

# Data Quality Rules Reference

## Contents:

- *Rule Types*
  - *Unique*
  - *Implies*
  - *Not Missing*
  - *Not Null*
  - *Valid*
  - *Match*
  - *Not Match*
  - *Starts With*
  - *Ends With*
  - *Equal*
  - *Not Equal*
  - *In Range*
  - *Greater Than*
  - *Less Than*
  - *In Set*
  - *Not In Set*
  - *Formula*
- *Metric Input Types*
  - *Average*
  - *Count Distinct*
  - *Maximum*
  - *Minimum*
  - *Sum*
  - *Standard Deviation*
  - *Variance*
  - *Count*
  - *Correlation*
  - *Z-Score*

---

**Feature Availability:** This feature is available in the following editions:

- Trifacta® Enterprise Edition
- Trifacta Professional Edition
- Trifacta Premium

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

- 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
<b>Unique</b>	Column values must be unique.
<b>Implies</b>	Source column values imply the values of a target column. For each unique source value, there should be exactly one implied target value.
<b>Not Missing</b>	Column values must not be missing. Null values and empty strings are not allowed.
<b>Not Null</b>	Column values must not be null. Empty strings are allowed.
<b>Valid</b>	Column values must be valid instances of a data type.
<b>Match</b>	Column values must match a pattern.
<b>Not Match</b>	Column values must not match a pattern.
<b>Starts With</b>	Column values must start with a pattern.
<b>Ends With</b>	Column values must end with a pattern.
<b>Equal</b>	Column values must equal a provided value.
<b>Not Equal</b>	Column values must not equal a provided value.
<b>In Range</b>	Column values must lie between provided minimum and maximum values.
<b>Greater Than</b>	Column values must be greater than a minimum value.
<b>Less Than</b>	Column values must be less than a maximum value.
<b>In Set</b>	Column values must be one of a set of acceptable values.
<b>Not In Set</b>	Column values must not be one of a set of unacceptable values.
<b>Formula</b>	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
<b>Average</b>	The average column value.

<b>Count Distinct</b>	The number of unique column values.
<b>Maximum</b>	The maximum column value.
<b>Minimum</b>	The minimum column value.
<b>Sum</b>	The sum of column values.
<b>Standard Deviation</b>	The sample standard deviation of column values.
<b>Variance</b>	The sample variance of column values.
<b>Count</b>	The number of rows.
<b>Correlation</b>	The Pearson correlation coefficient between two columns.
<b>Z-Score</b>	The distance from the mean, in units of standard deviations.