Free Training Transcript:
The Join Step

Welcome to this video on the Join 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 and movie adaptations.

Joins

Performing a join lets you combine data from two tables by bringing in new fields or columns based on a shared field or fields. For example, if we have book bestseller data and want to add fields for movie adaptations and their release dates, we could join the bestseller information with the movie information based on a shared field such as title.

Let's do exactly that and see how it works in Tableau Prep. We need to have two inputs or steps to join together. Here, we have the combined 2018 bestseller data input step, and a cleaning step for the Movies data.

To create the join, we can either use the plus icon menu and select Add Join, then drag the second step to it, or we can simply drag one step onto the other, and chose the New Join drop area. If we want to name the steps so the join configuration pane is easier to understand, we can do so right in the flow pane.

The Join Configuration Pane

When we create a join, the pane below shows us tons of information about the join. To the far left, we can see the join clause area. Tableau Prep will pick a join clause by default based on the fields in the data, but we can add another by clicking the plus and selecting the fields, and we can modify the operator if we choose. And we can remove join clauses simply by clicking the X (it appears when we hover).

Below that, we see the Join Type. The step that is dragged is considered the "right" table, as we can see here (we dragged Movies onto Books). We can click on parts of the diagram to change the join type--and the summary below updates to show what records are included or excluded based on the join clause and type. Here, for example, we see 9 movies are matching 93 rows of books (because books can appear in multiple lists and for multiple weeks), and 61 movie adaptations do NOT have books on the bestseller lists. If we switch to a left join, those books without movie adaptations, indicated by the hatched blue bar, move to the Included column and our results are now 1300 rows. Let's stick with this.

We can click on the summary of join results bars, such as the mismatched values, or the included ones, or the default, the results. This controls what we see over to the right.

Join Results

The join clases pane shows all the values in the join clause. Red values are not matched across data sources; black values indicate a match for the join clause. Further to the right we can see a profile pane-like view of the results. Looking at the red, mismatched values in the Join Clauses, I see that "A Wrinkle in Time" is in red...in both columns. Shouldn't that be a black, joined value, then? There must be something going on here.

Each title looks identical, which makes me think there must be a rogue space somewhere causing trouble. If we go back to the cleaning step for Movies, we can use the Clean option and Trim Spaces. Let's see if that fixed the join, now we have 11 movie results, and A Wrinkle in Time is matched. It's easy to identify errors with the visual join interface and verify the results.

If we add a step after the join, we see that the fields that share a name across both sides of the join are appended with a -1 to distinguish the new fields. We don't need the author and title from the Movies side of things, so we can remove those fields. Now we have tidy, joined data ready for the next step.

Conclusion

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