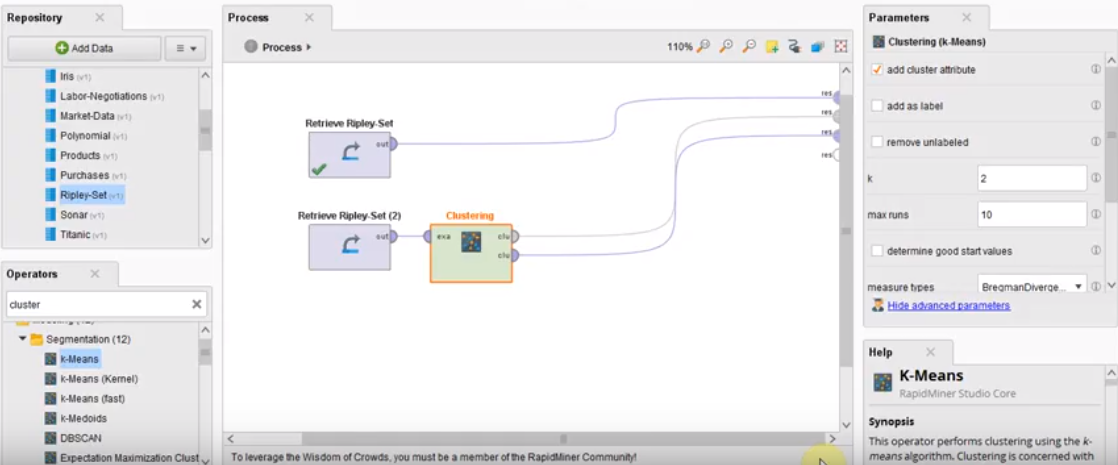
**Lab: K-Means Based Clustering**

1) Retrieve the Ripley Dataset from the Sample Repository into the Process window.

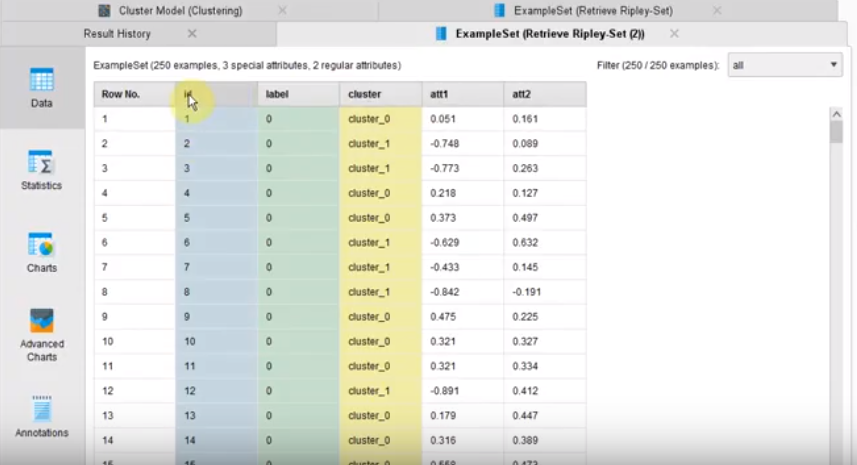
2) Connect to the output port and Run the process, It will show the data set in tabular format and the label attribute.

3) Select the Operator Clustering and then in Segmentation section, select “k-Means” clustering operator

4) Connect the Ripley operator ports to the “k-means” operator and then join both “k-Means” operator port to the output port as mentioned below

