

EXAMPLE - Percentile Functions

This example illustrates how you can apply the following percentile-related functions to your transformations:

- **MEDIAN** - Calculate the median value from a column of values. See *MEDIAN Function*.
- **PERCENTILE** - Calculate a specified percentile for a column of values. See *PERCENTILE Function*.
- **QUARTILE** - Calculate a specified quartile for a column of values. See *QUARTILE Function*.

The following functions use an approximation technique for calculating median, percentile, and quartiles. In some cases, these calculations can be computed faster across large datasets.

- **APPROXIMATEMEDIAN** - Calculate a close approximation of the median value from a column of values. See *APPROXIMATEMEDIAN Function*.
- **APPROXIMATEPERCENTILE** - Calculate a close approximation of a specified percentile for a column of values. See *APPROXIMATEPERCENTILE Function*.
- **APPROXIMATEQUARTILE** - Calculate a close approximation of a specified quartile for a column of values. See *APPROXIMATEQUARTILE Function*.

Source:

The following table lists each student's height in inches:

Student	Height
1	64
2	65
3	63
4	64
5	62
6	66
7	66
8	65
9	69
10	66
11	73
12	69
13	69
14	61
15	64
16	61
17	71
18	67
19	73
20	66

Transformation:

Use the following transformations to calculate the median height in inches, a specified percentile and the first quartile.

- The first function uses a precise algorithm which can be slow to execute across large datasets.
- The second function uses an appropriate approximation algorithm, which is much faster to execute across large datasets.
 - These approximate functions can use an error boundary parameter, which is set to 0.4 (0.4%) across all functions.

Median: This transformation calculates the median value, which corresponds to the 50th percentile.

Transformation Name	New formula
Parameter: Formula type	Single row formula
Parameter: Formula	median(heightIn)
Parameter: New column name	'medianIn'

Transformation Name	New formula
Parameter: Formula type	Single row formula
Parameter: Formula	approximateMedian(heightIn, 0.4)
Parameter: New column name	'approxMedianIn'

Percentile: This transformation calculates the 68th percentile.

Transformation Name	New formula
Parameter: Formula type	Single row formula
Parameter: Formula	percentile(heightIn, 68, linear)
Parameter: New column name	'percentile68In'

Transformation Name	New formula
Parameter: Formula type	Single row formula
Parameter: Formula	approximatePercentile(heightIn, 68, 0.4)
Parameter: New column name	'approxPercentile68In'

Quartile: This transformation calculates the first quartile, which corresponds to the 25th percentile.

Transformation Name	New formula
Parameter: Formula type	Single row formula
Parameter: Formula	quartile(heightIn, 1, linear)
Parameter: New column name	'percentile25In'

Transformation Name	New formula
Parameter: Formula type	Single row formula
Parameter: Formula	approximatequartile(heightIn, 1, 0.4)
Parameter: New column name	'approxPercentile25In'

Results:

studentId	heightIn	approxPercentile25In	percentile25In	approxPercentile68In	percentile68In	approxMedianIn	
1	64	64	64	67.1	66.92	66	6
2	65	64	64	67.1	66.92	66	6
3	63	64	64	67.1	66.92	66	6
4	64	64	64	67.1	66.92	66	6
5	62	64	64	67.1	66.92	66	6
6	66	64	64	67.1	66.92	66	6
7	66	64	64	67.1	66.92	66	6
8	65	64	64	67.1	66.92	66	6
9	69	64	64	67.1	66.92	66	6
10	66	64	64	67.1	66.92	66	6
11	73	64	64	67.1	66.92	66	6
12	69	64	64	67.1	66.92	66	6
13	69	64	64	67.1	66.92	66	6
14	61	64	64	67.1	66.92	66	6
15	64	64	64	67.1	66.92	66	6
16	61	64	64	67.1	66.92	66	6
17	71	64	64	67.1	66.92	66	6
18	67	64	64	67.1	66.92	66	6
19	73	64	64	67.1	66.92	66	6
20	66	64	64	67.1	66.92	66	6