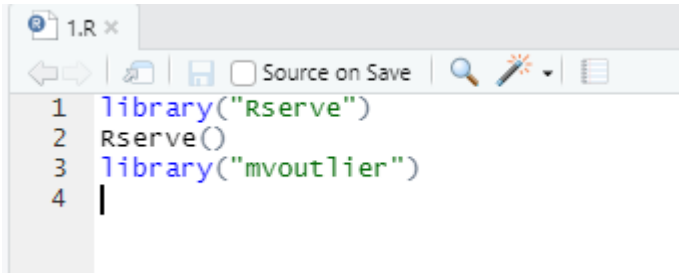
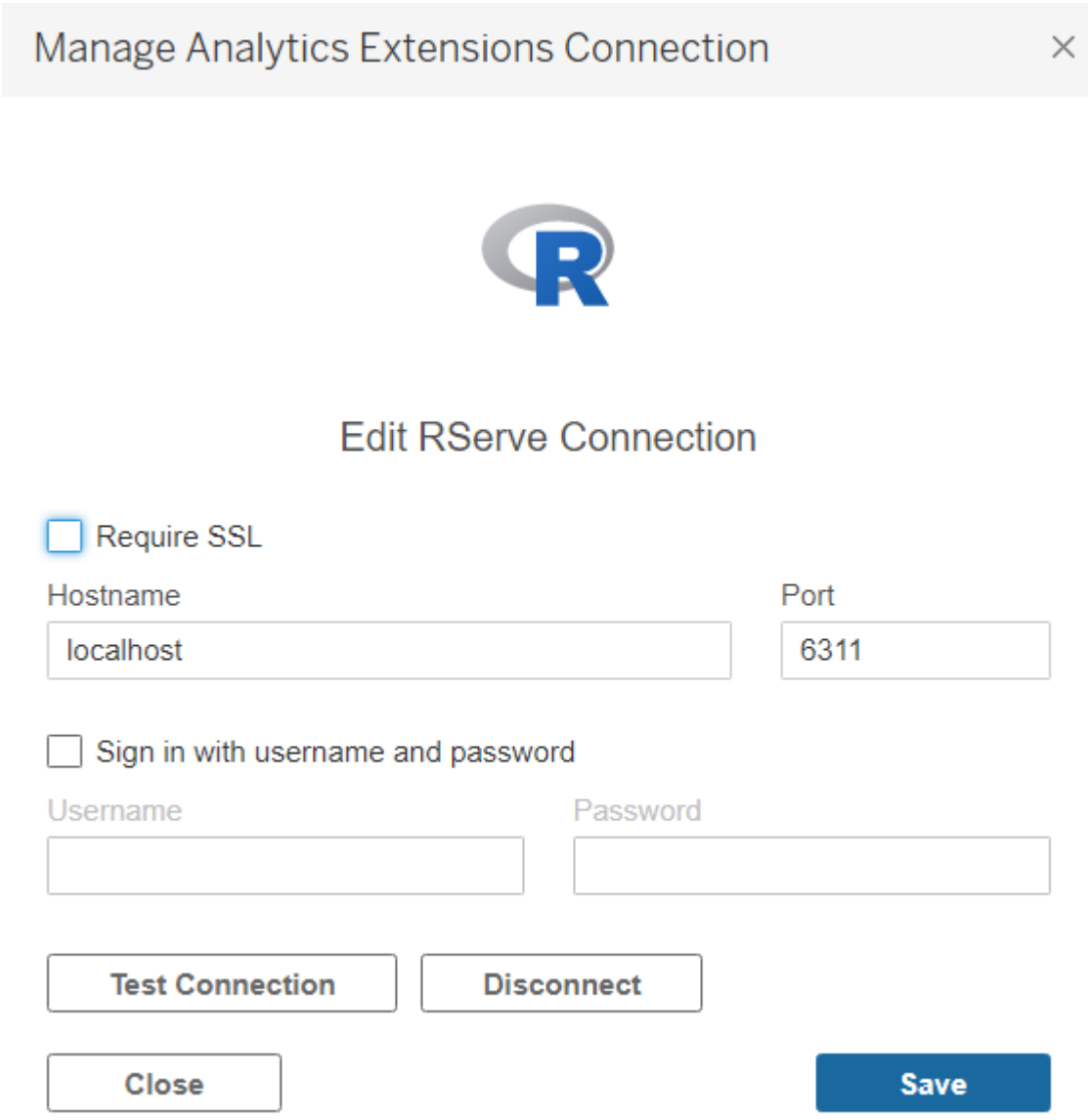
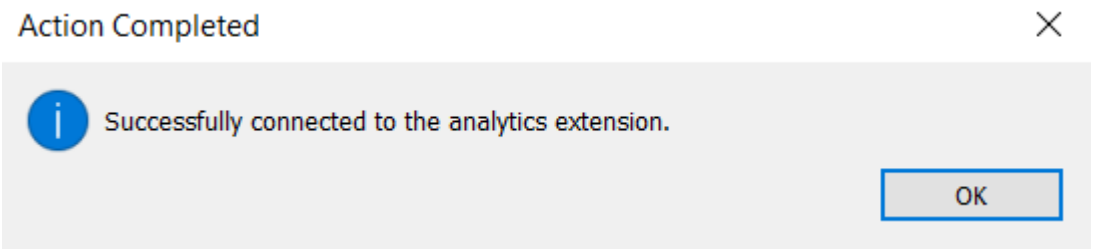
**Aryaman Mishra**

**19BCE1027**

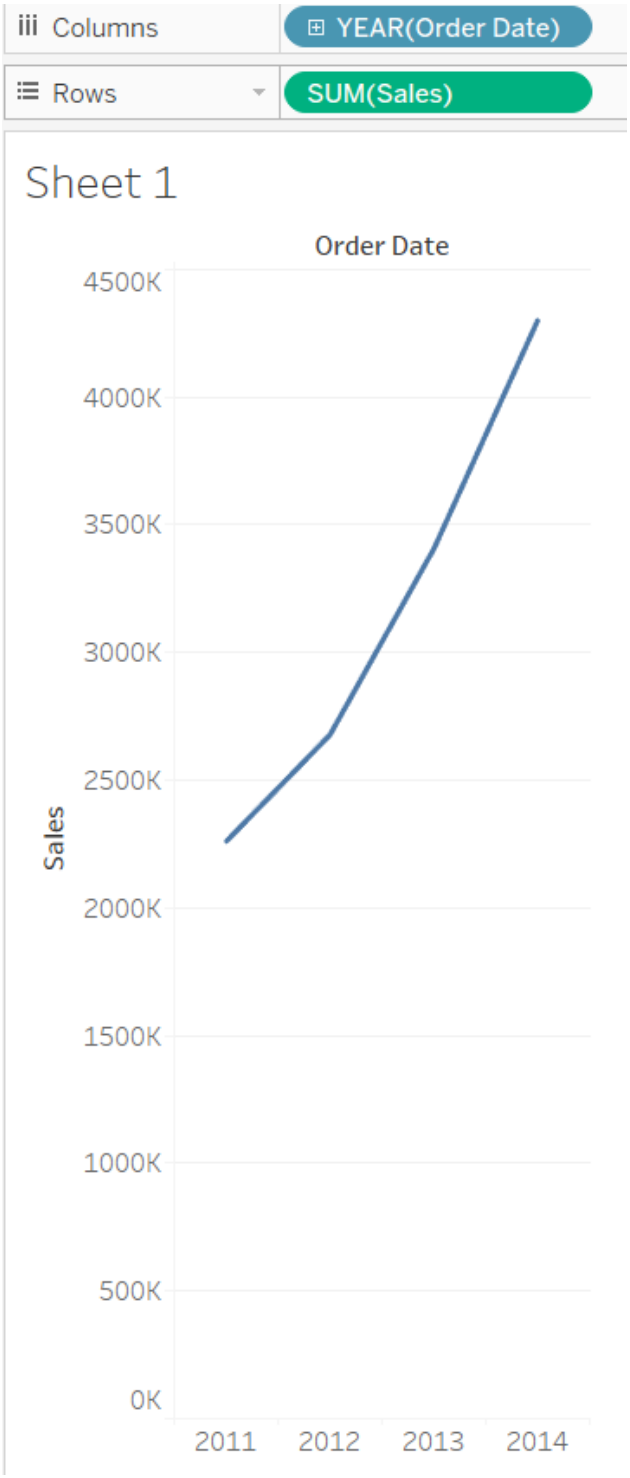
Find the outliers in superstore dataset using R in Tableau and write the steps for the following: o Connect to Rserve from Tableau

