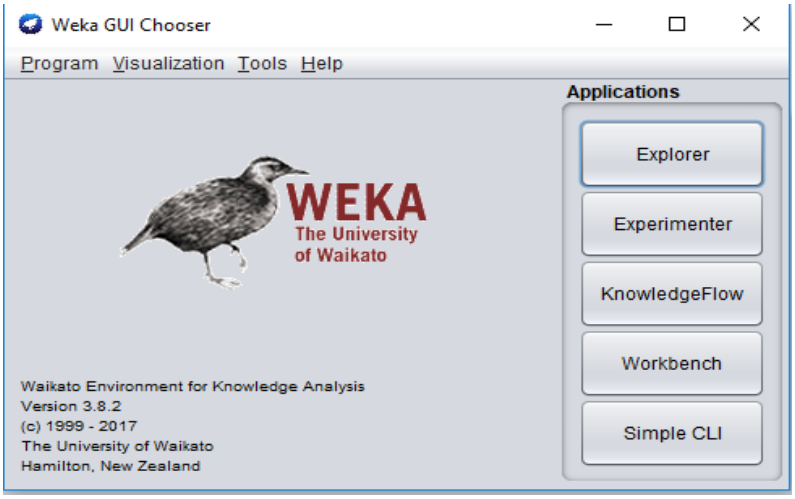
EX.No: 09
Date :

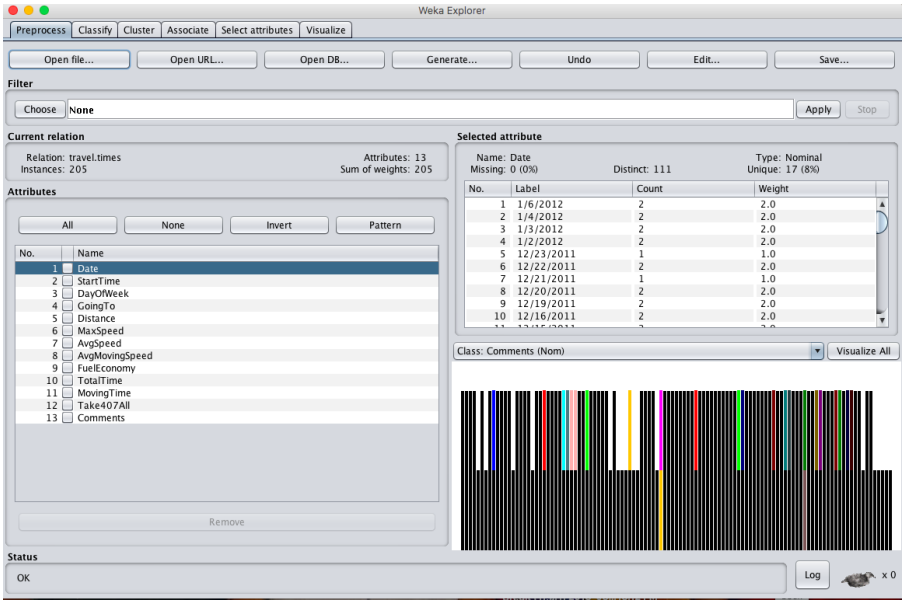
DATA PREPROCESSING AND ANALYSIS FOR DATASET USING WEKA

DESCRIPTION :

Consider a dataset of traveltimes.csv file where it contains the columns of (or) attributes as Date, StartTime, DayOfWeek, GoingTo, Distance, MaxSpeed, AvgSpeed, AvgMovingSpeed, FuelEconomy, TotalTime, MovingTime, Take407All comments.



PREPROCESS :



OBSERVATION :

A. ATTRIBUTE TYPE :

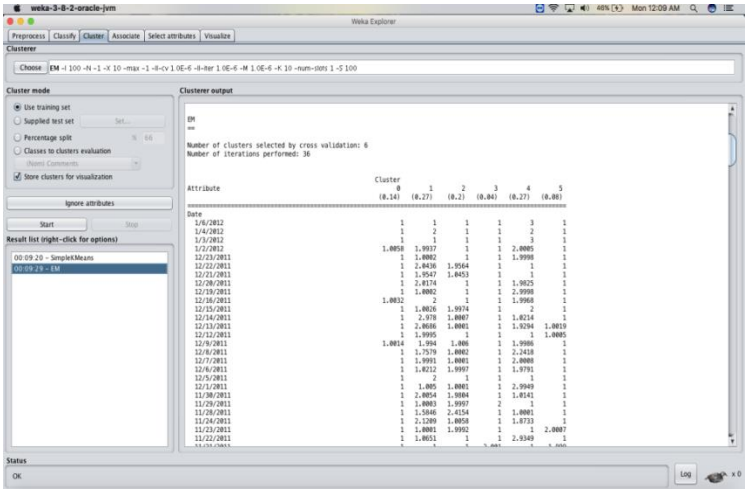
S.NO	ATTRIBUTE	TYPE
1.	Date	Nominal
2.	Start Time	Nominal
3.	Day Of Week	Nominal
4.	Going To	Nominal
5.	Distance	Numeric
6.	Max Speed	Numeric
7.	Avg Speed	Numeric
8.	Avg Moving Speed	Numeric
9.	Fuel Economy	Nominal
10.	Total Time	Numeric
11.	Moving Time	Numeric
12.	Comments	Nominal
13.	Take 407 All	Nominal

