‘Clustering using OPTICS’ visual by MAQ Software analyze a given dataset based on OPTICS algorithm and identifies clusters in your data. The algorithm works on density-based clustering, so you can also identify individual points that don’t belong to any of the groups and there is no need to define the number of clusters in advance.

Clustering is a powerful tool to optimize marketing campaigns based on customer clusters. For example, as an insurance company, we need to identify the insurance needs of states in the US. To create such products, we need to know specific attributes of states like income, life expectancy etc. which specifically cater to them. Since it is practically not possible to create each state-specific product or product which is successful in one state can be successful in a similar type of state as well, we will try to see the group of states having common attributes. Here we can cluster US states using the data of their residents. We chose to cluster by life expectancy and income as that is our interest area. Though if required we can add more dimensions as well.

From this dataset, we get 4 different type of clusters and some outliers. So, we can create a plan accordingly.

Here, Alaska (AK) is an outlier with the highest average income, and Mississippi is outlier having the lowest average income. So probably separate plan might be needed for them as well.

You can also perform hierarchical clustering on your data to identify more closely-knit groups within larger groups. Here we can see that two of the clusters are falling under a bigger cluster. So instead of creating a separate entire plan, the plan made for bigger cluster may be sufficient for them with little modifications. Hence efforts can be saved.

This visual also allows you to adjust the clustering parameters to control hierarchy depth and cluster sizes.

Let’s see how we can customize the chart. For that let’s go to formatting pane. In ‘Clustering settings’ user can define parameters as per requirement. The parameters include –

1. Data scaling, allows you to get your axis values. In case of random data with no axes values, the visual will show values from -2 to 3
2. Epsilon, this parameter provides the “maximum” distance or radius to consider while making clusters. That is, the entire range of distances is considered in the OPTICS algorithm, up to the epsilon value.
3. Minimum number of points in cluster
4. Steepness threshold classifies hierarchical clustering. Less the value higher the cluster classification within a cluster

There are other settings which would be helpful in labeling and coloring of the component of the chart like axis and plot background. Another thing to note is the availability of various options to visualize this chart, if I want to focus on certain clusters I can zoom in on those clusters. Also, I can download the image of the chart. It also provides tooltips on hover and highlighting a portion of the plot.

The visual is dependent on R packages like dbscan, plotly and ggplot2.