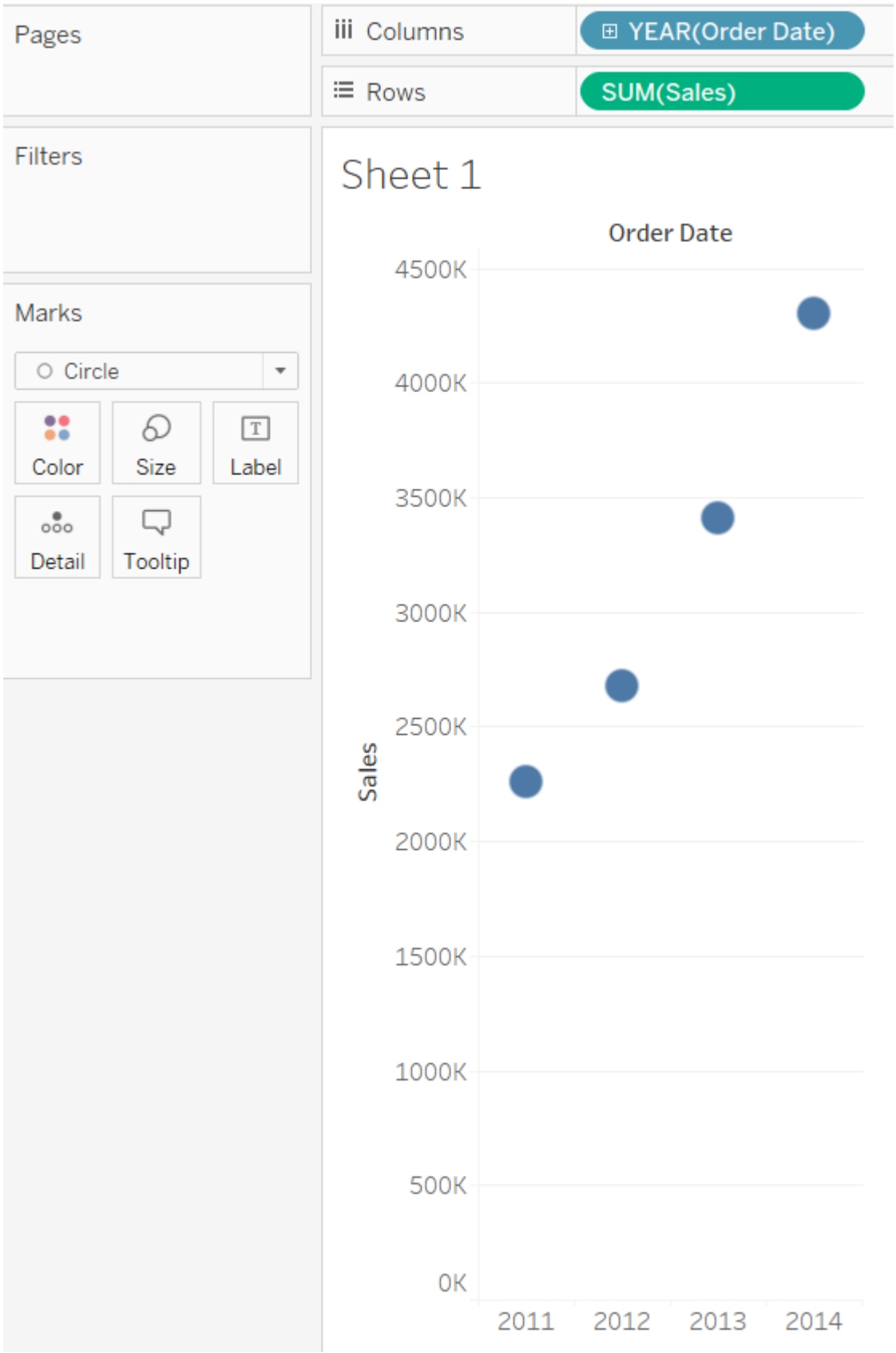


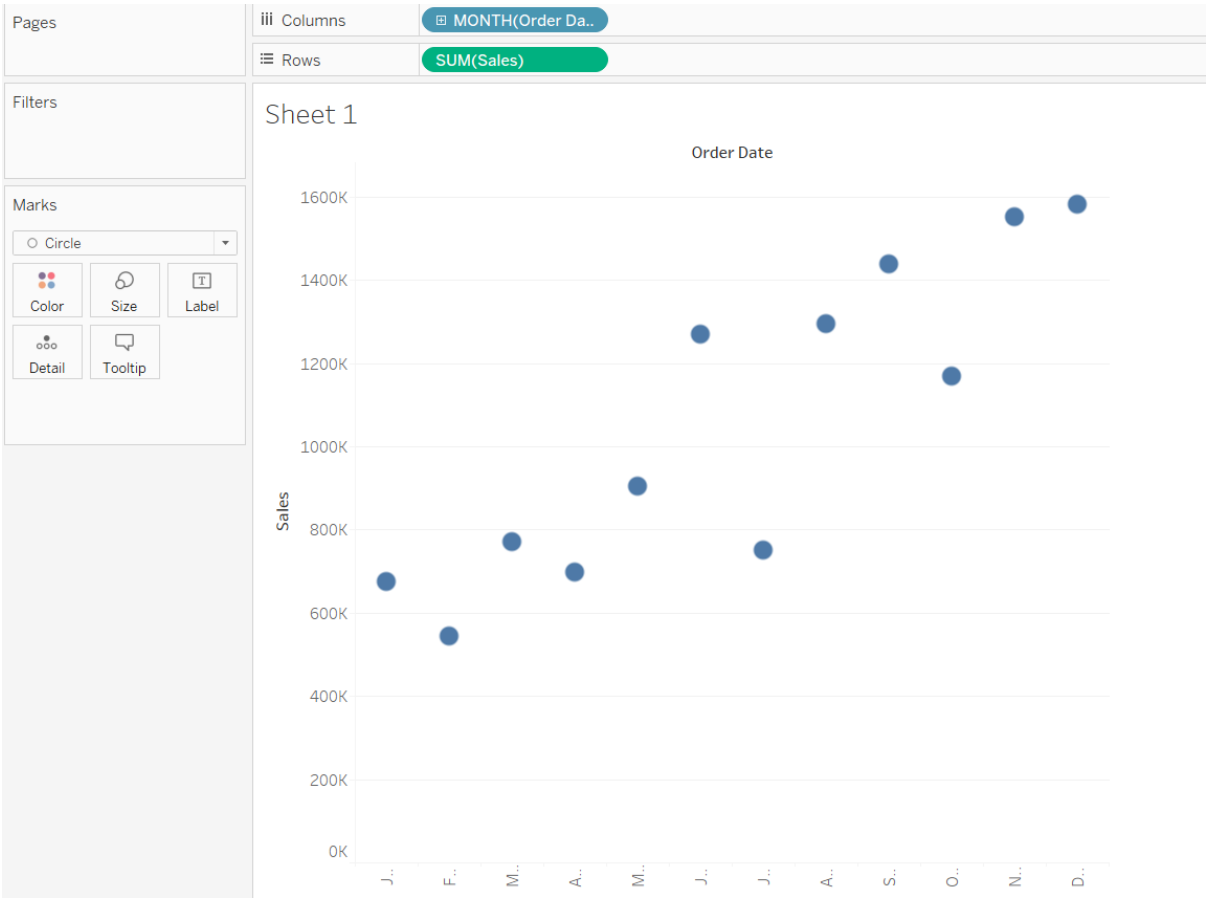





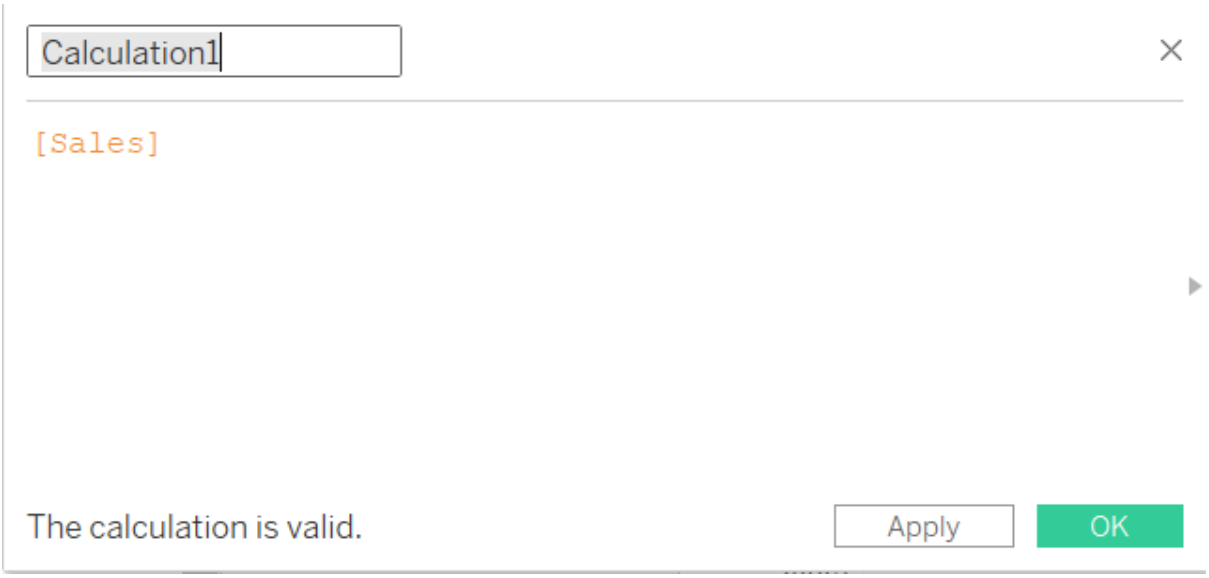

o Create a chart as per requirement in Tableau

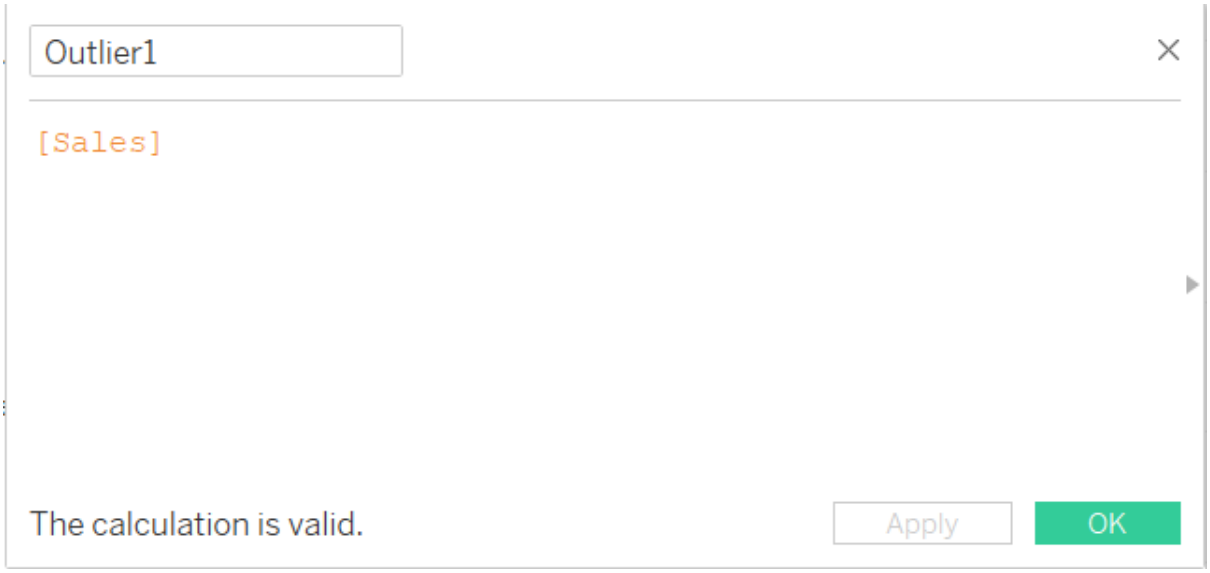



Pages
Columns
YEAR(Order Date)
Rows
SUM(Sales)
Filters
Marks
Circle
Color
Size
Label
Detail
Tooltip
Sheet 1
Order Date
4500K
4000K
3500K
3000K
2500K
2000K
1500K
1000K
500K
0K
Sales
2011
2012
2013
2014

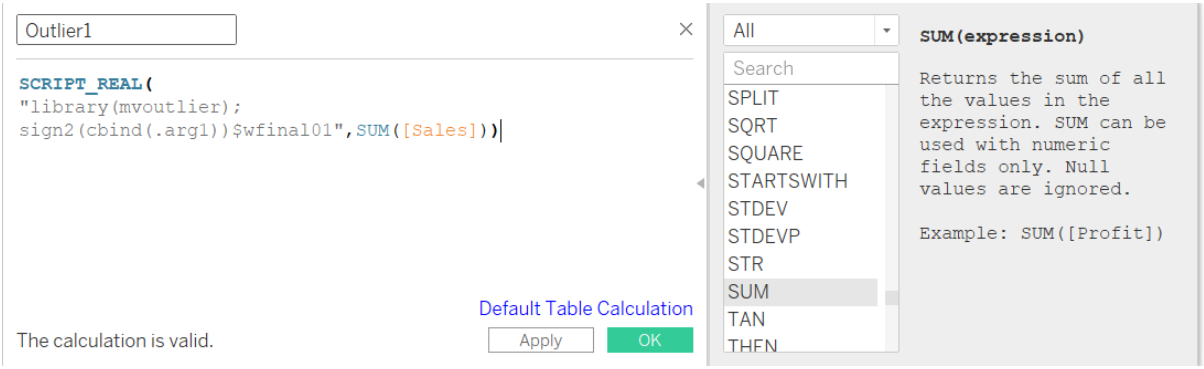



o Create a calculative field

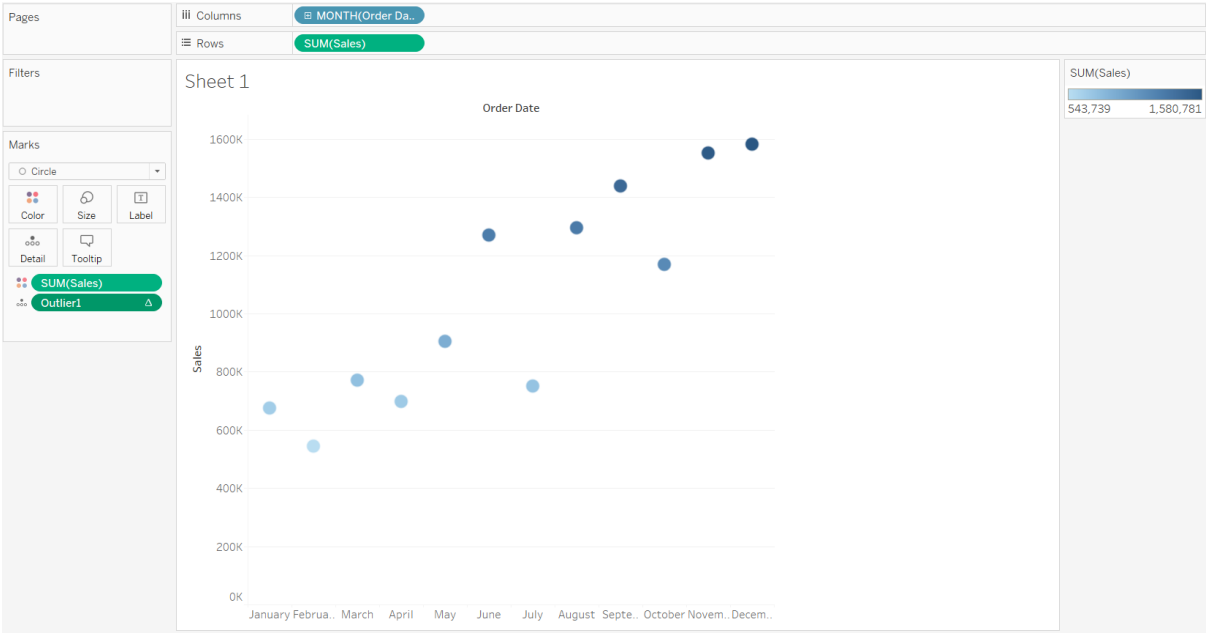





o Explanation to the calculative field