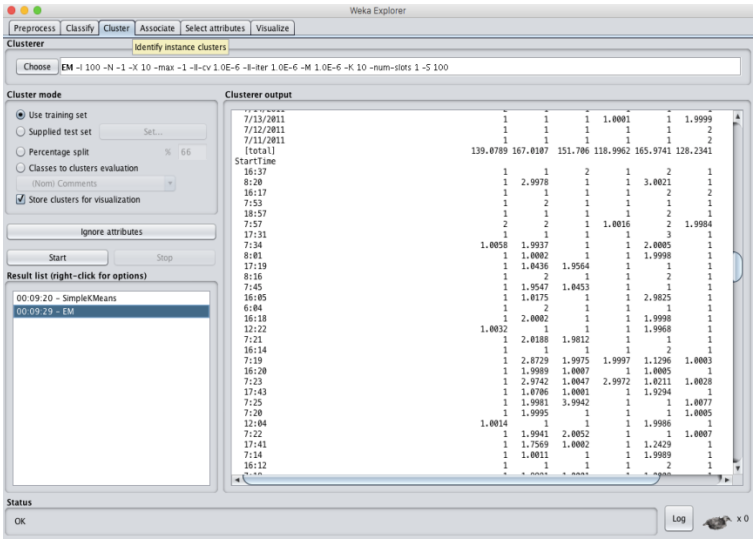
B. PERCENTAGE OF MISSING VALUES :

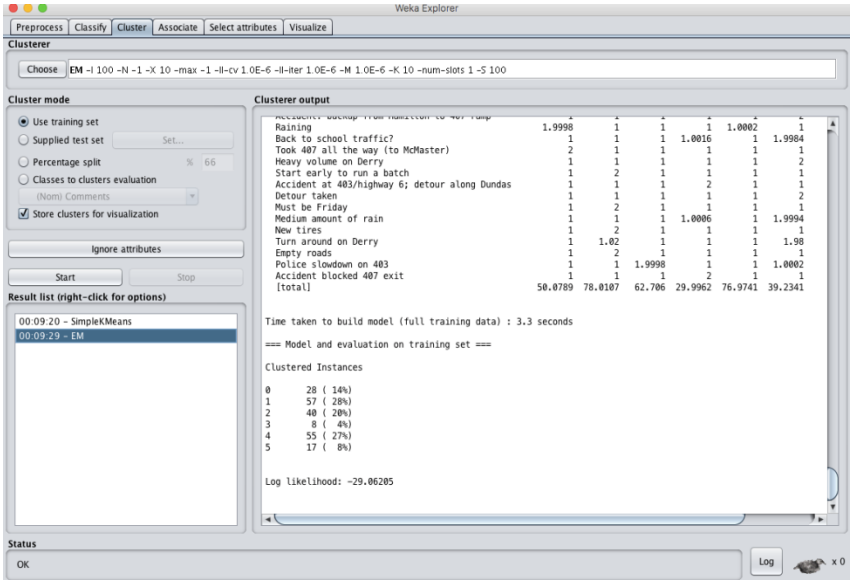
S.NO	ATTRIBUTE	Percentage Of Missing Values

1.	Date	0 %
2.	Start Time	0 %
3.	Day Of Week	0 %
4.	Going To	0 %
5.	Distance	0 %
6.	Max Speed	0 %
7.	Avg Speed	0 %
8.	Avg Moving Speed	0 %
9.	Fuel Economy	8 %
10.	Total Time	0 %
11.	Moving Time	0 %
12.	Comments	88 %
13.	Take 407 All	0 %

C. MIN, MAX, MEAN, STANDARD DEVIATION :







RESULT :

Thus, the dataprocessing and analysis for a dataset using weka tool has been successfully completed.

