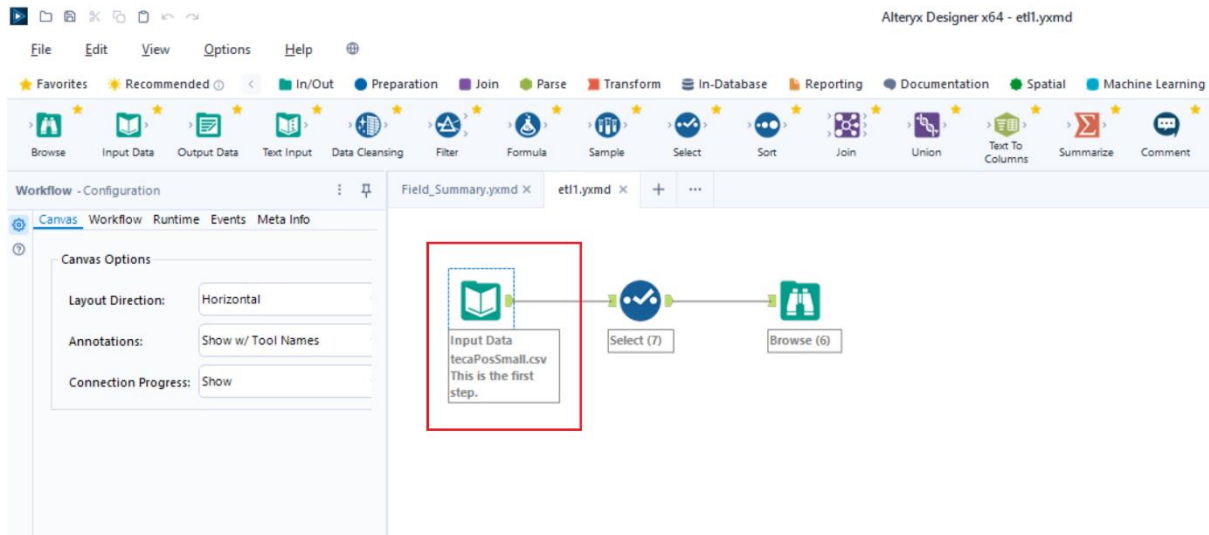


Alteryx – Extract, Transform and Load

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

The screenshot shows the Alteryx Designer interface with the 'Select (7)' tool highlighted by a red box. The 'Browse (6)' tool is also visible. The left sidebar shows the 'Profile' view for the 'Input Data' tool, displaying 3,000 records and 23 fields. The 'Results - Browse (6) - Input' table is shown at the bottom, displaying 3,000 records and 23 fields. The table has columns: unique_id, transaction_id, date, customer_id, product_id, product_name, category_id, category_name, and parent_id.

Record	unique_id	transaction_id	date	customer_id	product_id	product_name	category_id	category_name	parent_id
1	2612027	2018121915623024909048	3/14/19	6977.63	2179	HOT DOG PCS	280	Roller Grill Food	279
2	1281537	201707214971141771950	10/14/17	1240056	917	320Z KIOOLEE REFILL	158	Carbonated Drinks	234
3	2339438	2017062813022024410239	9/21/17	NA	2348	200Z COFFEE REFILL	147	Hot Bev-retills (305)	232
4	126643	20181015147320113513035	1/8/19	NA	594	CAMEL CRUSH MEN SLVR 85 BX	161	Cig-premium (331)	8

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.

Alteryx Designer x64 - etl2.yxmd

File Edit View Options Help

★ Favorites ★ Recommended ◀ ▶ In/Out Preparation Join Parse Transform In-Database Reporting Documentation Spatial Machine Learning Text Mining

DateTime (8) - Configuration

Select the format to convert

☐ Date/Time format to string

☒ String to Date/Time format

Select the string field to convert

unformatted_date

Specify the new column name

date

Specify your DateTime Language

English

Select the format that matches the incoming string field

yyyy-MM-dd hh:mm:ss
MM/dd/yyyy hh:mm:ss
MM/dd/yy hh:mm:ss
dd/MM/yyyy hh:mm:ss
dd/MM/yy hh:mm:ss
day, dd Month, yyyy
dd-MM-yy
dd-MM-yyyy
dd-Mon-yy
dd Month, yyyy
dd/MM/yyyy
dy, Month dd, yyyy
MM-dd-yy
MM-dd-yyyy
MM/dd/yy

DateTime (8) Convert unformatted_date From: MM/dd/yy

DateTimeYear([date])
month = DateTimeMonth([date])
day = DateTimeDay([date])

Results - DateTime (8) - Output

0 of 0 Fields

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

Alteryx Designer x64 - etl2.yxmd

File Edit View Options Help

★ Favorites ★ Recommended ◀ ▶ In/Out Preparation Join Parse Transform In-Database Reporting Documentation Spatial Machine Learning Text Mining Compute

Formula (9) - Configuration

Output Column Data Preview

1 year

DateTimeYear([date])

Data type: Int16 Size: 2

2 month

DateTimeMonth([date])

3 day

DateTimeDay([date])

4 quarter

IF [month] <= 3 THEN 1 ELSEIF [month] > 3 AND [month] <= 6 THEN 2 ELSEIF [month] > 6 and [month] <= 9 THEN 3 ELSE 4 ENDIF

Data type: Int16 Size: 2

6 gp_margin

Results - Formula (9) - Output

0 of 0 Fields

Use 'Select' tool to move columns Up or Down

The screenshot shows the Alteryx Designer interface. On the left, the 'Select (10) - Configuration' pane is open, displaying a list of fields with checkboxes. The 'Options' tab is selected, showing up and down arrows for column reordering. The main workflow area shows a sequence of tools: 'Input Data', 'Select (7)', 'DateTime (8)', 'Formula (9)', 'Select (10)', and 'Browse (6)'. The 'Select (10)' tool is highlighted with a red box. The 'Results - Select (10) - Messages' pane at the bottom shows 0 errors, 0 warnings, and 0 info messages.

ETL 3 - Outliers

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

Note the True and False connectors to the Browse tools.

The screenshot shows the Alteryx Designer interface. On the left, the 'Filter (11) - Configuration' pane is open, showing the 'Custom filter' section with the expression '[revenue] <= 88 AND [revenue] >= -5'. The main workflow area shows a sequence of tools: 'Input Data', 'Select (7)', 'DateTime (8)', 'Formula (9)', 'Select (10)', 'Filter (11)', 'Sort (13)', 'Select (14)', and 'Browse (12)'. The 'Filter (11)' tool is highlighted with a red box. The 'Results - Browse (12) - Input' pane at the bottom shows a table of data with columns: product_name, category_name, parent_name, revenue, and a 'Data Cleanse' column.

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

The screenshot shows the Alteryx Designer interface. On the left, the 'Sort (13) - Configuration' pane is open, showing the 'Sort by' section with 'revenue' selected and 'Ascending' chosen. The main workflow area shows a sequence of tools: 'Input Data', 'Select (7)', 'DateTime (8)', 'Formula (9)', 'Select (10)', 'Filter (11)', 'Sort (13)', 'Select (14)', and 'Browse (12)'. The 'Sort (13)' tool is highlighted with a red box. The 'Results - Browse (12) - Input' pane at the bottom shows a table of data with columns: product_name, category_name, parent_name, revenue, and a 'Data Cleanse' column.

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

The screenshot shows the Alteryx interface. At the top, a data preview table is visible with columns: longitude, site_status, revenue, gross_profit, gp_margin, and cost. Below this, the 'Sort (13) - Configuration' tool is highlighted with a red box. The configuration shows the 'Name' field set to 'revenue' and the 'Order' set to 'Ascending'. The main workflow area shows a sequence of tools: Input Data (localPosSmall.csv), Select (7), DateTime (8) - Convert unformatted_date From: MM/dd/yy, Formula (9) - year = DateTimeYear([date]), month = DateTimeMonth([date]), day = DateTimeDay([date]), Select (10), Filter (11) - [revenue] <= 88 AND [revenue] >= -5, Sort (13) - revenue - Ascending (highlighted with a red box), Select (14), and Browse (12).

Use the 'Select' tool to display only selected columns

The screenshot shows the Alteryx interface. On the left, the 'Field List' is visible, showing a list of fields with their types and sizes. The 'product_name' field is highlighted with a red box. The main workflow area shows a sequence of tools: Input Data (localPosSmall.csv), Select (7), DateTime (8) - Convert unformatted_date From: MM/dd/yy, Formula (9) - year = DateTimeYear([date]), month = DateTimeMonth([date]), day = DateTimeDay([date]), Select (10), Filter (11) - [revenue] <= 88 AND [revenue] >= -5, Sort (13) - revenue - Ascending, Select (14) (highlighted with a red box), and Browse (12). Below the workflow, the 'Results - Select (14) - Output' table is displayed, showing 4 fields: product_name, category_name, parent_name, and revenue. The first two records are visible:

Record	product_name	category_name	parent_name	revenue
1	LOTTERY WIN	Lottery-win (270)	Store Services (92)	-100
2	LOTTERY WIN	Lottery-win (270)	Store Services (92)	-100

ETL 4 – Missing Values

Data Cleansing

The screenshot shows the Alteryx Data Cleansing tool configuration. The 'Options' pane on the left includes settings for removing null data, selecting fields to cleanse, replacing nulls, and removing unwanted characters. The main workflow area shows a sequence of tools: 'Input Data', 'Filter', 'Formula', 'Sample', 'Select', 'Sort', 'Join', 'Union', 'Text To Columns', 'Summarize', and 'Output Data'. The 'Results' pane at the bottom displays a table of 29 fields with 2,970 records displayed.

Record	site_name	address	city	zip	latitude
1	562 Columbia	101 S Providence Rd	Columbia	65203	38.9509
2	497 Patton	1000 State Route 51	Patton	63662	37.5003

Filter out missing columns

The screenshot shows the Alteryx Filter tool configuration. The 'Select Basic or Custom Filter' pane on the left shows the 'Basic filter' selected with the condition 'gross_profit Is not null'. The main workflow area shows a sequence of tools: 'Input Data', 'Convert unformatted_date', 'Formula', 'Sample', 'Select', 'Sort', 'Join', 'Union', 'Text To Columns', 'Summarize', and 'Output Data'. The 'Results' pane at the bottom displays a message: 'Filter (12) 2899 records were True and 71 were False'.

Imputation

The screenshot shows the Alteryx Designer interface with the 'Filter (12) - Configuration' pane on the left. The 'Basic filter' option is selected, and the filter expression is set to `IsNotNull([gross_profit])`. The main workspace displays a workflow starting with 'Input Data' (3,000 records, 604KB), followed by a 'Convert unformatted_date' tool, and then a 'Filter (12)' tool (highlighted with a red box). The 'Filter (12)' tool output shows 2,899 records (633KB). The 'Results - Filter (12) - Messages' pane at the bottom indicates 0 errors, 0 conversions, 0 warnings, 1 info, and 0 files, with a message stating 'Filter (12) 2899 records were True and 71 were False'.

Imputation – Null values removed

The screenshot shows the Alteryx Designer interface with the 'Browse (2) - Configuration' pane on the left. The 'Profile' tab is selected, showing a summary of the 'gross_profit' field. The main workspace displays a workflow starting with 'Input Data' (3,000 records, 604KB), followed by a 'Filter (12)' tool, and then a 'Browse (2)' tool. The 'Browse (2)' tool output shows 2,970 records (356KB). The 'Results - Browse (2) - Input' pane at the bottom indicates 29 of 29 fields displayed, 2,970 records displayed, and 356 KB.

