

Using Alteryx to Join Data Streams

Alteryx is used to automate the ETL process. In this exercise, you will:

Extract data files from Excel.

Transform the data into a flat data set in which all the data for a specific sales order is on one row. Aggregate the Number of units sold, sales and shipping cost by salesperson and shipping costs by product and region.

Load the transformed data into Excel.

There are five streams of data in the Excel file, *Stuffed Animal.Join.Demo.xlsx*. The first stream is stored in a sheet titled “**Sales Data**”. Notice that the column headings are in row three.

	A	B	C	D	E	F
1	Cuddly Stuffed Animals					
2	Sales					
3	Sales Order #	Salesperson #	Region	State	Product #	Units Sold
4	35005	1303	West	WA	103	120
5	35006	1302	West	AZ	101	96
6	35009	1305	West	CA	102	156
7	35011	1305	West	CO	101	144
8	35014	1302	Midwest	IN	106	84
9	35017	1305	West	NV	103	72
10	35025	1301	Northeast	CT	106	84
11	35026	1306	West	WY	105	132
12	35027	1301	Midwest	IL	103	120
13	35032	1305	South	AL	103	132
14	35041	1305	Midwest	MI	103	144
15	35046	1302	Midwest	KS	102	132
16	35052	1303	South	LA	102	108

Notice that the last “Sales Order #” is 50150, and there are 5,017 rows of data (row 5020 – row 3 = 5017 rows).

	A	B	C	D	E	F
1	Cuddly Stuffed Animals					
2	Sales					
3	Sales Order #	Salesperson #	Region	State	Product #	Units Sold
5016	50138	1303	West	NM	102	108
5017	50141	1304	West	AK	102	96
5018	50143	1303	West	NV	102	120
5019	50147	1306	Northeast	NH	106	72
5020	50150	1305	Northeast	NY	106	156
5021						

The second stream is stored in a sheet titled **"Products"**. Notice that the column headings are in row fifteen.

The last "Product ID" is 106, and there are 6 rows of data.

	B	C	D	E	F	G
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						

Cuddly Stuffed Animals
Product Information

Product ID	Product	Unit Selling Price	Unit Manufacturing Cost	Profit Margin by Product
101	Stuffed Lamb	20.00	8.75	11.25
102	Stuffed Giraffe	22.00	9.75	12.25
103	Stuffed Elephant	24.00	10.75	13.25
104	Stuffed Unicorn	21.00	8.50	12.50
105	Stuffed Horse	23.00	10.75	12.25
106	Stuffed Pig	25.00	12.50	12.50

The third stream is stored in a sheet titled **"Salesperson"**. Notice that the column headings are in row fifteen.

	C	D	E	F	G	H
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						

Cuddly Stuffed Animals
Salesperson Information

Code	Salesperson	Title
1301	James Polk	Sales Associate I
1302	Ulysses Grant	Senior Sales Associate
1303	Thomas Jefferson	Senior Sales Associate
1304	James Madison	Sales Associate I
1305	James Monroe	Sales Associate I
1306	Grover Cleveland	Senior Sales Associate

The last "Code" is 1306, and there are 6 rows of data.

The fourth stream contains the shipping cost per unit by the product and region in which the sales were made is stored in a sheet titled **"Shipping Costs"**. Notice that the column headings are in row thirteen.

Cuddly Stuffed Animals Shipping Cost Per Unit				
Product ID	Region			
	Midwest	Northeast	South	West
101	2.15	2.12	2.08	2.02
102	2.11	2.08	2.04	1.98
103	2.14	2.11	2.07	2.01
104	2.12	2.09	2.05	1.99
105	2.07	2.04	2.00	1.94
106	2.02	1.99	1.95	1.89

The last "Product ID" is 106, and there are 6 rows of data.

The fifth stream contains data regarding the Regions and states in each region sold are stored in a sheet titled **"Region"**.

Cuddly Stuffed Animals Region I								
Code	Region	1	2	3	4	5	6	7
1	Midwest	IL	IN	IA	KS	MI	MN	MO
2	Northeast	CT	ME	MA	NH	NJ	NY	PA
3	South	AL	AR	DE	FL	GA	KY	LA
4	West	AK	AZ	CA	CO	HI	ID	MT

The deliverable consists of two reports exported to the original Excel file. The first report is a report by salesperson that shows the number of units sold, sales in dollars and shipping costs. The second report is a shipping cost report by product and region loaded back into Excel.

By Salesperson			
Salesperson	# of units	Sales	Shipping Costs
Ulysses Grant	36,636	\$818,052.00	\$77,266.44
George Washington	58,284	\$1,299,192.00	\$123,242.28
James Monroe	97,152	\$2,162,436.00	\$205,622.64
John Tyler	78,924	\$1,751,352.00	\$166,596.00
James Madison	38,388	\$854,412.00	\$81,105.24
James Polk	77,484	\$1,723,956.00	\$163,717.08
Total	386,868	\$8,609,400.00	\$817,549.68

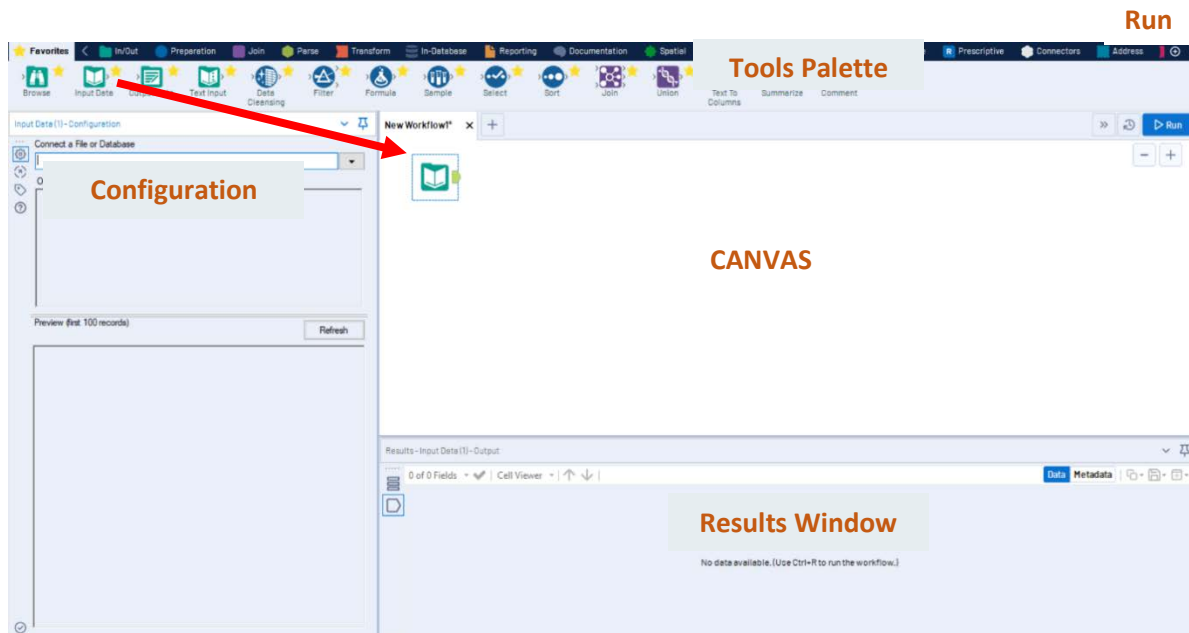
Automate the process so that the reports can be completed at the end of every month. The unit selling price and costs do not change during the year.

Shipping Costs by Product and Region					
	Midwest	Northeast	South	West	Grand Total
Stuffed Bear	\$ 26,904.00	\$ 43,048.44	\$ 22,066.20	\$ 41,647.80	\$ 133,666.44
Stuffed Elephant	\$ 54,810.00	\$ 72,060.84	\$ 48,486.60	\$ 86,971.80	\$ 262,329.24
Stuffed Frog	\$ 36,405.60	\$ 52,861.20	\$ 30,753.60	\$ 54,081.00	\$ 174,101.40
Stuffed Koala	\$ 14,094.00	\$ 20,699.28	\$ 14,572.80	\$ 25,364.40	\$ 74,730.48
Stuffed Panda	\$ 16,875.60	\$ 26,664.00	\$ 15,720.00	\$ 23,108.40	\$ 82,368.00
Stuffed Unicorn	\$ 13,648.20	\$ 30,604.32	\$ 18,144.00	\$ 27,957.60	\$ 90,354.12
Grand Total	\$ 162,737.40	\$ 245,938.08	\$ 149,743.20	\$ 259,131.00	\$ 817,549.68

Start Alteryx

Note: You cannot extract data from an open Excel file, so make sure that the Excel file is closed before starting to input the data.

