

Calculate Metrics across Columns

You can use a variety of mathematical and statistical functions to calculate metrics within a column.

To calculate metrics across columns, you can use a generalized version of the following example.

Source:

Your dataset tracks swimmer performance across multiple heats in a race, and you would like to calculate best, worst, and average times in seconds across all three heats. Here's the data:

Racer	Heat1	Heat2	Heat3
Racer X	37.22	38.22	37.61
Racer Y	41.33	DQ	38.04
Racer Z	39.27	39.04	38.85

In the above data, Racer Y was disqualified (DQ) in Heat 2.

Transformation:

To compute the metrics, you must bundle the data into an array, break out the array into separate rows, and then calculate your metrics by grouping. Here are the steps:

1. When the data is imported, you may need to create a header for each row:

Transformation Name	Rename columns with a row
Parameter: Option	Use row as header
Parameter: Row	1

2. The columns containing heat time data may need to be retyped. From the drop-down next to each column name, select Decimal type.
3. The DQ value in the Heat2 column is invalid data for Decimal type. You can use the following transformation to turn it into a missing value. For purposes of calculating averages, you may or may not want to turn invalid data into zeroes or blanks. In this case, replacing the data as 0.00 causes improper calculations for the metrics.

Transformation Name	Replace text or patterns
Parameter: Column	Heat2
Parameter: Find	'DQ'
Parameter: Replace with	' '

4. Use the following to gather all of the heat data into two columns:

Transformation Name	Unpivot columns
Parameter: Columns	Heat1,Heat2,Heat3
Parameter: Group size	1

5. You can now rename the two columns. Rename key to HeatNum and value to HeatTime.
6. You may want to delete the rows that have a missing value for HeatTime:

Transformation Name	Delete rows
Parameter: Condition	ISMISSING([value])

7. You can now perform calculations on this column. The following transformations calculate minimum, average (mean), and maximum times for each racer:

Transformation Name	New formula
Parameter: Formula type	Multiple row formula
Parameter: Formula	MIN(HeatTime)
Parameter: Group rows by	Racer
Parameter: New column name	'BestTime'

Transformation Name	New formula
Parameter: Formula type	Multiple row formula
Parameter: Formula	AVERAGE(HeatTime)
Parameter: Group rows by	Racer
Parameter: New column name	'AvgTime'

Transformation Name	New formula
Parameter: Formula type	Multiple row formula
Parameter: Formula	MAX(HeatTime)
Parameter: Group rows by	Racer
Parameter: New column name	'WorstTime'

8. To make the data look better, you might want to reformat the values in the AvgTime column to two decimal points:

Transformation Name	Edit column with formula
Parameter: Columns	AvgTime
Parameter: Formula	NUMFORMAT(AvgTime, '##.00')

Results:

After you use the Move transformation to re-organize your columns, the dataset should look like the following:

Racer	HeatNum	HeatTime	BestTime	WorstTime	AvgTime
Racer X	Heat1	37.22	37.22	38.22	37.68
Racer X	Heat2	38.22	37.22	38.22	37.68
Racer X	Heat3	37.61	37.22	38.22	37.68
Racer Y	Heat1	41.33	38.04	41.33	39.69
Racer Y	Heat3	38.04	38.04	41.33	39.69
Racer Z	Heat1	39.27	38.85	39.27	39.05
Racer Z	Heat2	39.04	38.85	39.27	39.05
Racer Z	Heat3	38.85	38.85	39.27	39.05