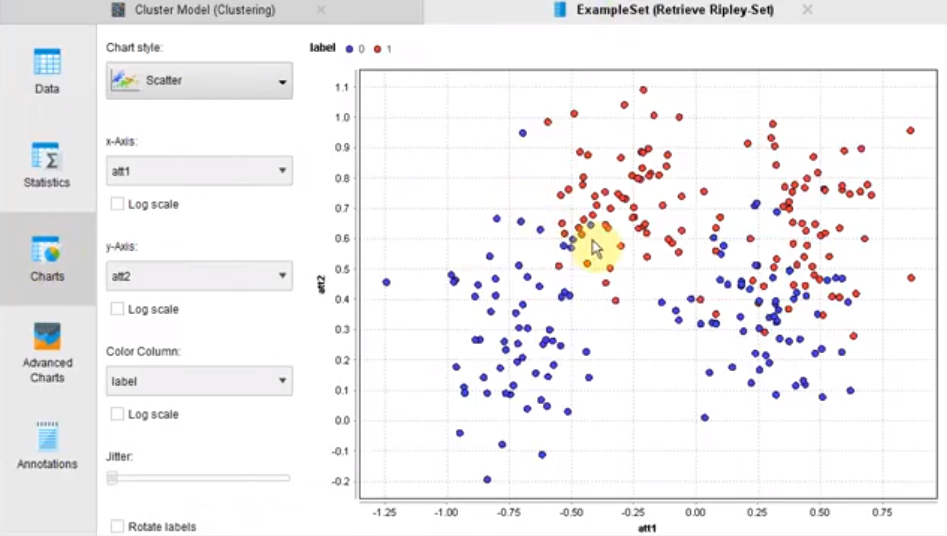


5) Run the program and the following output will be displayed as mentioned below

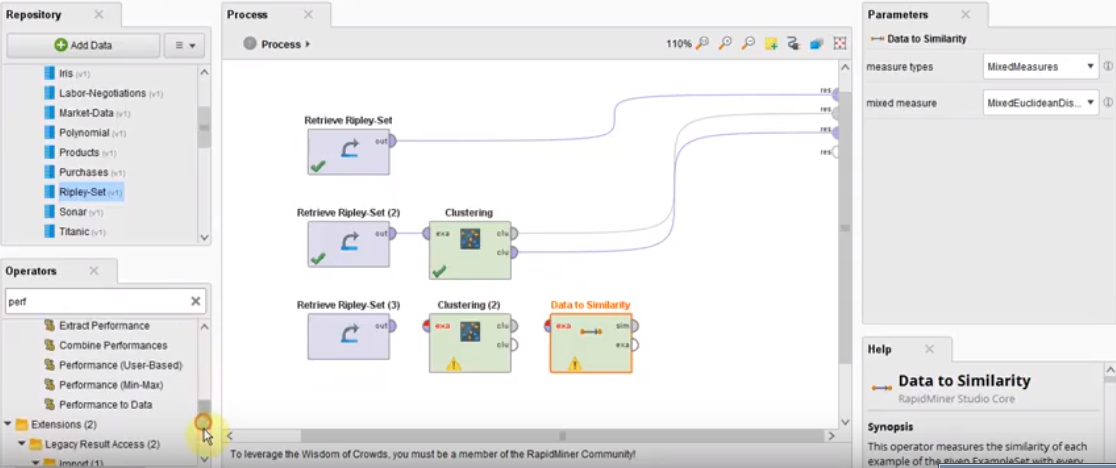


6) 0 represent the first cluster and 1 represents the second cluster as you scroll down the page.

