

Import Google Sheets Data

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

- Cloud Dataprep Standard by TRIFACTA® INC.
- Cloud Dataprep Premium by TRIFACTA INC.

Cloud Dataprep by TRIFACTA® INC. can directly import Google® Sheets® spreadsheets.

The sheets of a spreadsheets can be imported as:

- Individual datasets
- A single dataset

Limitations:

NOTE: This integration provides access to all Google Sheets in the connecting user's account. Access includes spreadsheets with disabled options for download, print, or copy as well as hidden sheets within spreadsheets.

- Import-only support

NOTE: After you import a Google Sheet into Cloud Dataprep by TRIFACTA INC., renaming the source Google Sheet or a tab in it can break your datasets and flows. Details are below.

- Creation of a dataset with parameters from Google Sheets is not supported.
- Connected sheets or embedded external datasources in your Google Sheets are not supported.
- A Google Sheet can contain up to 5,000,000 cells. Each cell can contain up to 50,000 characters.
 - Cloud Dataprep by TRIFACTA INC. supports a maximum of 25,000 characters in a cell.
- Filepath and source row number information is not available from original Sheets. These references return values from the CSV files that have been converted on the backend. For more information, see *Source Metadata References*.
- Source Sheets files with cells bracketed by single double quotes may not be properly ingested if any terminating quotes are missing.

Tip: You can check the data quality bars for mismatched values or, for strings, the data histogram bars for anomalous values to see if the above issue is present. If so, deselect Detect Structure on import. Then, use a Split rows transformation applied to the affected column to break up the column as needed.

- If loading your Sheets-based dataset in the Transformer page results in a blank screen, please take a new sample. The file requires conversion again with each generated sampling.
- Latest state of the spreadsheet may not be reflected in the Transformer page due to caching. When you run a job, the platform collects the latest version of the data and converts it to CSV for execution.

Process:

1. A spreadsheet can be read directly from your Google Drive.

NOTE: When you first use the Google Sheets connector, you must enable Cloud Dataprep by TRIFACTA INC. to read all of your Google Drive data. When the connector is used, it locates only the Google Sheets data, including any Sheets that have been shared with you. All other data in Google Drive, including any Microsoft® Workbooks®, is ignored. You can then select the Sheet or Sheets you wish to import.

2. Sheets in a worksheet are ingested and written to Google Cloud Storage in CSV format.
3. CSV files are available for selection.
4. These CSV files are the source from which the imported datasets are created.

Steps:

1. In the menu bar, click **Library**.
2. In the Library page, click **Import Data**. Select the Google Sheets connection.

Tip: You can paste links that you gather from Google to select spreadsheets. To access a Google Sheet, edit the path and paste the link. Use this method for publicly available Google Sheets, too.

See *Import Data Page*.

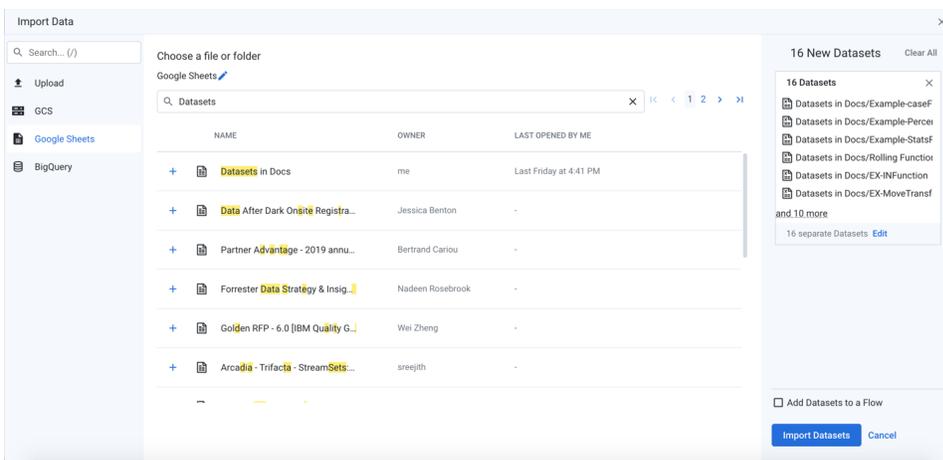


Figure: Import Google Sheets spreadsheet

3. After you select the spreadsheet, it is uploaded and converted to CSV format and stored by the platform. Depending on the size of the spreadsheet, this process may take a while.

