

**What is a real-time streaming ETL pipeline?**

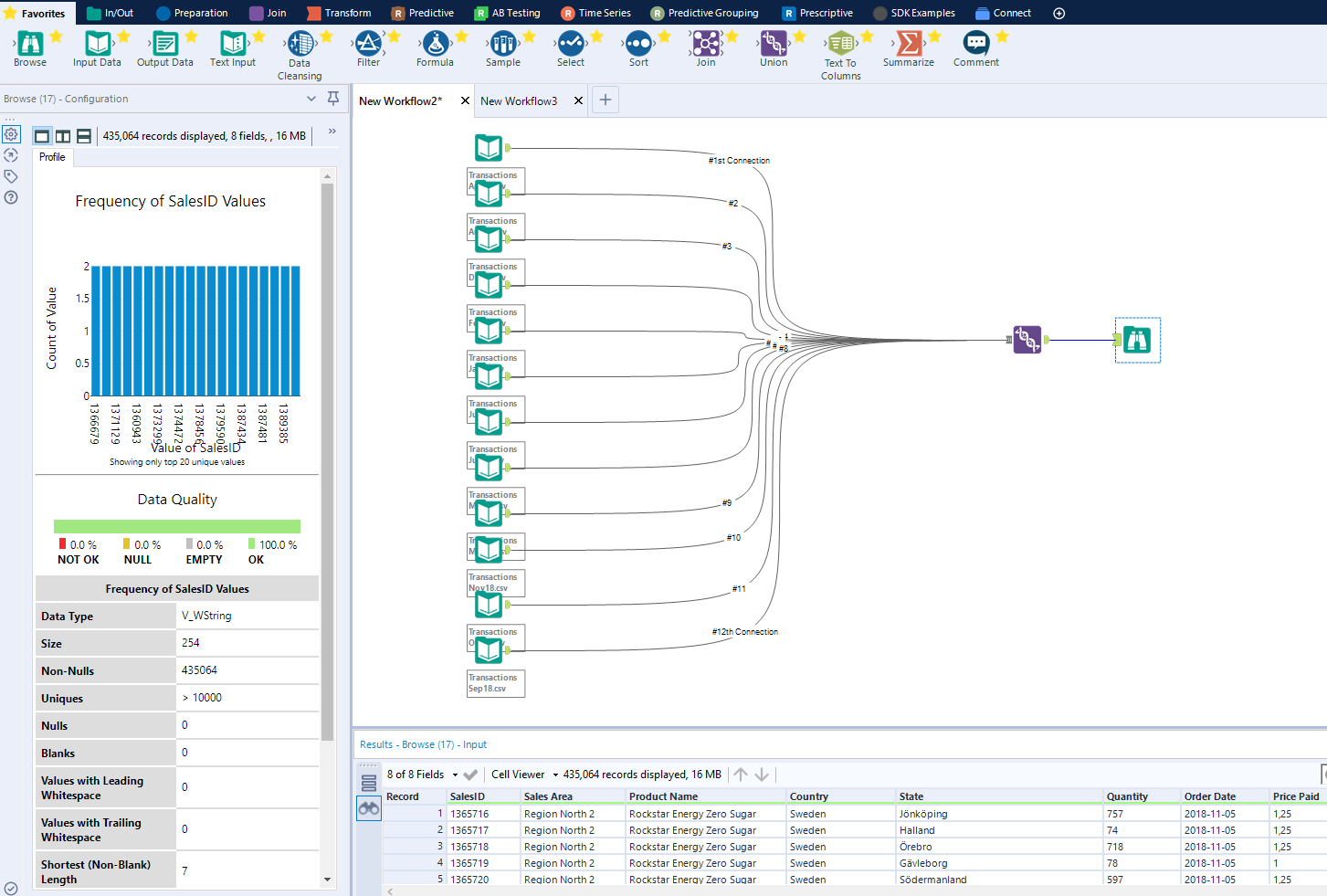
**What are the steps for data preparation in Alteryx?**

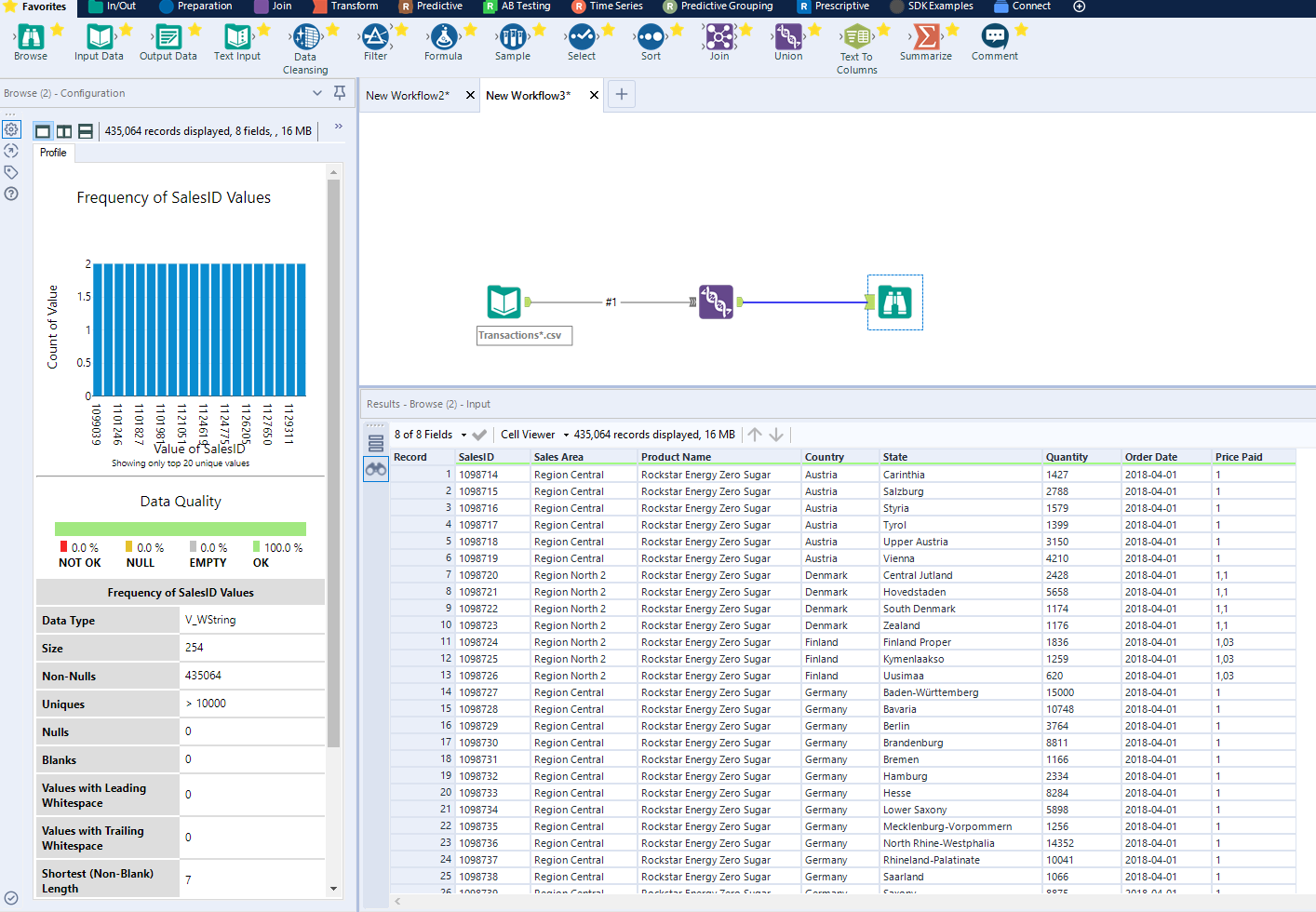
Data analytics and visual analytics are key competitive advantage for companies. The biggest obstacles companies face cleaning data where more than 80% of the time is usually spent by data analytics team. Data is usually stored in various databases and accessing them for visual analytics is challenging most of the time.