There are four major parts of the Alteryx window. The first is the Canvas, and that's where the workflow will be created. Second, the Tools Palette is where all the tools are stored. Third, the Configuration area is where the tool settings are modified. Lastly, the Results window shows the outcome after "Run" is selected to apply the changes.

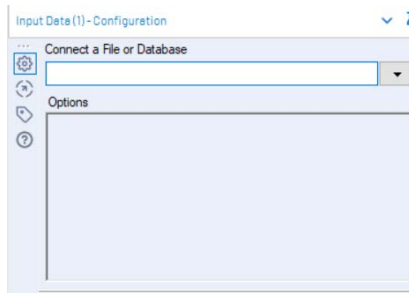


Requirement 1: Extract the data from the "Sales Data" sheet. Compare the values in the "Region" field to the values on the "Region" sheet and make any required changes.

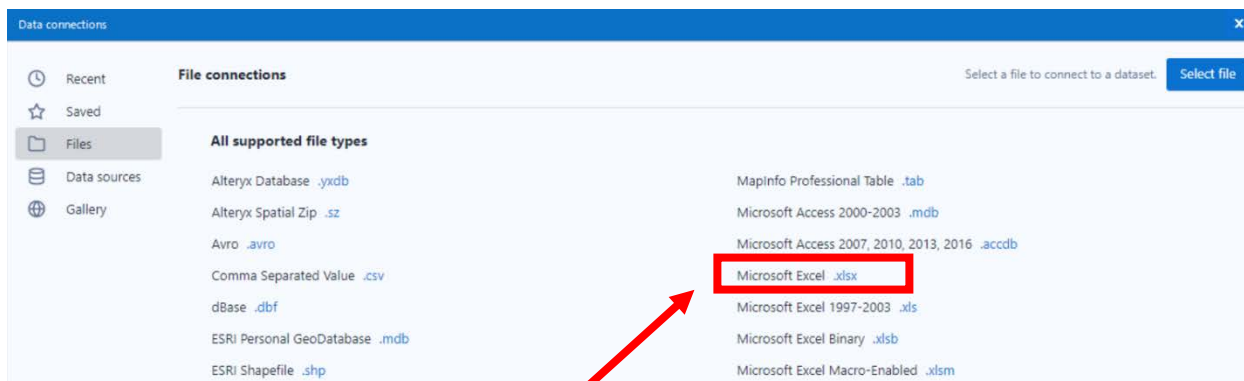
Tool 1: Input Data Tool – Input Excel file

Drag the “Input Data Tool”  from the Tools Palette to the Canvas.

In the Configuration area, select the “Connect a File or Database” drop-down arrow.



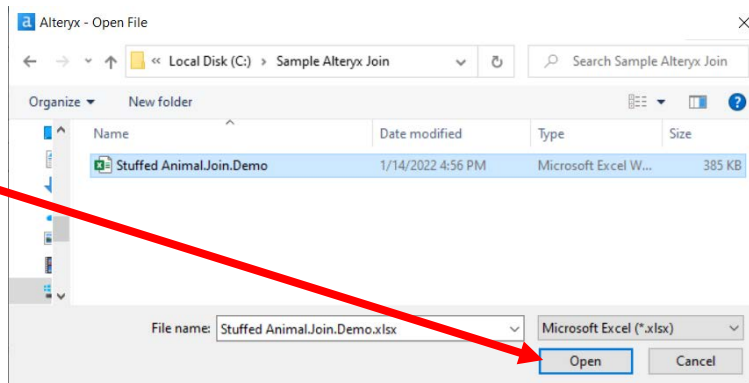
Select “File”.



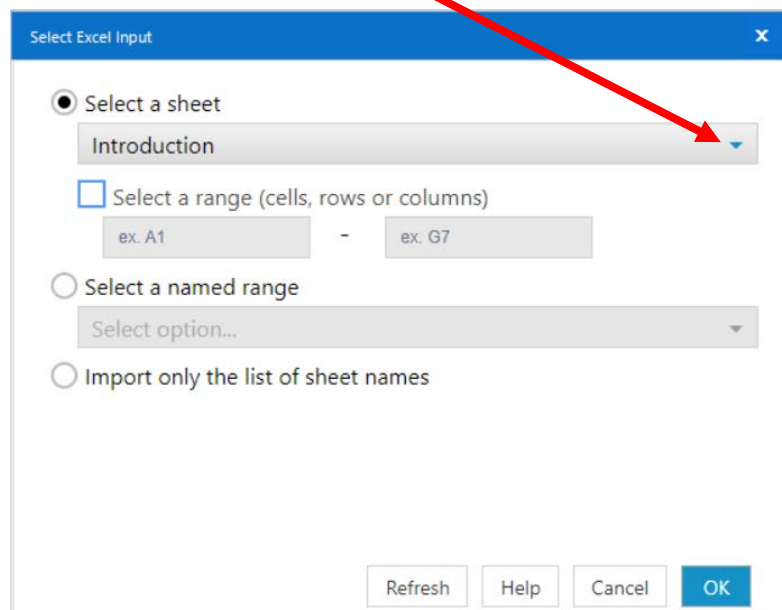
Select “Microsoft Excel .xlsx”.

Browse for the desired file and select the version with the .xlsx file extension.

Select "Open".



Select the drop-down arrow to the right of "Introduction".



Select "Sales Data".

Select "OK".

Select a sheet

- Introduction
- Introduction
- Products
- Region
- Sales Data
- Salesperson
- Shipping Costs

Import only the list of sheet names

Refresh Help Cancel OK

Remember that the field names are in the third row.
In the Configuration area, change the number in “7 Start Data Import on Line” to “3”.

Input Data (2) - Configuration

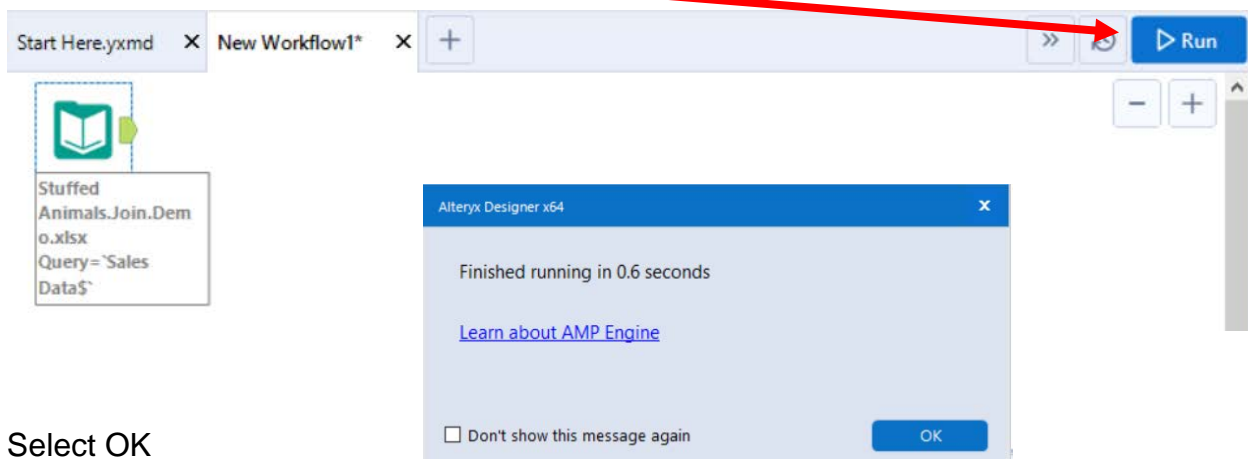
Connect a File or Database

C:\Sample Alteryx Join\Stuffed Animals.Join.Demo.xlsx

Options

Name	Value
2 File Format	Microsoft Excel (*.xlsx)
3 Table or Query	'Sales Data\$'
4 Search SubDirs	<input type="checkbox"/>
5 Output File Name as Field	No
6 First Row Contains Data	<input type="checkbox"/>
7 Start Data Import on Line	3

Select “Run”.



Select OK

The Results window will display the field headings and the data.

Notice that six fields were extracted

Results - Input Data (2) - Output

6 of 6 Fields ✓ Cell Viewer 5,017 records displayed

Record	Sales Order #	Salesperson #	Region	State	Product #	Units Sold
5,006	50122	1303	Midwest	IA	103	48
5,007	50124	1304	Northeast	NH	105	72
5,008	50126	1303	South	WV	106	60
5,009	50128	1302	South	KY	102	72
5,010	50129	1301	West	HI	101	96
5,011	50133	1303	South	KY	102	72
5,012	50137	1301	Northeast	ME	103	108
5,013	50138	1303	West	NM	102	108
5,014	50141	1304	West	AK	102	96
5,015	50143	1303	West	NV	102	120
5,016	50147	1306	Northeast	NH	106	72
5,017	50150	1305	Northeast	NY	106	156

There are 5,017 rows of data and the last one is “Sales Order #” 50150.

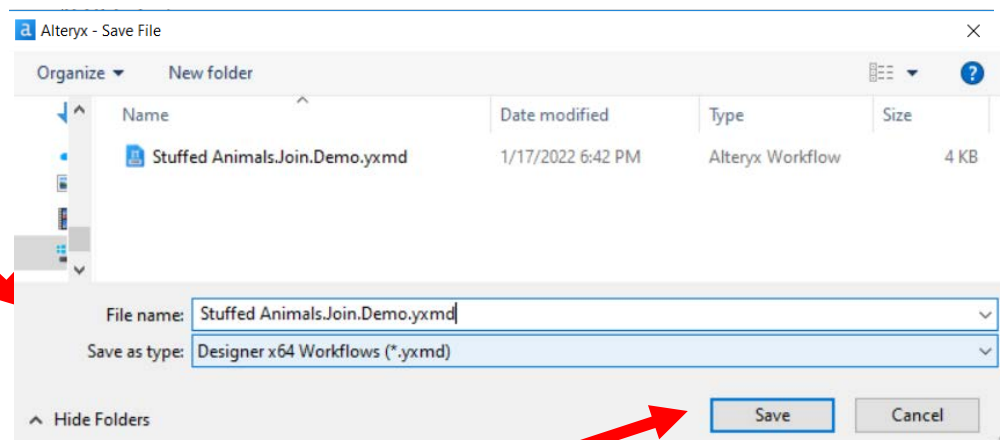
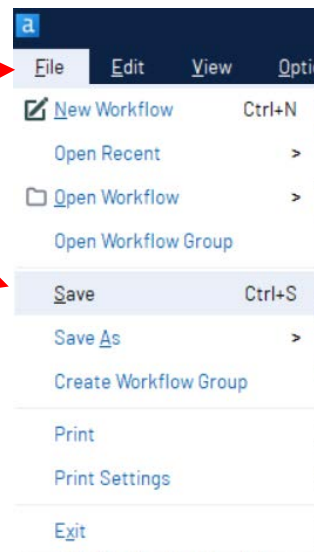
The workflow must be saved.

Select "File".

Select "Save".

Browse to the required directory.

Type in the name of the file that will be uploaded.

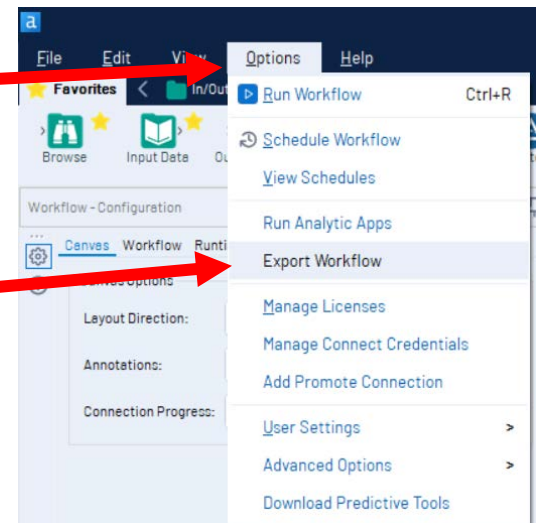


Select "Save".

Export the workflow and the related input files so that the project can be completed on another computer or uploaded to the web.

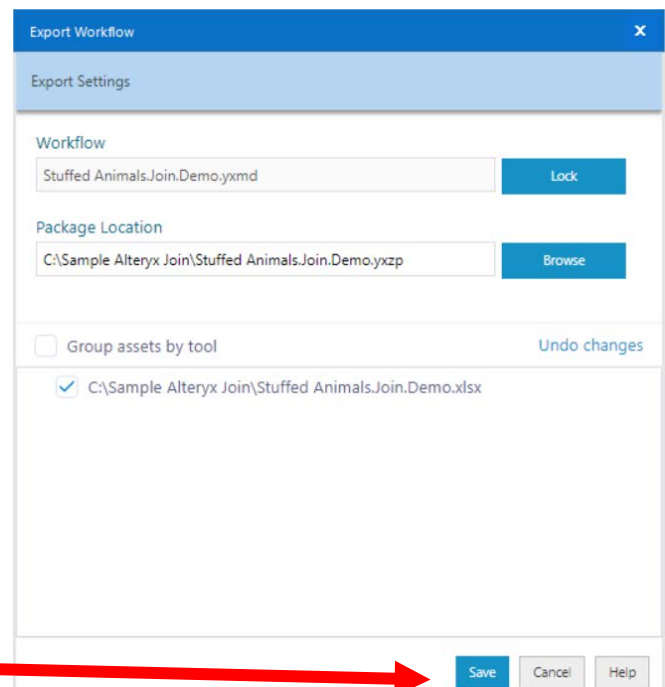
Select "Options".

Select "Export Workflow".

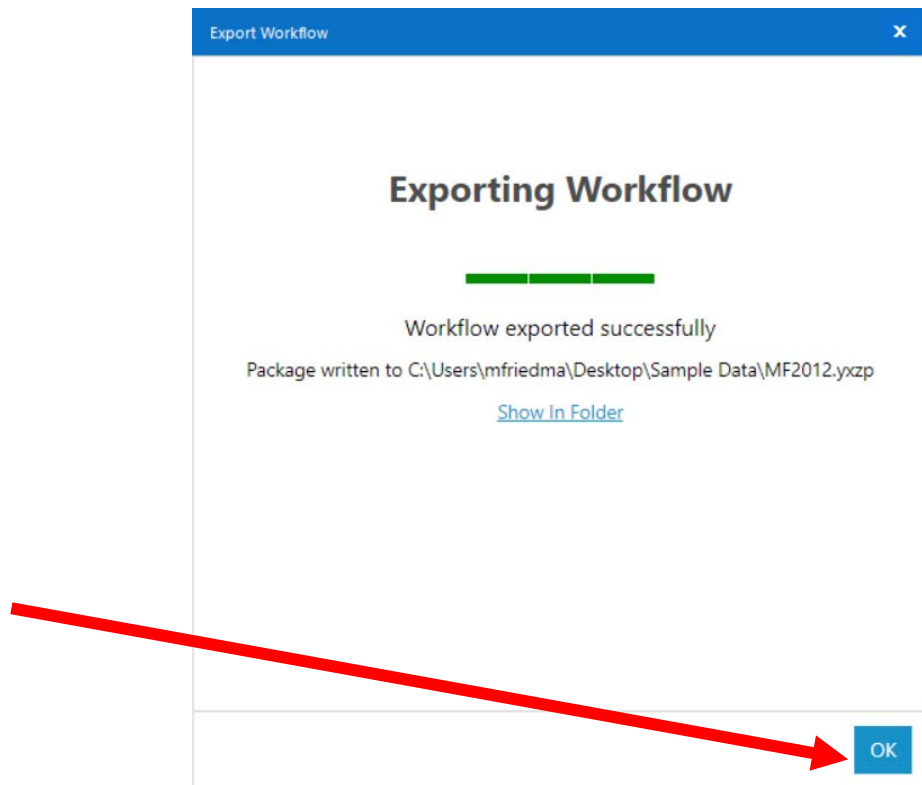


Verify the "Workflow".

Verify the "Package Location".



Select "OK"




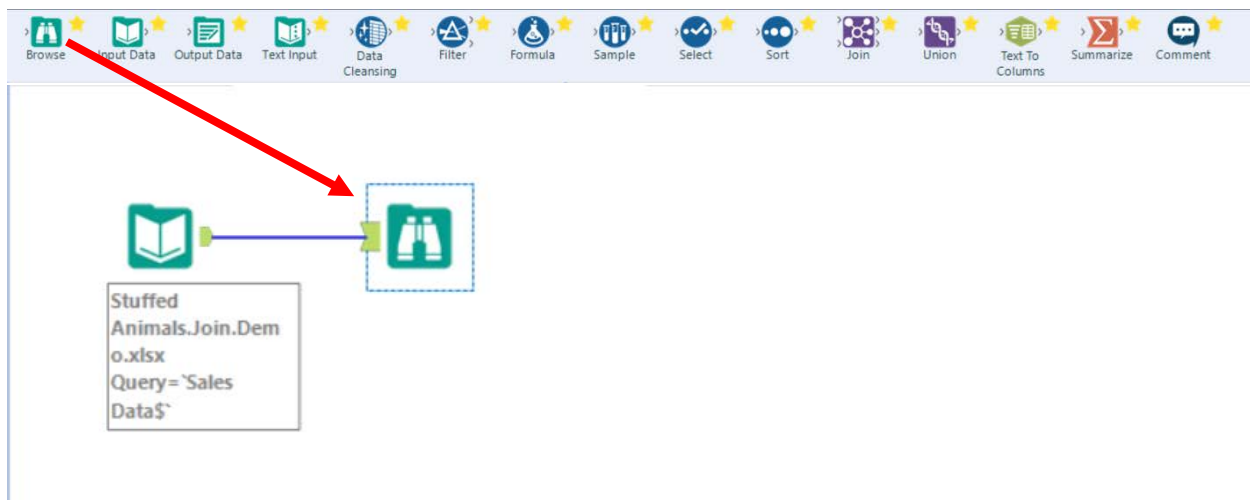
Temporary Tool 2: Browse Tool – The Data in the Fields Must Conform to Predefined Rules.

In the original Excel file, The Cuddly Stuffed Animals company provided a table with the only acceptable values for “Region”. Sometime data is entered incorrectly.

The “Region” field can only contain one of the four regions. Browse the data in the “Region” field to verify that all the data is correct. Make any changes that are required.

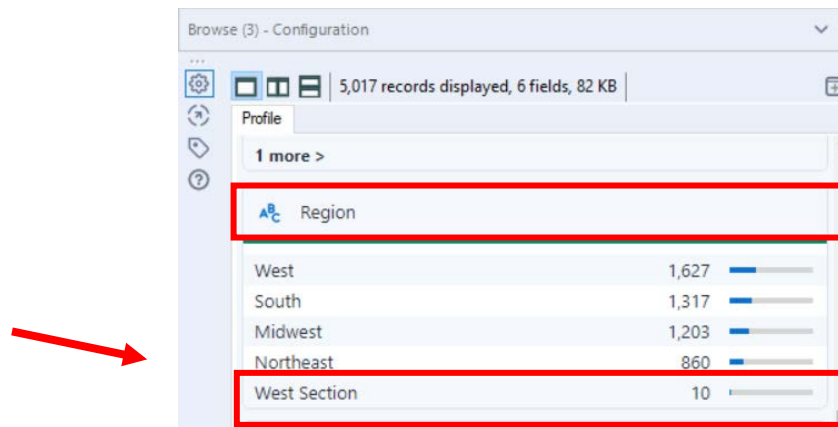
Code	Region
1	Midwest
2	Northeast
3	South
4	West

Drag the “**Browse Tool**”  from the Tools Palette to the output of the “**Input Data Tool**” on the Canvas. The two will be connected automatically.




Select Run.

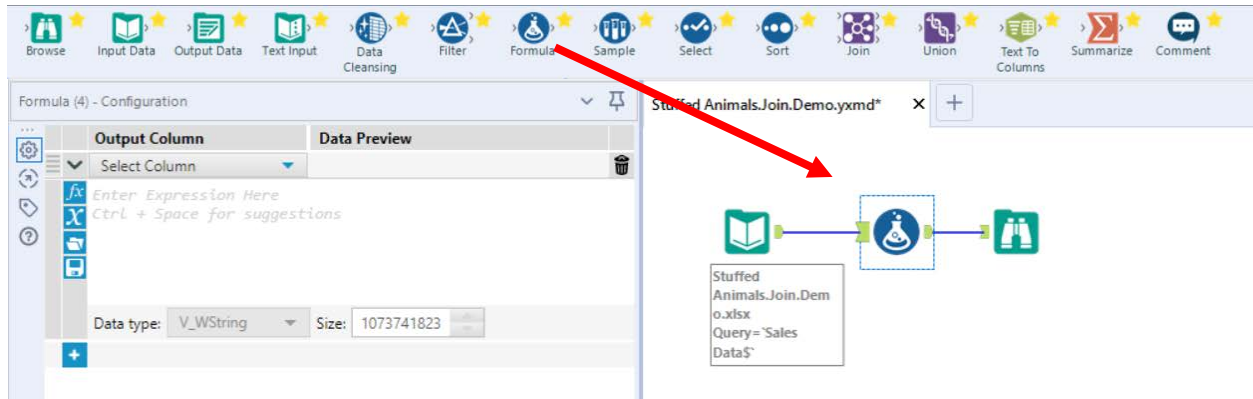
In the Configuration section scroll down to “Region”.



“West Section” must be replaced with “West” for a total of 1,637 distinct values.

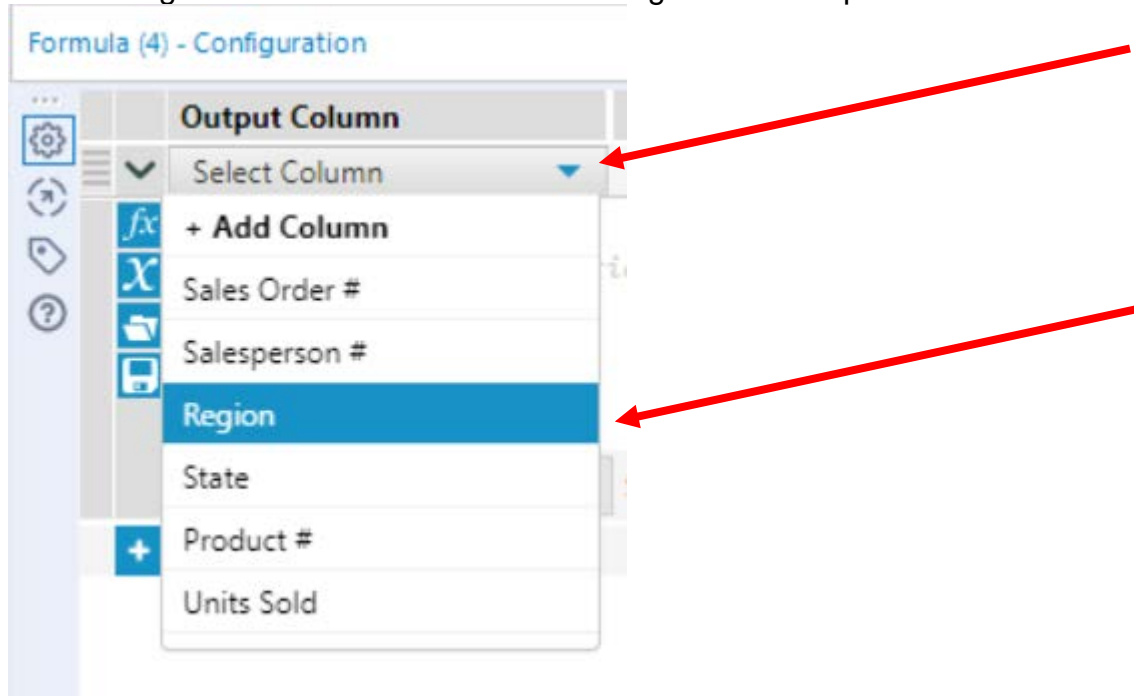
Tool 2: Formula Tool – The Data in the Fields Must Conform to Predefined Rules.

Drag the “**Formula Tool**”  from the Tools Palette between the output of the “**Input Data Tool**” and the “**Browse Tool**” on the Canvas. The three will be connected automatically.

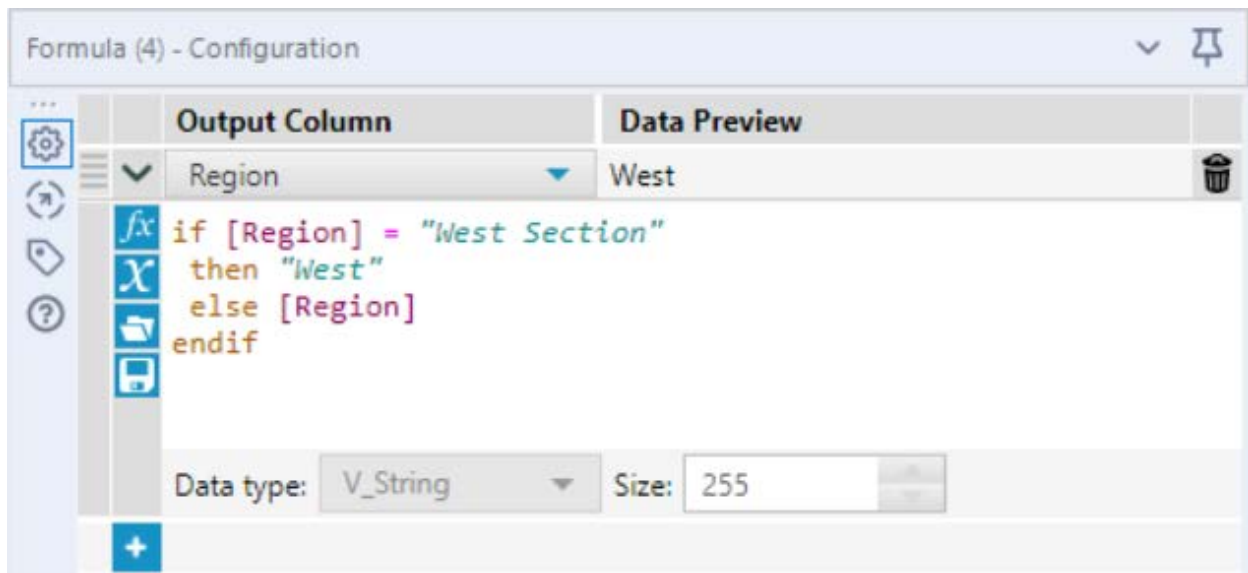


If the value in the field “Region” is “West Section” replace it with “West” otherwise leave the existing name in the field. Do not create a new field, put the output back into the “Region” field.

In the configuration section select the field Region from the pull-down list.



Type in the formula.



Select the Browse step.

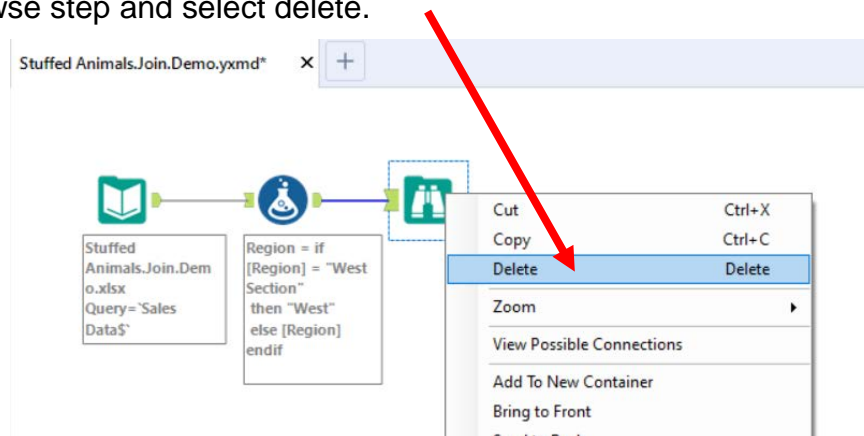
Select Run.

In the Configuration section scroll down to "Region". There are only four regions and West has 1,637 distinct values.

The screenshot shows the 'Data Preview' window for the 'Region' column. It displays a table with four rows: West, South, Midwest, and Northeast. The 'West' row has a count of 1,637, 'South' has 1,317, 'Midwest' has 1,203, and 'Northeast' has 860. Each row has a blue progress bar to its right.

Region	Count
West	1,637
South	1,317
Midwest	1,203
Northeast	860

Right click on the Browse step and select delete.



Requirement 2: Extract the data from the "Products" sheet. Review the data, join to Stream 1 and verify the number of rows.

Data regarding the Product's ID, name, selling price, manufacturing cost and profit margin are stored in an Excel sheet titled "Products". Notice that the column headings are in the fifteen row.

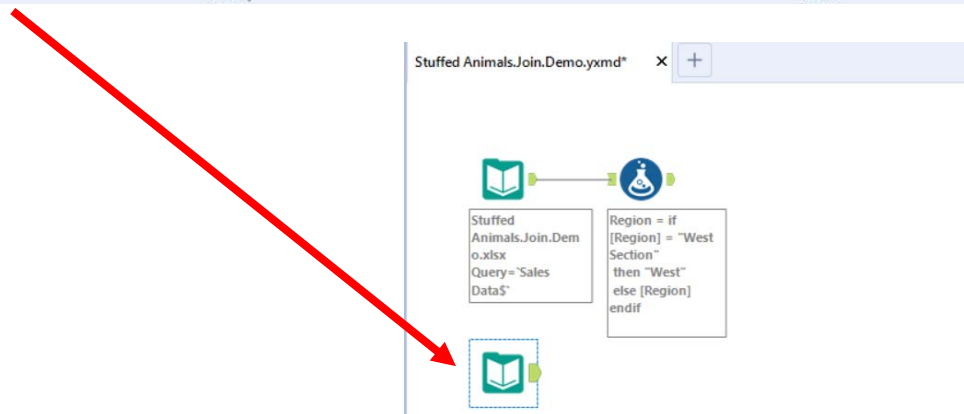
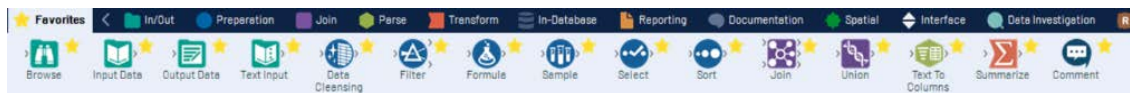
	B	C	D	E	F	G
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						

	B	C	D	E	F	G
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						

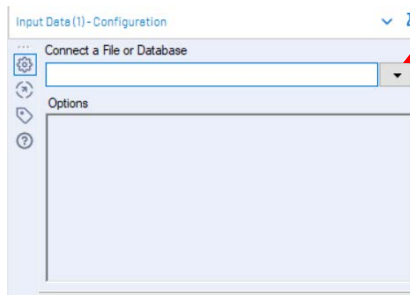
Tool 3: **Input Data Tool** – Input Excel file



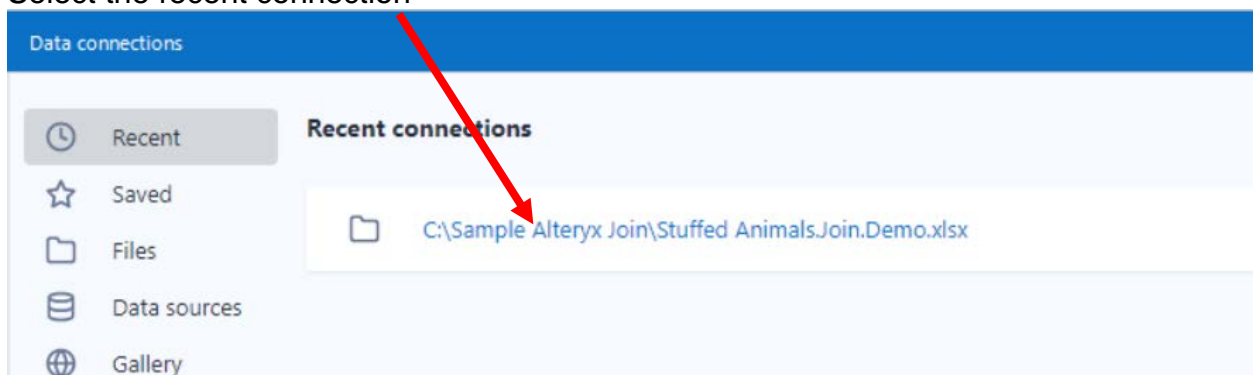
Drag the "**Input Data Tool**" from the Tools Palette to the Canvas.



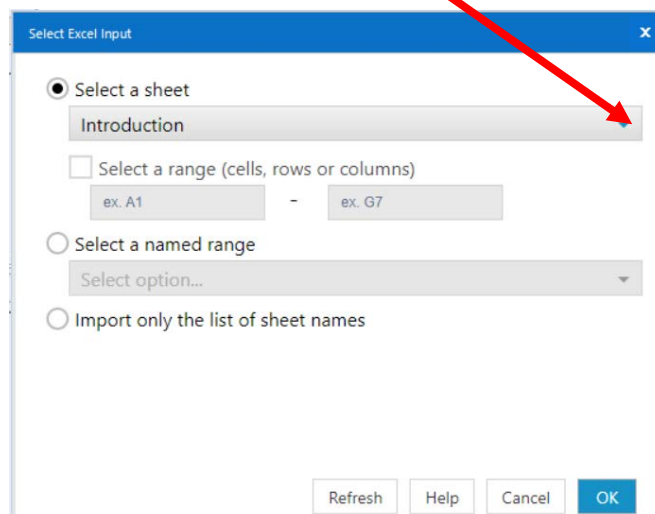
In the Configuration area, select the “Connect a File or Database” drop-down arrow.



Select the recent connection

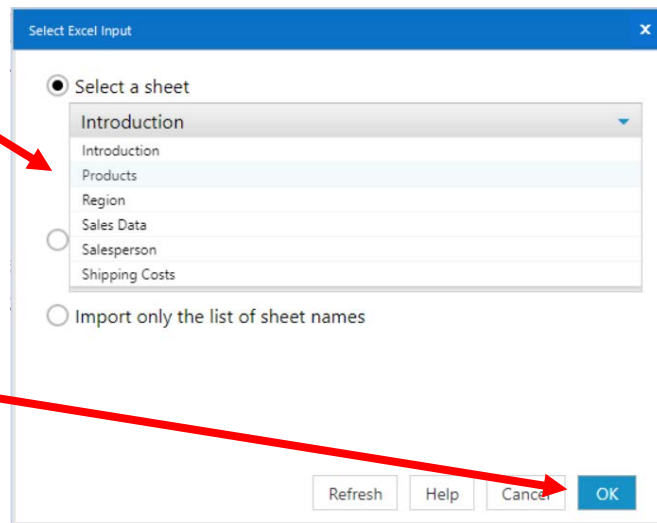


Select the drop-down arrow to the right of “Introduction”.



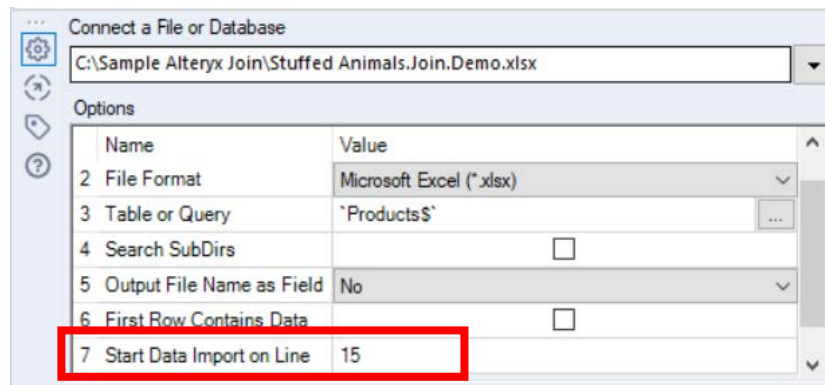
Select "Products".

Select "OK".



Remember that the field names are in the fifteenth row.

In the Configuration area, change the number in "7 Start Data Import on Line" to "15".



Select "Run".

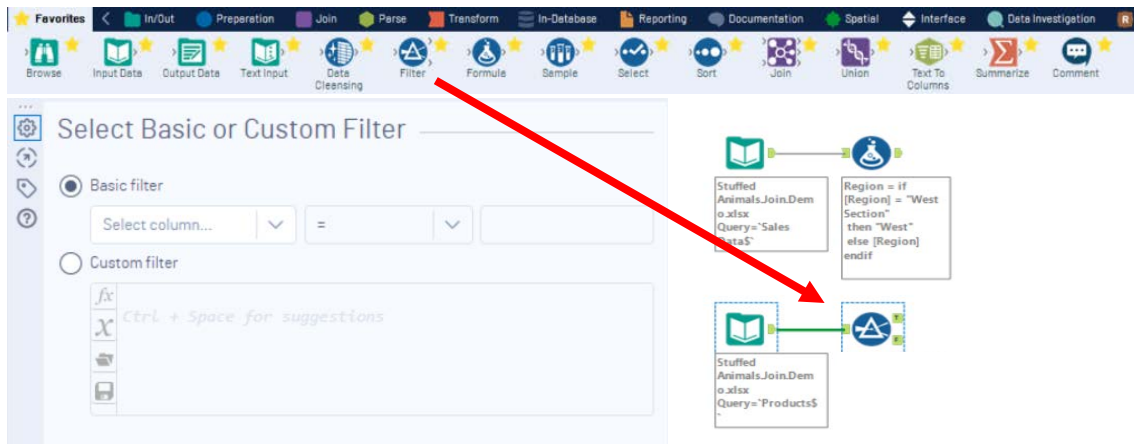
The screenshot shows the Alteryx interface. At the top, a workflow is visible with two tools: 'Stuffed Animals.Join.Demo.o.xlsx' and 'Region = if [Region] = "West Section" then "West" else [Region] endif'. A red arrow points from the text 'Select "Run"' to the 'Run' button in the top right corner. Below the workflow, the 'Results - Input Data (8) - Output' window is open. It shows a table with 5 fields and 11 records displayed. The table has columns: Record, Product ID, Product, Unit Selling Price, Unit Manufacturing Cost, and Profit Margin by Product. The data is as follows:

Record	Product ID	Product	Unit Selling Price	Unit Manufacturing Cost	Profit Margin by Product
1	101	Stuffed Lamb	20	8.75	11.25
2	[Null]	[Null]	[Null]	[Null]	[Null]
3	102	Stuffed Giraffe	22	9.75	12.25
4	[Null]	[Null]	[Null]	[Null]	[Null]
5	103	Stuffed Elephant	24	10.75	13.25
6	[Null]	[Null]	[Null]	[Null]	[Null]
7	104	Stuffed Unicorn	21	8.5	12.5
8	[Null]	[Null]	[Null]	[Null]	[Null]
9	105	Stuffed Horse	23	10.75	12.25
10	[Null]	[Null]	[Null]	[Null]	[Null]
11	106	Stuffed Pig	25	12.5	12.5

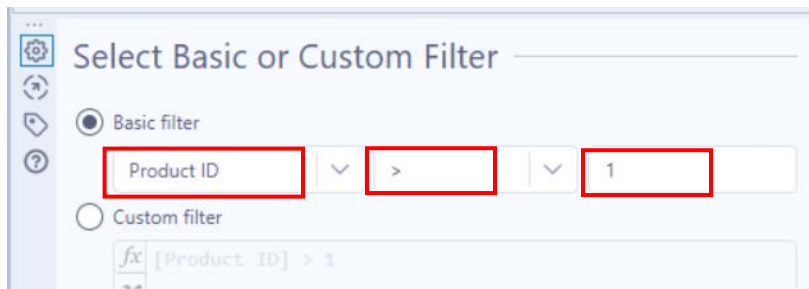
The Results window will display the field headings and the data. Notice that five fields were extracted and that there are eleven records. The extra rows that contain [Null] must be removed. The last product's ID" is 106.

Tool 4: Filter Tool – Exclude all rows that do not contain data

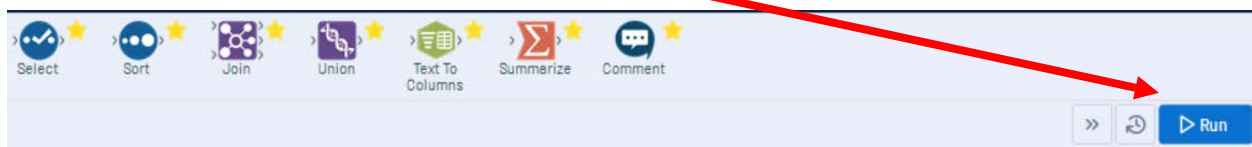
Drag the “**Filter Tool**” from the Tools Palette to the right of the “**Input Data Tool**” on the Canvas. The two will be connected automatically. The outflow from the tool on the left is the inflow to the one on the right.



With the “**Filter Tool**” selected, the “Basic” filter in the Configuration area can be modified to filter the field “Product ID” for those accounts greater than, “>” “1”.




To apply the latest change, select “Run”.



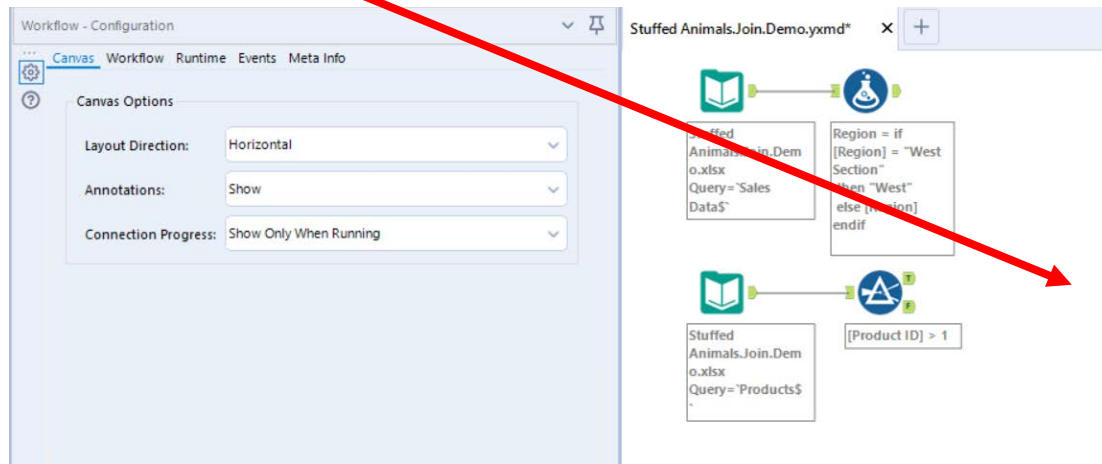
The records that meet the filtered conditions, “Product ID > 1”, are displayed as T (True). There are six records displayed and the last one is “Product ID” 106.

Results - Filter (10) - Out - True

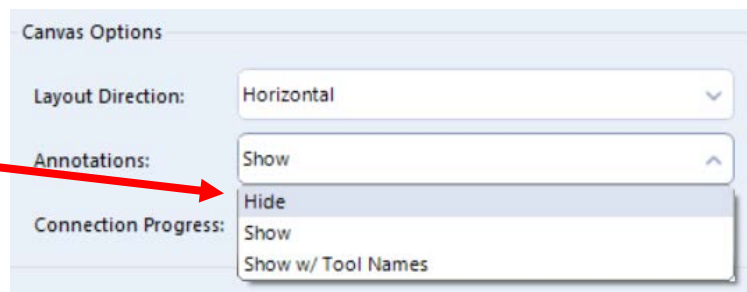
5 of 5 Fields ▾ ✓ Cell Viewer ▾ 6 records displayed | ↑ ↓ |

Record	Product ID	Product	Unit Selling Price	Unit Manufacturing Cost	Profit Margin by Product
	1 101	Stuffed Lamb	20	8.75	11.25
	2 102	Stuffed Giraffe	22	9.75	12.25
	3 103	Stuffed Elephant	24	10.75	13.25
	4 104	Stuffed Unicorn	21	8.5	12.5
	5 105	Stuffed Horse	23	10.75	12.25
	6 106	Stuffed Pig	25	12.5	12.5

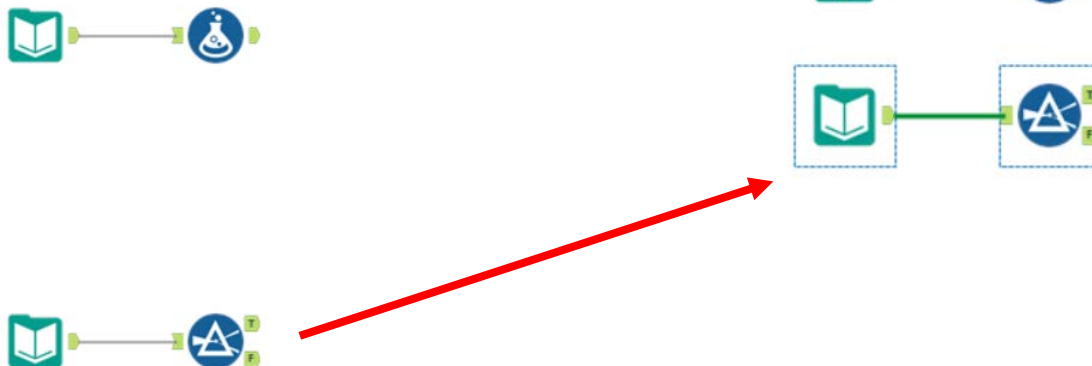
The Canvas is hard to follow because of the descriptions under the steps. Click on any blank space on the Canvas.



In the Configuration area select the down arrow for "Annotations:" and select hide.



Click and drag both second stream steps and drag them closer to the first stream's steps.



Tool 5: Join Tool – Combine Data From Two Streams On A Common Field

The first stream, “Sales Data”.

Results - Formula (4) - Output

6 of 6 Fields | 5,017 records displayed

Record	Sales Order #	Salesperson #	Region	State	Product #	Units Sold
1	35005	1303	West	WA	103	120
2	35006	1302	West	AZ	101	96
3	35009	1305	West	CA	102	156
4	35011	1305	West	CO	101	144

The second stream, “Products”.

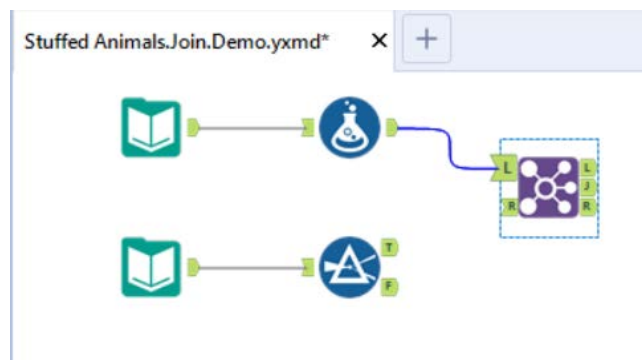
Results - Filter (10) - Out - True

5 of 5 Fields | Cell Viewer | 6 records displayed

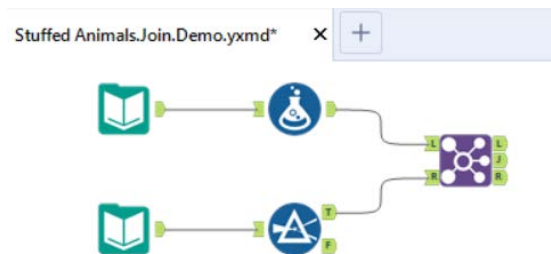
Record	Product ID	Product	Unit Selling Price	Unit Manufacturing Cost	Profit Margin by Product
1	101	Stuffed Lamb	20	8.75	11.25
2	102	Stuffed Giraffe	22	9.75	12.25
3	103	Stuffed Elephant	24	10.75	13.25
4	104	Stuffed Unicorn	21	8.5	12.5
5	105	Stuffed Horse	23	10.75	12.25
6	106	Stuffed Pig	25	12.5	12.5

The field “Product #” in the “Sales Data” stream can be joined, linked, to the “Product ID” field in the “Products” stream.

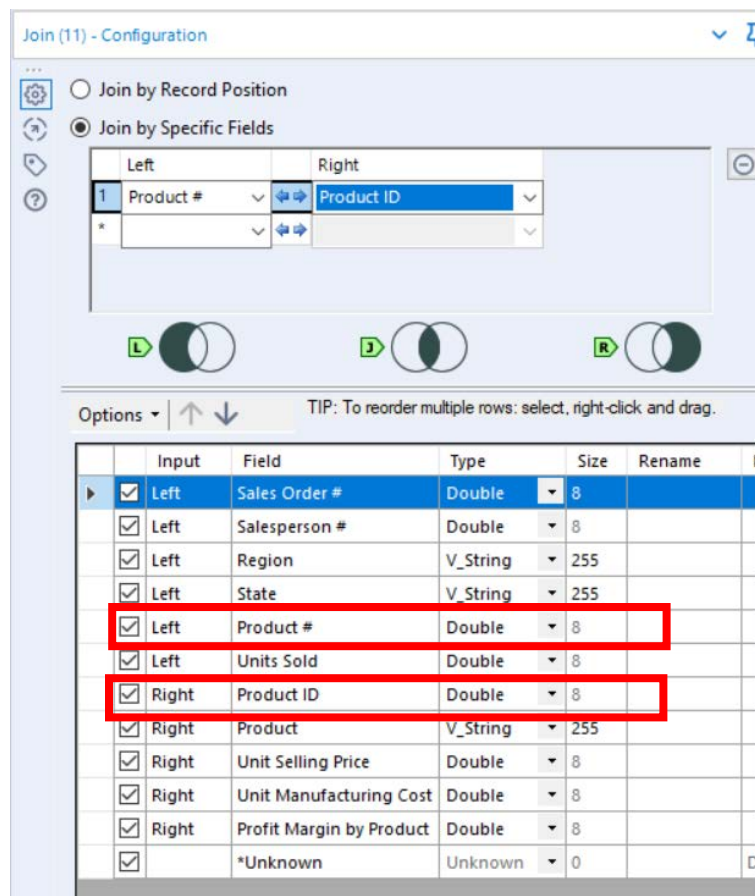
Drag the **Join Tool** from the Tools Palette to the right of the **Formula Tool** from the first data stream connecting to the L,(Left) inflow of the **Join Tool**. The outflow from the tool on the left is the inflow to the one on the right.



Connect the TRUE output of the “**Filter Tool**” from the second data stream to the R,(Right) inflow of the “**Join Tool**”.

