Free Training Transcript:
The Aggregate Step

Welcome to this video on the Aggregate Step in Tableau Prep. You can download the data sets or 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 Aggregation

An aggregate step in Tableau Prep is used to change the level of detail of data so it's less granular, often in preparation for being combined with other data at a higher level of aggregation. (For more information on what aggregation is and why it can be a necessary part of preparing data, check out the Aggregation video in the Basics of Data Prep section.)

In our example here, we have bestseller information about books-- in this data set, a row is a book's information for a given week. We also have data that's transactional data about sales from a bookstore. The sales data is a row per transaction. If we were to join this data, we'd wind up with replicated bestseller information for each transaction of a given book.

Instead, what we want is the bestseller information to be supplemented with a count of the copies sold and the average discount for that book. The data should maintain the one-row-per-book granularity of the bestseller data.

How do we get there? We'll aggregate the sales data to be book-level rather than transaction-level.

How to Aggregate

First, we need to add an aggregate step in the flow. Click the plus on the previous step and select Add Aggregate. The aggregate pane opens below, allowing us to configure this aggregation.

On the left we see the list of fields in this data source. On the right we have two drop areas, one for grouped fields and one for aggregated fields. In order for a field to "come out the other side" of an aggregation step, it must either be grouped or aggregated.

Grouped fields determine the granularity of the row. If we want our post-aggregation data to be the number of books sold per date, we would group by book and date. If we want our post-aggregation data to simply be book (which we do), we'd group only by book.

Whatever fields we want to be aggregated to the level of the grouped fields (in this case, number of books sold, which means the count of transactions, and the average discount)

will be brought to the Aggregated fields area and have their aggregation operation set.

So let's do it. We'll group by ISBN as we know that's in the other data set and it's more specific than using book and author (as ISBN accounts for different formats of the same title). Simply drag the ISBNfield to the grouped fields area.

Next, we'll drag Transaction ID to the Aggregated Fields area. Because this is a string data type, the default aggregation is Count. That's what we want, so we can leave it. Finally, we'll drag Discount to the Aggregated Fields area. As a numeric field, the default aggregation is SUM, but we can change this to AVERAGE by clicking where it says SUM and choosing AVERAGE from the dropdown.

To see the results of this step, let's add a Cleaning Step and look at the profile pane. We can see in the contextual toolbar that we have 3 fields and 19 rows. Only three fields came through--ISBN, Transaction ID, and Discount, since those were the only fields we included in the aggregation pane. We should rename these so they're more correct--instead of Transaction ID, we'll name this "Number of Books Sold" and for Discount, we'll add "Avg".

Modifying an Aggregate Step

If we wanted to keep more of the original fields--say, author and title--we can click back into the aggregate step and add them to the Grouped Fields area. When we click back into the Clean step, we're still at 19 rows. Those two fields, author and title, didn't change the granularity of the data because they're already accounted for in the ISBN.

If we go back and add Date, we jump to 121 rows, because the granularity of what makes up a row in the data is now ISBN and date. Let's go back to our original 3 fields and 19 rows.

Our data is now fully aggregated to the desired level of detail and we can continue on to the next step in the prep process, such as joining this with the bestseller data.

Conclusion

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