Alteryx is an advanced data preparation tool which can save most of the data cleansing time for analysts.

Alteryx is an ETL tool. AT first, we need to combine all the input data files and then we need to union or join them. We need to check the data types, check results and checking number of records.

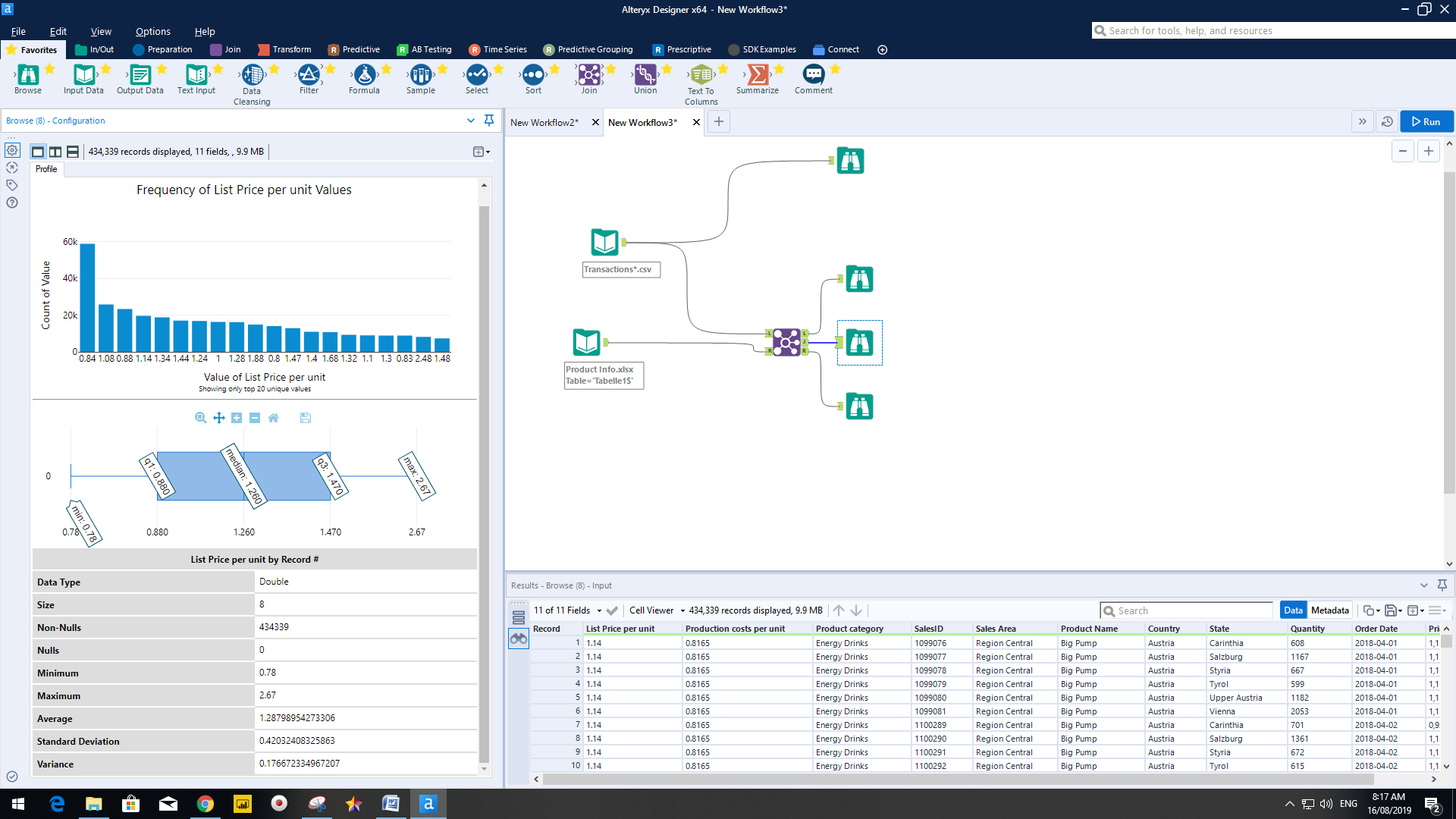
To join multiple tables into giant one table, you can drag them all onto the canvas and then change the delimiters to semicolon, drag the union field, connect to browse field and then run. This process can be bit tedious so you can use \* sign on your file name like E:\Alteryx\Transactions\*.csv. It will import all the tables’ files names start with Transactions, will make the union and populate the same result.