o Drop the field in color shelf to see the outliers

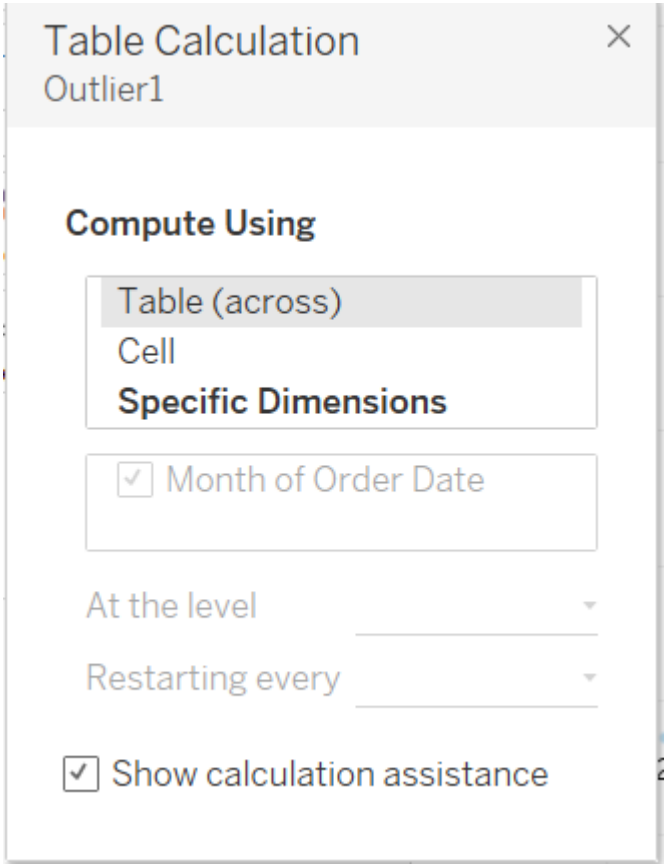



o Inference

Extreme values are aggregated marks that are outliers, based on a model of the visualized marks. The selected mark is considered to contain an extreme value if a record value is in the tails of the distribution of the expected values for the data.

An extreme value is determined by comparing the aggregate mark with and without the extreme value. If the mark becomes less surprising by removing a value, then it receives a higher score.

