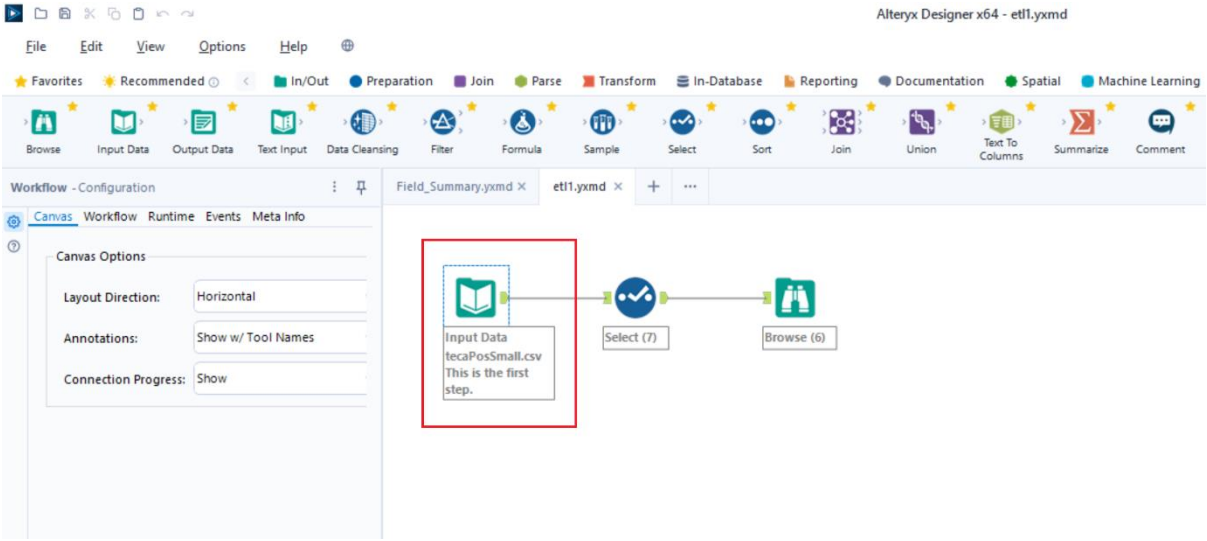
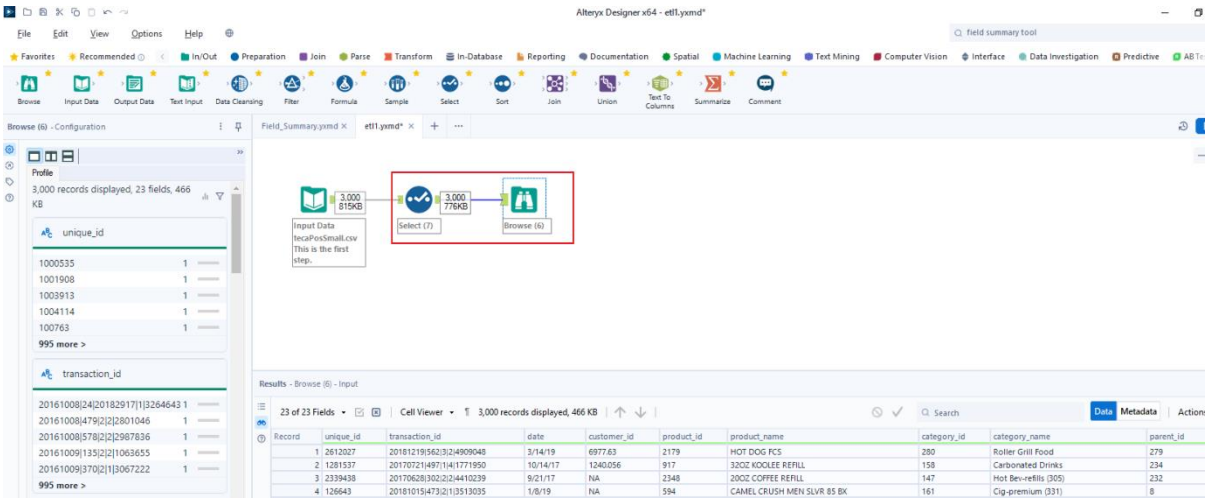
Alteryx – Extract, Transform and Load

