

Module 5: Advanced Visual Analytics

Demo Document IV

edureka!

edureka!

© Brain4ce Education Solutions Pvt. Ltd.

Demo IV

Using the “Student Academic Data.xlsx”, create segments of Good, Average and Bad performance in “Math test” and “Writing test”.

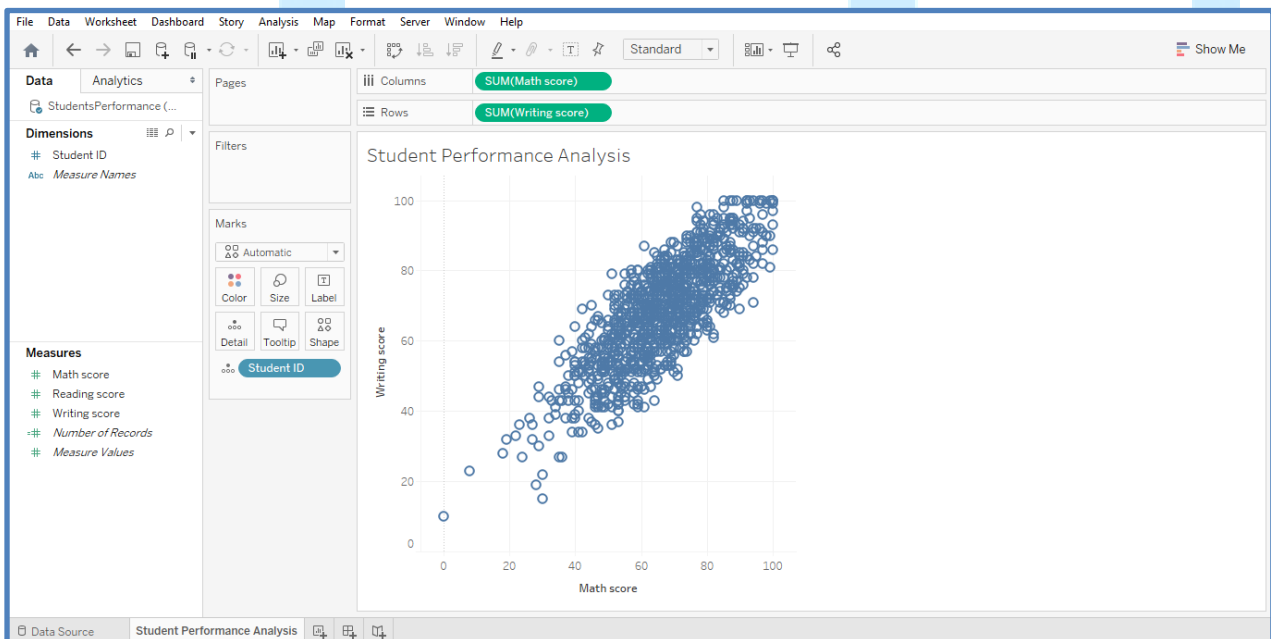
Demo IV – Solution

Clustering in Tableau helps you create statistically based segments which provides insights on how different groups are similar as well as how they perform when compared to each other. Tableau uses the k-means clustering method to create a cluster and Lloyd’s algorithm with squared Euclidean distances are used to compute the k-means clustering for each k.

Let us get started with the Student Performance report:

Step 1: Click on Connect → Microsoft Excel → Student Academic Data.xlsx
→ Sheet1 (Rename it to Student Performance Analysis)

Step 2: Data pane → Measures → Drag Maths score and Writing Score to Column and Row shelf respectively.



Step 3: To create cluster:

Data pane → Analytics → Models → Clusters → Drag and drop Cluster on view

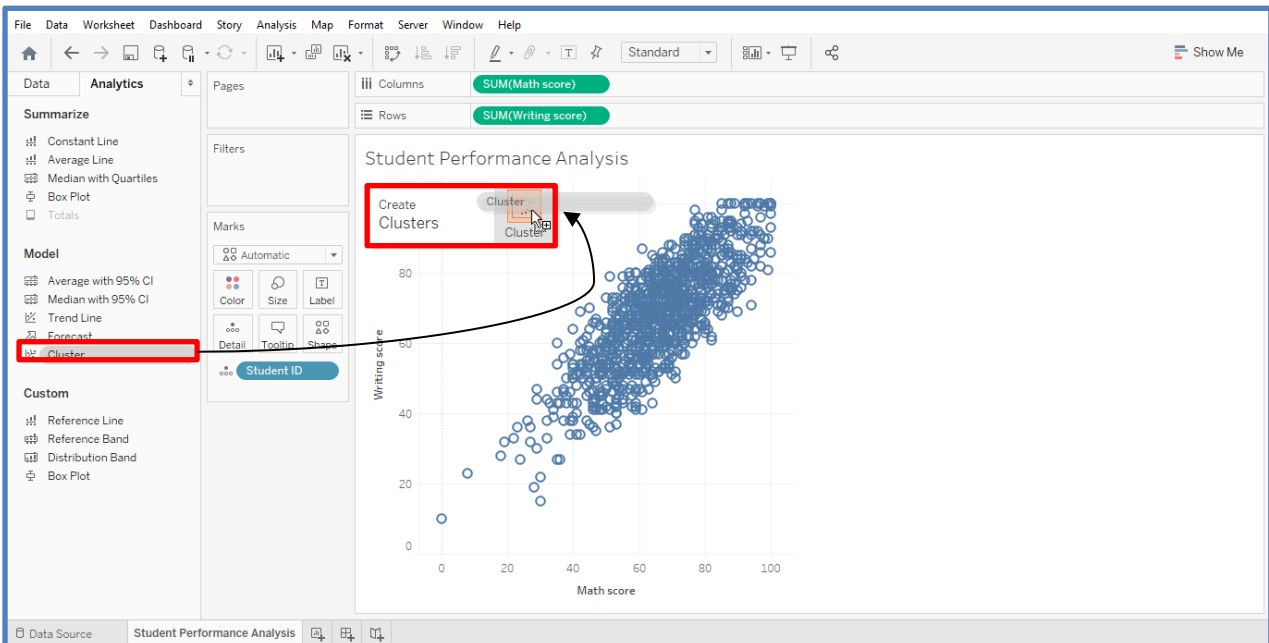


Tableau creates cluster groups on colors (marks card), and colors the marks in our groups by cluster. Tableau assigns each value in the view to one of the clusters and if some values do not fit well in clusters, they are assigned to “Not Clustered” cluster. In this view, red, blue and orange clusters represent high, average and low performance of students in Maths and Writing tests.

