



Free Training Transcript:
The Union Step

Welcome to this video on the Union Step in Tableau Prep. You can download the data set and packaged flow file underneath the video to follow along in your own copy of Tableau Prep. We're working with data for bestselling books.

Unions

Performing a union lets you combine data from multiple tables by bringing in new rows or records that match the structure of the original data (that is, the fields or columns are the same). For example, if we have multiple weeks of bestseller data as individual files and want to create a unified data source of all the weeks together, we can do a union.

Wildcard Union

If we want to union tables with the exact same structure or schema, a wildcard union allows us to perform the union during the input step (if we're using flat files). This keeps the flow tidy and simplified by adding a single data input to the flow, even if it's made up of lots of tables behind the scenes.

We'll add a data connection, to an Excel file, and we'll chose February 28th. Bring a table to the flow, and in the input step, click to the multiple files tab, and then chose wildcard union. Here, this file contains 7 sheet tabs for 7 different bestseller lists, as seen over in the connections pane.

A wildcard union requires all the files to be of the same type (all Excel or all CSV, for example). By identifying a pattern to define what content to include, multiple tables can be unioned automatically. Note that there is no ability to perform a merge on mismatched fields with a wildcard union.

We can set which folder to search in, including sub folders or not, set whether to include or exclude the files that match our pattern, and specify the pattern itself. If we want to include any files in the same folder that start with "ABA Bestsellers" we can simply remove everything but those words, and add an asterisk instead. Now, below, note that we're currently looking at 1 file and `7 sheets. But when I hit enter to apply this matching pattern, we now have 9 files and 37 sheets.

We can also set the pattern for sheets we want to include or exclude, such as only bringing in sheets for Trade Paperback Fiction or Nonfiction, by using the pattern Trade* (and again hitting enter or "Apply"). But we'll undo that and go back to the default, which

is just this file and all the sheets inside it.

Regardless of how broadly or narrowly we define the wildcard union, note that two fields are added to the list of fields automatically. Table Names will convey the table (or sheet tab) name, and File Paths will convey the original file path. This helps keep track of where data is coming from.

The Union Step

To perform a union outside of a wildcard union during the input step, we can add a union step to the flow. We need to have at least two inputs steps to union together. Here, we have four weeks worth of bestseller data. To perform a union in the flow, we can add a union from the plus menu, selecting “Add Union” and dragging a step to it, or we can drag one step onto another, choosing the drop area for New Union.

Unlike a join, which can only support two tables, a union can be made up of many tables. Simply drag another table onto the union, choosing the “add” drop area. Selecting “New Union” will add a subsequent union step in the flow.

The Union Configuration Pane

The union step in the flow opens the configuration pane below. The left side offers a summary--it shows the inputs and resulting fields. Here we see there are two mismatched fields. Over to the right, we see the union results. The colored bars on the field cards match the inputs and show us where the data in these fields is coming from. We can quickly see that the “Info” field is missing the week of the 14th, and we have an “Information” column from just that week. Clearly these should be the same field.

To make it a little easier to see, let’s check the “Show only mismatched fields” box. Now we see just those two fields, and we can resolve the issue by merging them--simply drag one card onto the other. I want the field to be named Info, so I’ll bring Information onto Info. Now there are no longer any mismatched fields. Unchecking the box shows us the full results again, and we can see each card has all 4 color bars.

The exception is this Table Names-1 field. Tableau Prep created this field for us so we can see where the data came from. This field is redundant information with the original Table Names and File Paths (from the wildcard unions in the input steps) so we can remove this field. We can’t do it from the union step, We need to add a cleaning step and then we can remove that field. Now we have clean, consistent data for all 4 data sources combined in one data set.

Conclusion

Thank you for watching this video on The Union Step in Tableau Prep. We invite you to continue with the free training videos to learn more about using Tableau.

