

Overview of Data Quality Rules

Contents:

- *Limitations*
- *Data Quality Rule Categories*
 - *Data quality types*
- *Creating Rules*
- *Reviewing Suggestions*
- *Rules Evaluation*

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

- Trifacta Enterprise Edition
- Trifacta Professional Edition
- Trifacta Premium

In the Transformer page, you can design data quality rules to apply to the displayed sample of your data. These data quality rules can be used to identify anomalies, completeness, uniqueness, and validity.

Additionally, rules can be defined to assess the quality of the data for its intended purpose in your data pipeline. In addition, you can use calculated metric type (derived metrics) as a source of data quality input types and create a metric-based data quality rule.

Examples:

- Check that all product identifiers fit a specified pattern
- Verify that there are no negative values for any count columns
- Validate that primary key columns contain unique values
- Specify metrics in the Column value for some rule types.
 - For example, instead of specifying a column name such as `OrderTotal` as the input for the data quality rule, you could specify for some rule types, `AVERAGE(OrderTotal)`.
 - For more examples, see *Add Data Quality Rule*.

NOTE: Metric-based rules are supported only for some metric types. For more information on the rules that support metrics, see *Data Quality Rules Reference*.

NOTE: Data quality rules are not transformation steps. They assess the current state of the sampled data in the Transformer page and can be used to assist in constructing transformation steps to improve data quality.

NOTE: As you apply transformation steps to the data, the state of your data quality rules is automatically updated to reflect the changes. If you delete columns or other elements referenced in the data quality rules, errors are generated in the Transformer page.

Limitations

- Rules cannot be included in macros.
- Rules cannot be parameterized.
- Sets of rules are created for each recipe. Rules cannot be shared between recipes.

Data Quality Rule Categories

Rules break down into the following categories:

Category	Description
Integrity Constraints	Rule types in this category assess the validity of a column's data and any implied relationships between the data (e.g., City + State implies Zip Code)
Pattern Matching	These rule types test whether the data in your column matches patterns that you define.
Column Values	These rule types compare column values to limits or sets of acceptable values. In addition to column references, you can specify metric-based values. For example, you can create a constraint that the sales quantity should be within a specific range.
Other Rules	You can also create data quality rules based on custom Wrangle formulas.

Data quality types

Within each of the above categories, you can explore and define a variety of types of data quality rules. These rule types provide a template for creating the rule, which accepts one or more input parameters that you specify. For more information on the set of available rule types, see *Data Quality Rules Reference*.

Creating Rules

For each recipe, you can create individualized sets of rules from within the Transformer page. In the Data Quality Rules panel, you build your data-specific rules and can review the quality bars of each rule as you continue to build your recipe.

For more information on creating rules, see *Add Data Quality Rule*.

For more information, see *Data Quality Rules Panel*.

Reviewing Suggestions

Through the Data Quality Rules panel, you can review a set of suggested data quality rules that are applicable to your dataset. These rules are generated based on heuristics applied to your sampled data. For more information, see *Data Quality Rules Panel*.

Rules Evaluation

In the Transformer page:

1. **Rules are evaluated and displayed for the current location in the recipe.** For example, if you change the location of the recipe cursor to a point earlier in the recipe, all of the defined rules are evaluated for the state of the dataset sample at that point in the recipe.
2. **The data quality rules defined in the Transformer page are applied to the displayed sample.** If your sample is not the full dataset, you should consider taking additional samples to validate the rules across other parts of your dataset.

In job results:

After job execution, these rules are applied across the entire dataset and available when visual profiling is enabled. For more information, see *Job Details Page*.