

Tips and Troubleshooting

Try these tips and techniques during your data preparation process.

Using Tableau Prep Builder

Opening and Saving Flows

- If you open a flow (.tfl) file created on a different computer or the data source has moved, Tableau Prep Builder may not be able to locate the data source. To correct the error:
 - In the flow, select the input step.
 - In the **Connections** pane, click on the drop-down menu for the connection, and choose **Edit Connection**.
- Browse to the file's location.
- Double-click to reset the connection location.
- If you have local files you want to share with your flow, save the flow as a Packaged Tableau Flow (.tflx) file, which also includes the local data.

Input

- If you know that you do not need a field in your data source, exclude it from your flow: in the **Input** pane, click the check box next to the field in the metadata profile to deselect it. Excluding fields on input results in Tableau Prep Builder bringing in less data from the source database and improves flow processing time, compared to removing the field in a cleaning step after input.



- Files in a wildcard union on input must have the same data source type and data types for fields. If you need more flexibility or have files with different data source types, add a union step anywhere in the flow. Use **Show Mismatched Fields** and merge fields.
- If your database connection supports custom SQL, you can enter queries from the **Connections** pane to select and manipulate the data before it is added to your flow.
- If the data in your data source has changed, click the Refresh button on the toolbar to refresh your data.

Flows

- Rename the steps in your flow by right-clicking on the step or double-clicking on the name of the step, so you and others can understand their purpose.
- To add a description to a step, right-click on the name of the step and choose **Add Description**.
- To change the order of steps in a flow, disconnect and reconnect steps.
- To disconnect a step in a flow, right-click on the line connecting the step and click **Remove**.
- To reconnect a step in a flow, drag the previous step to the step you want to connect it to, and release when **Add** is highlighted.
- Click **Add Clean Step** to insert a cleaning step at any point in your flow.
- To see the data in your flow before any changes, insert a cleaning step after the input step.
- There are no limits to the number of steps you can have in a flow, but if no actions are performed in a cleaning step, consider removing it to unclutter the flow and improve readability.
- Add a cleaning step between steps to branch data in order to perform steps that you want to include in a different output file.

Changes Pane

- Click on an operation in the **Changes** pane to see the data at that point in the cleaning changes, and to more fully understand the effect that each transformation has on the data.
- If you need to change the order of operations in a step, drag and drop the operations in the **Changes** pane.
- Edit an operation in the **Changes** pane by right-clicking on the operation and choosing **Edit**.
- Add a description to an operation in the **Changes** pane by right-clicking on the operation and choosing **Add Description**.

Cleaning Operations

- Edit an operation for a field in the profile pane by right-clicking on the annotation in the field header and choosing **Edit**.
- It is recommended that you check the group and replace results that are generated by the algorithms for pronunciation, spelling and common characters. The results are a starting point that you will likely need to adjust manually, especially if you are working with a sample.
- If the data source has changed, the group and replace algorithms for pronunciation, spelling and common characters will not automatically find and group new misspellings or data entry errors. You will need to run the group and replace algorithms again and then make manual adjustments as you did previously.
- In addition to the built-in Tableau Prep Builder cleaning operations, you can use RegExp functions in a calculated field to perform more complex pattern matching cleaning operations.
- Hover over the **Multiple Cleaning Operations** annotation icon (an ellipsis) above steps in the flow to see which cleaning operations are performed in that step.
- Remember that any time you are editing, excluding, or filtering specific values, all the rows that contain those values are changed/removed.

Combining Data

- When trying to decide between a join and a union, recall the common use cases for each:
 - Use a union to combine data with the same structure, so that rows are added for each table.
 - Use a join to combine related data using common fields. Typically, a join adds fields (columns) to the data structure.
- Combine data before cleaning operations when possible, so that you do not need to duplicate the cleaning operations.

Unions

- If you want to rename the **Table Names** field, which is automatically created in a union, add a cleaning step after the union.
- You can merge mismatched fields in a union step or in a cleaning step at any time in the flow.