ETL1 – Read in Data and Examine Data Quality

Connect file tecaPosSmall.csv using Input Data tool



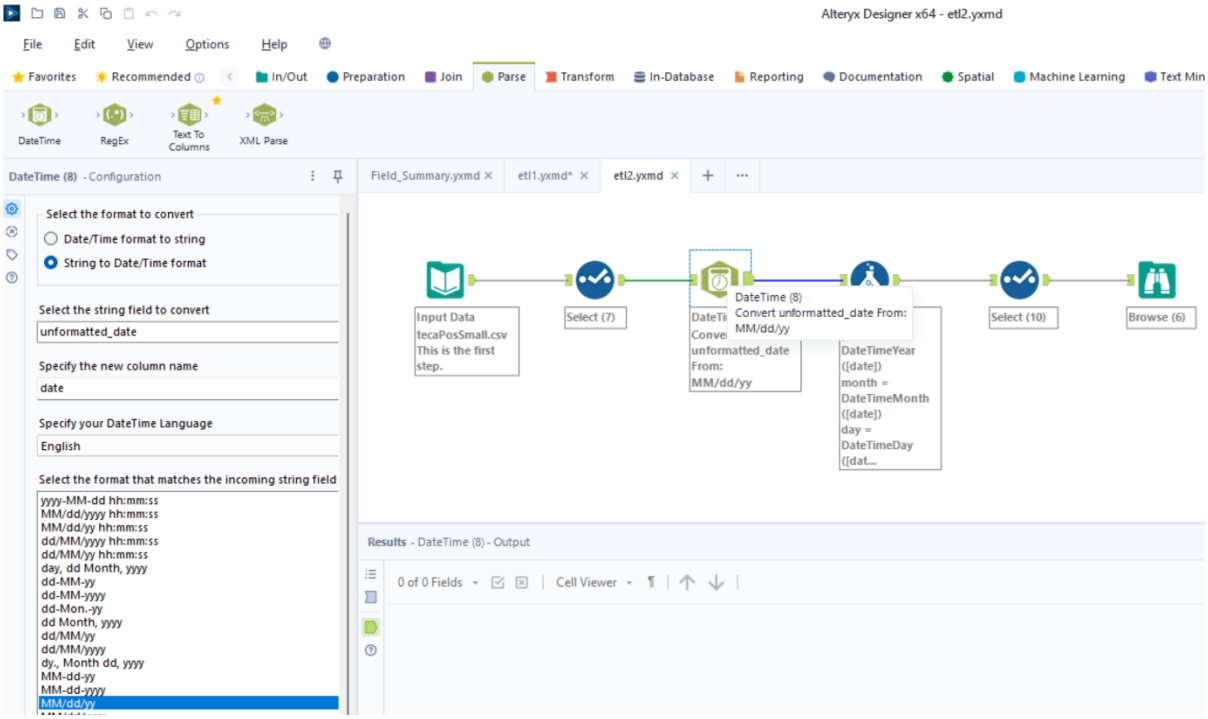
'Select' the files fields and then 'Browse' the data



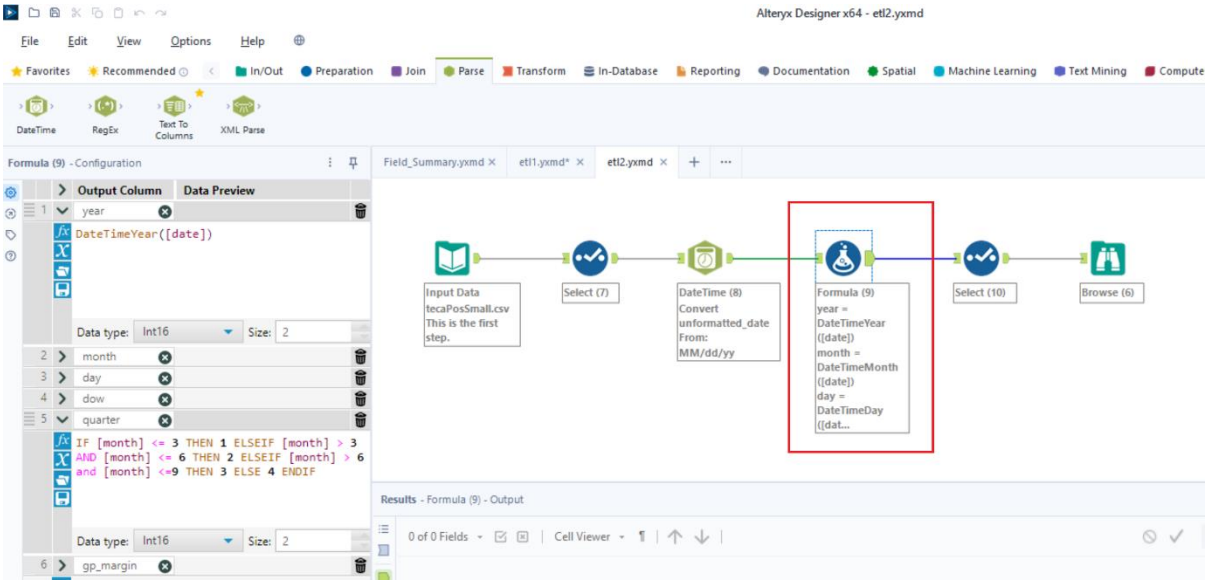
ETL 2 – Dates and Calculated Columns

The date string needs to be in the ISO 8601 format: YYYY-MM-DD.