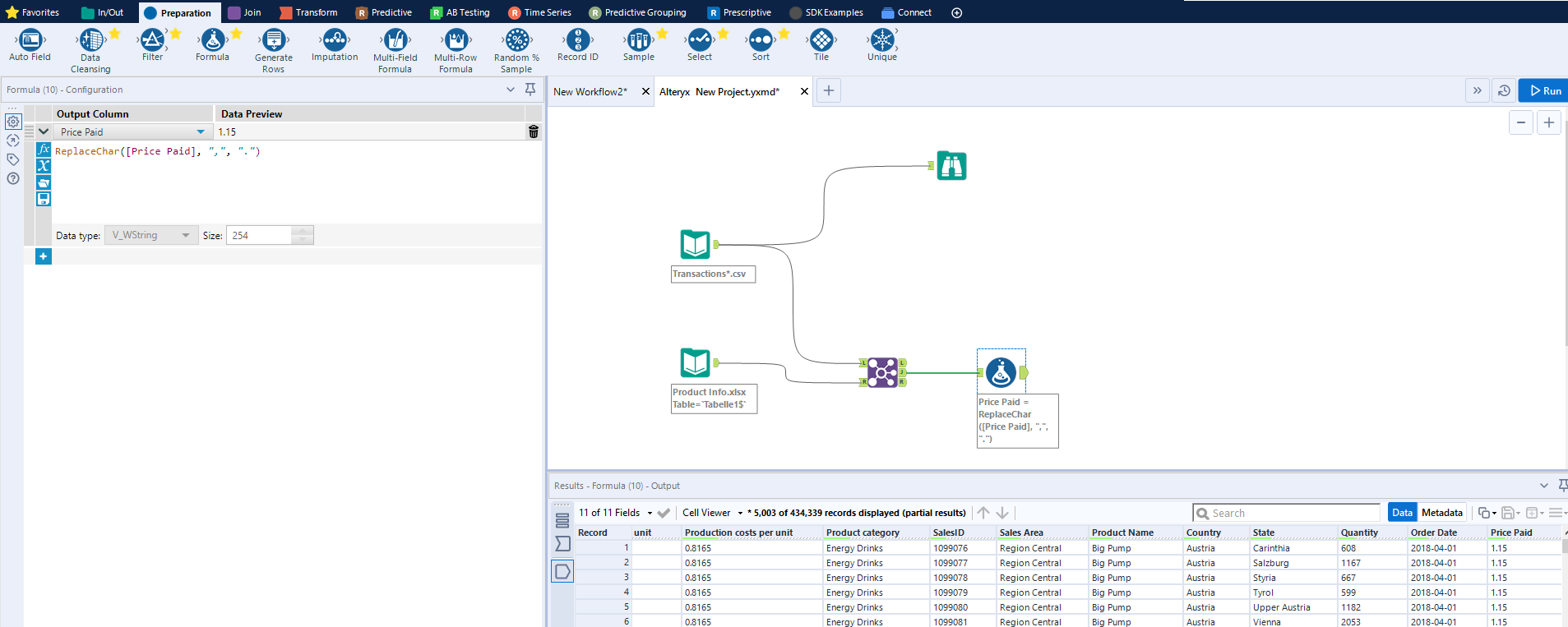


You can join them two different files and select the type of join by left or right. If you are looking for add additional column then, you need to check the ‘Unknown’ box. In the below example, we need to join and show the output where product name and country both matches in the both tables.

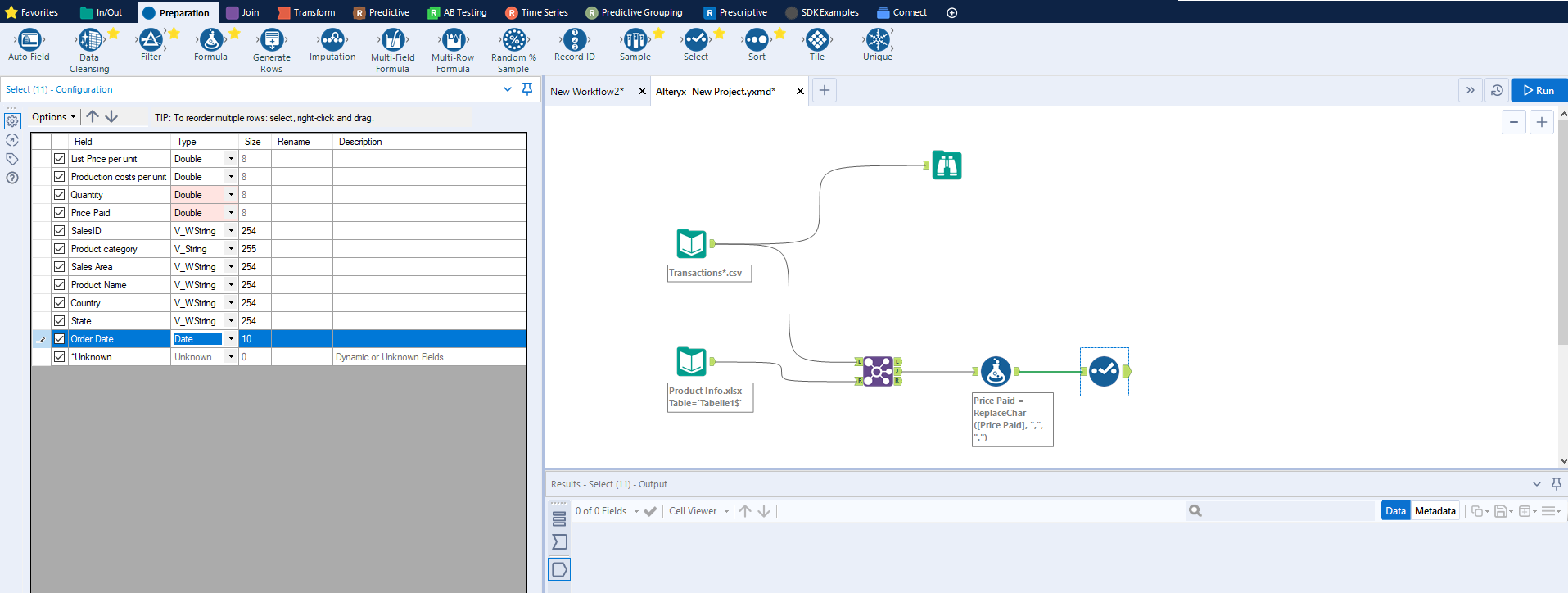
You can set three browses for right join, left join and just for the join if you want to see which fields do not have common product name and country.