Joins

- Consider the aggregation level of both tables before joining. If they are not at the same level, consider adding an aggregate step before the join.
- If tables are joined and they are at different aggregation levels, you could be overcounting values in the joined data set. You can investigate this by looking at the number of rows in the join results under **Join Summary**. Make sure that you do not double-count the same values that are in multiple joined rows.
- You can experiment with your join configuration to get exactly the join results you want.

Pivots

- For some analytic questions, you need your data to be tall and thin, for example analyzing years or survey questions. If the data you want to analyze together is in different columns, pivot your columns to rows in Tableau Prep Builder.
- You can also pivot from rows to columns.

Aggregates

- When adding an aggregate step, remember that the data after the aggregation cannot be shown at a more detailed level; for example, if you aggregate rentals by month, you can no longer see details by day.
- In addition to aggregating numeric values, you can also aggregate fields that are dates or strings, for example, to get a distinct count of states by year for the orders data.
- You can experiment with your aggregate configuration to get exactly the aggregate results you want.

Split, Custom Split and Date Fields

- **Split** and **Custom Split** only work on **String** fields. If you want to split a **Date** or a **Date & Time** field, change the field's data type to a **String**.

Previewing in Tableau Desktop

- You can preview your data sample from any step in your flow, other than an input or an output step.
- Every time you select the **Preview in Tableau Desktop** operation, new files for the Tableau extract (.hyper) and Tableau data source (.tds) are created by Tableau Prep Builder in the **Datasources** folder of your Tableau repository. These files are not deleted by Tableau Prep Builder, so when you are done using them, delete them manually.



Creating Output Files

- An output will contain all cleaned rows, regardless of which sampling options are configured.
- If fields have been excluded or rows filtered in the **Input** pane, they will not be included in an output file. If you want to include them in your output, you must change the input configuration.
- Use an output step to branch in order to produce an output at a certain stage in your flow, such as before a filter, join, or aggregation.
- By default, output files are saved to the **Datasources** folder in your Tableau Prep Repository.
- If an output file cannot be found, click **Browse** and edit the location to use the correct file path and name.
- You must have appropriate permission to write to the output file.

Finding "Missing" Data

If you do not see data that you expect to see, check the following:

Item	Suggestions
Sampling configuration in the Input pane	<p>Check the amount of sample data included in the flow.</p> <p>Check the sampling method for the sample data included in the flow.</p> <p>Increase the sample size to include more data or choose random sampling to decrease the sampling bias. However, both will degrade the application performance.</p>
Fields excluded in the Input pane	Look for fields with their check boxes cleared in the metadata profile to check which fields are excluded.
Field values filtered in the Input pane	Look for fields with a Filter annotation icon in the metadata profile to check which field values are filtered.
Field values filtered in a flow step	Look for the Filter annotation icon above steps in the flow to check which field values are filtered.
Field values edited or grouped and replaced in a flow step	Look for the Group and Replace annotation icon above steps in the flow to check which field values are edited or grouped and replaced.
Fields merged in a flow step	Look for the Merge Fields annotation icon above steps in the flow to check which fields are merged.
Fields removed in a flow step	Look for the Remove Field annotation icon above steps in the flow to check which fields are removed.



Item	Suggestions
Renamed fields	Look for the Rename Field annotation icon above steps in the flow to check which fields are renamed.
Field values in summary distribution bins	Use the detailed distribution view to see all the values in the data sample for that field. The summary distribution bins the values automatically, sometimes in a bin size that you did not expect; some values are not obvious.
Fields merged in a union step	Click on field headers in the Union Results pane to check which fields are merged.
Data excluded from a join	<p>Check for values in the Join Clauses pane listed in red. These are values without a match in the join clause.</p> <p>Click on the Excluded bars in the Summary of Join Results to see the profile and values for rows excluded because of mismatched join clause values.</p> <p>Look at the join type diagram to check the join type used. Click on the join type diagram to test different join types.</p>
Pivoted fields	Look in the Pivoted Fields pane to check which fields are pivoted.
Fields not included in an aggregate step	Look in the Grouped Fields and the Aggregated Fields panes to check which fields are included in the aggregate results.

