

Name: Alpanshu Kataria

Reg. No.: 18BCE1267

Faculty: Dr. Maheshwari R

Lab 11

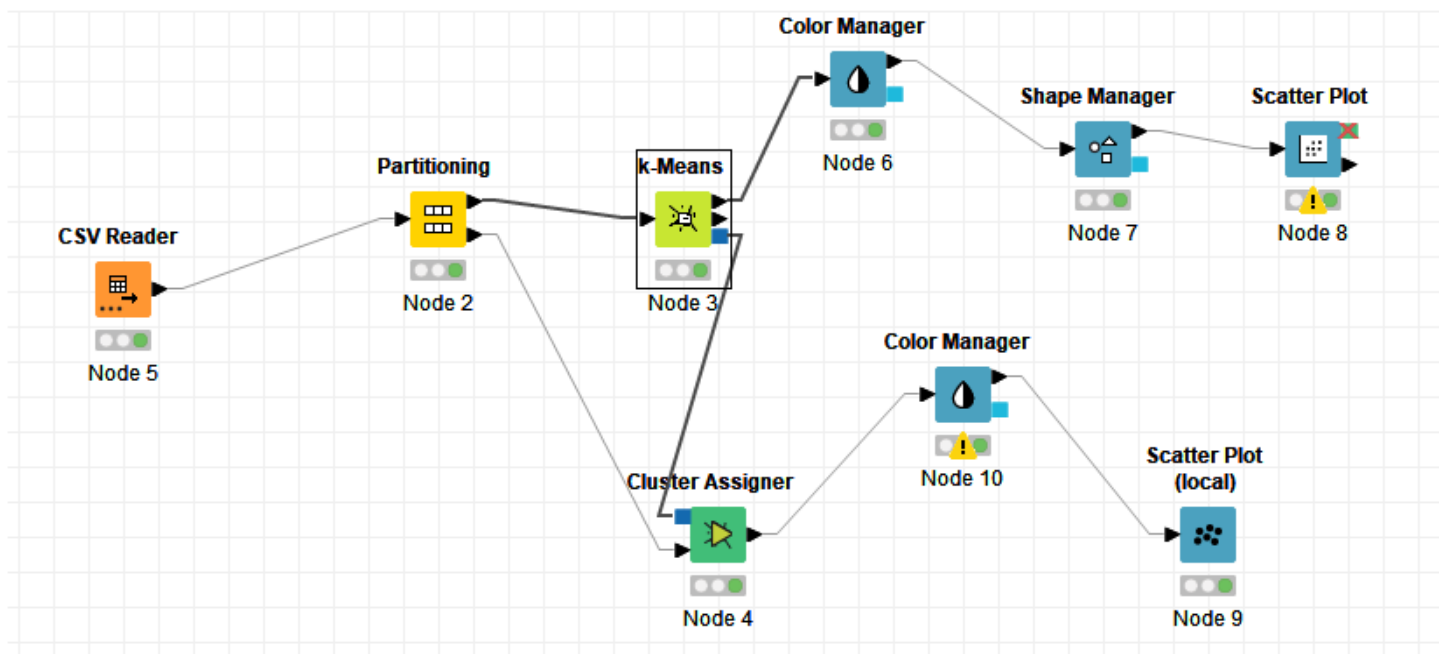
Task:

Machine Learning (Unsupervised Learning) in Knime analytics platform for applying clustering on corona virus related dataset

Dataset:

https://www.kaggle.com/sudalairajkumar/covid19-in-india?select=covid_19_india.csv

Workspace:



All nodes and their configuration:

CSV reader


Dialog - 3:5 - CSV Reader

File

Settings Transformation Advanced Settings Limit Rows Encoding Flow Variables Memory Policy

Input location


Read from Custom/KNIME URL Timeout 1,000

Mode ☒ File ☐ Files in folder  Custom/KNIME URL does not support listing/browsing of files.

URL file:/E:/VIT/Sem%206/IoT/Lab/Lab%2011/covid_19_india.csv Browse...

Reader options

Format

Autodetect format 


Column delimiter , Row delimiter
 Quote char " Quote escape char \"

Comment char #

☒ Has column header ☐ Has row ID

☐ Support short data rows ☐ Prepend file index to row ID

Preview

 The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.

