|  |  |
| --- | --- |
|  |  |

|  |
| --- |
| Clustering using OPTICS user guide |
|  |

# Clustering using OPTICS a brief introduction:

# Visual Description

# Use density-based clustering algorithm OPTICS to analyse groups and hierarchies within group in a dataset, identify outliers and interact with the plot by using hover tooltips.

# Overview

Analyse 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. You can also perform hierarchical clustering on your data to identify more closely-knit groups within larger groups. This visual also allows you to adjust the clustering parameters to control hierarchy depth and cluster sizes.

Key features include:

· Ability to identify outliers based on convex hull plots.

· Data is scaled and pre-processed automatically so no need to do so externally.

· Add your own dataset for tooltips

· Ability to identify sub clusters/hierarchal clusters in your data.

· Ability of hover tooltips and zoom effects.

· Capability to manually adjust the parameters of the clustering model.

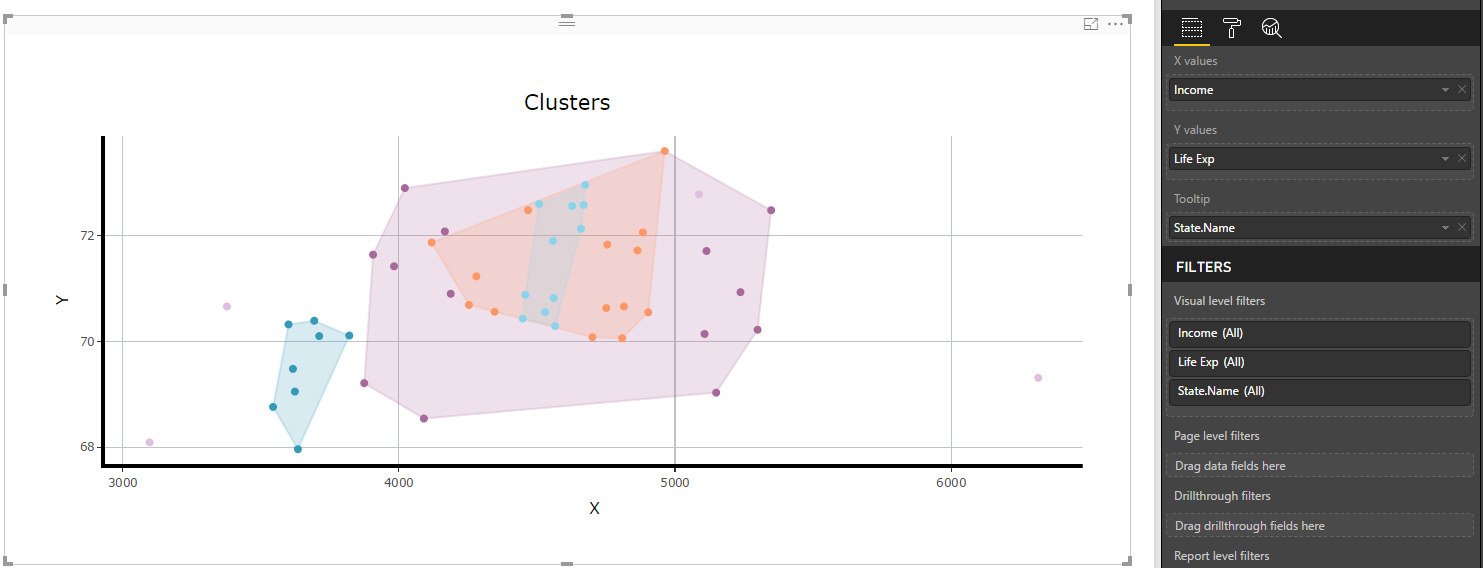
· No need to define the number of clusters.

# Example

Let’s assume as an insurance company, we need to identify the insurance needs of states in the US. To create products which specifically cater to a particular state, we need to know specific attributes of states like income, life expectancy etc. To get the clusters, we must simply supply the visual with two values –

* X value – In this case it is ‘Income’
* Y value – In this case it is ‘Life Expectancy’

Both the fields are essential to the visual’s functionality. We can also provide additional fields in Tooltip.



From this dataset, we get 4 different type of clusters and some outliers. So, we can create a plan accordingly. As visible there are some outliers in this visual so probably a separate plan might be needed for them.

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.

There are various settings which allows user to customize the visual -

1. ‘**Clustering settings**’ user can modify parameters provided to algorithm as per requirement. The parameters include –

* **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
* **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.
* **Minimum number of points in cluster**
* **Steepness threshold** classifies hierarchical clustering. Less the value higher the cluster classification within a cluster

1. **Plot settings:** The user can also control the appearance of the plot such as the color of plot background and title.
2. **X-axis settings:** The user can control the behavior of chart elements on x-axis such as x labels, grid lines for the x-axis, their colors etc. by using “X-axis settings”.
3. **Y-axis settings:** The user can control the behavior of chart elements on x-axis such as y labels, grid lines for the y-axis, their colors etc. by using “Y-axis settings”.