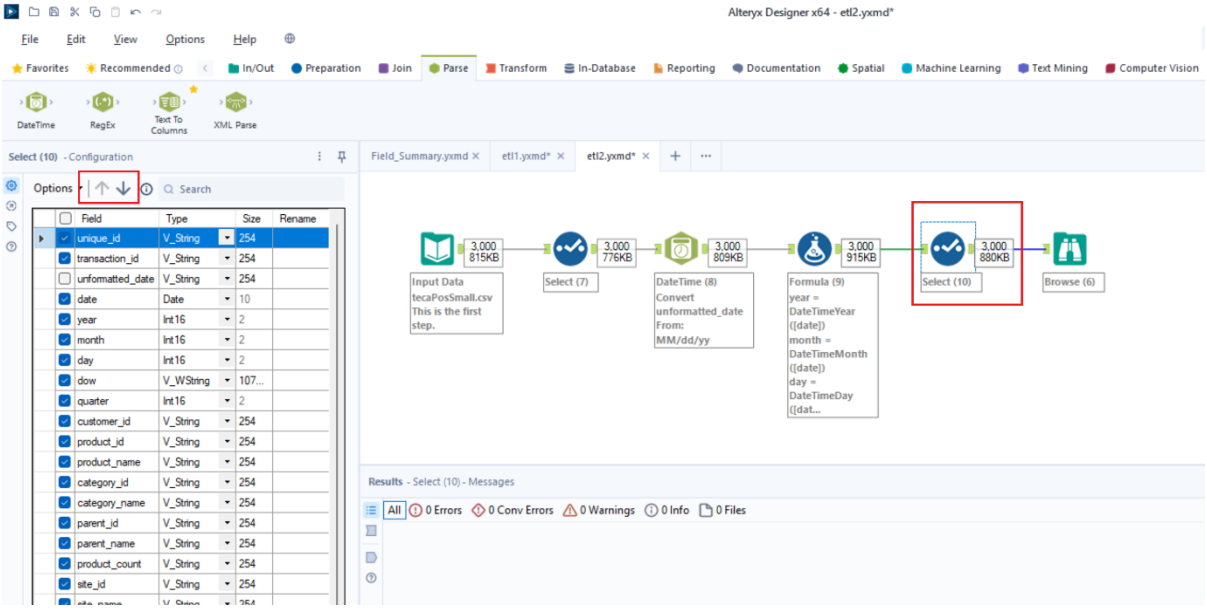
Use the ‘Parse’ DateTime tool to do the conversion.



Create columns for each component of the date, year, month, quarter, day of week using formulas



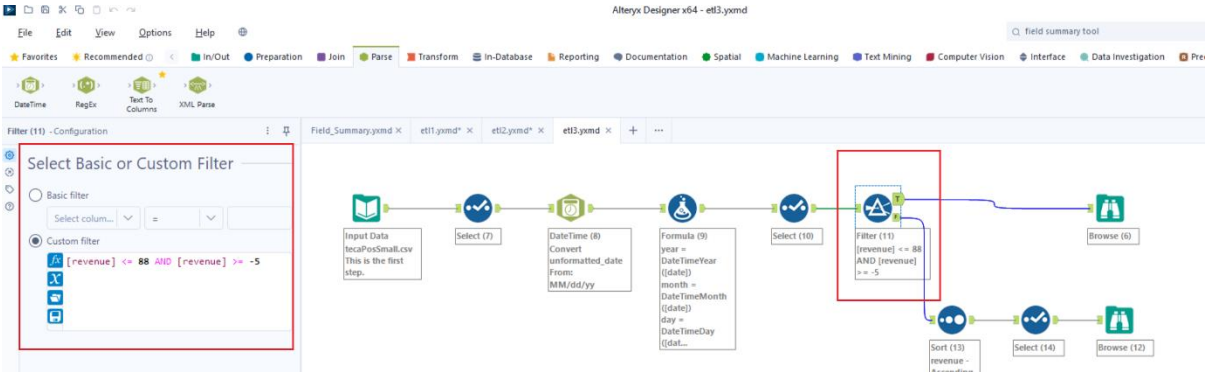
Use ‘Select’ tool to move columns Up or Down