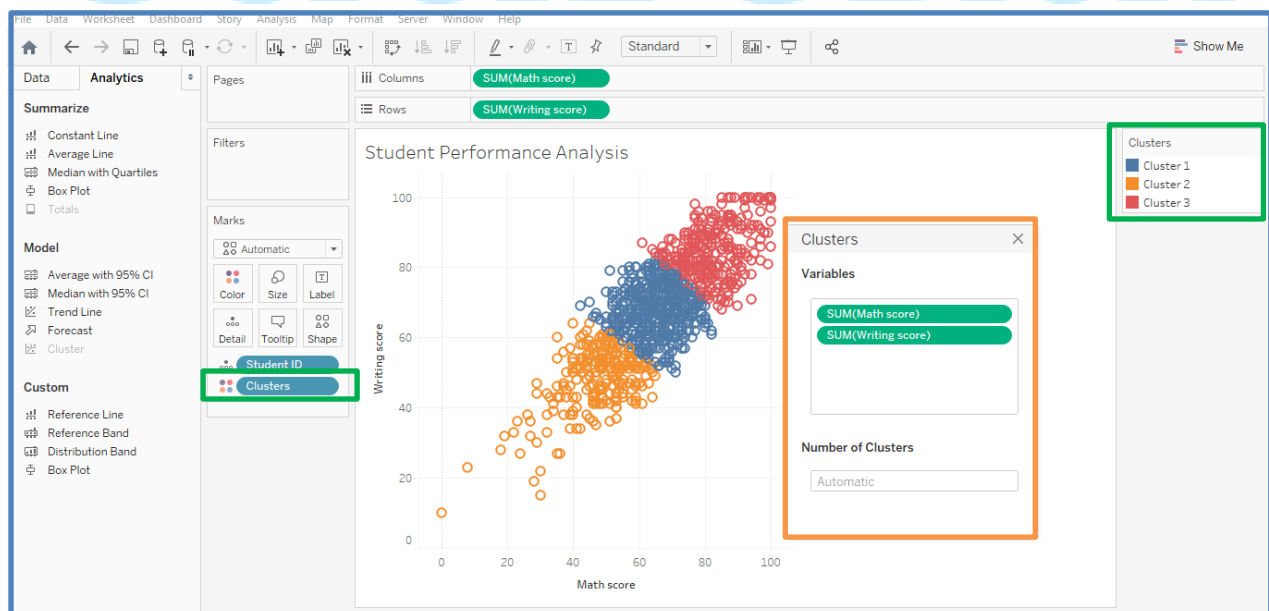
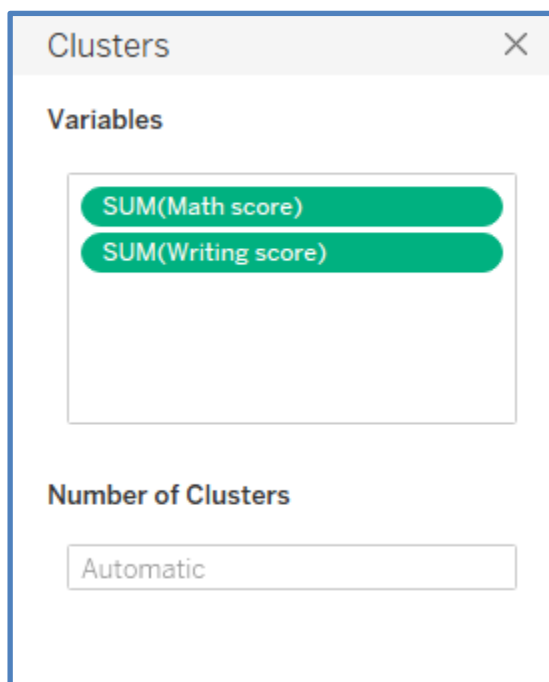


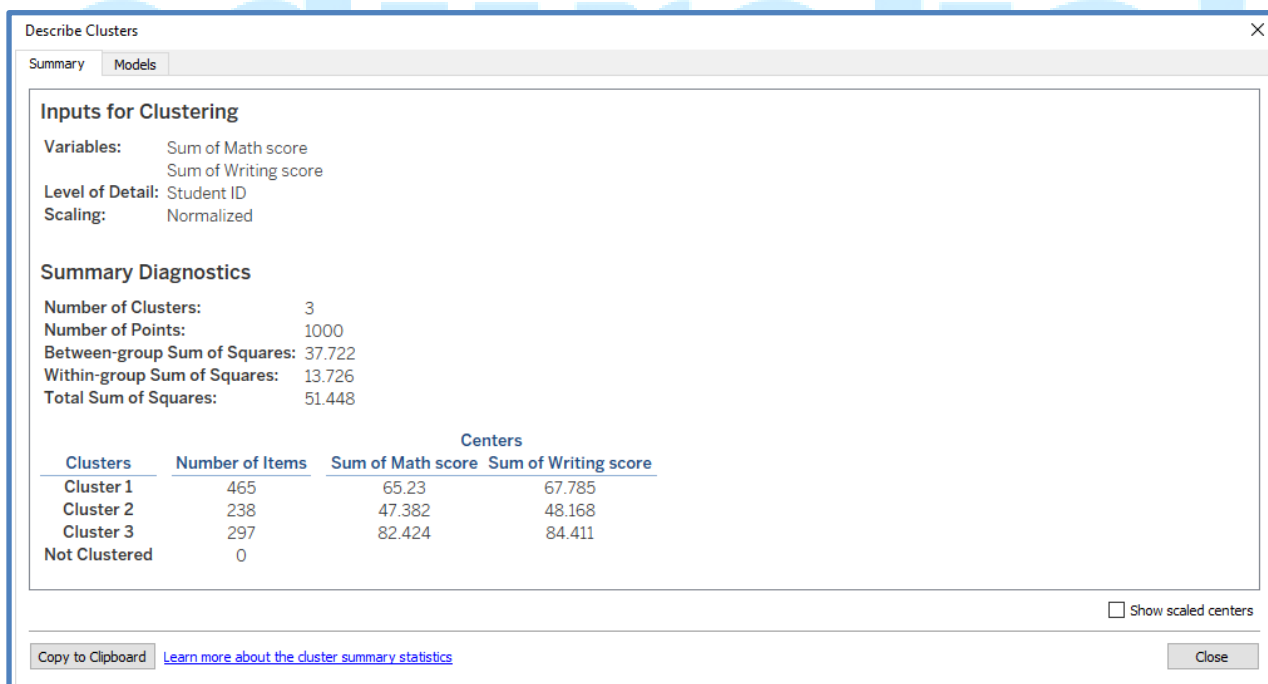
Tableau displays the Clusters dialog box, which can be customised adding variables from the data pane to the cluster dialog box in variables section.

