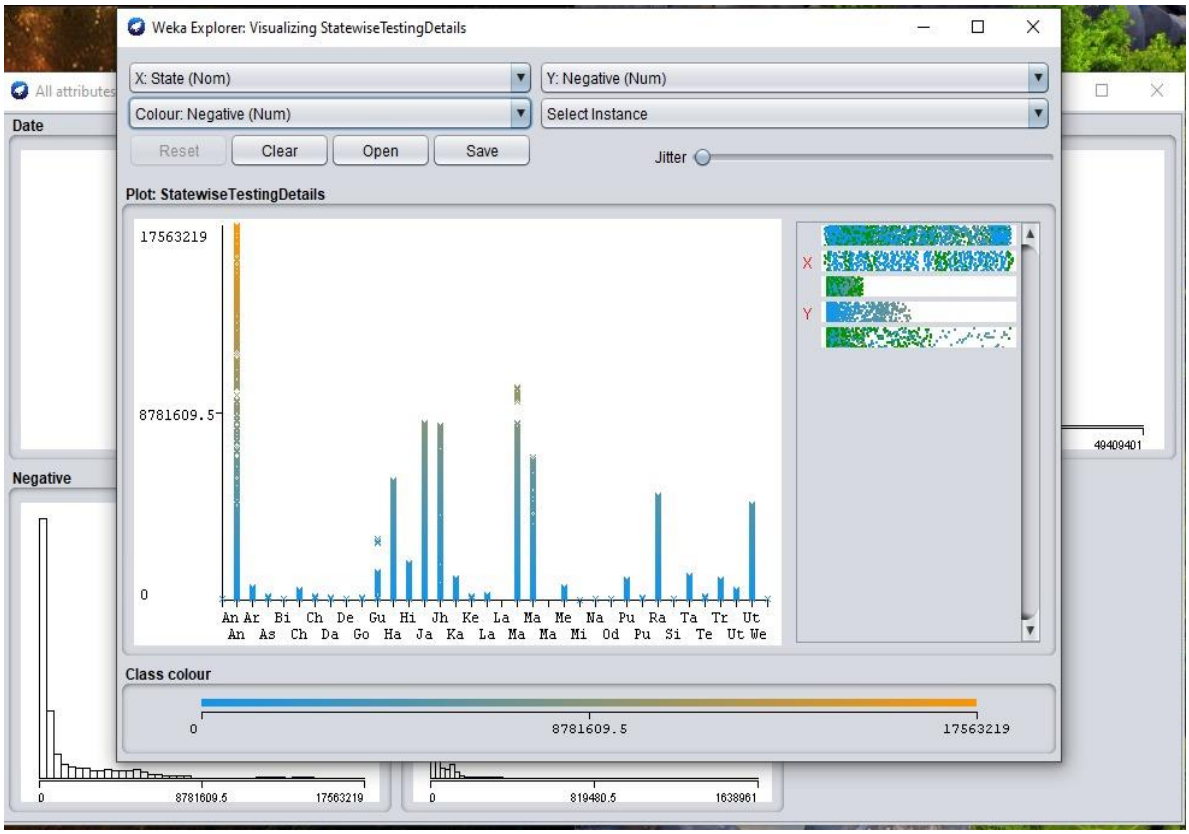


The graph below demonstrates the visualization of State vs Negative cases:

X label:State

Y label: Neagtive cases

Color label: Negative

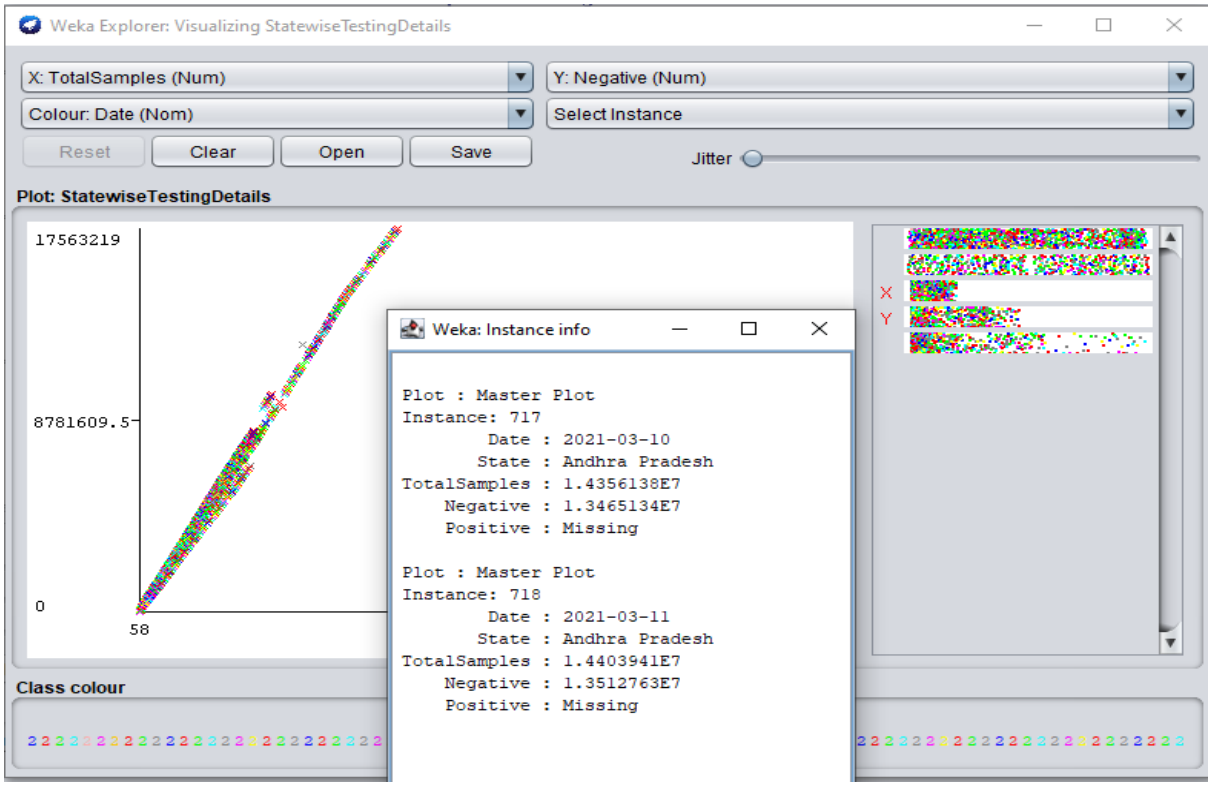


The graph below demonstrates the visualization of Total samples vs negative:

X label:Total samples.

Y label: Negative cases, Color label: Date

Also, if we click on one of the plots in the graph we can see total details of the plot.

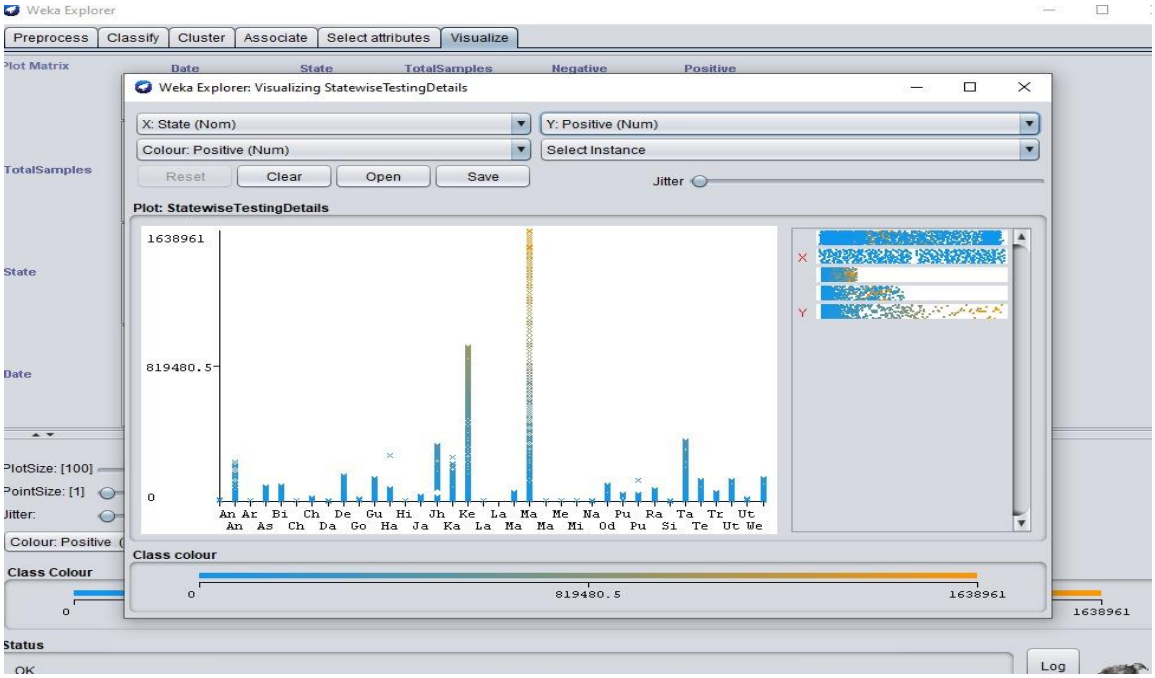


The graph below demonstrates the visualization of State vs Postitive:

X label:State

Y label: Positive cases

Color label: Positive



Special Pattern/ Interesting information on Statewise covid detials dataset is:

1. Highest number of cases are observed only in state Maharashtra compared all other states.

2. Highest number of total samples taken is Uttar Pradesh State

3. Highest number Negative cases have been observed in Andaman and Nicobar State

4. We can see that the positive, negative cases and total cases are co-related.

Higher the total cases, higher the postive cases is seen in Mahrashtra state.

Higher the total cases, higher the neagtive cases in Andaman and Nicobar Island.