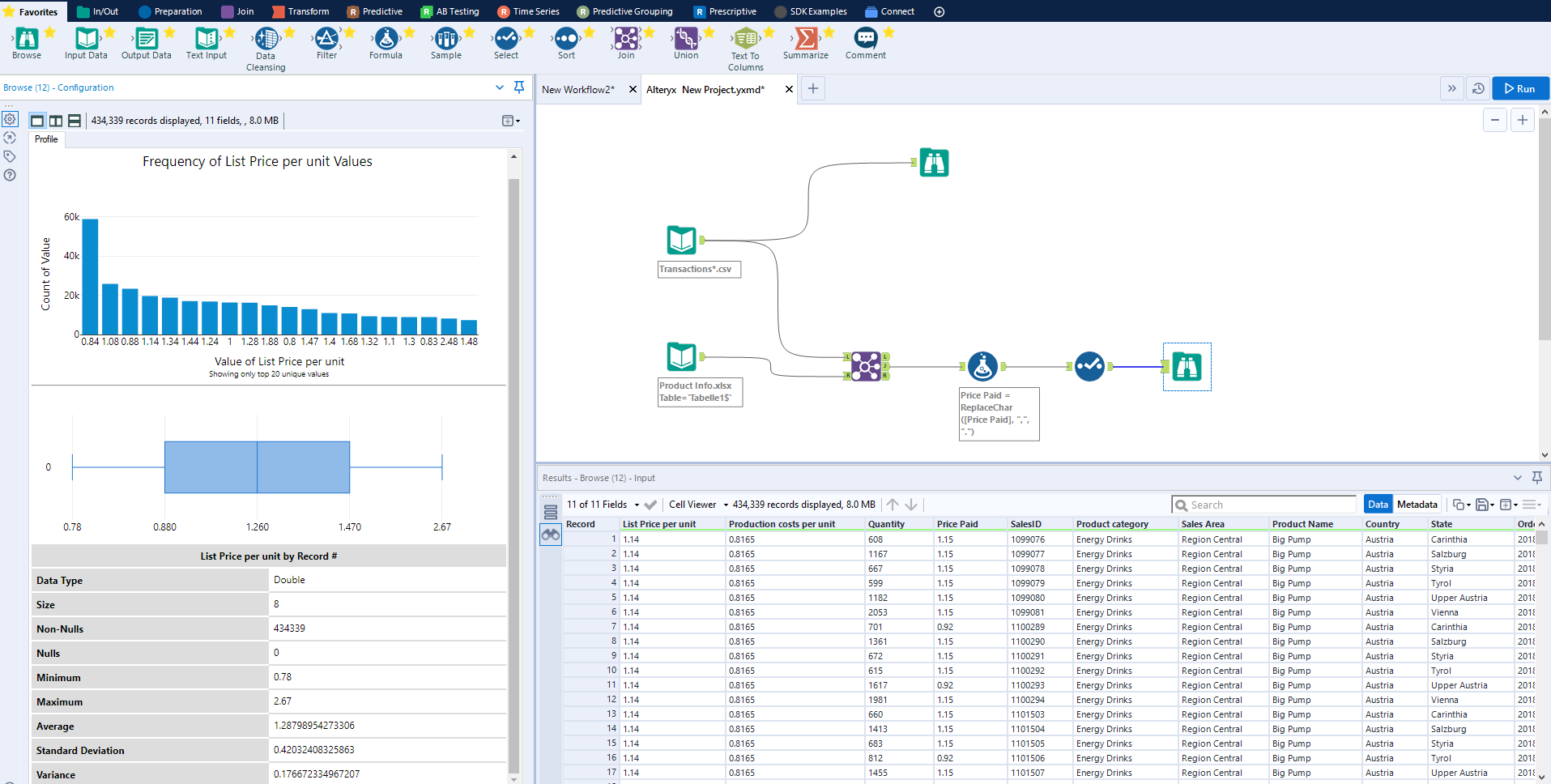
The formula to change the string value of the price paid from comma to dot through ‘Replacechar’ formula. Go to the ‘Preparation’, then ‘Formula’, drag this formula tool on your canvas, join and run. You will be able to see the difference.



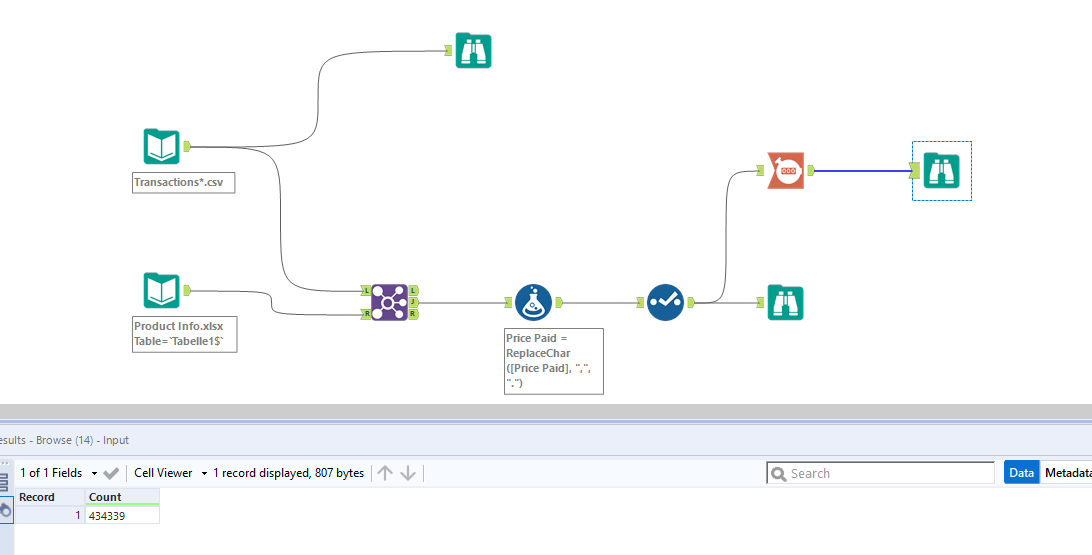
The ‘Select’ tool can help you to change the sequence of your column and you can also change the file character type of your each column. You need to make sure the file type is correct.



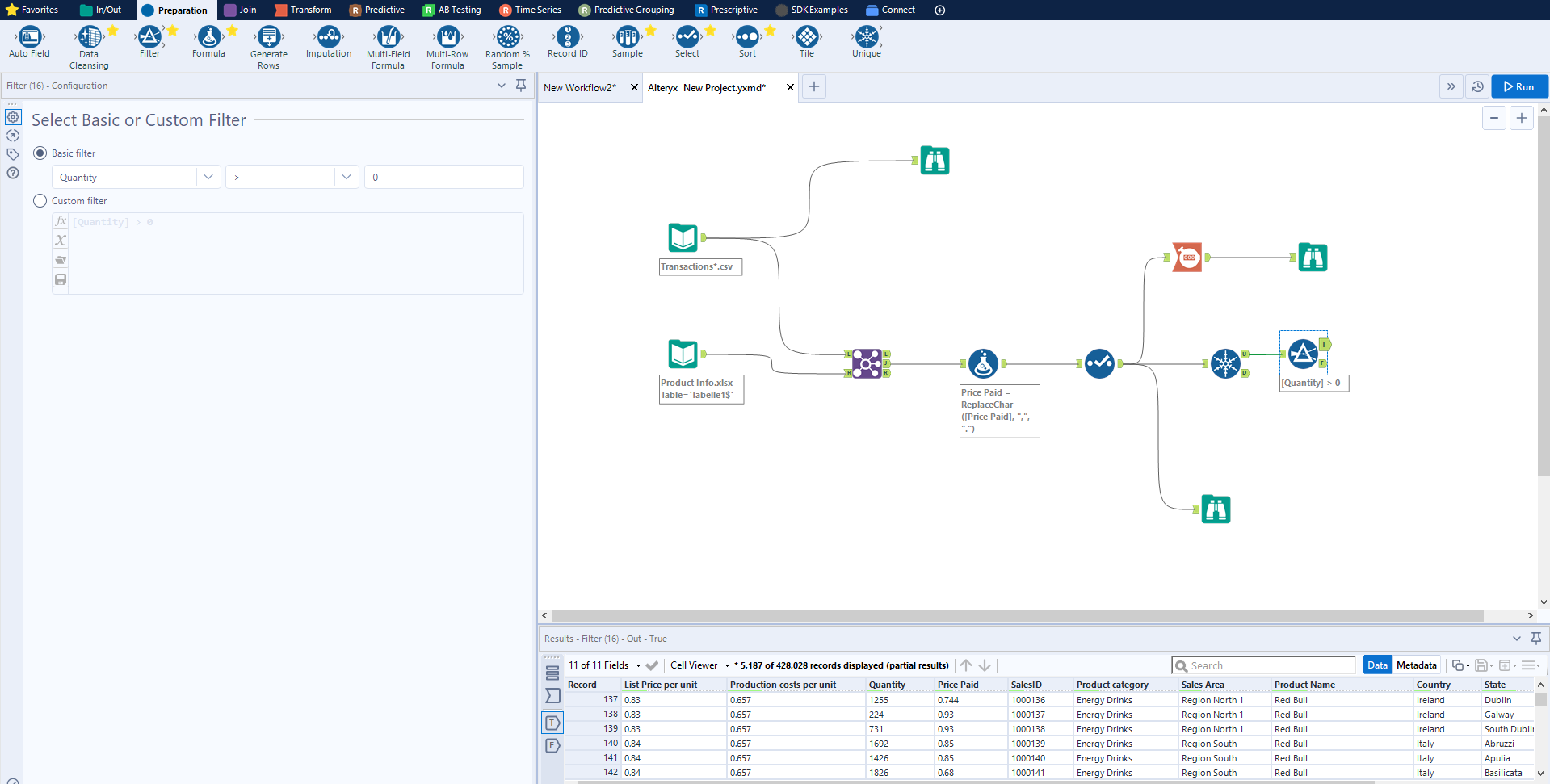
One all file types are changes, Run again to see the new table.



