

Tableau Prep Hands-On

Reference Booklet



Welcome!

Thank you for taking the time to attend our Virtual Prep Test Drive! Test Drives serve to give you a short introduction to the product so that you feel confident to continue the exploration with your own data.

This booklet will provide you with resources needed for the test drive, details the concepts covered so that you have something to refer to after the session, as well as give you an avenue to provide your feedback and questions.

1. Pre-Requisites and Important Links

Important: What you need so that you can follow along in the session.

Item	Link	Tips
Downloaded version of Tableau Prep Trial	https://tabsoft.co/prep	If you already have it, you're encouraged to download the latest version from the link provided
Download the datasets	http://bit.ly/preptest20	This is the dataset that will be used during the session
Question form	https://tabsoft.co/prepqquestions	In case you need to leave the session earlier, we'd request you to fill in this form if you have follow-up questions, feedback or would like to be part of our growing prep community!



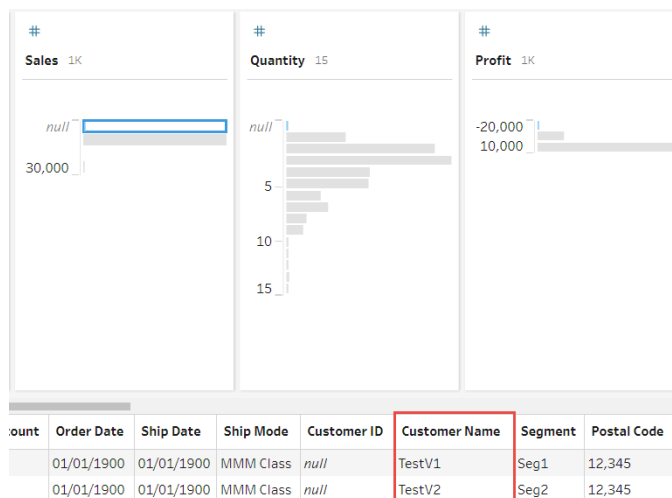
2. Concepts covered during the Hands-On (Recap)

We recommend reviewing this section **after** completing the Hands-On.

During the Hands-on, you played the role of an analyst at an organization called APAC superstore and you were tasked to bring together datasets from 4 different regions. Each of these datasets had their own cleaning challenges and some of the data was stored in different databases. The scenarios and the concepts covered are as below:

2.1. Data Cleaning Operations:

- Dealing with Null Values (*South-East Asia Dataset*):



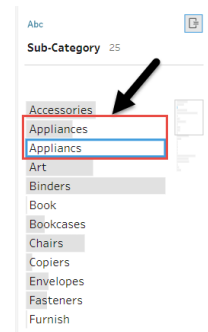
The key challenge here was the presence of null values in the Sales and Quantity column. We often notice this in datasets where there have been certain transformations done (like joins or unions).

Leveraging on the flow, profile and grid pane, we were able to understand what was behind the null values and could remove them from our analysis. You can [read about these three panes here](#).

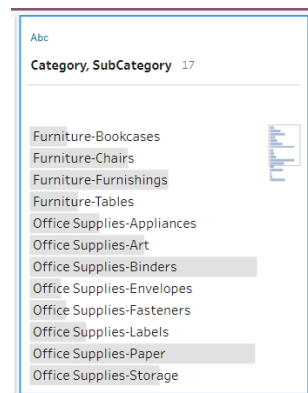


– **Dealing with misspelt and invalid entries (South Asia Dataset):**

Here we came across the challenge of misspelt values in the column 'Sub-Category'. We often notice this come up in areas where we rely on manual user input (like surveys). By leveraging on some of the built-in clean up functions available in the toolbar like removing spaces and punctuations and some smarter machine-learning based algorithms like 'Group and Replace', we were able to remove the invalid entries. You can read more about [the toolbar options here](#).



– **Separating out combined values (Oceania dataset):**



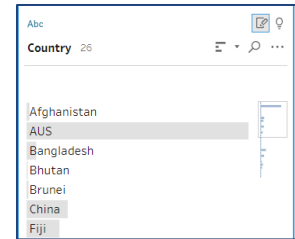
Here we came across a challenge where a single column had combined values that needed to be separated out. This is generally very commonly observed in situations where data is brought in from web-based sources like Facebook ad reports or Google Analytics. To clean this up, we leverage on the split function, which is also accessible from the toolbar (*please reference the link above*).