When a mark has extreme values, it doesn't automatically mean it has outliers, or that you should exclude those records from the view. That choice is up to you depending on your analysis. The explanation is simply pointing out an interesting extreme value in the mark. For example, it could reveal a mistyped value in a record where a banana cost 10 dollars instead of 10 cents. Or, it could reveal that a particular sales person had a great quarter.



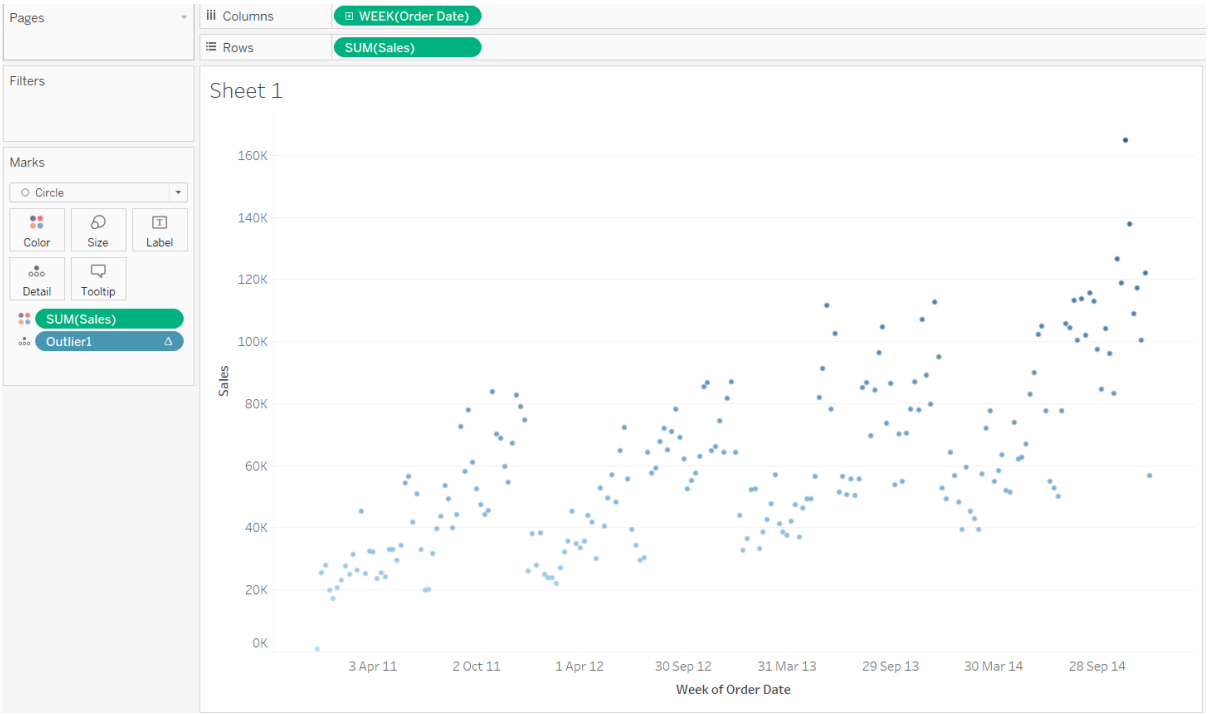

Outlier1

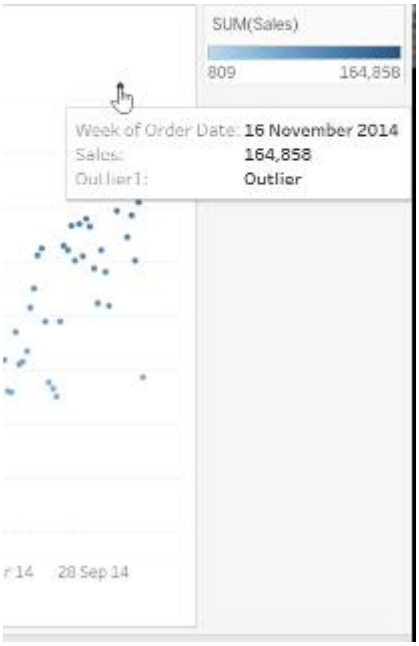
Results are computed along Table (across).

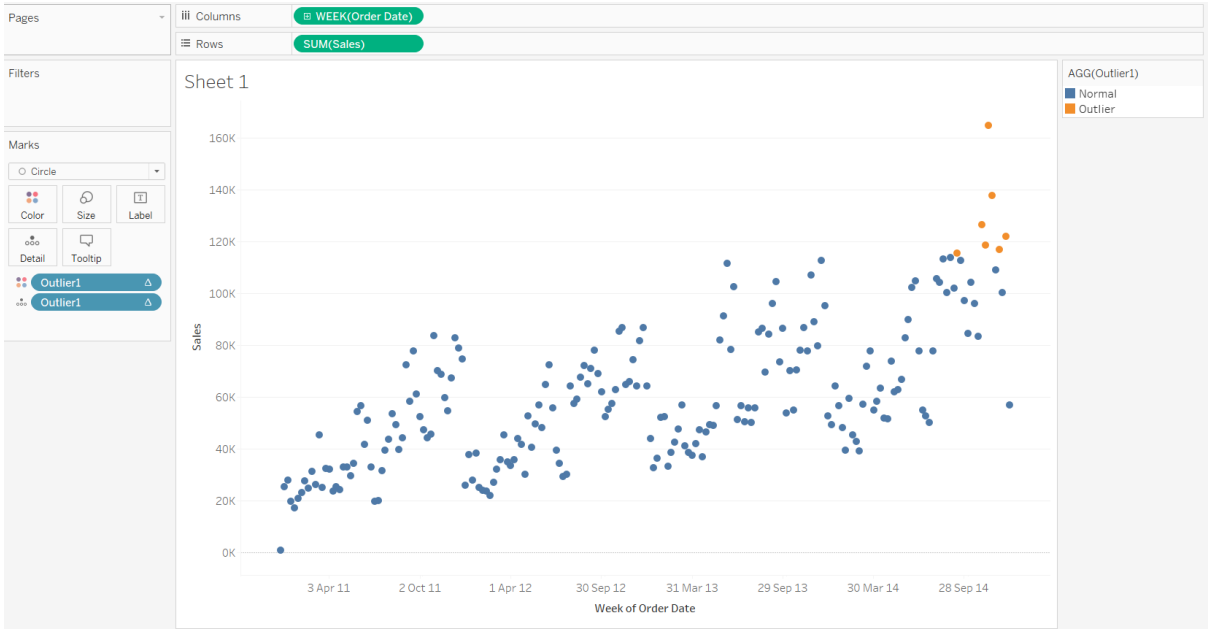
```
if SCRIPT_REAL(
              "library(mvoutlier);
              sign2(cbind(.arg1))$wfinal01",
              sum([Sales]))==0
then "Outlier"
else "Normal"
END
```

Default Table Calculation

The calculation is valid. 1 Dependency Apply OK


Pages
Columns
WEEK(Order Date)
Rows
SUM(Sales)
Filters
Marks
Circle
Color
Size
Label
Detail
Tooltip
SUM(Sales)
Outlier1
Sheet 1
160K
140K
120K
100K
80K
60K
40K
20K
0K
Sales
3 Apr 11
2 Oct 11
1 Apr 12
30 Sep 12
31 Mar 13
29 Sep 13
30 Mar 14
28 Sep 14
Week of Order Date

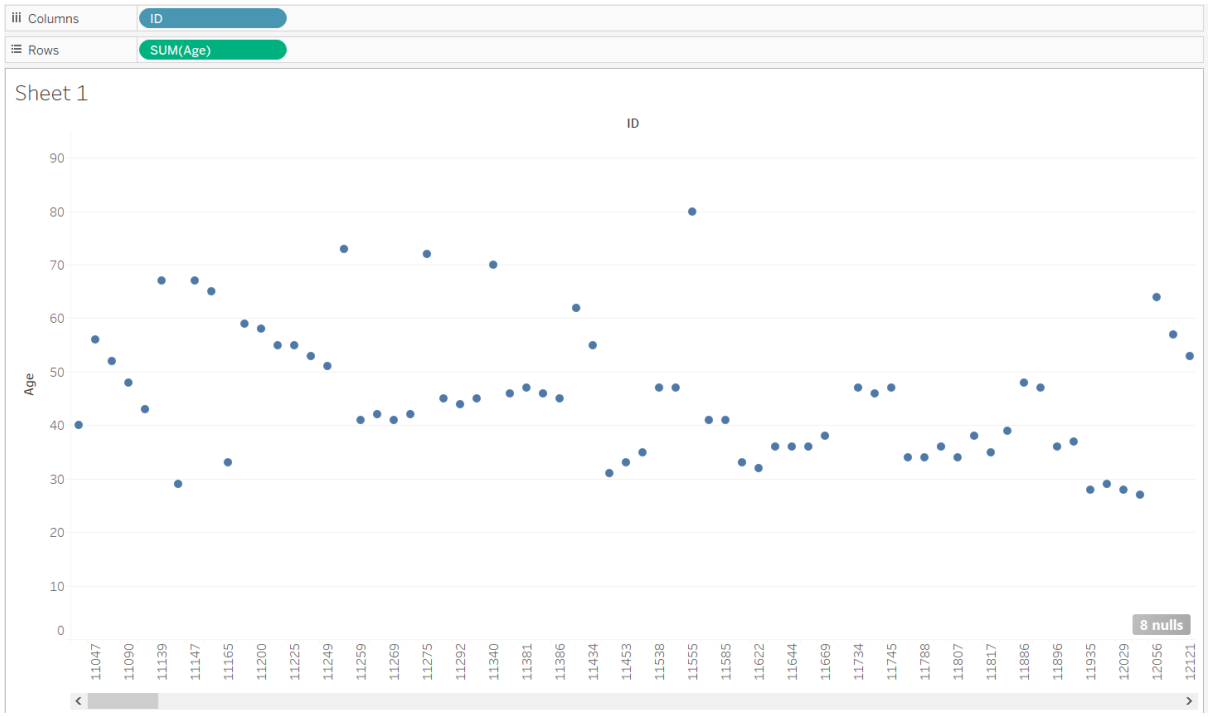

SUM(Sales)
809
164,858
Week of Order Date: 16 November 2014
Sales: 164,858
Outlier1: Outlier
28 Sep 14