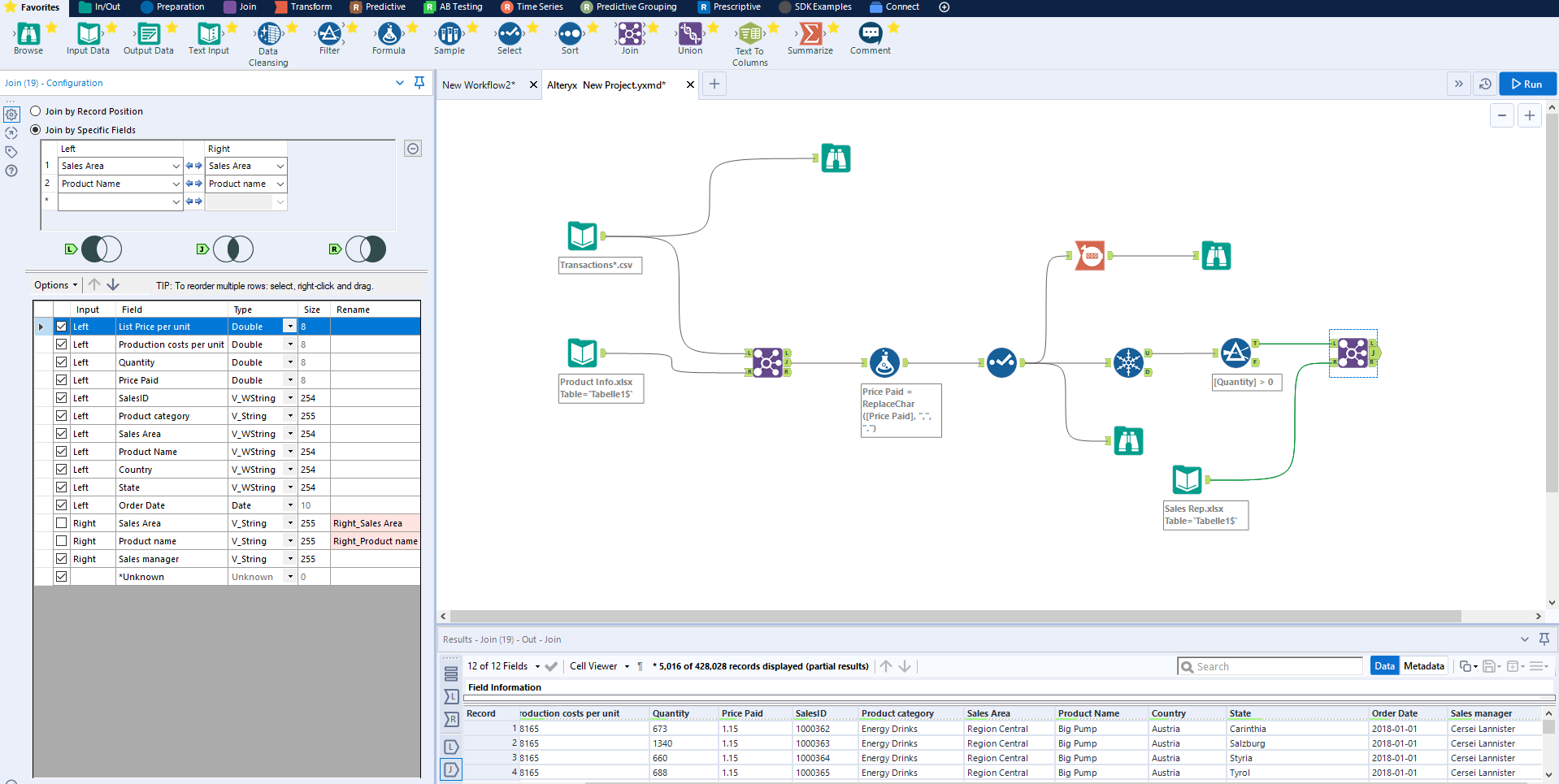
Now you need to check your records they are same as before. We will count the number of fields in the below workflow. Go to Transform and drag the ‘Count Records’ on your canvas and run.