– **Cleaning invalid or mistyped entries that fit a certain format (Cleaning countries and customer name values)**

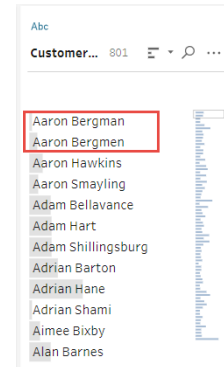
Post the union step (below) we observed concerns with mistyped entries. This often comes up in situations like surveys or registration forms where we ask users to type in their responses. Breaking this up further, there are two categories of mistyped entries:



1. **Mistyped entries that fit a generic convention:** This would include email addresses, website links, geographic information, etc. In the session, we came across mistyped entries in the 'Country' column. We used 'Data Roles' to identify these mistyped values.



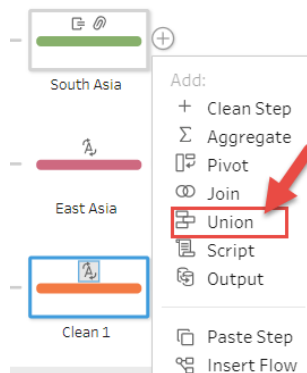
2. **Mistyped entries that should align to a custom list:** This is generally when we expect people to type in entries which we would like to cross reference with an internal dataset. For example, when people add in their organization names to a survey or a registration form. In this case, it was customer names that we wanted to check against a pre-defined list of customers where mistyped customer names are flagged as invalid. Here we used something called 'Custom Data Roles' along with the group and replace step to flag out and replace any invalid entries.



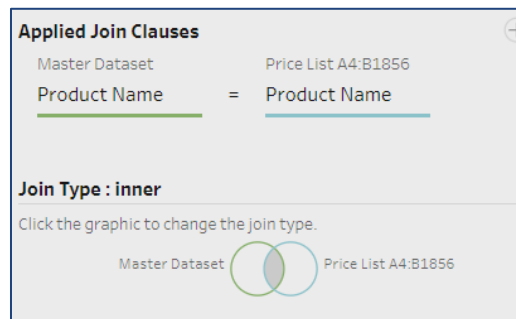
You can read about [general data roles and custom data roles here](#).

2.2. Creating a Master APAC Dataset (Union the 4 datasets):

The last challenge was bringing the 4 individual regional datasets together into one master dataset. We often see this come up in organizations where data is generated at different points in time or from different regions, and it needs to be stacked one after the other. We brought this together using a 'Union Step'. You can read more about [the 'Union' Step here](#).



2.3. Adding in additional price information (Joining Additional Data)



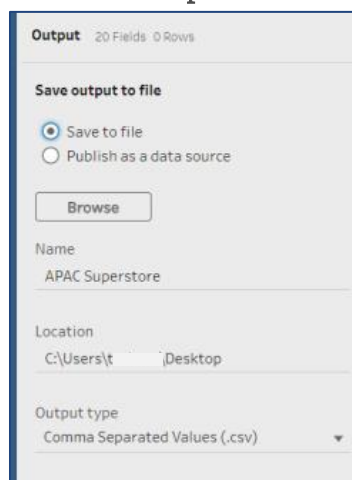
As a bonus exercise we needed to add on price data to our existing orders dataset. This often comes up when you need to bring in additional information from different data sources together. For example, orders data from a SQL database with budget data from an excel sheet. We brought this data together using a 'Join' Step. You can read more about [the Join step here](#).

2.4. Sharing the final output dataset

After doing all those transformations, we were ready to share the final data output with different audiences.

You can read more about [saving and sharing the output here](#).

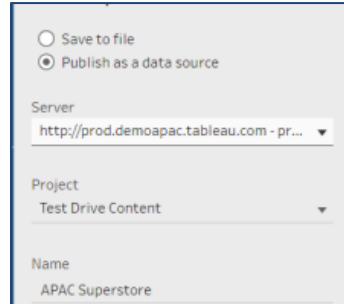
- Saving the data as a file in the local computer



With audiences that need to access the data outside the Tableau platform, we generated the data as a text file. You could also have created this as a Tableau recognized extract. These files would be saved locally to your computer.