Row ID	I IdSno	S Date	S Time	S State/U...	S Confirm...	S Confirm...	I Cured	I Deaths	I Confirmed
Row0	1	2020-01-30	6:00 PM	Kerala	1	0	0	0	1
Row1	2	2020-01-31	6:00 PM	Kerala	1	0	0	0	1
Row2	3	2020-02-01	6:00 PM	Kerala	2	0	0	0	2
Row3	4	2020-02-02	6:00 PM	Kerala	3	0	0	0	3
Row4	5	2020-02-03	6:00 PM	Kerala	4	0	0	0	4

Partitioning Node:

File

First partition

Flow Variables

Memory Policy

Choose size of first partition

☐ Absolute

100

☒ Relative[%]

80

☐ Take from top

☐ Linear sampling

☒ Draw randomly

☐ Stratified sampling

S

ConfirmedForeignNational

▼

☐ Use random seed

1,621,966,417,2

OK

Apply

Cancel

?

K-means Node:

Dialog - 3:3 - k-Means

File

K-Means Properties

Flow Variables

Memory Policy

Clusters

Number of clusters: 36

Centroid initialization:
☐ First k rows
☒ Random initialization ☒ Use static random seed 0

New

Number of Iterations
Max. number of iterations: 99

Column Selection

Exclude

Filter

i>zSno

Cured

Deaths

>

>>

<

<<

Include

Filter

Confirmed

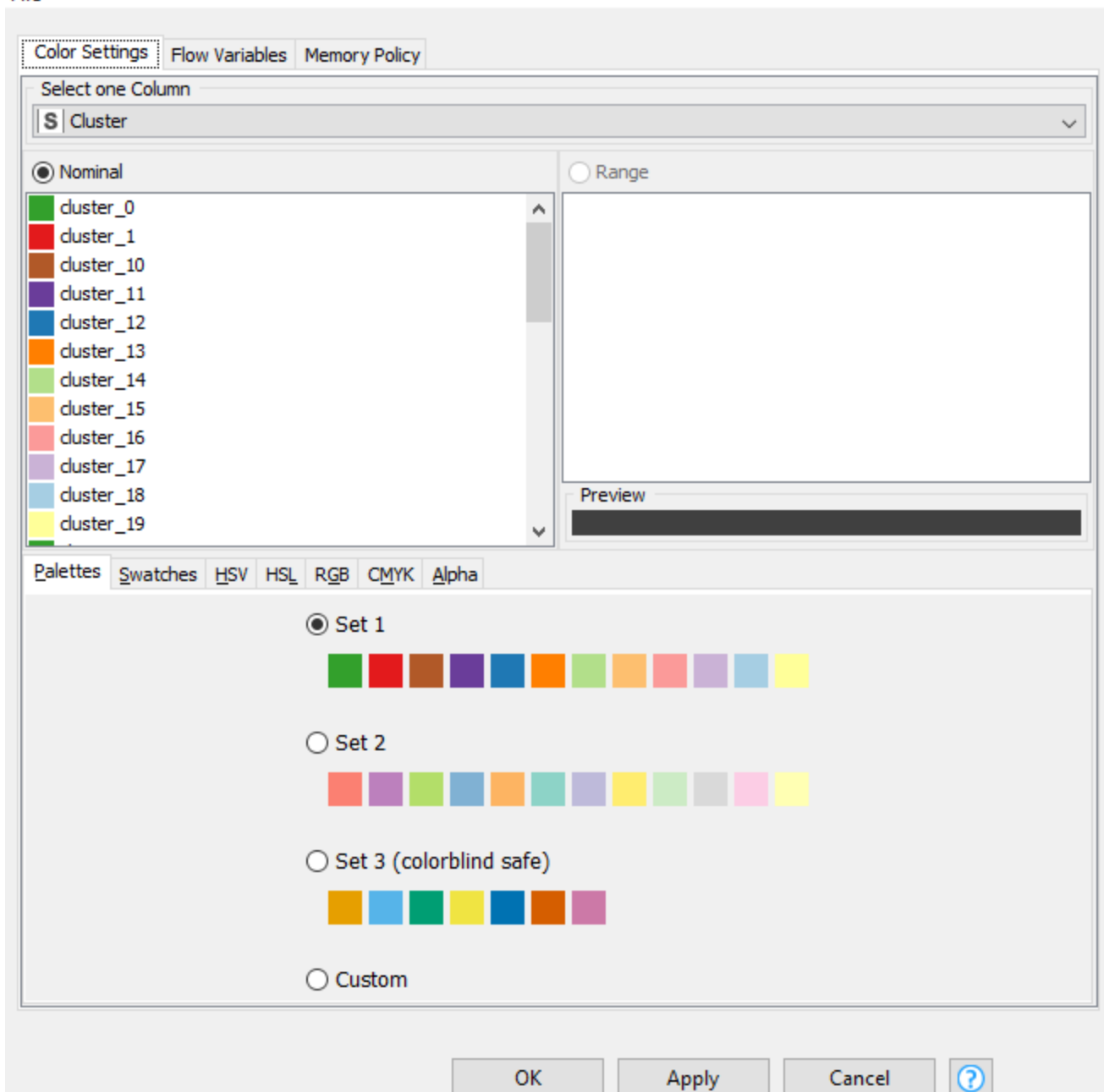
☐ Always include all columns

Hilite Mapping

☐ Enable Hilite Mapping

Color Manager Node:

File