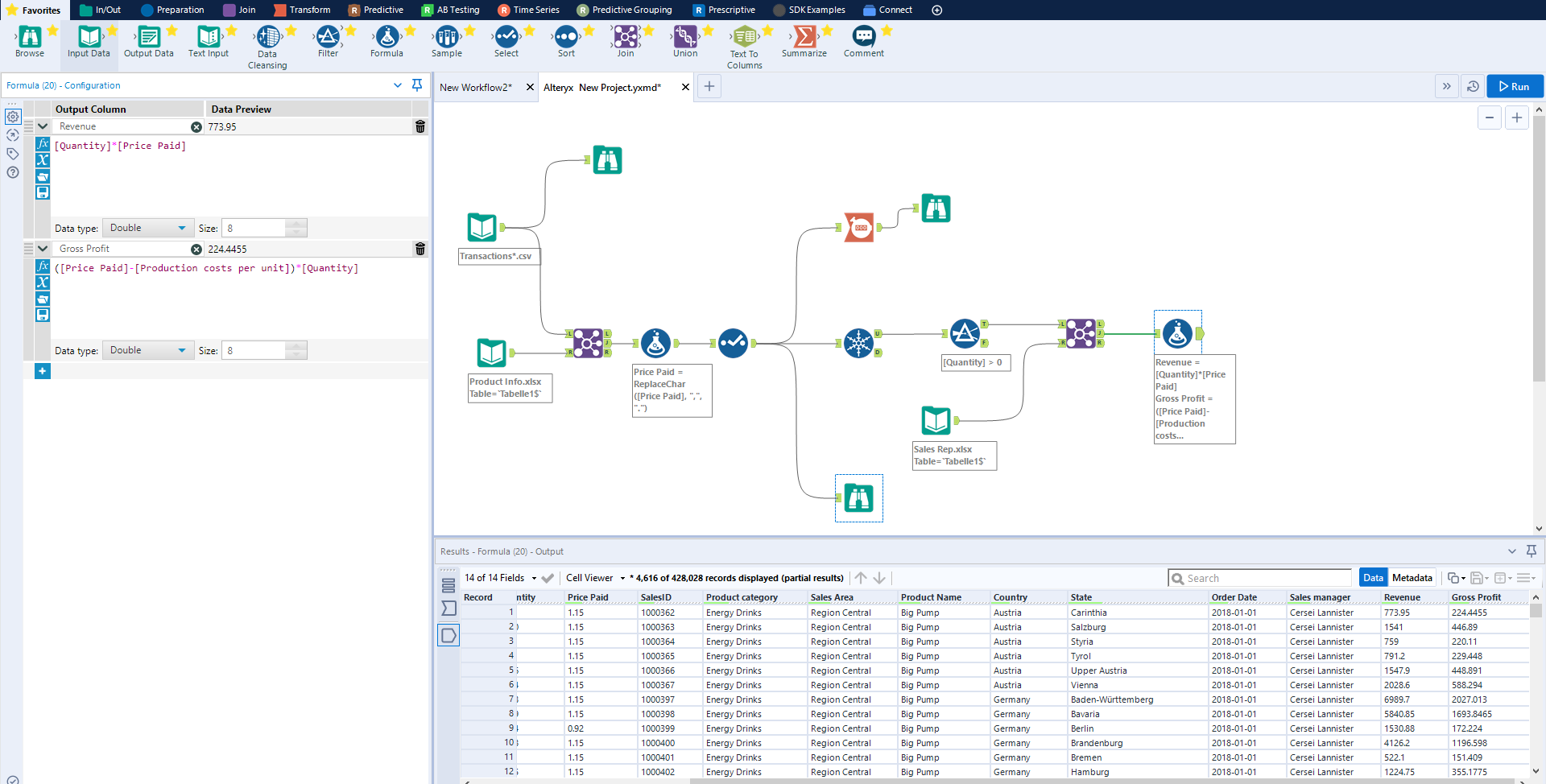


Now we will do the data validation process to see the unique values or duplicate values just dragging the ‘Unique’ tool on the canvas and run. In the below example we wanted to see the unique sales ID. Once we can see the unique vales, you can use ’Filter’ tool to filter out data based on your criteria. In the below example, we wanted to see all the order quantities which are more than zero. You can choose your own formula to get the required output.

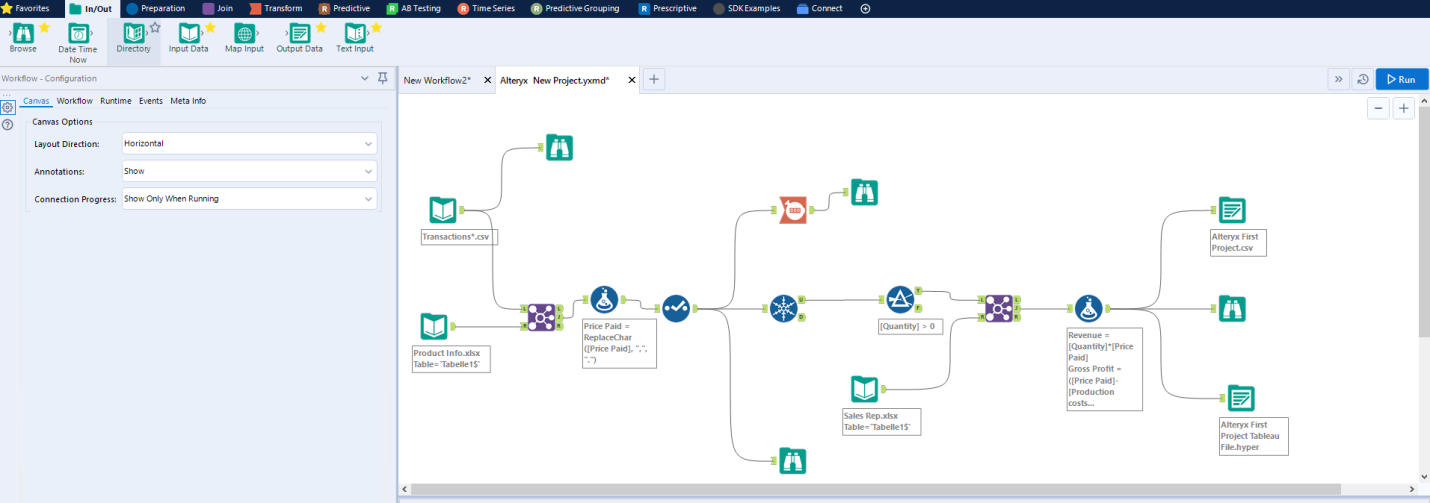


Now, we are going to drag the sales manager excel file on the canvas and joined to create a column of ‘sales manager’ to the existing unique table. The output will be displayes based on fully join, left join or right join.