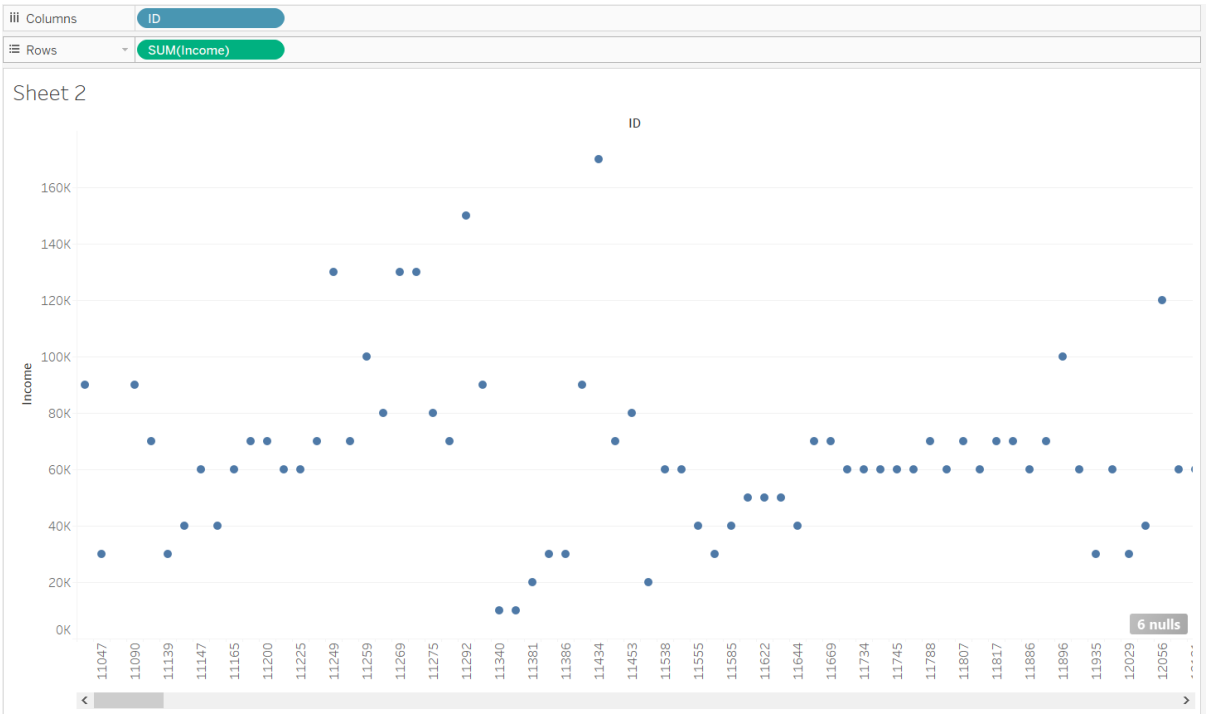


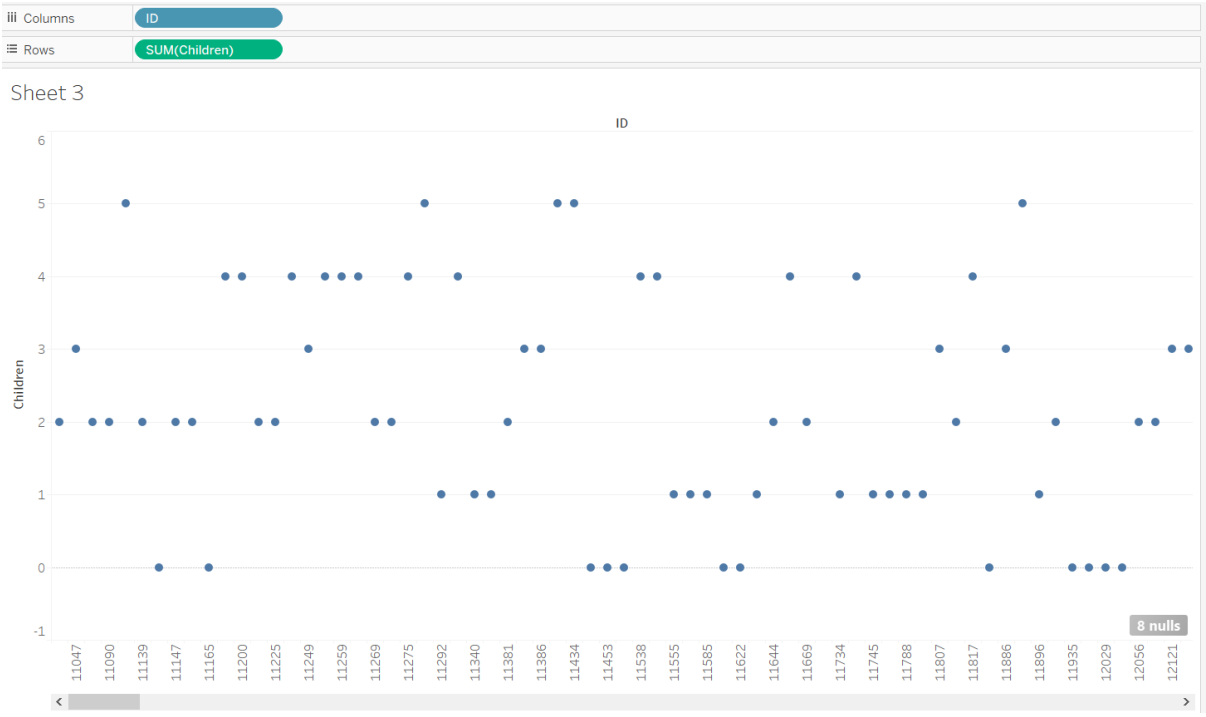

• Formulate the clusters based on age, income, children and cars from Bike buyers dataset using R in Tableau and write the steps for the following:

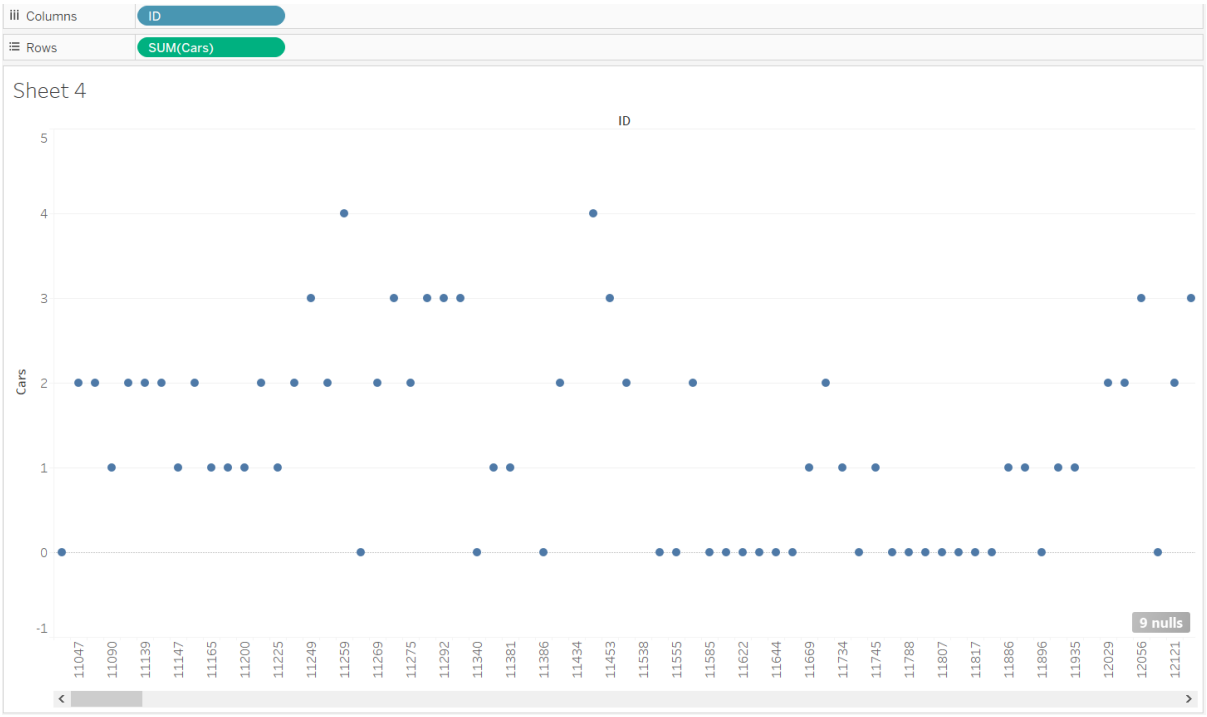
o Connect to Rserve from Tableau

o Create a chart as per requirement in Tableau




Columns
ID
Rows
SUM(Income)
Sheet 2
ID
Income
160K
140K
120K
100K
80K
60K
40K
20K
0K
6 nulls
11047
11090
11139
11147
11165
11200
11225
11249
11259
11269
11275
11292
11340
11381
11386
11434
11453
11538
11555
11585
11622
11644
11669
11734
11745
11788
11807
11817
11886
11896
11935
12029
12056