Shape Manager Node:

Dialog - 3:7 - Shape Manager

File

Shape Settings | Flow Variables | Memory Policy

Select nominal column:
S State/UnionTerritory

Shape Mapping

Values of State/UnionTerritory	Shapes
Kerala	Rectangle
Telangana	Circle
Delhi	Triangle
Rajasthan	Reverse Triangle
Uttar Pradesh	Diamond
Haryana	Asterisk
Ladakh	Cross
Tamil Nadu	X-Shape
Karnataka	Horizontal Line
Maharashtra	Vertical Line
Punjab	Rectangle
Jammu and Kashmir	Circle
Andhra Pradesh	Triangle
Uttarakhand	Reverse Triangle
Odisha	Diamond
Puducherry	Asterisk
West Bengal	Cross
Chhattisgarh	X-Shape
Chandigarh	Horizontal Line
Gujarat	Vertical Line
Himachal Pradesh	Rectangle
Madhya Pradesh	Circle

OK Apply Cancel ?

Scatter Plot Node:

Dialog - 3:8 - Scatter Plot

File

Options Axis Configuration General Plot Options View Controls Flow Variables Memory Policy

☐ Create image at output

Maximum number of rows: 2,500

Selection column name: Selected (Scatter Plot)

Choose column for x axis

I » Sno

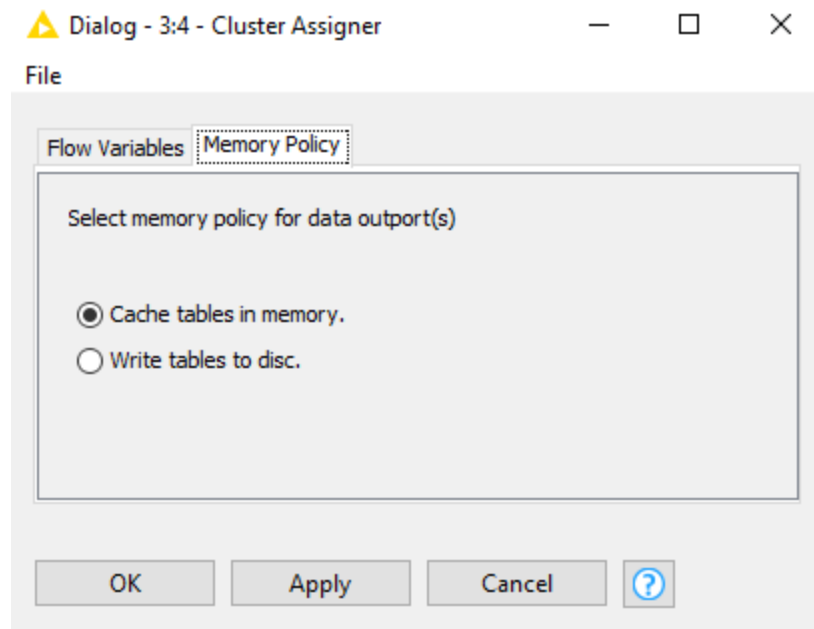
Choose column for y axis

S Date

☒ Report on missing values

OK Apply Cancel ?

Cluster Assigner Node:



Scatter Plot Output:

