Data Quality Rules Reference

Contents:

This section contains reference information on the data quality rule types and input types that are available in Trifacta®.

- Data quality rules can be applied to your dataset through the Transformer page. See *Data Quality Rules Panel*.
- Input types identify the calculated metric types that can be used as inputs for a data quality rule.

For more information on data quality, see *Overview of Data Quality*.

Rule Types

name	description
Unique	Column values must be unique.
Implies	Source column values imply the values of a target column. For each unique source value, there should be exactly one implied target value.
Not Missing	Column values must not be missing. Null values and empty strings are not allowed.

Not Null	Column values must not be null. Empty strings are allowed.
Valid	Column values must be valid instances of a data type.
Match	Column values must match a pattern.
Not Match	Column values must not match a pattern.
Starts With	Column values must start with a pattern.
Ends With	Column values must end with a pattern.
Equal	Column values must equal a provided value.
Not Equal	Column values must not equal a provided value.
In Range	Column values must lie between provided minimum and maximum values.
Greater Than	Column values must be greater than a minimum value.
Less Than	Column values must be less than a maximum value.
In Set	Column values must be one of a set of acceptable values.
Not In Set	Column values must not be one of a set of unacceptable values.
Formula	Apply a custom data quality rule formula.

Metric Input Types

The following metric input types can be selected as the source of a data quality rule.

> **NOTE:** These input types are available for selection from the Column drop-down.

Metric input types are supported for the following rules:

- In Range
- Greater Than
- Less Than
- Equals
- Not Equals
- In Set
- Not In Set

name	description
Average	The average column value.
Count Distinct	The number of unique column values.
Maximum	The maximum column value.
Minimum	The minimum column value.
Sum	The sum of column values.
Standard Deviation	The sample standard deviation of column values.

Variance	The sample variance of column values.
Count	The number of rows.
Correlation	The Pearson correlation coefficient between two columns.
Z-Score	The distance from the mean, in units of standard deviations.