Summary

Type	Records	Data Type	Size
Float	2,970	4	

Value Statistics

Grouped Values	Count	Percentage
Ok	2,970	100.00%
Unique	2,325	78.28%
Null	0	0.00%
Not Ok	0	0.00%
Empty	0	0.00%

ETL 5 – Data Joins

Join with states.csv using zip code

The screenshot shows the Alteryx Designer interface for a Join tool configuration. The tool is set to 'Join by Specific Fields'. The left dataset has a field 'zip' and the right dataset has a field 'zip'. The join is configured as an inner join. The results pane shows that 2970 records were joined with 0 un-joined left records and 264 un-joined right records.

Input	Field	Type	Size
Left	unique_id	V_String	25-
Left	transaction_id	V_String	25-
Left	date	Date	10
Left	year	Int16	2
Left	month	Int16	2
Left	day	Int16	2
Left	dow	V_WString	10-
Left	quarter	Int16	2
Left	customer_id	V_String	25-
Left	product_id	V_String	25-
Left	product_name	V_String	25-

ETL 6 – Workflow

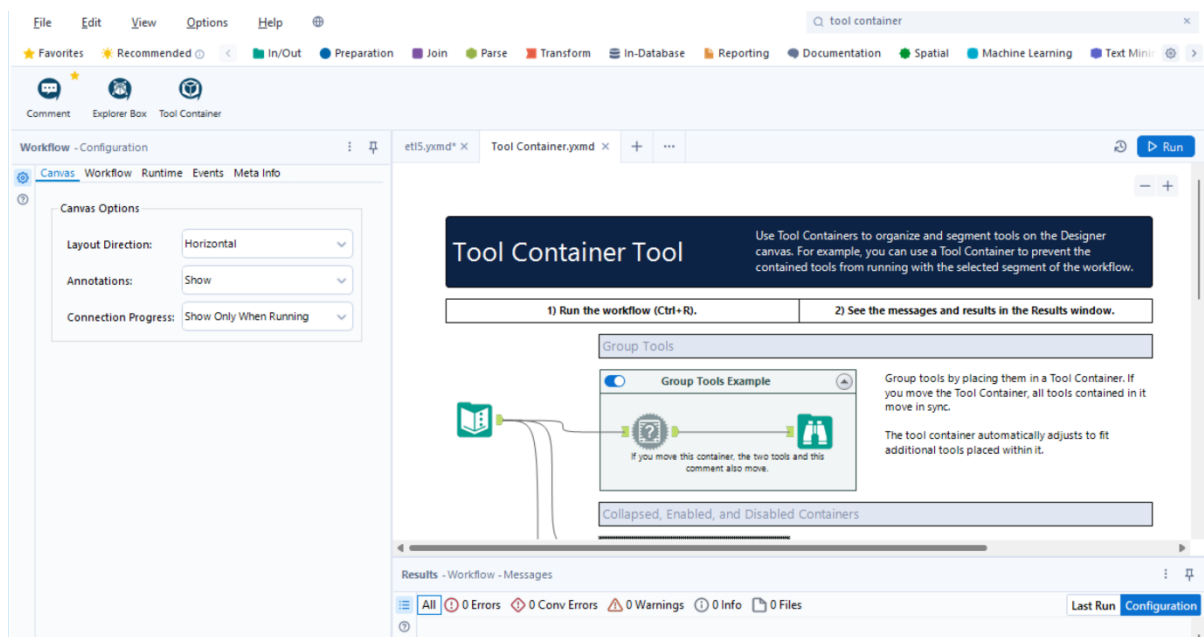
Write data to file to visualise in Power BI or R or Tableau

Write to output file -dataForEDA.csv

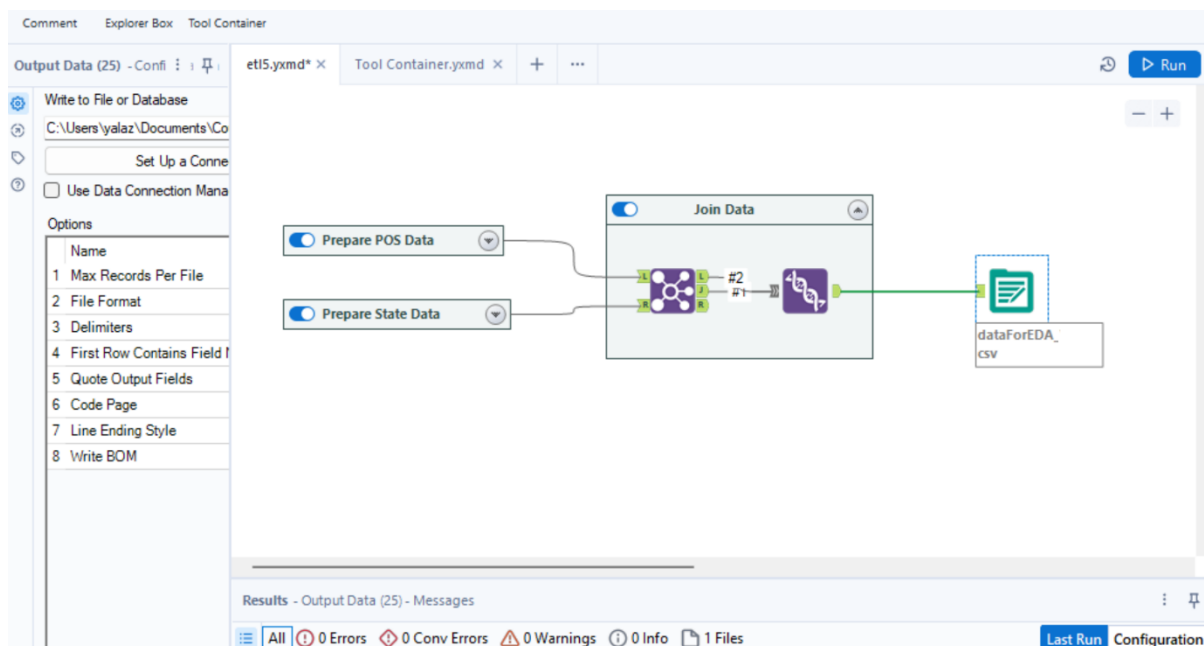
The screenshot shows the Alteryx Designer interface for a workflow configuration. The workflow includes a Join tool, a Formula tool, and an Output Data tool. The Output Data tool is configured to write data to a file named 'dataForEDA_YAA.csv'. The results pane shows that 33 conversion errors, 5 warnings, and 5 info messages were generated during the workflow execution.

Input	Field	Type	Size
Left	unique_id	V_String	25-
Left	transaction_id	V_String	25-
Left	date	Date	10
Left	year	Int16	2
Left	month	Int16	2
Left	day	Int16	2
Left	dow	V_WString	10-
Left	quarter	Int16	2
Left	customer_id	V_String	25-
Left	product_id	V_String	25-
Left	product_name	V_String	25-

Using Tool Container



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



Adding a comment

The screenshot displays the Alteryx software interface. On the left, the 'Comment (29) - Configuration' window is open, showing a text box with the content 'This is all data prep'. Below the text box, the 'Shape' is set to 'Rounded Rectangle', the 'Font' is 'Cambria, 15.75pt, Style = Bold', the 'Text Color' is 'Black', and the 'Background Color' is 'Blue'. The main workspace shows a data flow diagram with a green background. The diagram includes a 'Join Data' tool, a 'Prepare POS Data' tool, and a 'Prepare State Data' tool. The output of the 'Join Data' tool is connected to a 'dataForEDA.csv' file. The bottom status bar indicates 'Results - Comment (29) - Messages' with 'All', '0 Errors', '0 Conv Errors', '0 Warnings', '0 Info', and '0 Files'.