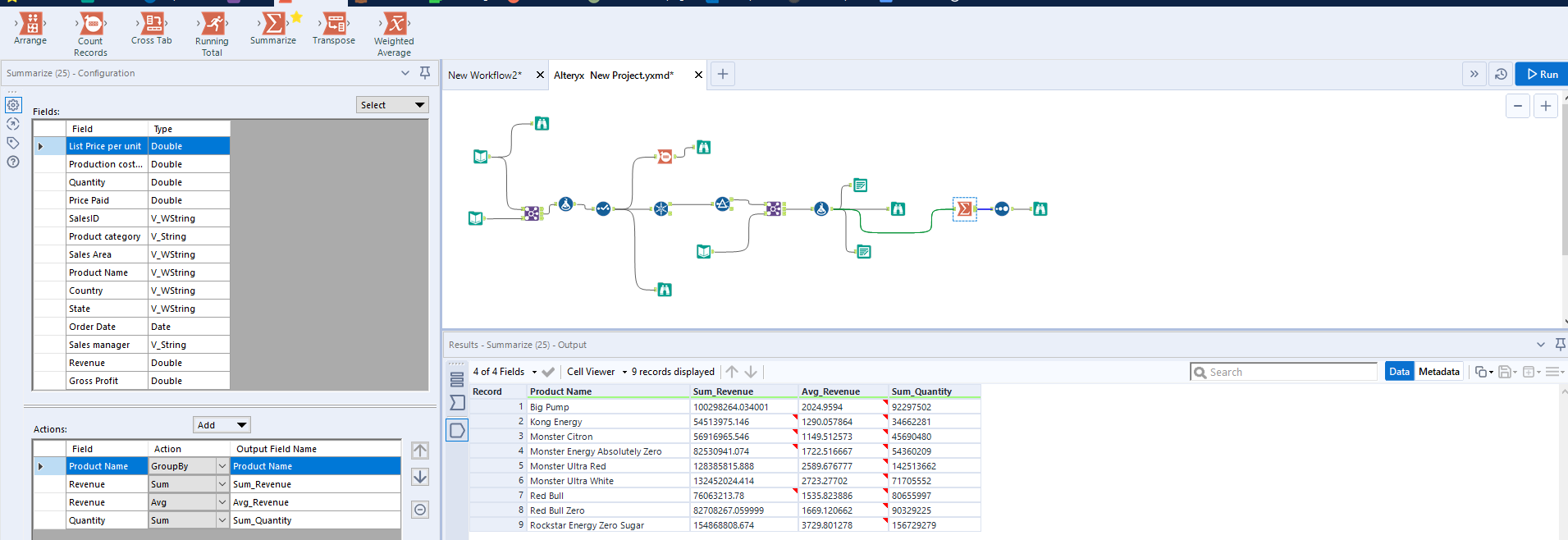


Next, we will create couple of additional columns with formula such as revenue or gross profit. 

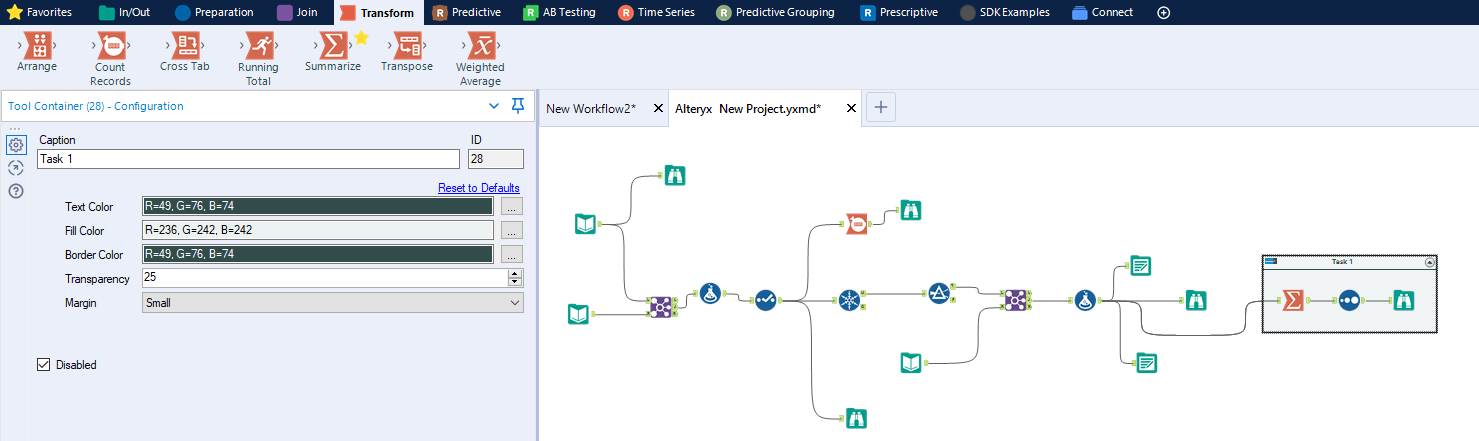
Once you have completed your data preparation, as part of the final data enrichment steps, save your data in the default file format .yxdb file or tableau extract file format as a .tde or for faster result .hyper

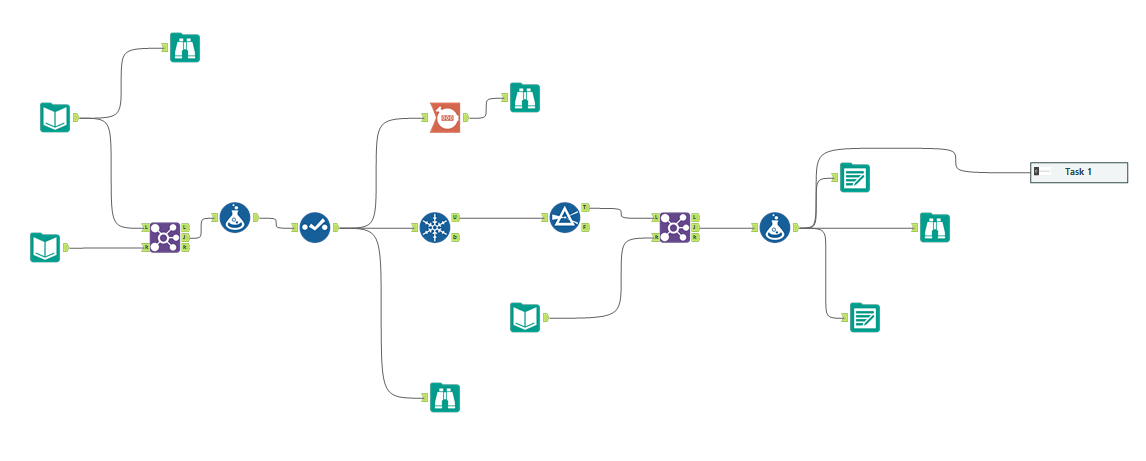


Now, we can create some queries such as grouping our category, summing sales per category, average per category or number of quantities sold per category. We use the ‘Summarize’ tool to do this function and join the ‘Sort’ tool for sorting your output value in an ascending or descending order. You can add multiple criteria on your sorting rules.

