

EXAMPLE - ROLLINGKTHLARGEST Functions

This example describes how to use rolling kthlargest functions for calculating ranking of values within a defined window of rows.

Functions:

Item	Description
ROLLINGKTHLARGEST Function	Computes the rolling <i>kth</i> largest value forward or backward of the current row. Inputs can be Integer, Decimal, or Datetime.
ROLLINGKTHLARGESTUNIQUE Function	Computes the rolling unique <i>kth</i> largest value forward or backward of the current row. Inputs can be Integer, Decimal, or Datetime.
ROWNUMBER Function	Generates a new column containing the row number as sorted by the <code>order</code> parameter and optionally grouped by the <code>group</code> parameter.

The following dataset contains daily counts of server restarts across three servers over the preceding week. High server restart counts can indicate poor server health. In this example, you are interested in knowing for each server the rolling highest and second highest count of restarts per server over the previous week.

Source:

Date	Server	Restarts
2/21/18	s01	4
2/21/18	s02	0
2/21/18	s03	0
2/22/18	s01	4
2/22/18	s02	1
2/22/18	s03	2
2/23/18	s01	2
2/23/18	s02	3
2/23/18	s03	4
2/24/18	s01	1
2/24/18	s02	0
2/24/18	s03	2
2/25/18	s01	5
2/25/18	s02	0
2/25/18	s03	4
2/26/18	s01	1
2/26/18	s02	2
2/26/18	s03	1
2/27/18	s01	1
2/27/18	s02	2
2/27/18	s03	2

Transformation:

First, you want to maintain the row information as a separate column. Since data is ordered already by the Date column, you can use the following:

Transformation Name	New formula
Parameter: Formula type	Single row formula
Parameter: Formula	ROWNUMBER()
Parameter: New column name	'entryId'

Use the following function to compute the rolling *k*th largest value of server restarts per server over the previous week. In this case, you can use the ROLLINGKTHLARGEST function, setting *k*=1. Uniqueness doesn't matter for calculating the highest value:

Transformation Name	New formula
Parameter: Formula type	Multiple row formula
Parameter: Formula	rollingkthlargest(Restarts, 1, 6, 0)
Parameter: Sort Rows by	Server
Parameter: Group Rows by	Server
Parameter: New column name	'rollingkthlargest_1'

Use the following function to compute the rolling second highest value. In this case, you can use ROLLINGKTHLARGESTUNIQUE:

Transformation Name	New formula
Parameter: Formula type	Multiple row formula
Parameter: Formula	rollingkthlargestunique(Restarts, 2, 6, 0)
Parameter: Sort Rows by	Server
Parameter: Group Rows by	Server
Parameter: New column name	'rollingKthLargestUnique_2'

Results:

entryId	Date	Server	Restarts	rollingKthLargestUnique_2	rollingkthlargest_Restarts
3	2/21/18	s02	0	0	0
6	2/22/18	s02	1	0	1
9	2/23/18	s02	3	1	3
12	2/24/18	s02	0	1	3
15	2/25/18	s02	0	1	3

18	2/26/18	s02	2	2	3
21	2/27/18	s02	2	2	3
4	2/21/18	s03	0	0	0
7	2/22/18	s03	2	0	2
10	2/23/18	s03	4	2	4
13	2/24/18	s03	2	2	4
16	2/25/18	s03	4	2	4
19	2/26/18	s03	1	2	4
22	2/27/18	s03	2	2	4
2	2/21/18	s01	4	4	4
5	2/22/18	s01	4	4	4
8	2/23/18	s01	2	2	4
11	2/24/18	s01	1	2	4
14	2/25/18	s01	5	4	5
17	2/26/18	s01	1	4	5
20	2/27/18	s01	1	4	5