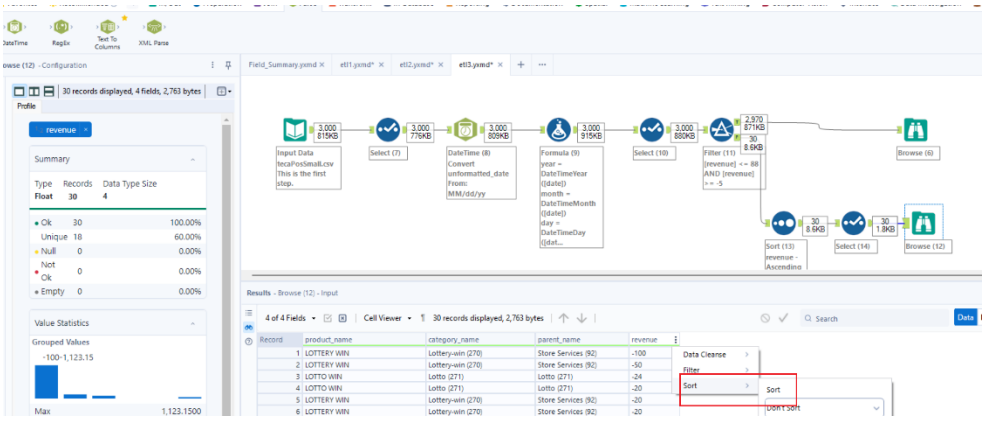
ETL 3 - Outliers

Remove outliers (such as negative values for revenue) using ‘Filter’ tool

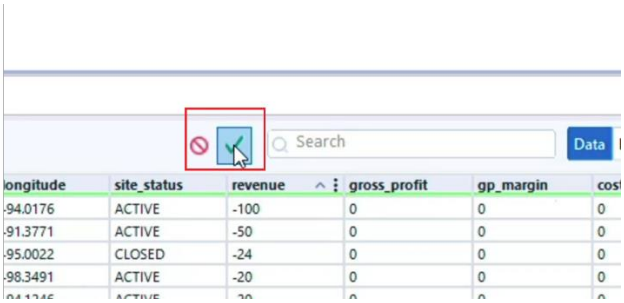
Note the True and False connectors to the Browse tools.

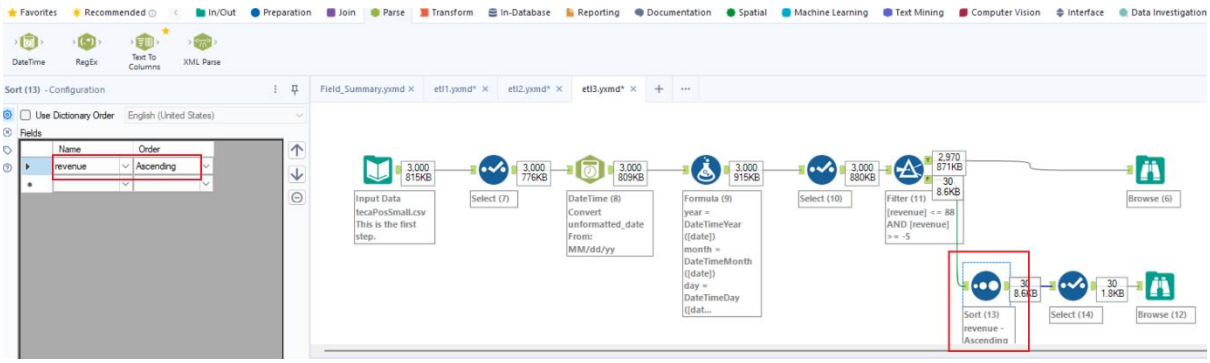


A Sort, Ascending or Descending can be applied to the data

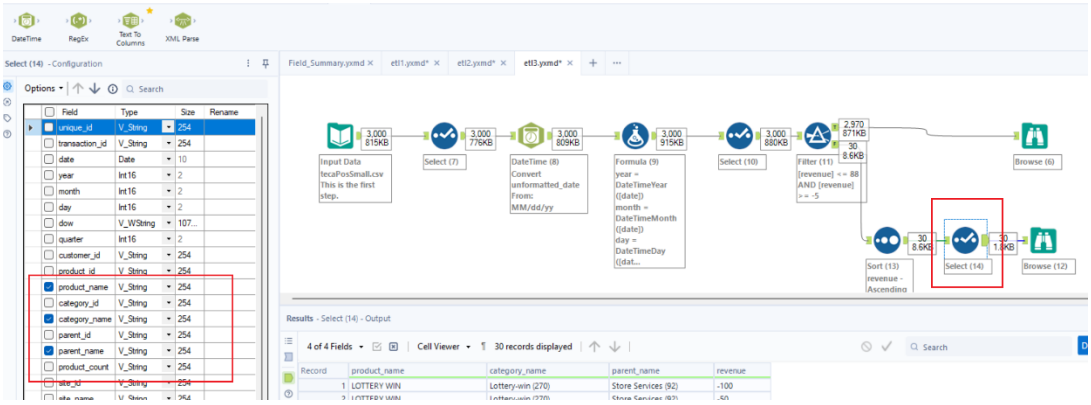


Clicking on the green tick (below), automatically adds the ‘Sort’ tool to the Workflow.



