# Column Statistics Reference

This page describes the statistical information available for individual columns of data.

- Statistics may vary depending on the column's data type. For example, the statistics retained for states may be different from the statistics for strings.
- Most of these statistics are available in the Column Details panel, which can be opened from the left side of the Transformer page.

Below, you can review general statistics maintained for each data type, followed by breakdowns of statistics for each specific type of data.

> **NOTE:** Before your job is run, profiling information such as column statistics are exact counts of the sample that is currently loaded. After the job is run, profiled results in the Job Results page might include estimates for some metrics and counts, depending on the scale of the dataset.

## General Column Counts

For any selection of values in a column, the following counts are generally available.

| Count Name | Description |
|---|---|
| Valid Values | Count of values that are valid for the column's data type |
| Unique Values | Count of unique values. Duplicate values are not counted. |
| Outlier Values | Count of values that qualify as outliers. An **outlier** value is either:<br>• < (25th percentile) - (2 * IQR)<br>• > (75th percentile) + (2* IQR)<br>• **IQR (interquarterile range)** is the range of values between the two middle quarters, which is equivalent to the range between the 25th and 75th percentiles. Thus, in the above computations, the IQR factor ensures that the outliers are at the extremes of the entire range. |
| Mismatched Values | Count of values that do not confirm to the column's data type. For example, an Integer column with a value of "MISSING" results in a mismatched value. |
| Missing Values | Count of values that are not populated |

## General Column Statistics

These statistics are available for most types of data through the Column Browser.

- For string types (String, Phone Number, Social Security Number, Boolean, Email Address, Credit Card Number, Gender, IP Address, URL, HTTP Code, Date/Time), these stats measure string length.
  - For structured string types (Phone Number, Social Security Number, Boolean, Gender, IP Address, HTTP Code, Date/Time), any variation in these numbers indicates data problems.
- Does not apply to: State

| Statistic Name | Description |
|---|---|
| Minimum | Lowest value in the column |
| Lower Quartile | The median of the lower half of values (25th percentile) |

| | |
|---|---|
| Median | The middle value of the selected set. For example, in a set of 21 values, the median value is the 11th value in ascending order.<br>- For datasets with an even number of values, the median is the mean of the two middle values. |
| Upper Quartile | The median of the upper half of values (75th percentile) |
| Maximum | Highest value in the column |
| Average | Average value in the column |
| Standard Deviation | The computed standard deviation for the selected values. |