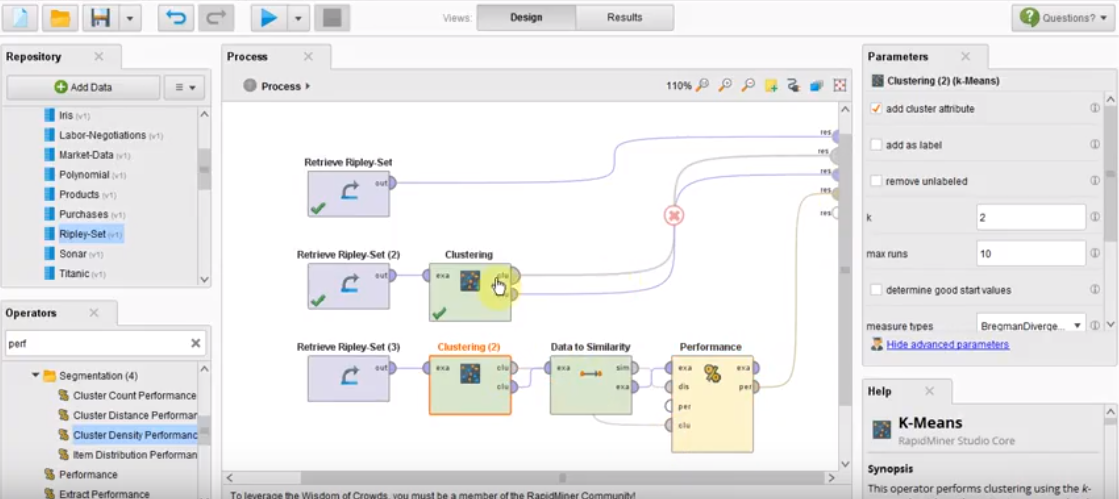
7) Click on the tab, “Chart” and you can see the groups of clusters as mentioned blow



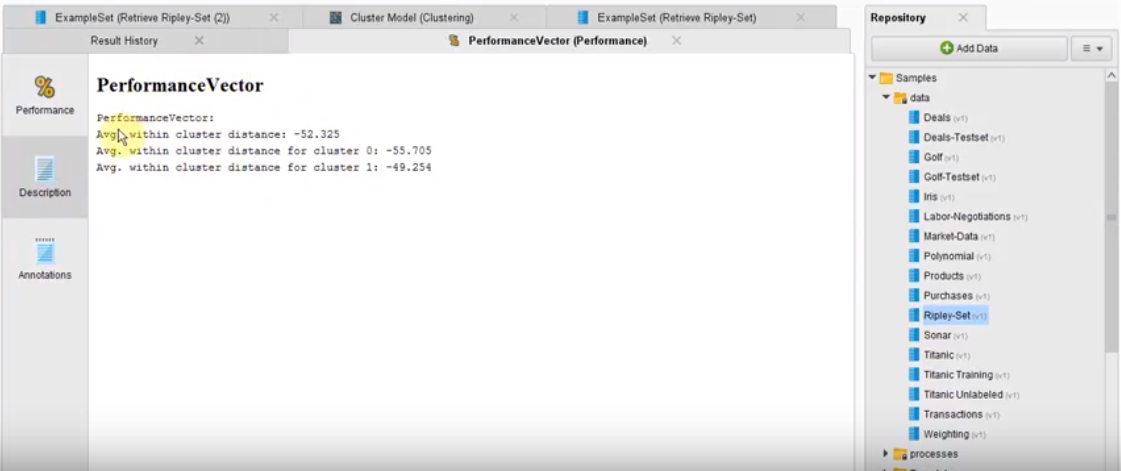
8) Again, Retrieve the Ripley data set third time and “k-Means” clustering. The purpose of “Data to Similarity” is described as *“The Data to Similarity operator calculates the similarity among examples of an ExampleSet. Same comparisons are not repeated again e.g. if example x is compared with example y to compute similarity then example y will not be compared again with example x to compute similarity because the result will be the same. Thus if there are n examples in the ExampleSet, this operator does not return n^2 similarity comparisons. Instead it returns (n)(n-1)/2 similarity comparisons. This operator provides many different measures for similarity computation. The measure to use for calculating the similarity can be specified through the parameters.”* Then Select the operator “Data to Similarity” operator as shown below



9) Drag another operator “Performance” of category “Cluster Density Performance” into the Process window. Connect the ports of the Operators as mentioned below



10) Run the process and you can check the performance vector as mentioned below



11) You can perform analysis based on the results obtained from this tutorial.

12) For complex data set analysis, Follow the instructions as mentioned on the youtube video link

<https://www.youtube.com/watch?v=lZho66YQEIM>