

Chrystelle Drouin – Solutions Consultant

Tableau Prep Builder Hands-on Workshop for Beginners

Follow-along Guide

Follow-along Guide

Before you start

Tableau Prep Builder

- Installed (BNPP Package from you Enterprise Store or from the Tableau website <https://www.tableau.com/products/prep/download>)
- License activated

Dataset

- Superstore Sales Orders - US.xlsx
- Product Database.xlsx
- Sales Amer 2011.csv, Sales Amer 2012.csv, Sales Amer 2013.csv, Sales Amer 2014.csv
- Sales Target.xlsx


This document is meant to be used as a help during the Workshop, not as a separate training document. You do not need to study this guide before, it can help as a guide during the session, or when you want to retrace your steps afterwards. The steps will be shown step-by-step during the session on the screen as well.




Step 1 : New data connection and first cleanings

- Open Tableau Prep Builder
- On the green button, click Connect to Data \ To a file \ Microsoft Excel
- Navigate to Superstore Sales Orders - US.xlsx and Product Database.xlsx
- On US Sales Transactions, verify that you can see sample data values in the lower half of the screen



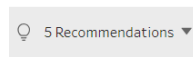
- Click on the following '+' icon , then Add Step
- Have a look at your data distribution on various fields; you can interact and select some values to see detailed data on the bottom panel
- Select the most recent Order Date bin and have a look at the detailed data related to these 'future' orders (where Order Date is between 2090 and 2100) – note that these 3 orders are related to the 'Test' Customer Name
- Select related Customer Names on the bottom panel, and Exclude the 'Test' value

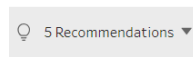


- Expand the 'Changes' menu on the left ; it will track all your changes done on this cleaning step
- Have a look to the 'Ship Mode' column; then right click on the ... symbol



- Select Group and Replace \ Pronunciation algorithm. Then check values in the 'Ship Mode' column



- Review Tableau Prep Recommendations: , Select City \ Change Data Role to City \ then click Apply
- Right click on step 'Clean 1' and rename it to 'Clean US' and optionally add a description

Here you are with a first data connection and some cleaning steps.

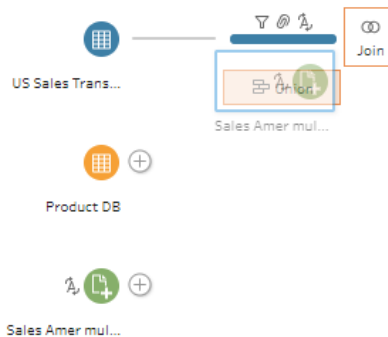


Step 2: Data Union

- Add a new connection on the left of your screen:
- Choose Text File \ then select the Sales 2014.csv file
- Review the Text Settings with 'Locale' setting to 'French (France)' and Switch to the 'Multiple Files' tab
- Select 'Wildcard union', and in the Matching Pattern (xxx*) field write: Sales*.csv and validate with the 'Enter' touch

So, all csv files with the Sales prefix will be combined into a single view

- Then click 'Apply'. This will add a new 'File Paths' column, with the name of the related csv file.
- Find the 'Revenue' field and click on the 'Abc' icon
- Rename the step 'Sales Amer 2014' into 'Sales Amer multi years'
- Drag 'Sales Amer multi years' onto the 'Clean US' step and drop it into the



orange zone 'Union'

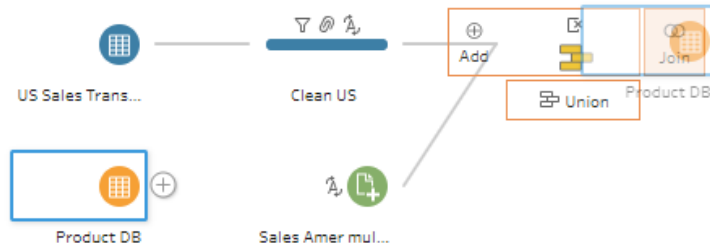
- Click on 'Show only mismatched fields'
- Select 'File Paths' column, right click \ Remove; same thing for 'Table Names' column (since these details are already included in the 'Country' column)
- Then Drag and Drop 'Revenue' column onto 'Sales'. You can then deselect 'shown only mismatched fields' and rename the 'Union 1' step as 'Sales Union'

You quickly combined as a unified view both US Sales and the historical/multi years view of Sales Amer.




Step 3: Data Join and Create an Output

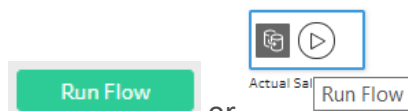
- Drag and Drop 'Product DB' onto 'Sales Union' and choose Join



- Review the join key (automatically done with Product ID) and analyze join results: you have 4 excluded rows coming from Sales Union
- Select 'show only mismatched values' option to understand what the Product ID values are
- Rename FUR-BO-10001798-US as FUR-BO-10001798, and note the impact on the join results : you do not have any more an exclusion from the Sales Union ... all your sales rows have a product ID, and related Product Name, Sub-Category and Category

Join Clauses ☐ Show only mismatched values ▼

- Click on the following '+' icon , then Add Output
- Choose 'Save to file', name it 'Actual Sales' and select the Output type you prefer (csv, tde or hyper)
- Rename the step as 'Actual Sales'



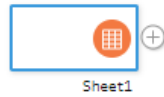
- Run the flow by clicking on **Run Flow** or **Run Flow**
- Then open the output with Tableau Desktop
- Alternative: directly Preview your data on any step in Tableau Desktop, via a right click 'Preview in Tableau Desktop'

You enriched your Sales data with more details coming from a different source and generated a file ready for analysis.