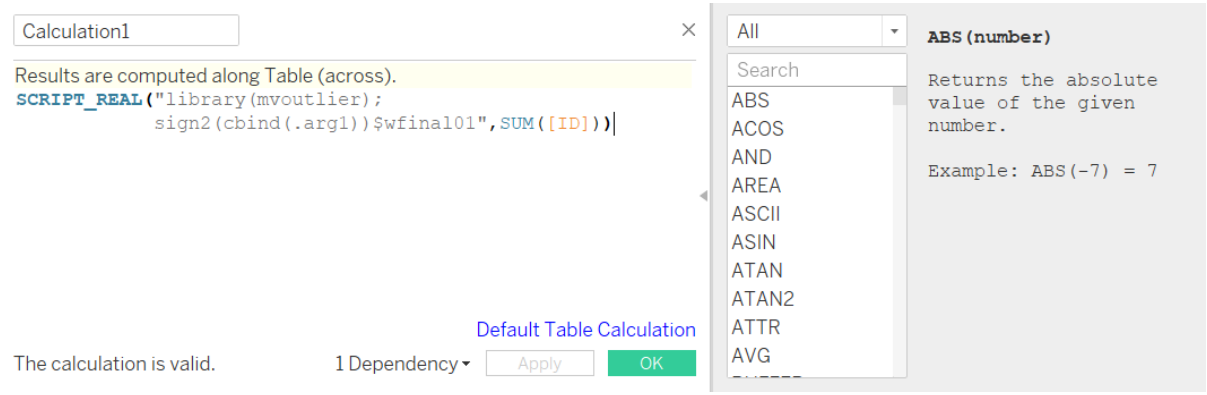

Columns
ID
Rows
SUM(Children)
Sheet 3
ID
Children
6
5
4
3
2
1
0
-1
8 nulls
11047
11090
11139
11147
11165
11200
11225
11249
11259
11269
11275
11292
11340
11381
11386
11434
11453
11538
11555
11585
11622
11644
11669
11734
11745
11788
11807
11817
11886
11896
11935
12029
12056
12121



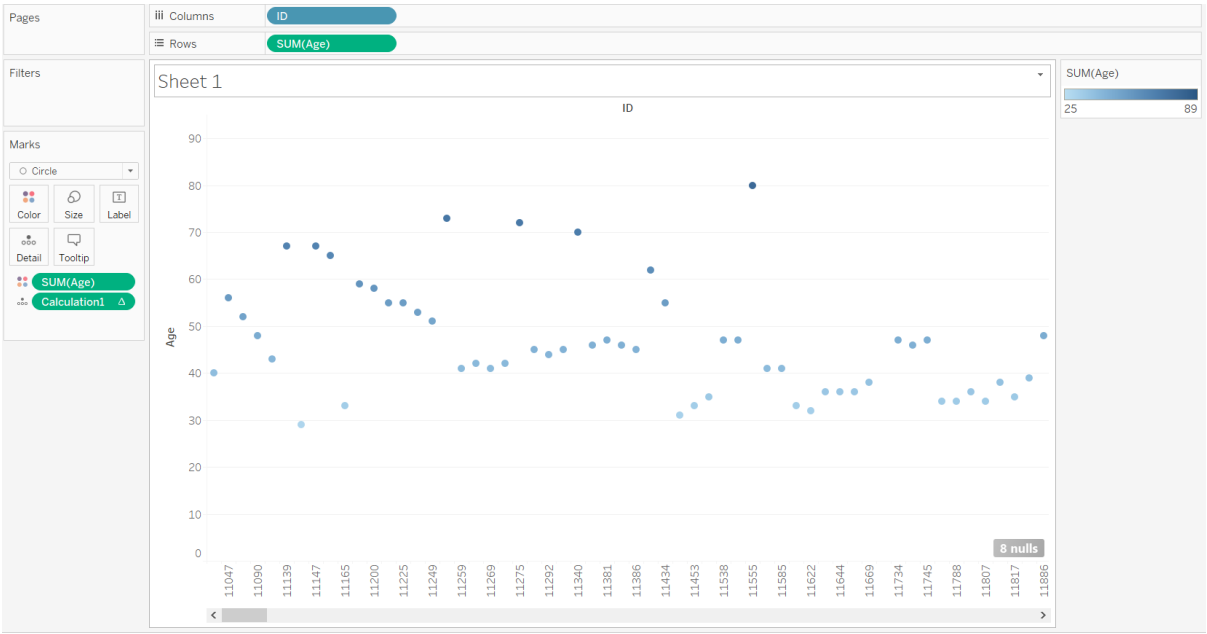

o Create a calculative field

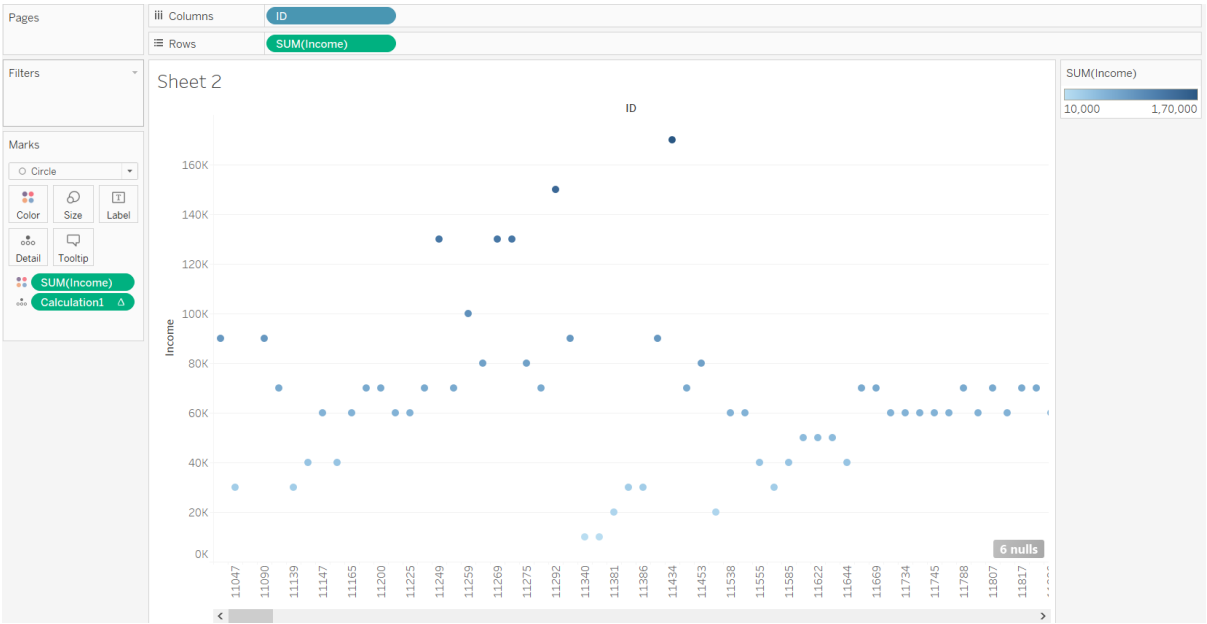
SCRIPT_REAL("library(mvoutlier);