EX.No: 10
Date :

DATA SEGMENTATION BYK- MEANS CLUSTER USING WEKA AND R-TOOL

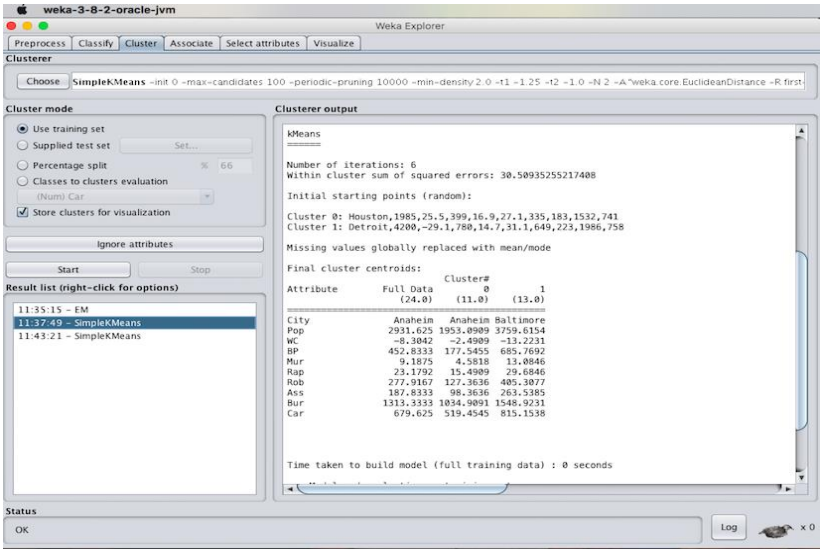
DESCRIPTION :

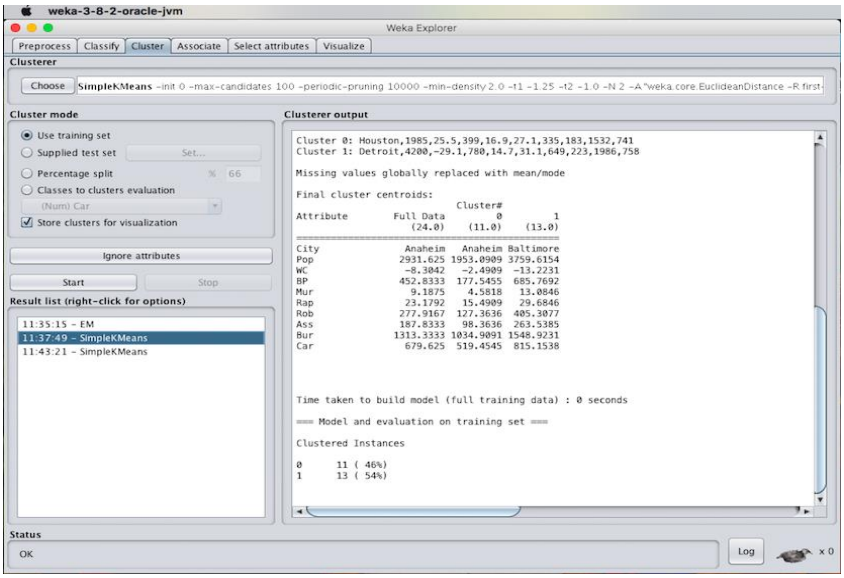
Consider a dataset of citycrimes.csv file of which it contains the attributes are City, Pop, WC, BP, Mur, Rap, Rob, Ass, Bus and car for the performance of the dataset by applying the K-means algorithm in weka and as well using R- tool.

❖ USING WEKA TOOL :

STEPS INVOLVED :

- Choose a set of attributes for clustering and for giving a motivation.
- Choose the dataset and import the dataset into Weka tool.
- Cluster the dataset and choose simple K-means algorithm and give the motivation.

