

Add Lookup Data

Contents:

- *Set up Your Lookup Data*
- *Perform the Lookup*
- *Example - Lookup for Timezones*

You can integrate data from other sources into your current dataset. Based on a key column that you identify in the lookup dataset, you can insert the corresponding values in other columns of the lookup dataset as new columns in your source dataset.

Tip: Column lookups are useful for adding reference data based on a column's values.

For example, your data contains the two-letter abbreviations for U.S. states, yet the target system is expecting the full name of each state. You need to replace the XY state abbreviation with the full name of each state in each row.

Set up Your Lookup Data

Your data table should like the following:

State-2Letter	State
AL	Alabama
AK	Alaska
AZ	Arizona
AR	Arkansas
CA	California
CO	Colorado
CT	Connecticut
DE	Delaware
DC	District of Columbia
FL	Florida
GA	Georgia
HI	Hawaii
ID	Idaho
IL	Illinois
IN	Indiana
IA	Iowa
KS	Kansas
KY	Kentucky
LA	Louisiana

ME	Maine
MD	Maryland
MA	Massachusetts
MI	Michigan
MN	Minnesota
MS	Mississippi
MO	Missouri
MT	Montana
NE	Nebraska
NV	Nevada
NH	New Hampshire
NJ	New Jersey
NM	New Mexico
NY	New York
NC	North Carolina
ND	North Dakota
OH	Ohio
OK	Oklahoma
OR	Oregon
PA	Pennsylvania
RI	Rhode Island
SC	South Carolina
SD	South Dakota
TN	Tennessee
TX	Texas
UT	Utah
VT	Vermont
VA	Virginia
WA	Washington
WV	West Virginia
WI	Wisconsin
WY	Wyoming

Tip: You can download a version of this table, which also includes some timezone information. See *Dict-TimezoneByState.csv*.

This data table must be uploaded as a new dataset. See *Import Data Page*.

Perform the Lookup

Steps:

1. In the Transformer page, click the drop-down on the column that contains your two-letter state abbreviations. Select **Lookup**
2. In the Lookup Wizard, select the dataset to use for your lookup.
3. For the lookup key, select the column in the dataset to use as the key value. In the above example, it is `State_2Letter`.
4. Click **Execute Lookup**.
5. The lookup key value is used to locate all of the other column values in the reference dataset. These values are inserted in separate columns to the immediate right of the source column.
6. You might need to delete some of the imported columns. In the above case, you might decide to delete the two-letter state identifier column, which has been replaced by the full state name column.

See *Lookup Wizard*.

Example - Lookup for Timezones

The CSV linked above also contains timezone information for each state, which you can use to provide higher fidelity information on timestamps.

U.S. timezones are not consistently demarcated by state lines. Some states are split across multiple timezones. For more accurate representation of timezones, you should download and use a zipcode database, many of which are freely available online. This CSV is provided for demonstration purposes only.

In this case, you are working with a dataset that contains timestamps, which are stored in different timezones based on the location where an event or transaction occurred. However, the timestamps do not contain any timezone information.

You can use an external source of timezone information to insert timezones into your dataset. In the following example, timezones are derived based on two-letter abbreviations for U.S. state. A more accurate representation would be based on zipcode data.

Steps:

1. Complete steps 1-5 in the previous section.
2. Delete all columns except the one containing timezone information. The `Time Offsets` column identifies the predominant timezone in each state as an offset of the UTC timezone (Greenwich Mean Time).
3. Move this column to the right of the column containing your timestamps.

NOTE: Depending on the requirements of your target system, you can use the `split` transform to break up column data so that only the numerical offset (e.g. `-6:00`) is present. Then, you can use the `DATEDIFF` function to apply the timezone offset to your timestamps. In this manner, you can convert timestamps to the source timezone before they are consumed by the target system.