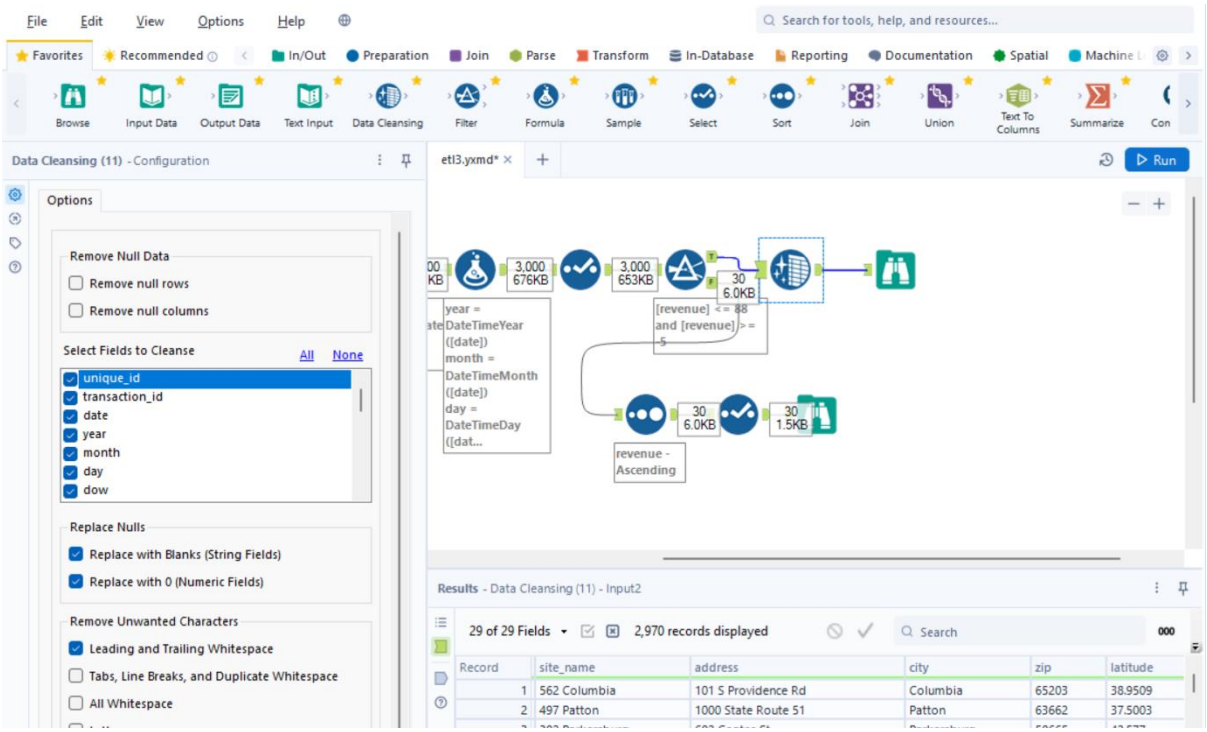


Use the ‘Select’ tool to display only selected columns

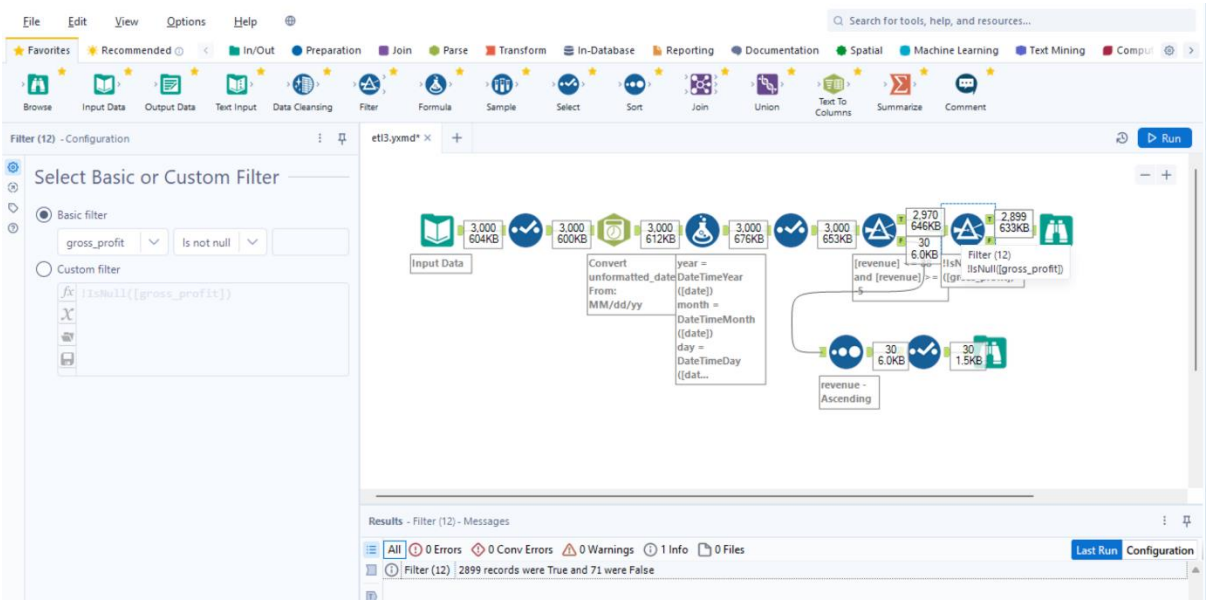


ETL 4 – Missing Values

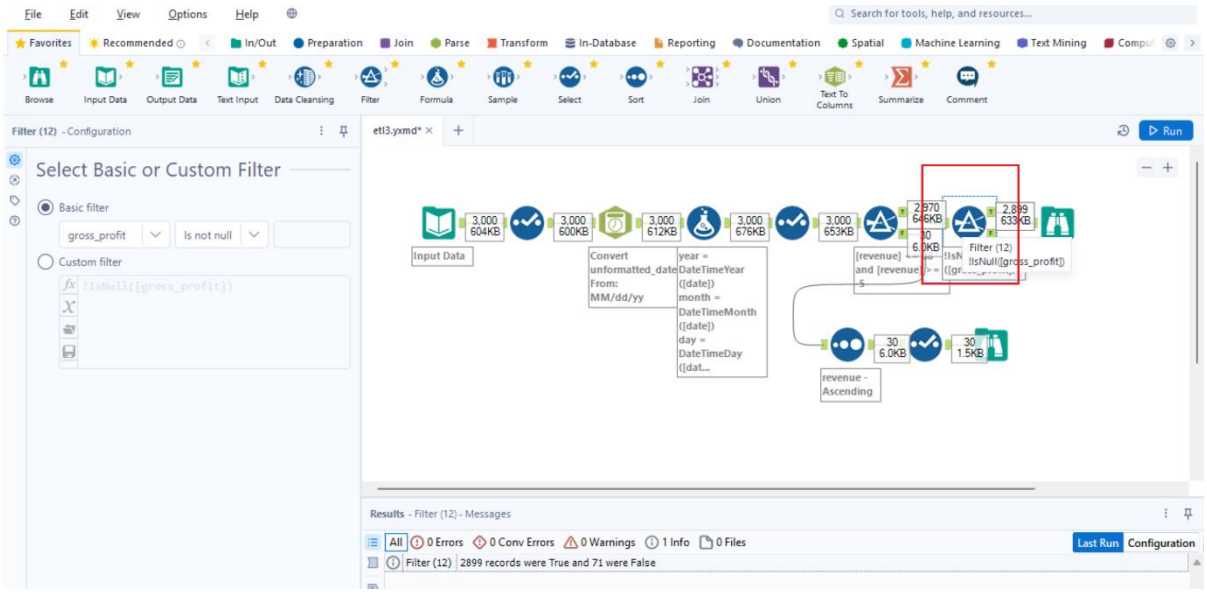
