

APPROXIMATEQUARTILE Function

Contents:

- *Basic Usage*
- *Syntax and Arguments*
 - *function_col_ref*
 - *num_quartile*
 - *dec_error_bound*
- *Examples*
 - *Example - Percentile functions*

Computes an approximation for a specified quartile across all row values in a column or group. Input column can be of Integer or Decimal.

- If a row contains a missing or null value, it is not factored into the calculation. If the entire column contains no values, the function returns a null value.
- When used in a `pivot` transform, the function is computed for each instance of the value specified in the `group` parameter. See *Pivot Transform*.
- The approximate percentile functions utilize a different algorithm for efficiently estimating quantiles for streaming and distributed processing, depending on the running environment where the function is computed.

Tip: Approximation functions are suitable for larger datasets. As the number of rows increases, accuracy and calculation performance improves for these functions.

- For an exact calculation of this function, see *QUARTILE Function*.

Quartiles are computed as follows:

| Quartile | Description |
|----------|----------------------------|
| 0 | Minimum value |
| 1 | 25th percentile |
| 2 | Median value |
| 3 | 75th percentile and higher |

Wrangle vs. SQL: This function is part of Wrangle , a proprietary data transformation language. Wrangle is not SQL. For more information, see *Wrangle Language*.

Basic Usage

```
approximatequartile(myScores, 3)
```

Output: Computes the approximate value that is at the 3rd quartile across all values in the `myScores` column.

Syntax and Arguments

```
approximatequartile(function_col_ref,num_quartile) [group:group_col_ref] [limit:  
limit_count]
```

| Argument | Required? | Data Type | Description |
|------------------|-----------|-----------|---|
| function_col_ref | Y | string | Name of column to which to apply the function |
| num_quartile | Y | integer | Integer value (0-3) of the quartile to compute |
| dec_error_bound | N | decimal | Error factor for computing approximations. Decimal value represents error factor as a percentage (0.4 is 0.4%). |

For more information on the `group` and `limit` parameters, see *Pivot Transform*.

For more information on syntax standards, see *Language Documentation Syntax Notes*.

function_col_ref

Name of the column the values of which you want to calculate the quartile. Column must contain Integer or Decimal values.

- Literal values are not supported as inputs.
- Multiple columns and wildcards are not supported.

Usage Notes:

| Required? | Data Type | Example Value |
|-----------|---------------------------|-----------------|
| Yes | String (column reference) | precipitationIn |

num_quartile

Integer literal value indicating the quartile that you wish to compute. Input value must be between 0 and 3, inclusive.

- Column or function references are not supported.
- Multiple columns and wildcards are not supported.

Usage Notes:

| Required? | Data Type | Example Value |
|-----------|-----------|---------------|
| Yes | Integer | 3 |

dec_error_bound

As needed, you can insert an error boundary factor as a parameter into the computation of this approximate value.

- This value must be a Decimal literal value.
- This decimal value represents the percentage error factor. By default, this value is 0.5 (0.5%).

Usage Notes:

| Required? | Data Type | Example Value |
|-----------|-----------|---------------|
|-----------|-----------|---------------|

| | | |
|----|-------------------|--------|
| No | Decimal (literal) | 0 . 01 |
|----|-------------------|--------|

Examples

Tip: For additional examples, see *Common Tasks*.

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 |