- By default, all sheets in the spreadsheet are imported as individual datasets. To change how the data is imported, click **Edit** in the right panel.

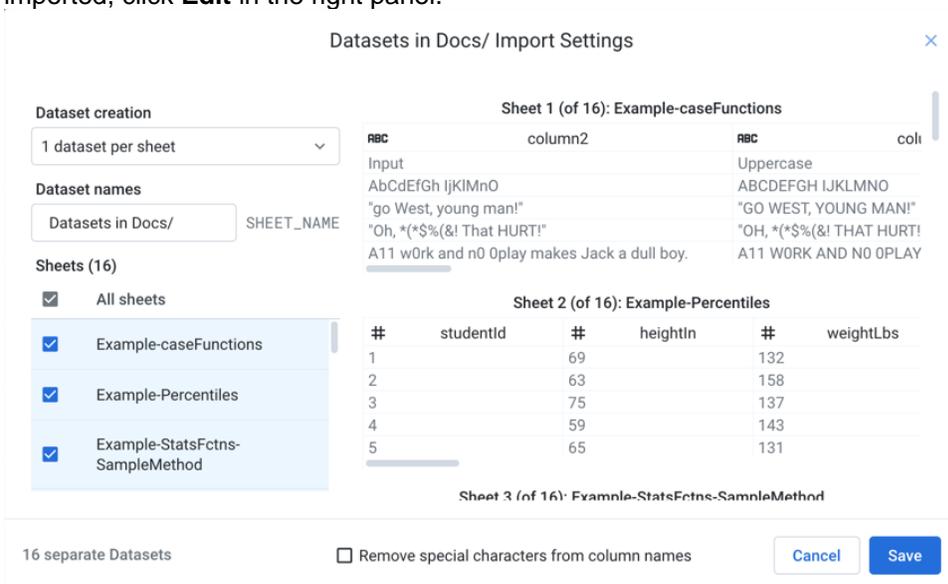


Figure: Import settings for Google Sheets datasets

- Dataset creation:
 - 1 dataset per sheet:** (Default) Each selected sheet in the spreadsheet is imported as a separate dataset. Specify the base name of the datasets that you are creating. If you are creating a single dataset, the name of the spreadsheet is used.
 - Selected sheets into 1 dataset:** All selected sheets in the spreadsheet are combined and imported as a single dataset.

NOTE: The schemas of each dataset must match. Columns must be listed in the same order in each dataset. The column headers are taken from the first selected dataset.

- All and future sheets into 1 dataset:** If the spreadsheet is updated periodically with new sheets that you would like to add in the future, select this option. After initial selection of sheets, all sheets that are added to the spreadsheet in the future are automatically added as part of the imported dataset.
- Selected sheets:
 - You can select the sheets to import.

NOTE: Special characters in sheet names are filtered out.

- To preview the data of an individual sheet, mouse over a dataset and click **Jump to**.
- Remove special characters from column names: Select this option to remove any special characters from the inferred column headers during import.
 - To save changes, click **Save**.
 - After your datasets have been added, you can edit the name and description information for each in the right navigation panel.
 - Optionally, you can assign the new dataset(s) to an existing flow or create a new one to contain them.

After import:

After you have imported the Google Sheet, you should avoid renaming the Google Sheet or any tab in it that is part of the imported datasets. If you rename a datasource, you can see one or more of the following issues in Cloudfare Dataprep by TRIFACTA INC.:

- When you open a recipe using the dataset in the Transformer page, you may receive an error that the Google Cloud Storage path cannot be loaded.
- Collecting samples in the Transformer page returns a generic error message.
- Jobs executed on Cloud Dataflow fail with generic error messages, such as `Creating the Dataflow job failed unexpectedly`.