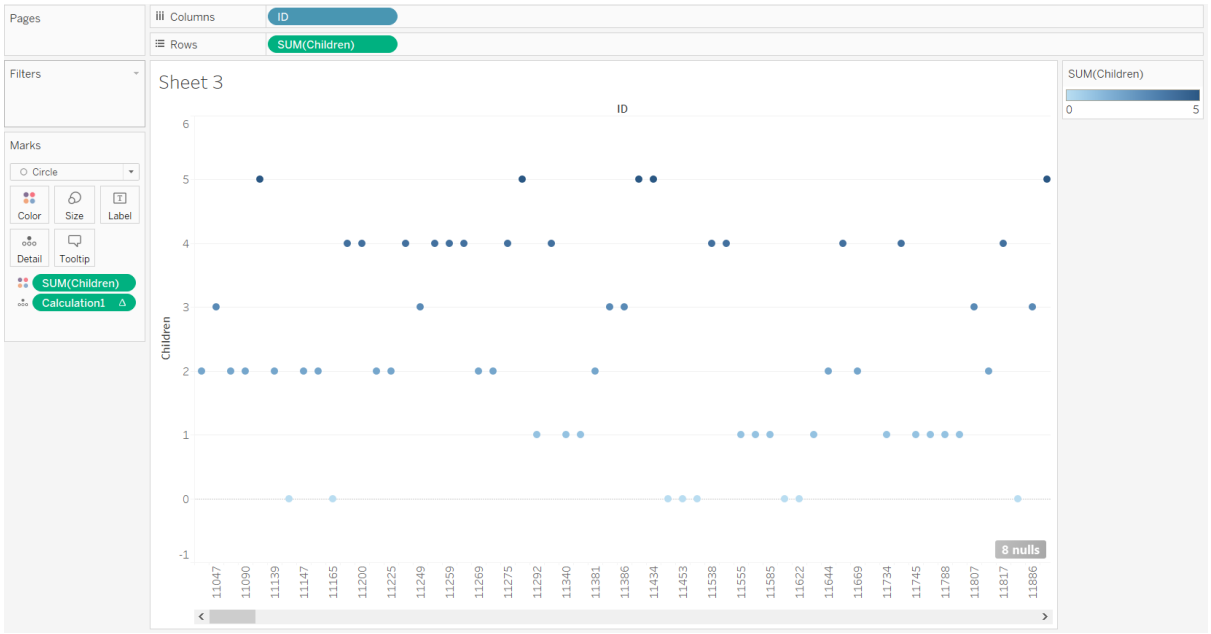
sign2(cbind(.arg1))$wfinal01",SUM([(field name]))

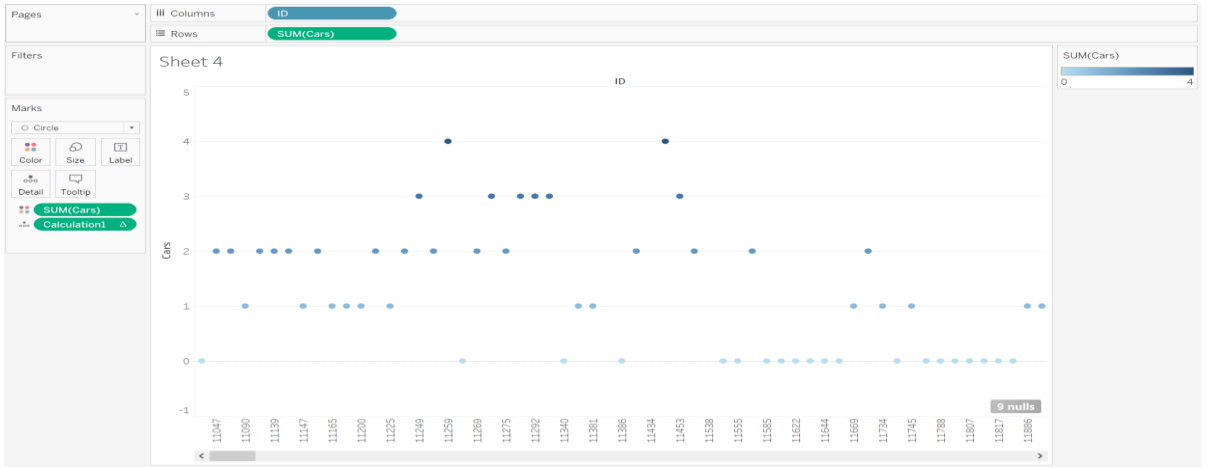


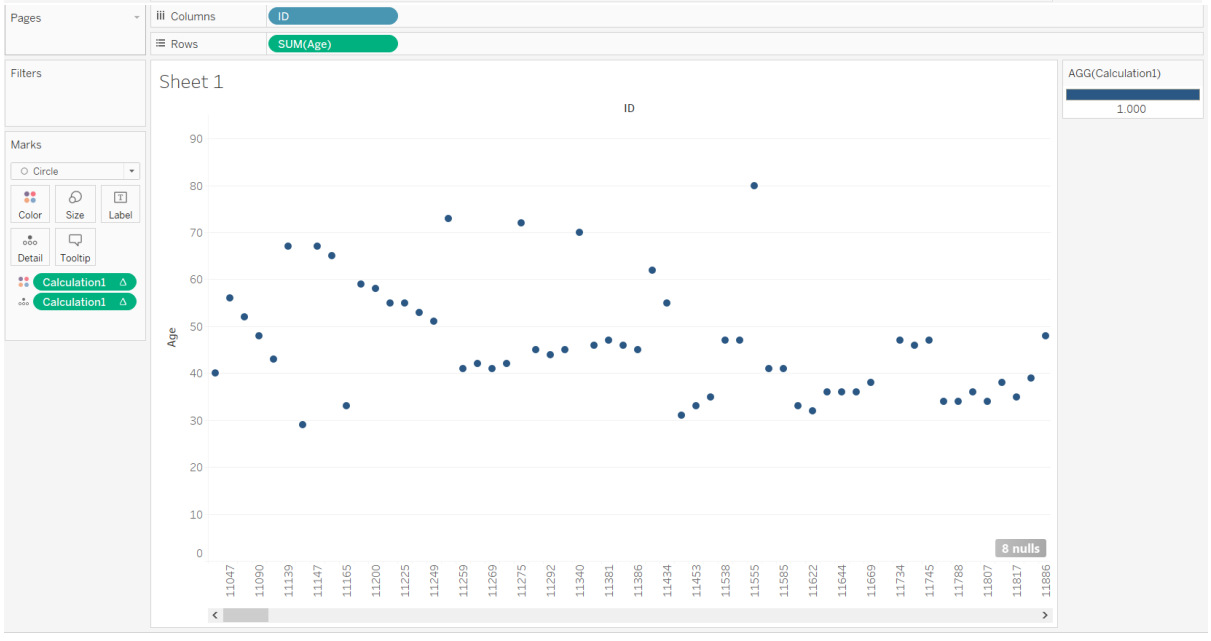

o Drop the field in color shelf to see the clusters

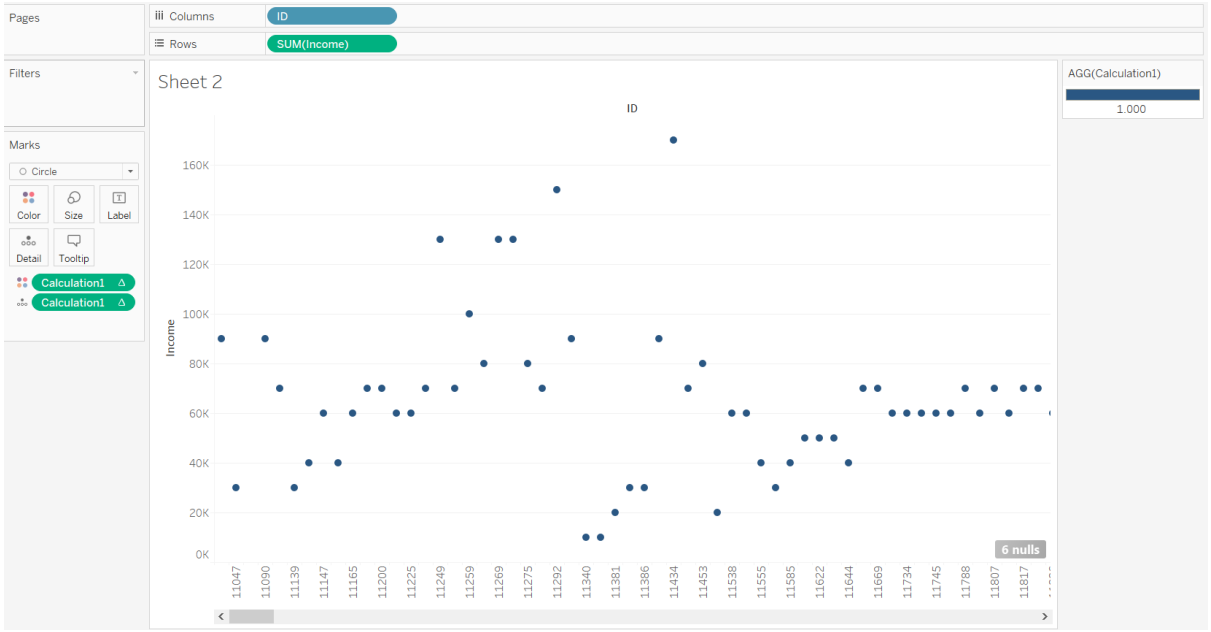

Pages
Columns
ID
Rows
SUM(Age)
Filters
Sheet 1
ID
SUM(Age)
25
89
Marks
Circle
Color
Size
Label
Detail
Tooltip
SUM(Age)
Calculation1
Age
90
80
70
60
50
40
30
20
10
0
8 nulls
11047
11090
11139
11147
11165
11200
11225
11249
11259
11269
11275
11292
11340
11381
11386
11434
11453
11538
11555
11585
11622
11644
11669
11734
11745
11788
11807
11817
11886