Data Cleansing



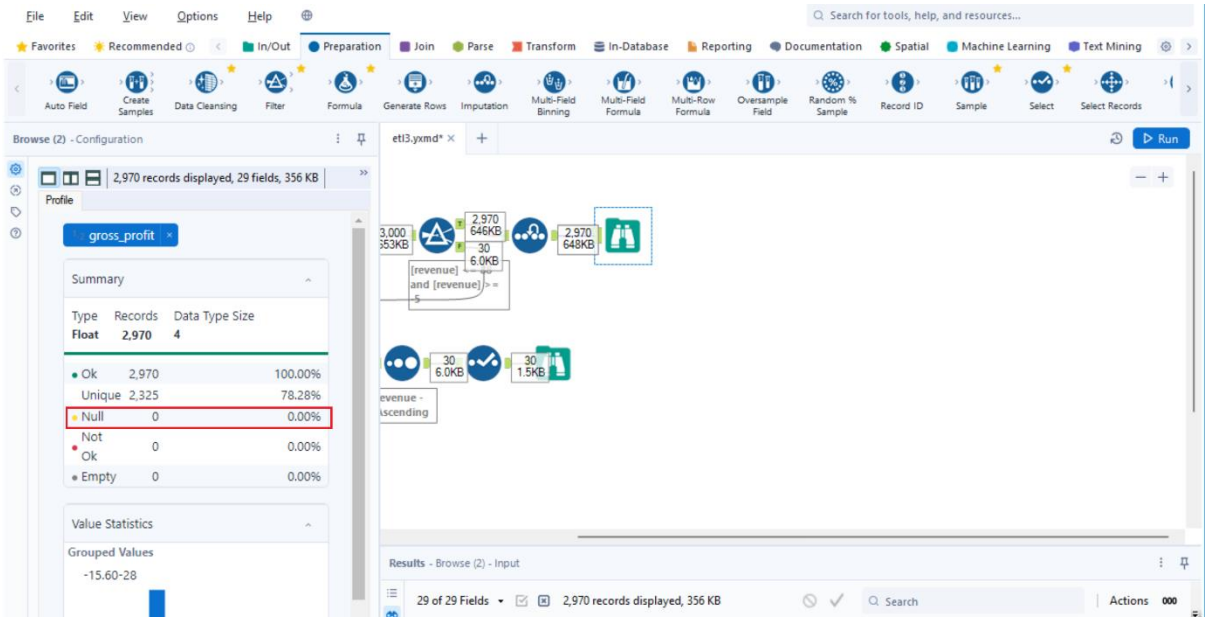
Filter out missing columns



Imputation

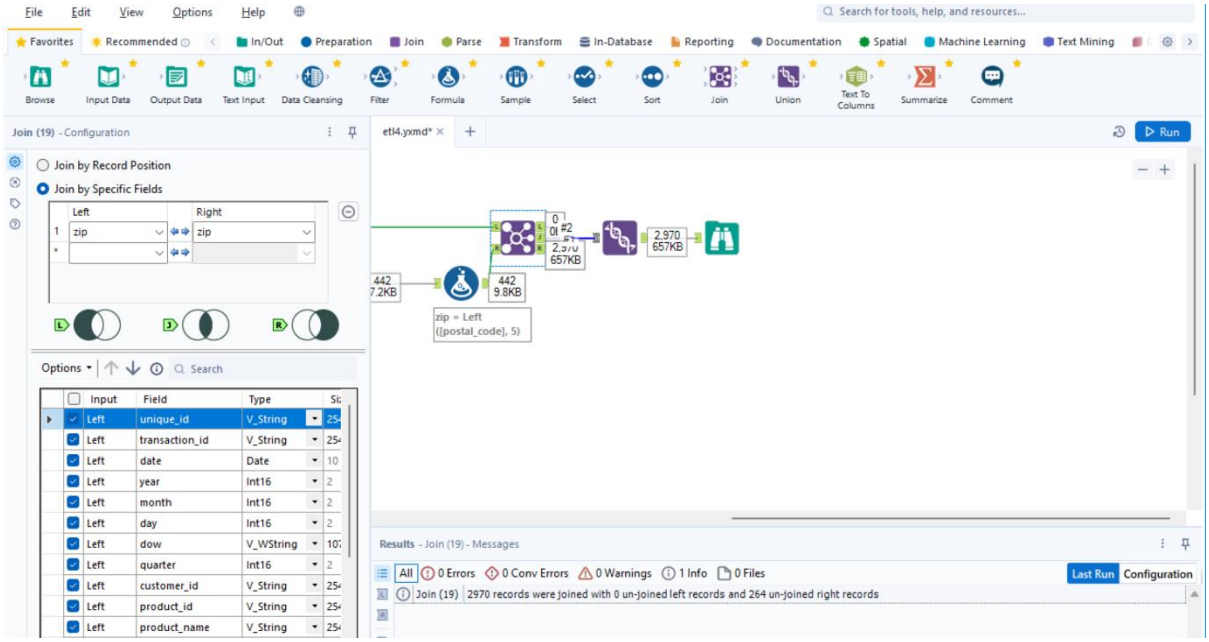


Imputation – Null values removed



ETL 5 – Data Joins

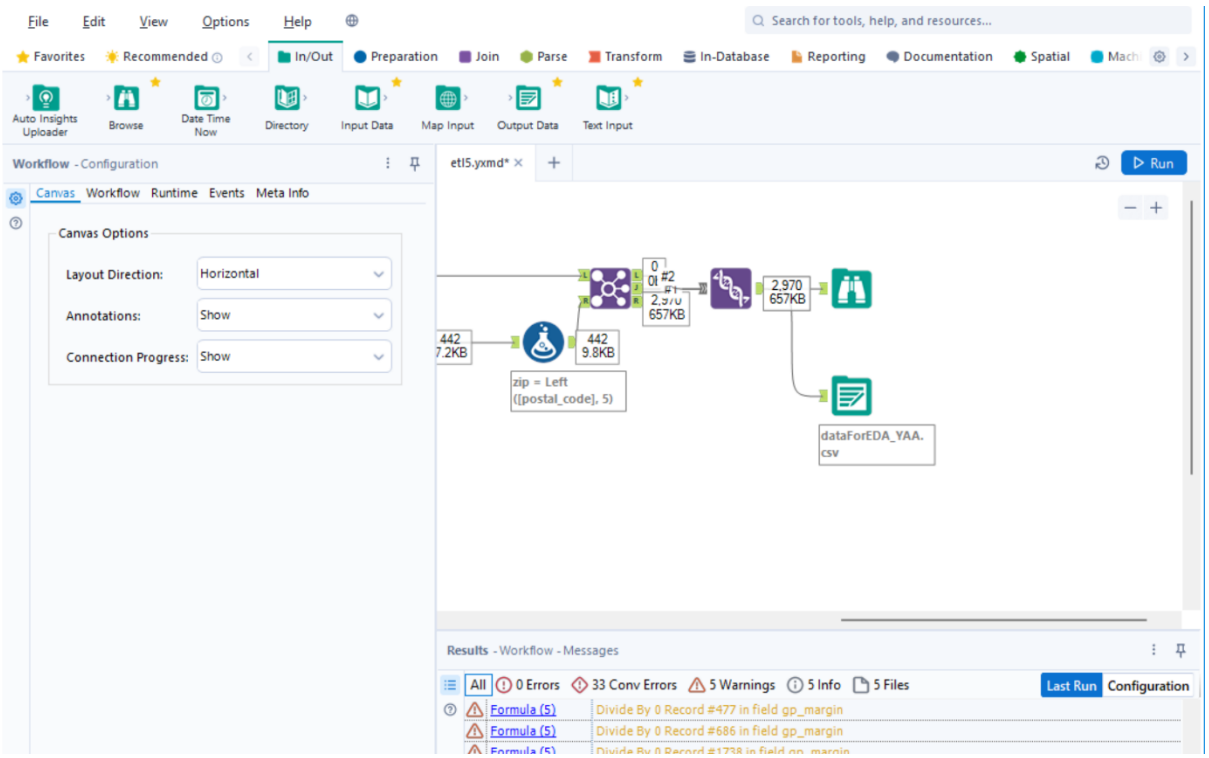
Join with states.csv using zip code



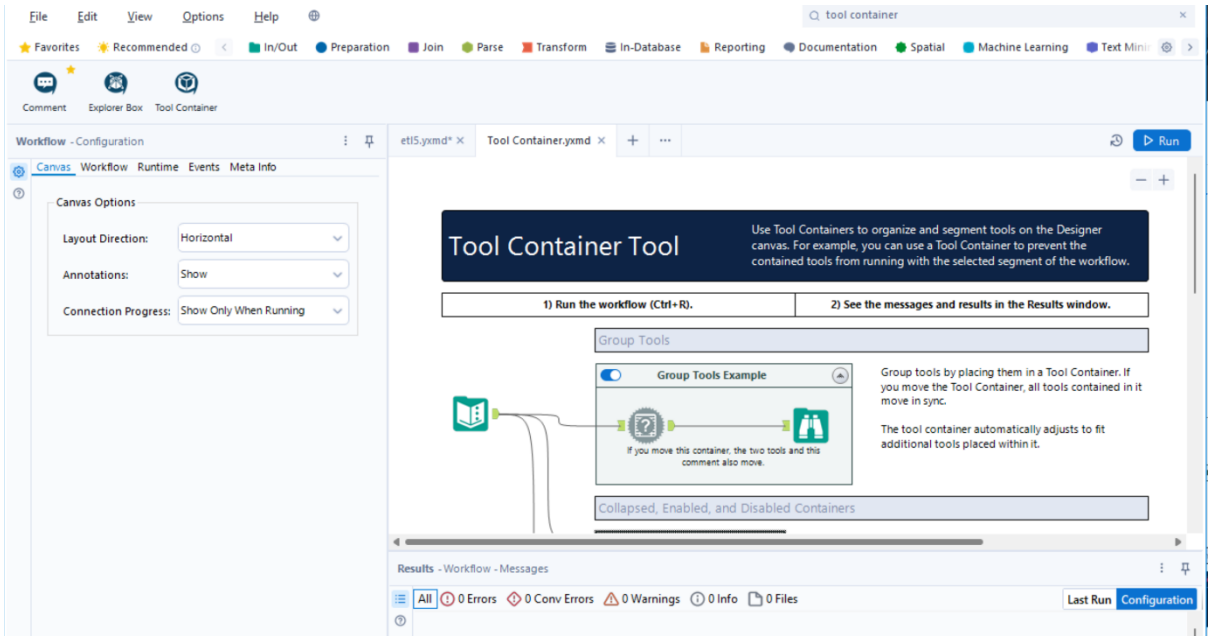
ETL 6 – Workflow

Write data to file to visualise in Power BI or R or Tableau