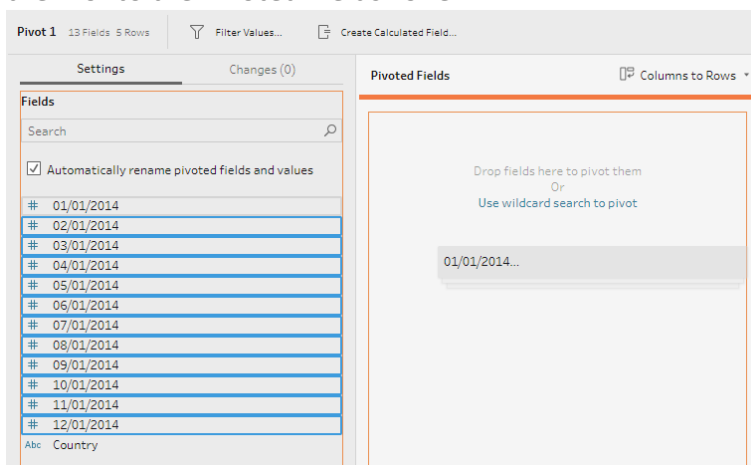


Step 4: Data Pivot and Aggregation

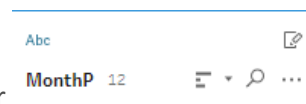
- Add a new connection on the left of your screen:
- Choose Microsoft Excel \ then select the 'Sales Target.xlsx' file



- Add a step **Sheet1** to understand the file structure
- Then add a pivot. CTRL + Select columns from 01/01 to 12/01 and Drag&Drop them onto the Pivoted Fields zone

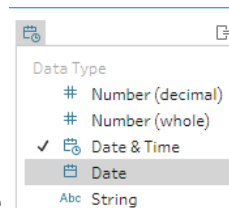


- Rename 'Pivot1 Names' column as MonthP and 'Pivot1 Values' column as Budget



- Click on the Abc icon of the MonthP header to switch it as a Date
- On step 'Join 1', click on the + icon and Insert a step.

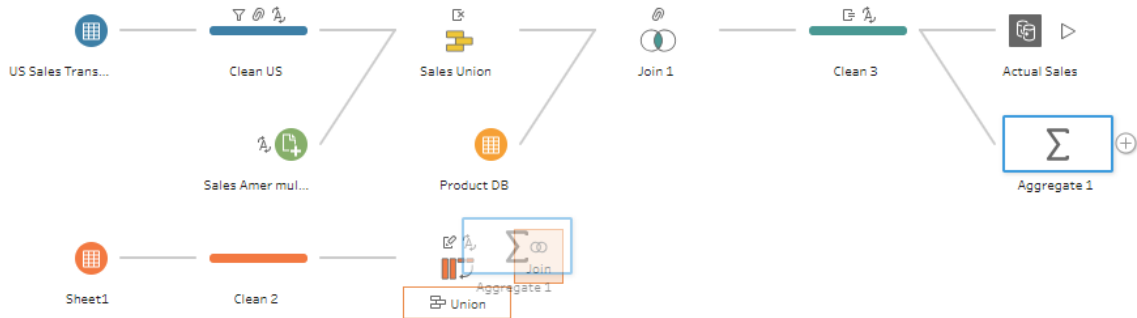
- Create a calculated field **Create Calculated Field...** 'MonthP' with the following formula DATETRUNC('month',[Order Date])



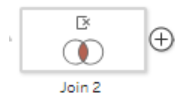
- Change MonthP format to 'Date' type
- On Clean 3 step, click on the + icon and Add Aggregate



- Drag and Drop 'MonthP' and 'Country' to Grouped Fields zone and 'Sales' to Aggregated Fields zone
- Drag&Drop this step onto Pivot 1, and choose Join



- Click on the + icon and add a 2nd join criteria 'MonthP'='MonthP'
- Observe the mismatched values, then on the Pivot 1 step rename 'Columbia' as 'Colombia' and go back to the Join 2 step to observe there is no more mismatched value
- Remove one of these 2 'MonthP' columns, idem for 'Country'



- Click on the following '+' icon, then Add Output
- Choose 'Save to file', name it 'Actual Sales' and select the Output type you prefer (csv, tde or hyper)
- Rename the step as 'Actual vs Budget'
- Save the Flow, so you can easily retrace your steps!

Congratulations! You just have a complete flow with 2 different outputs, one on detailed actual Sales data combined from various countries and periods, and another one with only the aggregated view for Sales vs Budget analysis ☺