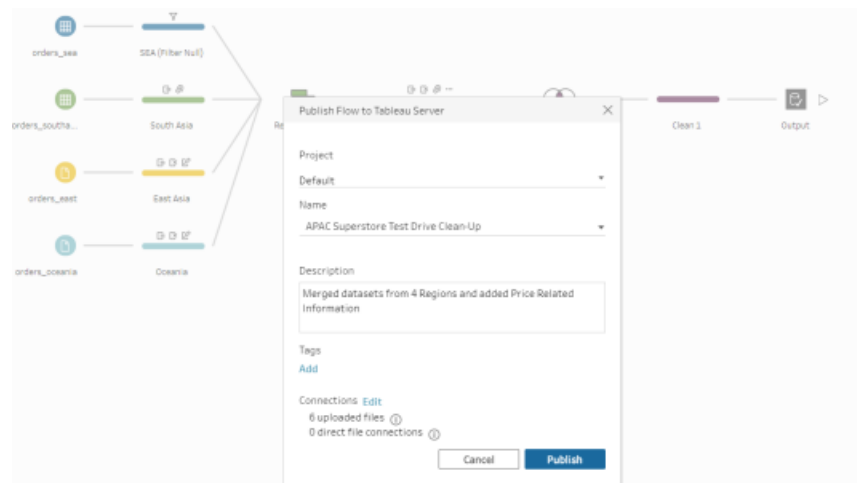


– Publishing the Data as a Data Source:



With audiences that are using Tableau, you can save the data as a published data source. This would enable your end users to explore this cleaned dataset using Tableau's regular drag and drop capabilities or by using its natural language capabilities like Ask Data.

2.5. Keeping the dataset refreshed securely



To ensure this dataset stays refreshed, it can be scheduled and set-up to give end-users details about the dataset itself, the entire flow can be published to Tableau Server or Tableau Online. This leverages the capabilities of Tableau Prep Conductor. You can read more on [‘Publishing a Flow’](#) here and on [‘Tableau Prep Conductor’](#) here.



3. Additional Use-Cases

If the topics covered today seemed relatively straightforward, you can explore some deeper use cases and features that our customers frequently use Tableau Prep for:

1. Alternative outputs for alternative teams at different levels of aggregations:

Very often business users have different requirements from a similar data source. For example, from the same sales data source, the customer success team may want to understand the sales per customer so that they can analyze their customer spend, the product team may request for the sales per product SKU so that they know what was the best-selling product variety. In that case, customers often leverage the [‘Aggregate’ function](#), which you can read more about here.

2. Making data structures more ‘analytics friendly’

We also often see datasets where the data is formatted in a way which is perfect for humans to read but is often very hard to use in an analytics software. It often needs to be transformed such that it’s tall and long as opposed to short and wide. To do this, customers often leverage the ‘Pivot’ function in Tableau Prep. You can read more about the [Pivot Function here](#).

3. Leveraging on models from R or Python

There are times when we noticed customers would like to extend Prep’s native capabilities and leverage on advanced data models from their R or Python environments. For example, let’s say the data science team has developed an advanced forecasting model built in Python and the Sales team wants to bring that into their Sales Output data. In that case, customers often leverage the [‘Script’ function](#) that you can read about here.

4. Bringing together published data sources

Very often the first cleaning operation is only the beginning and sometimes we need to bring together data sources that are published to Tableau Server or Tableau Online. One of the data connectors in Tableau Prep is a Tableau Server connector. This means that end-users with the appropriate permissions can bring in end-data sources and clean and transform it for themselves. We have observed our customers really unlock the idea of self-service data preparation by leveraging this. You can read more about [connecting to ‘Published Data Sources’ here](#).



4. Resources to continue your Learning Journey

As shared during the test drive, we are looking forward to you continuing your exploration and today's session was just to give you a flavor of the product. Here are some resources for you to continue your learning journey with Tableau Prep.

Question form	https://tabsoft.co/prepquestions	If you have questions, would like to share some feedback or would like to be part of our growing Prep Community, fill in this form.
Free eLearning Courses	https://tabsoft.co/elearning	If you'd like to explore a structured training course, we encourage you take this step-by-step e-learning course which is currently available for free for 90 days.
Free Online Tutorials	https://www.tableau.com/learn/training/20201#prep	If you'd like some short tutorial videos, do check out our online-help videos.
Level Up through weekly challenges	https://preppindata.blogspot.com/	If you're looking for some weekly challenges, we recommend checking out the Preppin Data blog.
50% off Desktop Specialist Exam	https://tabsoft.co/desktop50	Although not linked to Tableau Prep, for those of you who would like to get certified in Tableau Desktop, we would encourage you to check out the certification exam.

We trust that the session proved useful to you. For any follow up, do leverage the Question form or reach out to your account manager.

We wish you a fun Data Preppin Journey!

