Import Excel Data

In addition to CSV and other formats, Trifacta Wrangler can directly import Microsoft® Excel® workbooks and folders containing workbooks. The worksheets of a workbook can be imported as:

- Individual datasets
- A single dataset
- A dataset with parameters

> **NOTE:** When importing as a parameterized dataset, all selected worksheets are imported into a single dataset.

Limitations

- XLSX and XLS format are supported. Other Excel-related formats, such as XLSM format, are not supported.
- Filepath and source row number information is not available from original Excel files. These references return values from the CSV files that have been converted on the backend. For more information, see *Source Metadata References*.
- Source Excel 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.

- Macros in your Excel files are not imported.
  - During import, cell formulas are applied, and the output values are used in the imported dataset.
- You cannot import password-protected Excel files.
- Import of Excel files with protected columns or cells is not supported.
- Compressed Excel files are not supported.

- If loading your Excel-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 Excel file may not be reflected in the Transformer page due to caching. When you run a job, the platform always collects the latest version of the data and converts it to CSV for execution.

Use

When Excel data is imported into Trifacta Wrangler, each sheet in an imported file must be converted to a CSV and then ingested for use.

Steps:

1. In the menu bar, click **Library**.
2. In the Library page, click **Import Data**. Select the connection to use. See *Import Data Page*.

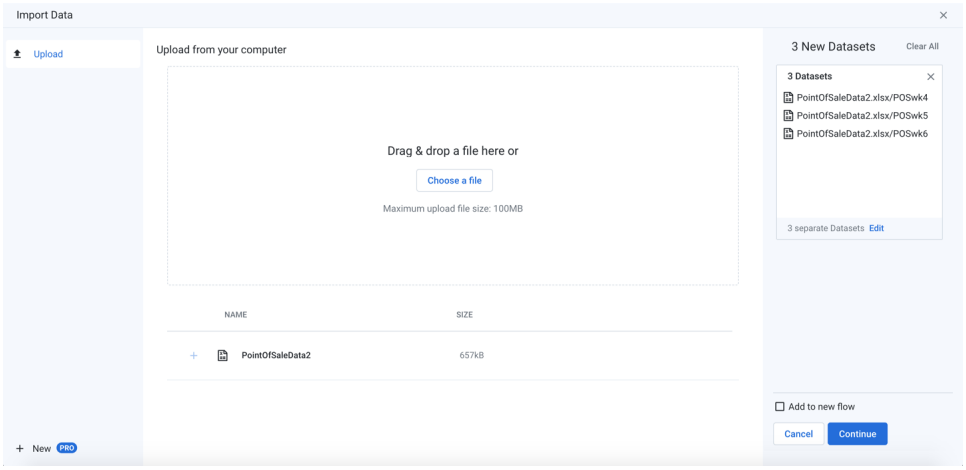

Figure: Import Excel workbook

> **Tip:** If you experience issues uploading large XLS/XLSX files, you can convert the files to CSV files and then upload them.

3. After you select the workbook, it is uploaded and converted to CSV format and stored by the platform. Depending on the size of the workbook, this process may take a while.
4. By default, all worksheets in the workbook are imported as individual datasets. To change how the data is imported, click **Edit** in the right panel.

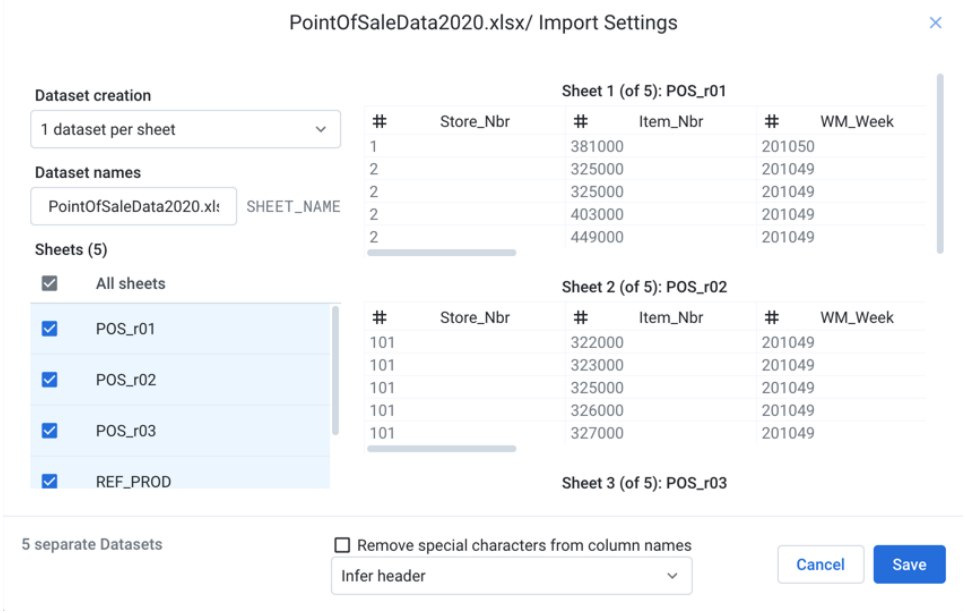

Figure: Import settings for Excel datasets

5. Dataset creation:
    a. **1 dataset per sheet:** (Default) Each selected sheet in the workbook 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 workbook is used.
    b. **Selected sheets into 1 dataset:** All selected sheets in the workbook 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.

6. Selected sheets:
    a. You can select the sheets to import.

> **NOTE:** If you are importing a folder of Excel files, data preview and initial sampling are executed against the first file found in the folder.

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