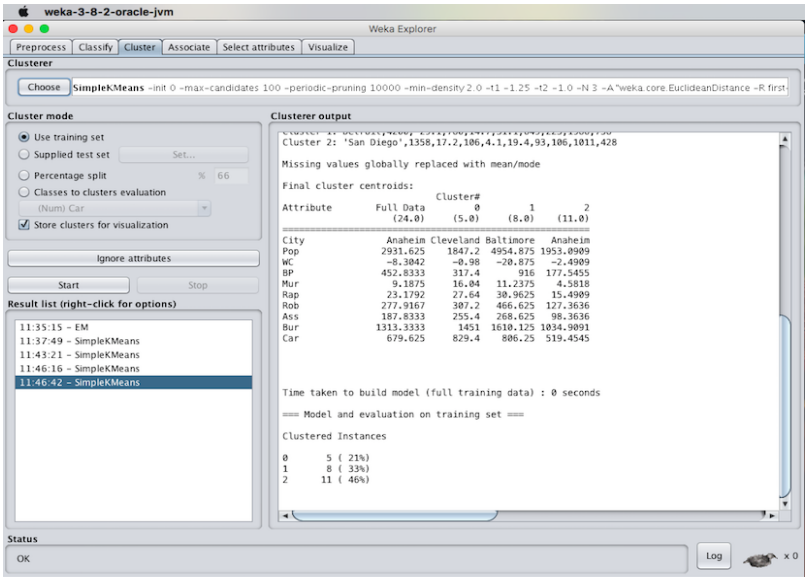


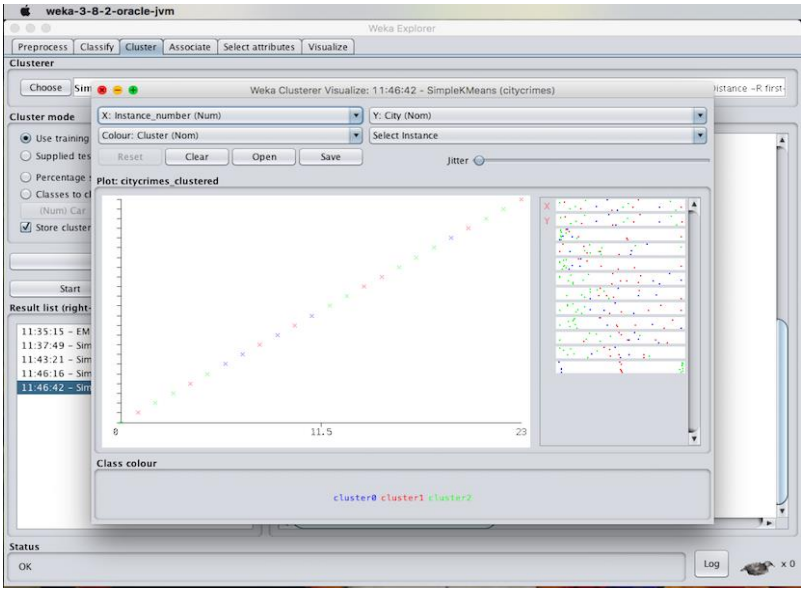


A. Experiment with atleast 2 different number of clusters but with same seed values:

STEPS INVOLVED :

- Compare the two different clusters but with the same seed values.
- Change the number of clusters value and need not to change the seed value.
- Apply the K-means algorithm and start executing the algorithm.





RESULT :
Thus, the K-means clustering analyzing using the weka tool has been successfully completed. In case of weka tool, the change in seed values lead to the decrease in the number of iterations.

EX.No: 11
Date :

DATASEGMENTATION BY EXPECTATION MAXIMISATION ALGORITHM THROUGH WEKA

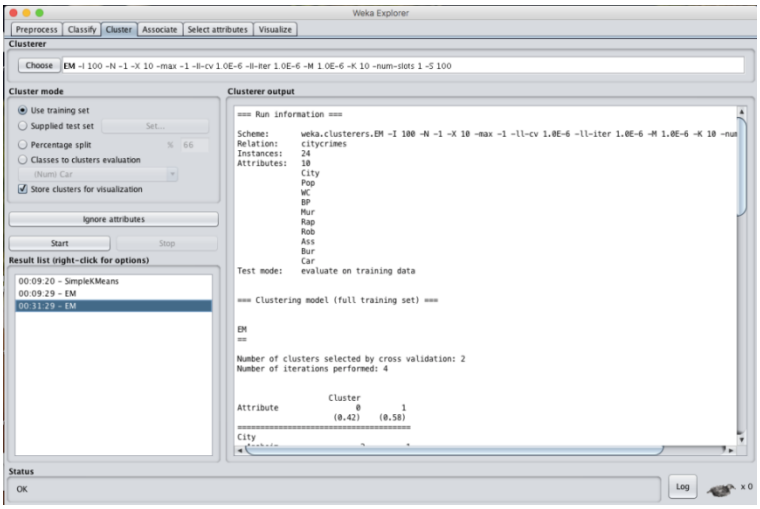
DESCRIPTION :