Using Tableau Desktop with Tableau Prep Builder

- When working iteratively cleaning data in Tableau Prep Builder and validating your work in Tableau Desktop, if the data you are looking at in Desktop is not as you expect, check that you selected the correct step in the flow to preview your data, or that you are using the correct output extract.
- When validating your work in Desktop, you can keep using the visualizations you have already created.
 - After you have performed more cleaning, preview in Tableau Desktop to generate a new extract.
 - Close the new blank workbook that just opened.
 - In the Tableau Desktop workbook with the visualization that you have open, click the **New Data Source** button on the toolbar.
 - Browse to the **Datasources** folder of your Tableau Repository in your Documents folder and open the .hyper extract for the sample data that you just created.
 - Go back to the worksheet, make sure the original data source (with a blue check mark) is selected, and from the **Data** menu, click **Replace Data Source**, and then click **OK**.
 - You can close data sources that you are no longer using.
 - NOTE: If you have visualizations that reference fields that have changed, those fields will be shown in red and you will need to update the field references.
- To create an output file in Tableau Prep Builder after creating visualizations using the sample data, use the same steps described above, but point to the newly created output extract file in the location you specified.
- If you make metadata changes in Tableau Desktop, such as renaming a field, Tableau Prep Builder will not be aware of those changes unless you create an extract in Desktop and use that as a data source in Prep Builder.
- Extracts created in Tableau Prep Builder are connected to Tableau Desktop, live, as flat files. To update an extract, re-run the flow in Tableau Prep Builder.

Actions on a Server

Tableau Prep Conductor

Tableau Prep Conductor is a separately-licensed add-on feature for Tableau Server or Tableau Online. It enables users with appropriate permissions to:

- Publish flows from Tableau Prep Builder.
- Publish custom data roles from Tableau Prep Builder.
- Publish reusable steps as flows from Tableau Prep Builder.
- Add or embed credentials for cloud connectors from Tableau Prep Builder.
- Leverage parameters in an Initial SQL statements when connecting to databases that support them.
- NOTE: TableauServerUser and TableauServerUserFull are not supported for flows.
- See alerts in Tableau Prep Builder when features within Prep Builder are incompatible with the currently installed version of Tableau Server. Choose to use or to disable incompatible features.
- Create schedules for flow tasks on Tableau Server.
- Check data-connection credentials for the data sources used by a flow.
- Send alerts if a flow fails.
- Monitor user activity and performance of flows using Administrative views.

Tableau Prep Conductor is licensed through the Data Management Add-on. To use Tableau Prep Conductor, you must activate a Data Management Add-on product key on your server.



Add a Script

Starting in version 2019.3.1, users can use R and Python scripts to perform more complex cleaning operations or to incorporate predictive modeling data into a flow. Data is passed from the flow as input through the R or Python script step, then returned as output data that users can continue cleaning using the features and functions of Tableau Prep Builder.

- To include Python scripts in a flow, users need to configure a connection between Tableau Prep Builder and a TabPy server. Then users can use Python scripts to apply supported functions to data from a flow using a pandas dataframe. When users add a script step to a flow and specify the desired configuration details, file, and function, data is securely passed to the TabPy server, the expressions in the script are applied, and the results are returned as a table that users can clean or output as needed.

NOTE: To run flows that include script steps on Tableau Server (version 2019.3 and later), Tableau Server must also have a connection to a TabPy server.

- To include R scripts in your flow, users need to configure a connection between Tableau Prep Builder and an Rserve server. Then users can use R scripts to apply supported functions to data from a flow using R expressions. After users enter the configuration details and point Tableau Prep Builder to the desired file and function, data is securely passed to the Rserve server, the expressions are applied, and the results are returned as a table (R data.frame) that users can clean or output as needed.

NOTE: To run flows that include script steps on Tableau Server (version 2019.3 and later), Tableau Server must also have a connection to an Rserve server.