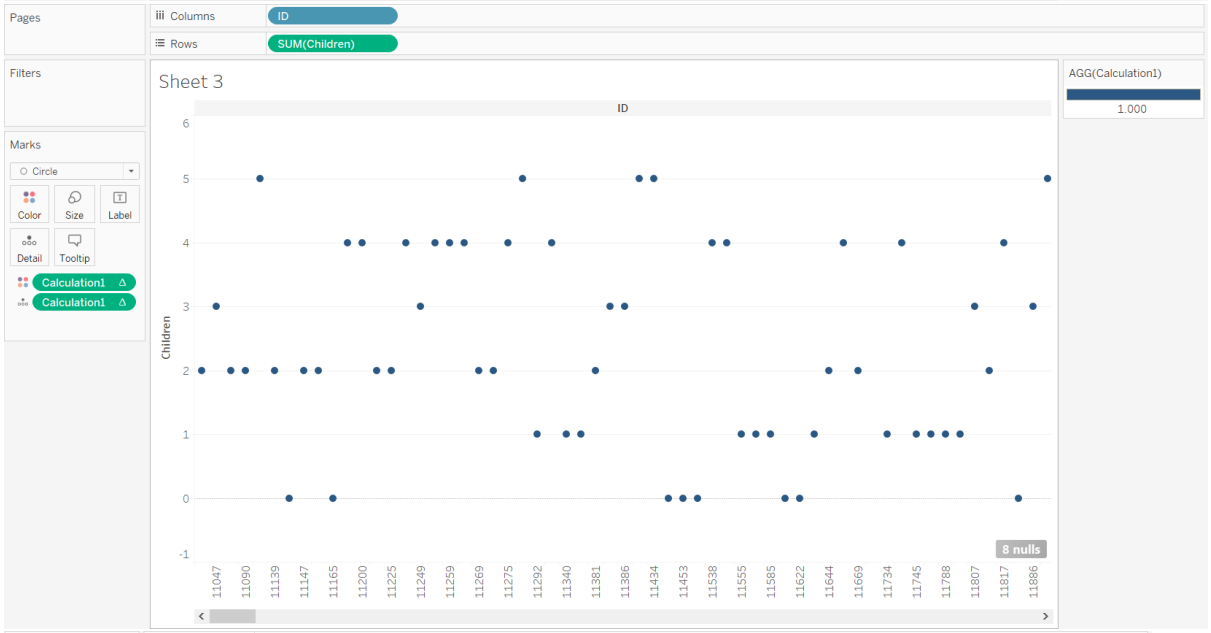

Pages
Columns
ID
Rows
SUM(Income)
Filters
Sheet 2
ID
SUM(Income)
10,000
1,70,000
Marks
Circle
Color
Size
Label
Detail
Tooltip
SUM(Income)
Calculation1
Income
160K
140K
120K
100K
80K
60K
40K
20K
0K
6 nulls
11047
11090
11139
11147
11165
11200
11225
11249
11259
11269
11275
11292
11340
11381
11386
11434
11453
11538
11555
11585
11622
11644
11669
11734
11745
11788
11807
11817

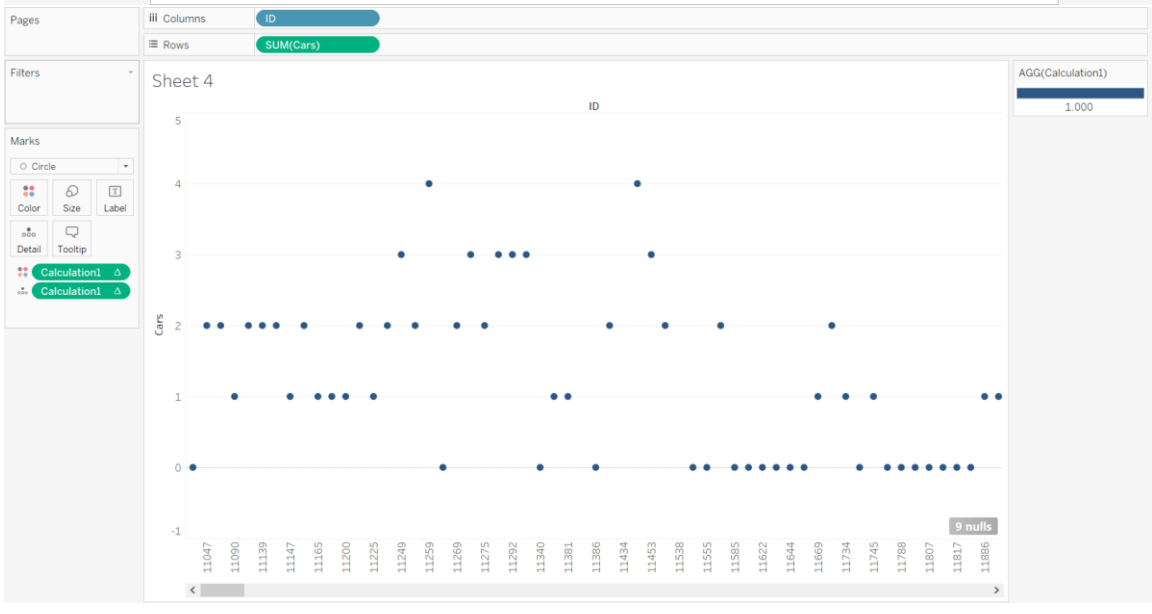

Pages
Columns
ID
Rows
SUM(Children)
Filters
Sheet 3
SUM(Children)
0
5
Marks
Circle
Color
Size
Label
Detail
Tooltip
SUM(Children)
Calculation1
ID
Children
6
5
4
3
2
1
0
-1
8 nulls
11047
11090
11139
11147
11165
11200
11225
11249
11259
11269
11275
11292
11340
11381
11386
11434
11453
11538
11555
11585
11622
11644
11669
11734
11745
11788
11807
11817
11886


Pages
Columns
ID
Rows
SUM(Cars)
Filters
Sheet 4
SUM(Cars)
0
4
Marks
Circle
Color
Size
Label
Detail
Tooltip
SUM(Cars)
Calculation1
ID
Cars
5
4
3
2
1
0
-1
9 nulls
11047
11090
11139
11147
11165
11200
11225
11249
11259
11269
11275
11292
11340
11381
11386
11434
11453
11538
11555
11585
11622
11644
11669
11734
11745
11788
11807
11817
11886


Pages
Columns
ID
Rows
SUM(Age)
Filters
Sheet 1
AGG(Calculation1)
1.000
Marks
Circle
Color
Size
Label
Detail
Tooltip
Calculation1
Calculation1
ID
Age
90
80
70
60
50
40
30
20
10
0
8 nulls
11047
11090
11139
11147
11165
11200
11225
11249
11259
11269
11275
11292
11340
11381
11386
11434
11453
11538
11555
11585
11622
11644
11669
11734
11745
11788
11807
11817
11886


Pages
Columns
ID
Rows
SUM(Income)
Filters
Sheet 2
AGG(Calculation1)
1.000
Marks
Circle
Color
Size
Label
Detail
Tooltip
Calculation1
Calculation1
ID
Income
160K
140K
120K
100K
80K
60K
40K
20K
0K
6 nulls
11047
11090
11139
11147
11165
11200
11225
11249
11259
11269
11275
11292
11340
11381
11386
11434
11453
11538
11555
11585
11622
11644
11669
11734
11745
11788
11807
11817


Pages
Columns
ID
Rows
SUM(Children)
Filters
Sheet 3
AGG(Calculation1)
1.000
Marks
Circle
Color
Size
Label
Detail
Tooltip
Calculation1
Calculation1
ID
Children
6
5
4
3
2
1
0
-1
8 nulls
11047
11090
11139
11147
11165
11200
11225
11249
11259
11269
11275
11292
11340
11381
11386
11434
11453
11538
11555
11585
11622
11644
11669
11734
11745
11788
11807
11817
11886


Pages
Columns
ID
Rows
SUM(Cars)
Filters
Sheet 4
AGG(Calculation1)
1.000
Marks
Circle
Color
Size
Label
Detail
Tooltip
Calculation1
Calculation1
ID
Cars
5
4
3
2
1
0
-1
9 nulls
11047
11090
11139
11147
11165
11200
11225
11249
11259
11269
11275
11292
11340
11381
11386
11434
11453
11538
11555
11585
11622
11644
11669
11734
11745
11788
11807
11817
11886

o Inference

Clustering is available in Tableau Desktop, but is not available for authoring on the web (Tableau Server, Tableau Online). Clustering is also not available when any of the following conditions apply:

- When you are using a cube (multidimensional) data source.
- When there is a blended dimension in the view.
- When there are no fields that can be used as variables (inputs) for clustering in the view.
- When there are no dimensions present in an aggregated view.

When any of those conditions apply, you will not be able to drag **Clusters** from the Analytics pane to the view.

In addition, the following field types cannot be used as variables (inputs) for clustering:

- Table calculations
- Blended calculations
- Ad-hoc calculations
- Generated latitude/longitude values
- Groups
- Sets
- Bins
- Parameters
- Dates
- Measure Names/Measure Values