You can assign the numbers of clusters or just choose the automatic option.



The 'Clusters' dialog box has a title bar with a close button. It contains two sections: 'Variables' and 'Number of Clusters'. The 'Variables' section has a list box containing 'SUM(Math score)' and 'SUM(Writing score)'. The 'Number of Clusters' section has a text box containing the word 'Automatic'.

Step 4: To view the clustered data in detail:
Marks Card → Cluster → Drop down → Describe



The 'Describe Clusters' dialog box has a title bar with a close button. It has two tabs: 'Summary' (selected) and 'Models'. The 'Summary' tab contains the following information:

Inputs for Clustering

Variables: Sum of Math score
Sum of Writing score
Level of Detail: Student ID
Scaling: Normalized

Summary Diagnostics

Number of Clusters: 3
Number of Points: 1000
Between-group Sum of Squares: 37.722
Within-group Sum of Squares: 13.726
Total Sum of Squares: 51.448

Clusters	Number of Items	Centers	
		Sum of Math score	Sum of Writing score
Cluster 1	465	65.23	67.785
Cluster 2	238	47.382	48.168
Cluster 3	297	82.424	84.411
Not Clustered	0		

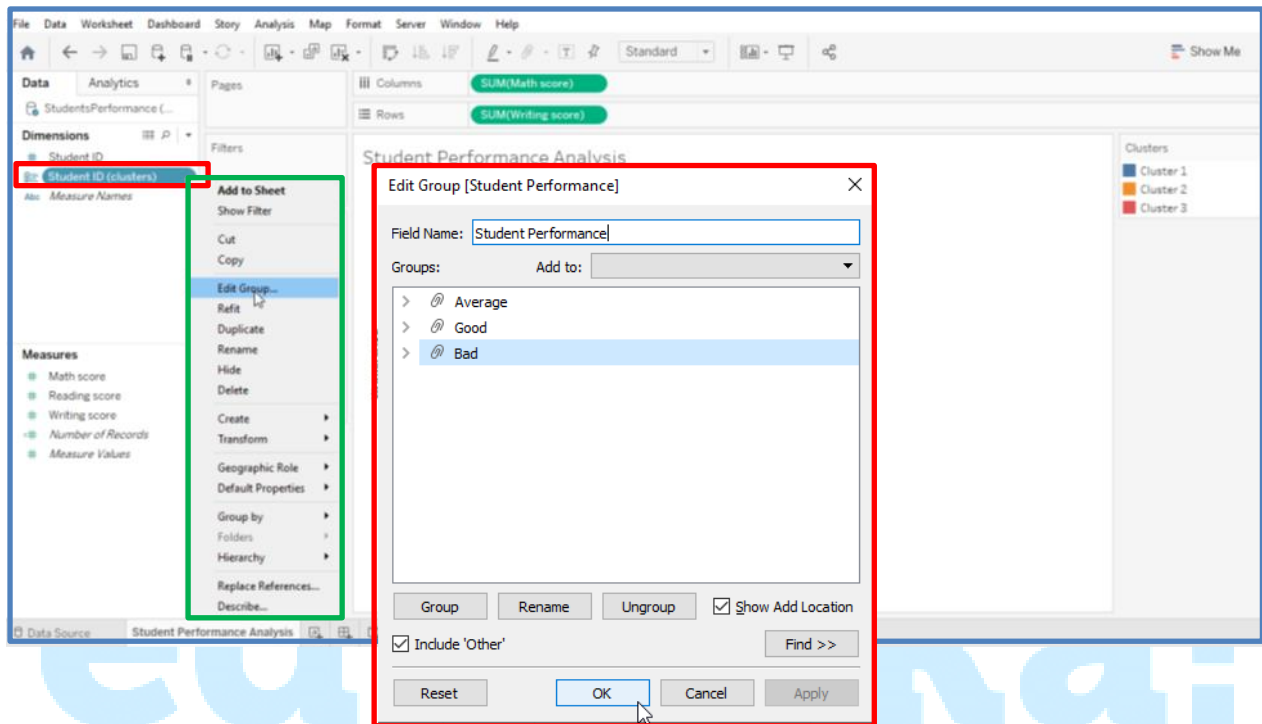
At the bottom right, there is a checkbox labeled 'Show scaled centers' which is unchecked. At the bottom left, there is a 'Copy to Clipboard' button and a link 'Learn more about the cluster summary statistics'. At the bottom right, there is a 'Close' button.