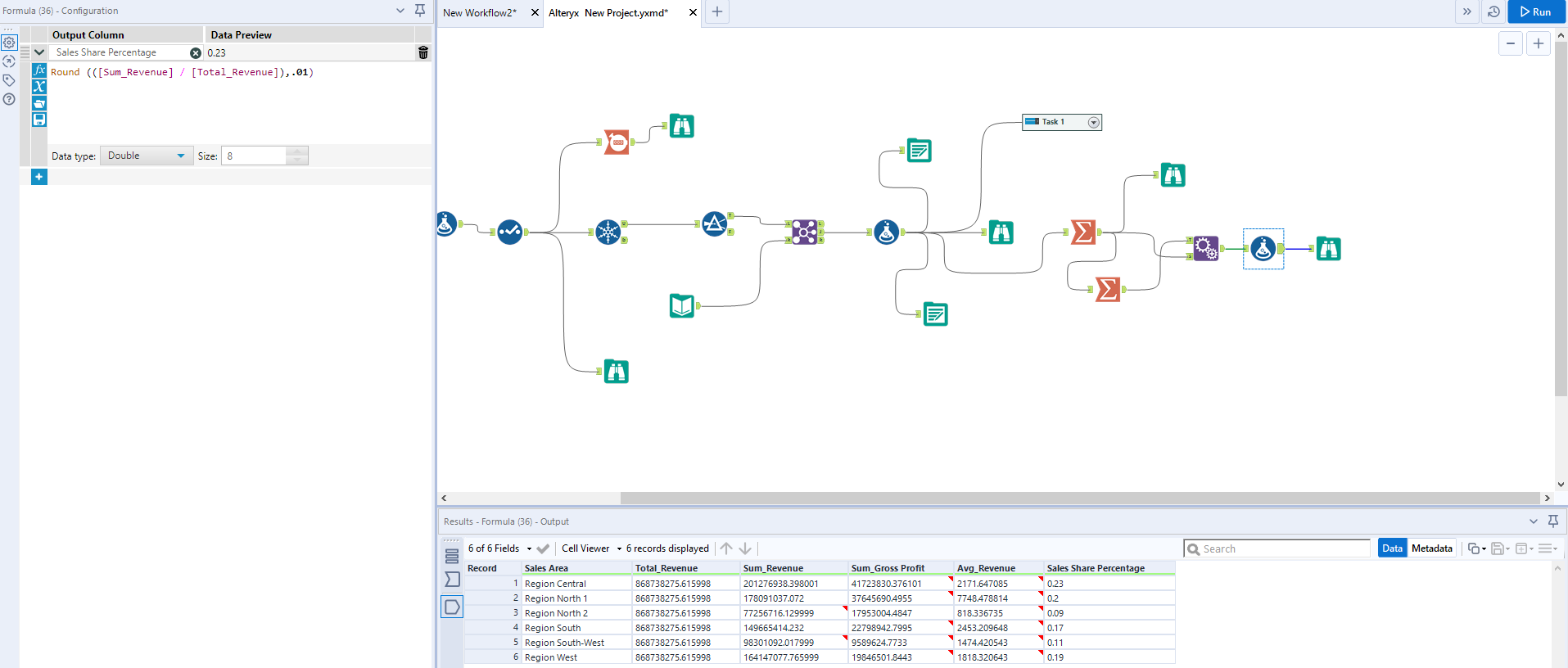


You can create various container to save the running time for your query. You can select those tools to put in a container and you can select ‘Disable’ if you do not need to run the same query all the time.

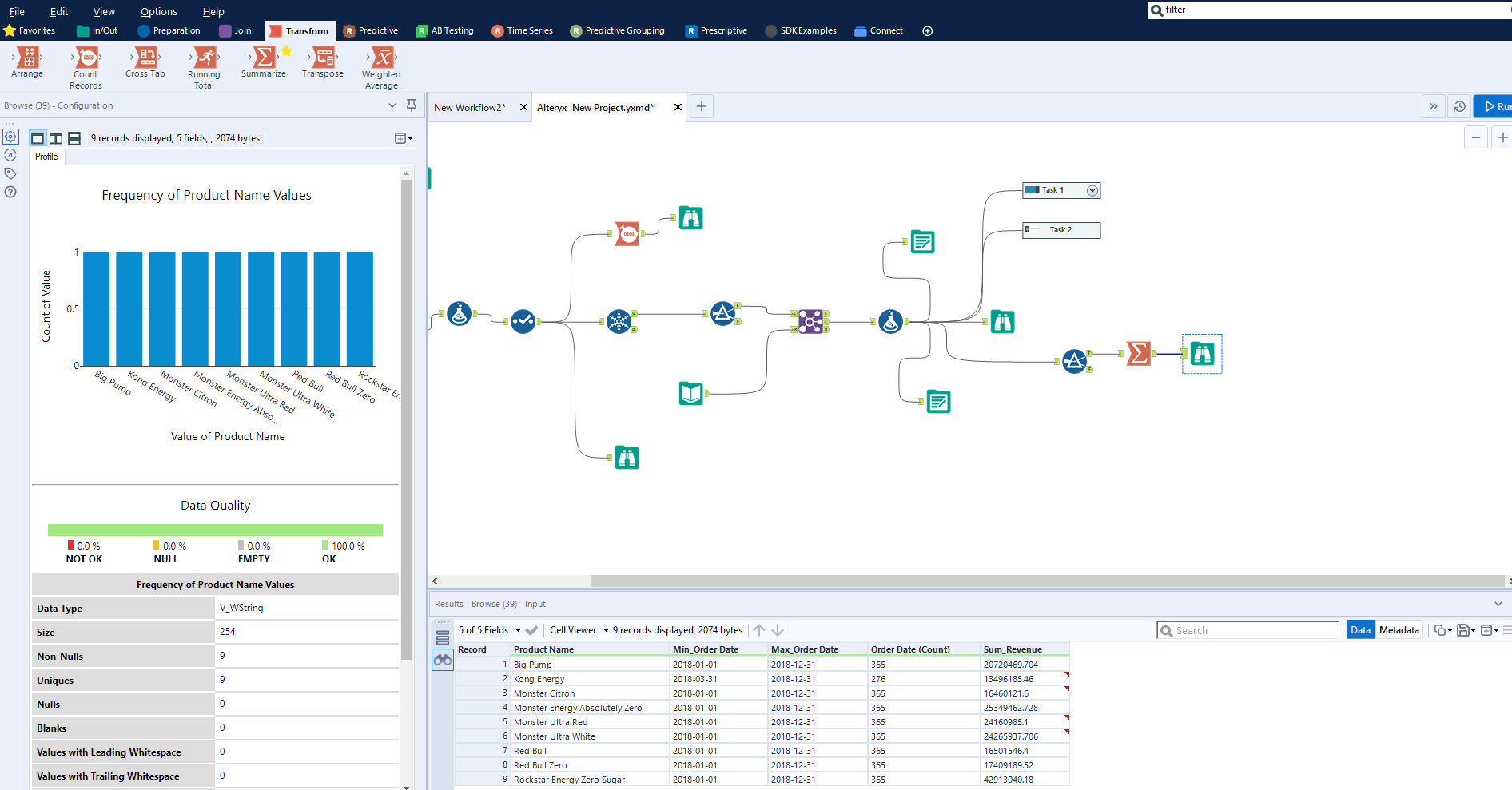




We have created the table for sales revenue with gross profit and average revenue. Another table we have created only total revenue. Now, we need to choose the ‘Append’ tool if we want to see the total value in each row of the table. Next, we need to see the sales revenue value relative to the total sales revenue in percentage (Gross profit margin). We have chosen the ‘Formula’ tool for that calculation and the output created a column called ‘Sales share percentage’.



Now we have created an another task which is to show the minimum order date, maximum order date, counting distinct dates and revenue by the product name in the only Central region area.



Now, in the below exercise we have created top three revenue generates states per each sales area. We have sorted the revenue by descending order and the sales area by ascending order. Then, we have used ‘Sample’ tool and selected three as number of values to display on the table by sales area.