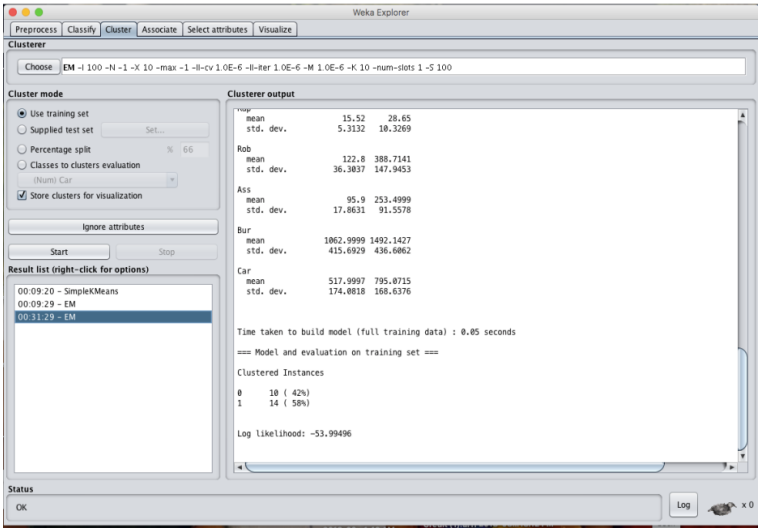
Consider a dataset of citycrimes.csv file of which it contains the attributes are City, Pop, WC, BP, Mur, Rap, Rob, Ass, Bus and car for the performance of the dataset by applying the K-means algorithm in weka and as well using R- tool.

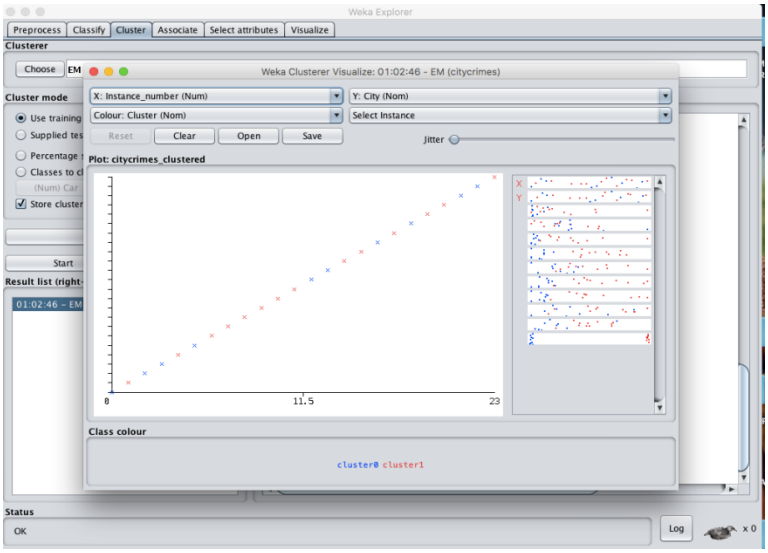
When the clustering is been made through the expectation maximization algorithm by setting minimum standard deviation values then the results will be of the following :

❖ Steps Involved :

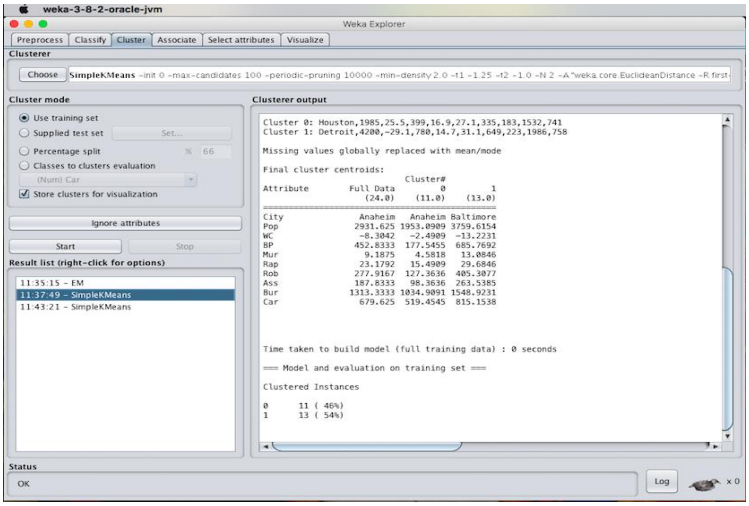
- Initially, load the dataset into the weka tool and check for all the attributes present in the dataset.
- Then move to cluster panel and apply the EM algorithm technique for the datasheet.
- Finally, Observe the results that are obtained.







❖ K- MEANS ALGORITHM:



RESULT :

Thus, the data analysis by the expectation maximization algorithm using weka has been analyzed and observed properly .