

Customer Dataset Upload Requirements



The ID that you use in a dataset upload must match the case of the value collected in an on-site ID.

General Requirements

You can upload existing data to use in experiences. All files you upload to the platform must adhere to the following requirements:

- All files must contain comma-separated values in CSV, TXT, or TSV format. You can compress the file with a ZIP or GZ extension.
- All files must use UTF-8 character encoding and without the byte order mark (BOM).
- Files cannot exceed 10 GB uncompressed or 100 million rows.
- Files cannot contain more than 30 columns and must contain at least one column.
- Include only a single row per unique identifier.



If a file contains multiple rows with the same unique identifier, only the last row is used.

- Omit thousandths separators from numeric values (for example, 9,000 should be `9000`).
- Omit leading currency symbols.
- You must put a set of double quotation marks (") at the beginning of and at the end of the content of a field if that content includes a comma (for example, `"Bldg 2, Ste 29"`).
- Dataset names and headers are limited to the [Latin-1 \(ISO-8859-1\)](#) character set.

If you have any questions about dataset requirements or increasing an account's row limit, contact your Account Manager.

You can include the following data types in a file:

- Strings (text)
- Numbers
- Dates
- Boolean (true/false)

The platform treats the following values as empty:

- `None`
- `NULL`
- An empty value



All empty values are case-insensitive and work regardless of capitalization.

Header vs. Headerless Rows

When you create a new dataset, the initial file must contain a header row. The header rows that you include in subsequent files are used to identify new and updated attributes. Verify that the column headers remain the same in subsequent files that you upload to maintain the integrity of the columns in the dataset.

If you want to update an existing dataset without adding new columns, you can exclude the header row from the file. Add `NOHEADER` to the end of the filename (for example, `123_filename_NOHEADER.csv`) to indicate that the file doesn't contain a header row.



If you exclude a header row from a dataset update, ensure that the columns in the file match the column order in the existing dataset. If the column order doesn't match, then the updated data cannot be placed into the correct columns in the dataset.

Date and Time Requirements

Use the [ISO 8601 standard](#) for all dates and times.

Columns that contain `date` values must adhere to one of the following formats:

- YYYYMMDD
- YYYY-MM-DD
- YYYY-MM



Week dates and ordinal dates are not currently supported.

Columns that contain `time` values must adhere to one of the following formats:

- hh
- hhmm
- hh:mm
- hhmmss
- hh:mm:ss

You can provide fractions of a second, but you must separate them with a period or a comma. You can include up to six digits. Any more than six digits are ignored.

You can use a space or `T` to delimit the date from the time. For example, submitting `YYYY-MM-DDT11:50:00` results in `YYYY-MM-DD 11:50:00`.

Unless you specify a different time zone, Coordinated Universal Time (UTC) is applied. When you specify other time zones, the following formats are supported:

- Z
- ±hh

- `±hh:mm`
- `±hhmm`

The following examples demonstrate time zones that have been correctly formatted:

- `2000-01-01T04:50:00+3`
- `2000-01-01T04:50:00+30`
- `2000-01-01T11:50:00+0300`
- `2000-01-01T04:50:00+3050`
- `2000-01-01T04:50:00+30:50`



If you do not follow these guidelines for columns with date or time values, then those values parsed as text and may cause a processing error.

Boolean Requirements

The platform interprets the following values for Boolean data as true or false:

- **True**
 - `yes`
 - `true`
 - `1`
- **False**
 - `no`
 - `false`
 - `0`

The platform converts text in Boolean data types to lowercase before comparing it to these values.