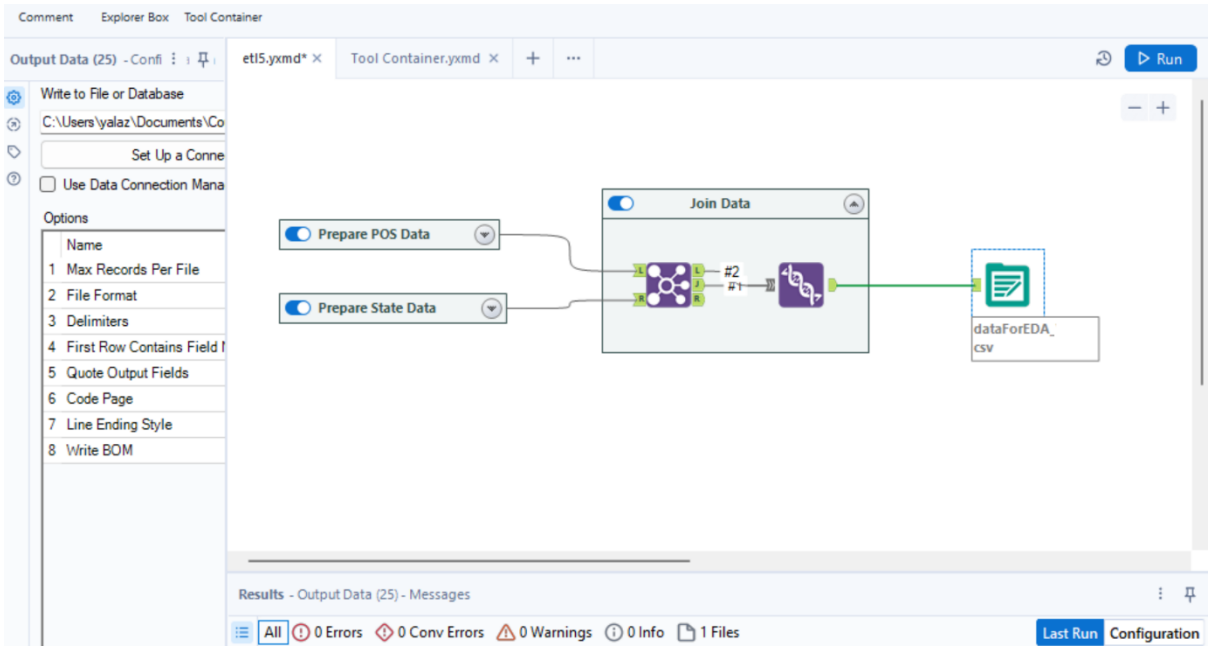
Write to output file -dataForEDA.csv



Using Tool Container



Clean-up by creating containers for different groups of processes and collapsing them



Adding a comment