



Free Training Transcript:
The Cleaning Step

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

The Context for Cleaning

Data can be unready for analysis for a lot of reasons. If you need to do things that have nothing to do with shaping the data (like pivoting or aggregating) or combining it (through joins or unions), then it's likely you'll need to perform some cleaning operations. Cleaning in Tableau Prep covers everything from removing fields, changing data types, creating calculated fields, and more.

How to Clean

Cleaning data in Tableau Prep is done in a Cleaning step in the flow. From the previous step, click the plus icon and choose "Add Step". The "default" step is a cleaning step, which appears in the flow as a bar. When we're on a clean step, below we see the profile pane and data grid.

Each field in the data is represented as a card in the profile pane. We can rename fields or change their data type by interacting directly on a card. If we open a card's menu, we see many other options. Let's go through brief examples of each of these.

Remove Field: First, this "Table Names-1" field is an artifact of the union. We don't need it—simply select the card and go to the menu, choosing "Remove Field".

Rename Field: Next, we'll rename a field—here, "Table Names" is actually the bestseller list, so we'll double click on the name and enter "List" instead.

Split Values: The information about which week's worth of data we're looking at is contained in the File Paths field, (this is another field generated automatically by Tableau Prep). We want to keep only the date portion of these file names. Let's use do some cleaning operations to get there.

We'll open the menu, and select "Custom Split". For the separator, we'll enter a period, and we'll choose to only "split off" the first field. This will look for the period and only keep what occurs before it—essentially trimming the file extension for us.

We now have a new field—"File Paths – Split 1"—that lacks that file extension. Let's

remove the original field, “File Paths” (note that in Tableau Prep, we can remove fields even if they’re used for calculations or splits.) and now we just have the new split field.

Clean: However, this new field still says “ABA Bestsellers” before the date. We can use some of the built in quick cleaning options to address this. We’ll open the menu and choose clean. These are the single-click options available to speed up these common cleaning processes. We’ll chose “Remove letters” and once again go back to the menu, choose clean, and we’ll follow up with “Trim spaces”. Now we’re left with just the numeric date, exactly as we want. Let’s name this “Week”.

Change Data Type: Tableau Prep hasn’t recognized this as a date yet, but we can set the data type here, by clicking on the icon for String (Abc) and choosing Date instead.

Let’s also change the data type for the Price field to be a decimal number—note how the view updates from showing the details (grey bars with text) to summary view (blue bars, binned on an axis).

Filtering: Next, let’s take a look at filtering. For some data types, such as date, we have the ability to filter to a range of dates or relative date using a filtering interface. Numeric fields, like Rank, also have a filtering interface for a range of values. For most field types, filtering is done simply as a calculation. For example, we could filter the List field to only values containing “nonfiction”. But we’ll undo that.

Group and Replace: Over in the Title column, we realize there are some similar values—for example, “12 Rules for Life” and “12 Rules for Life” with a subtitle as well. We know these should actually be considered the same value. We can use the Group and Replace feature to combine these values and replace them with a single value.

In the menu, choose Group and Replace, and we’ll chose “Manual Selection”—for more information on the other options, check out the video on Group and Replace. In the editor, select the first value we’d like to group—this will become the value that overrides the others.

Now, to the right, we’ll pick the value that should be grouped in with that value. We can repeat this process with “Enlightenment Now”, and any other titles that should be the same. The paperclip icon indicates which values have been grouped. When we click done, those values are combined and replaced with the primary value we chose. All records containing those values will be updated when we run the flow.

Calculated Fields: Finally, let’s say our standard is that for books with two authors, or an author and illustrator, we use the word “and” instead of an ampersand. Let’s use a calculation to fix this. We’ll open the menu for the Author field card and select Create

Calculated Field. If we name this “Author”, it will simply update the field rather than creating a new field entirely--very handy. We’ll enter our calculation and click save. If we search again, we see there are no ampersands.

Edit Value: We can also directly modify a value--simply right click and choose Edit Value and then make the desired modifications directly inline. This will update the one value (on every row where that value appears). It’s great for fixing specific things, but more programmatic changes, a calculation might work best.

Annotations and Changes

It can be easy to lose track of what we’ve done in a cleaning step since it can contain so many types of operations. Annotations are added to cards that we’ve worked on, and the cleaning step in the flow gets annotations to show what has been done, with icons for each of the main types of cleaning.

Even more powerfully, the Changes, here to the left, shows a log of everything that was done in that step. If we want to undo a specific change, such as removing this field, we can delete it--without losing everything we did afterwards.

If we want to modify a calculation, we can right click on it in the Changes and edit it directly. Note that when we click back into a change, we’re going back to the data as it was at that point. To see the entirety of what we did in a cleaning step, we need to click onto the last change.

To keep the flow organized, we can create a cleaning step at any point. This starts with a clean slate of no annotations and no changes--allowing us to perform cleaning operations that make sense together in a single step, with other steps for other sets of operations.

Conclusion

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