You can copy this data to clipboard and use it for further in-depth analysis.

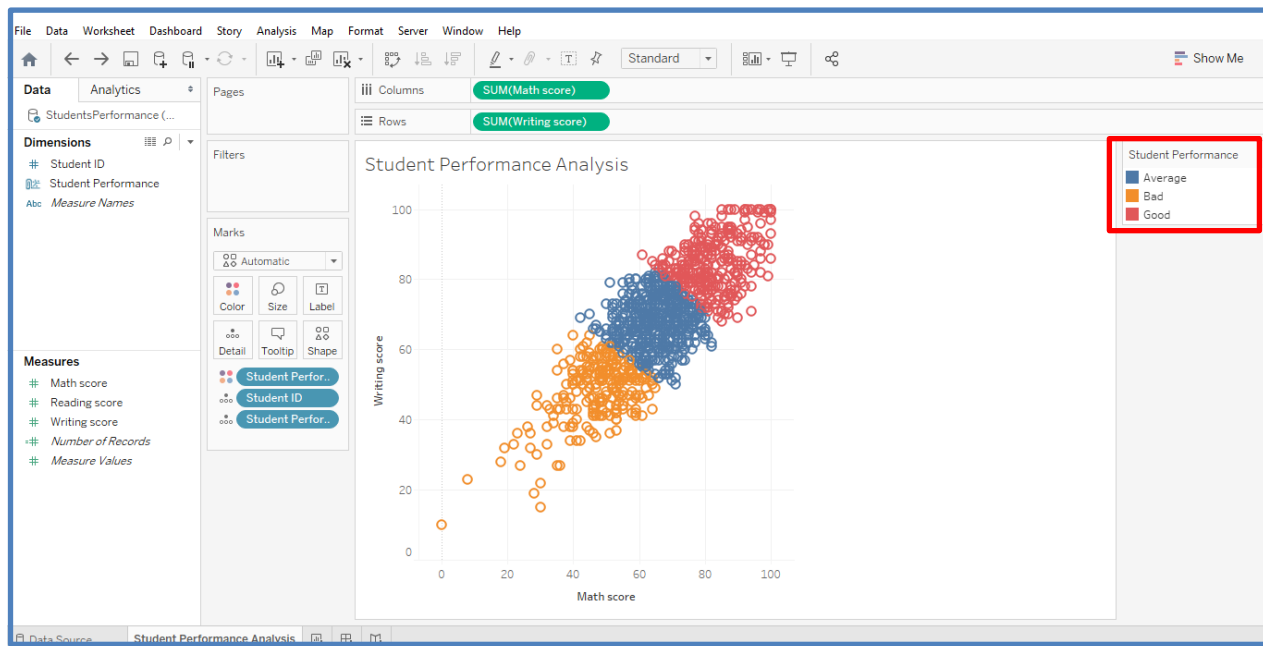
Step 5: Let us make our visualization more interactive by renaming the clusters:

Marks card → Drag "Cluster" to Data pane → Student ID cluster → Drop down
Drop down → Edit group → Edit Group dialog box → Change:

- Field Name: Student Performance
- Groups: Cluster 1 → Rename → Average
Cluster 2 → Rename → Bad
Cluster 3 → Rename → Good
- Apply → OK



Here, we have Students performance in Maths and Writing tests categorized into Good, Average and Bad clusters.



edureka!