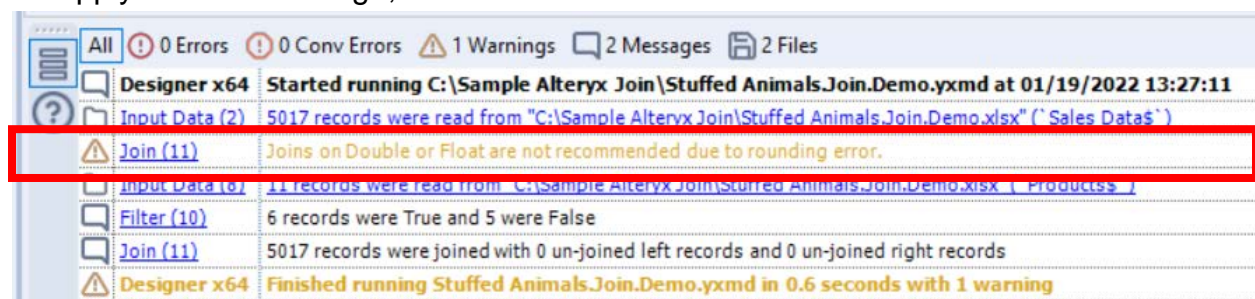


With the “**Join Tool**” selected, in the Configuration area select “Product #” on the left and “Product ID” on the right.



The two fields must be the same type; in this case “Double”, and the same size, in this case 8.

To apply the latest change, select “Run”.



There is a possible error Join should not use fields defined as “Double” or “Float”.

Highlight the **Join Tool** and change the field type of the “Product ID #” and “Product ID” from Double to Integer to Int 16.

	Input	Field	Type	Size	Rename
<input checked="" type="checkbox"/>	Left	Sales Order #	Double	8	
<input checked="" type="checkbox"/>	Left	Salesperson #	Double	8	
<input checked="" type="checkbox"/>	Left	Region	V_String	255	
<input checked="" type="checkbox"/>	Left	State	V_String	255	
<input checked="" type="checkbox"/>	Left	Product #	Double	8	
<input checked="" type="checkbox"/>	Left	Units Sold	Bool		
<input checked="" type="checkbox"/>	Right	Product ID	Int16		
<input checked="" type="checkbox"/>	Right	Product	Int32		
<input checked="" type="checkbox"/>	Right	Unit Selling Price	Int64		
<input checked="" type="checkbox"/>	Right	Unit Manufacturing Cost	FixedDecimal		
<input checked="" type="checkbox"/>	Right	Profit Margin by Product	Float		
<input checked="" type="checkbox"/>	Right	Profit Margin by Product	Double		
<input checked="" type="checkbox"/>	Right	Profit Margin by Product	Double: Forced		
<input checked="" type="checkbox"/>	Right	Profit Margin by Product	String		
<input checked="" type="checkbox"/>	Right	Profit Margin by Product	WString		
<input checked="" type="checkbox"/>	Right	Profit Margin by Product	V_String		
<input checked="" type="checkbox"/>	Right	Profit Margin by Product	V_WString		

<input checked="" type="checkbox"/>	Left	State	V_String	255	
<input checked="" type="checkbox"/>	Left	Product #	Int16	2	
<input checked="" type="checkbox"/>	Left	Units Sold	Double	8	
<input checked="" type="checkbox"/>	Right	Product ID	Int16	2	
<input checked="" type="checkbox"/>	Right	Product	V_String	255	

Select Run to apply.

In the result area, select J, Join, and review the output.

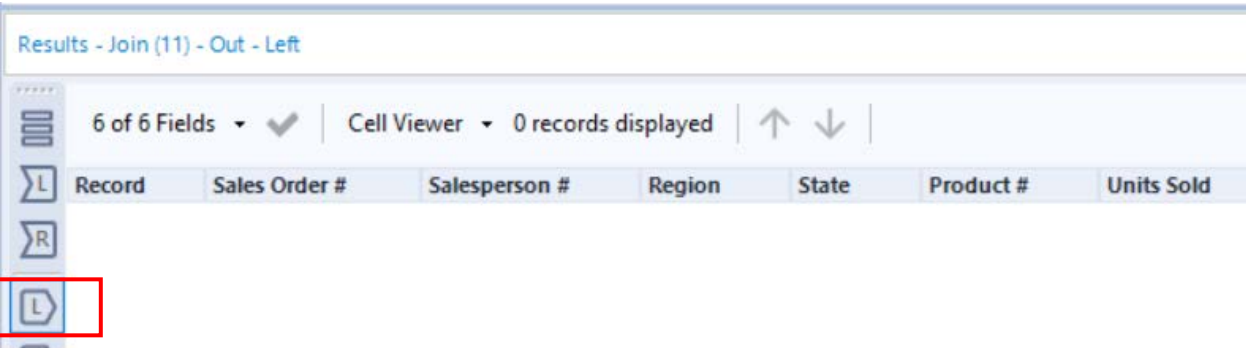
Results - Join (11) - Out - Join

11 of 11 Fields | Cell Viewer | 5,017 records displayed

Record	Sales Order #	Salesperson #	Region	State	Product #	Units Sold	Product ID	Product	Unit Selling Price	Unit Manufacturing Cost	Profit Margi
1	35006	1302	West	AZ	101	96	101	Stuffed Lamb	20	8.75	11.25
2	35011	1305	West	CO	101	144	101	Stuffed Lamb	20	8.75	11.25
3	35058	1303	Midwest	NE	101	180	101	Stuffed Lamb	20	8.75	11.25
4	35165	1305	South	VA	101	108	101	Stuffed Lamb	20	8.75	11.25
5	35167	1303	West	WY	101	108	101	Stuffed Lamb	20	8.75	11.25
6	35204	1305	Midwest	IL	101	96	101	Stuffed Lamb	20	8.75	11.25
7	35278	1303	South	TX	101	60	101	Stuffed Lamb	20	8.75	11.25

There are now 11 fields and all 5,017 rows we joined.

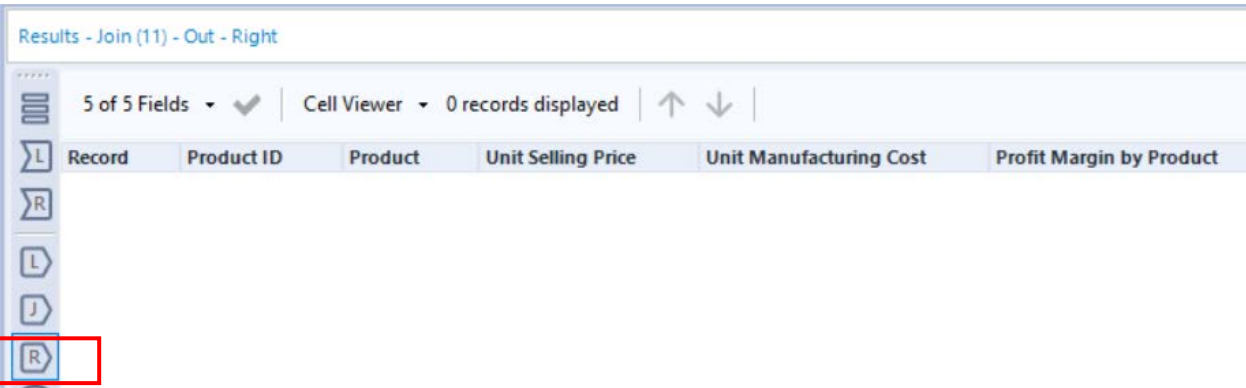
Verify that nothing was unmatched from L, Left data set. Sales with an invalid "Product #" number would be unmatched.



The screenshot shows the 'Results - Join (11) - Out - Left' window in Alteryx. The interface includes a toolbar with icons for fields, cell viewer, and record navigation. The main table has the following columns: Record, Sales Order #, Salesperson #, Region, State, Product #, and Units Sold. The 'Record' column is highlighted with a red box.

Record	Sales Order #	Salesperson #	Region	State	Product #	Units Sold
--------	---------------	---------------	--------	-------	-----------	------------

Look at R, Right data set. If anything is not matched, there would be "Product ID" that was not sold in any of the recorded sales.



The screenshot shows the 'Results - Join (11) - Out - Right' window in Alteryx. The interface includes a toolbar with icons for fields, cell viewer, and record navigation. The main table has the following columns: Record, Product ID, Product, Unit Selling Price, Unit Manufacturing Cost, and Profit Margin by Product. The 'Record' column is highlighted with a red box.

Record	Product ID	Product	Unit Selling Price	Unit Manufacturing Cost	Profit Margin by Product
--------	------------	---------	--------------------	-------------------------	--------------------------

Requirement 3: Extract the data from the "Salesperson" sheet. Review the data, join to combined Stream 1 and Stream 2 and verify the number of rows.

Data regarding the Salesperson's Code, name and Title are stored in an Excel sheet titled "Salesperson". Notice that the column headings are in the fifteenth row.

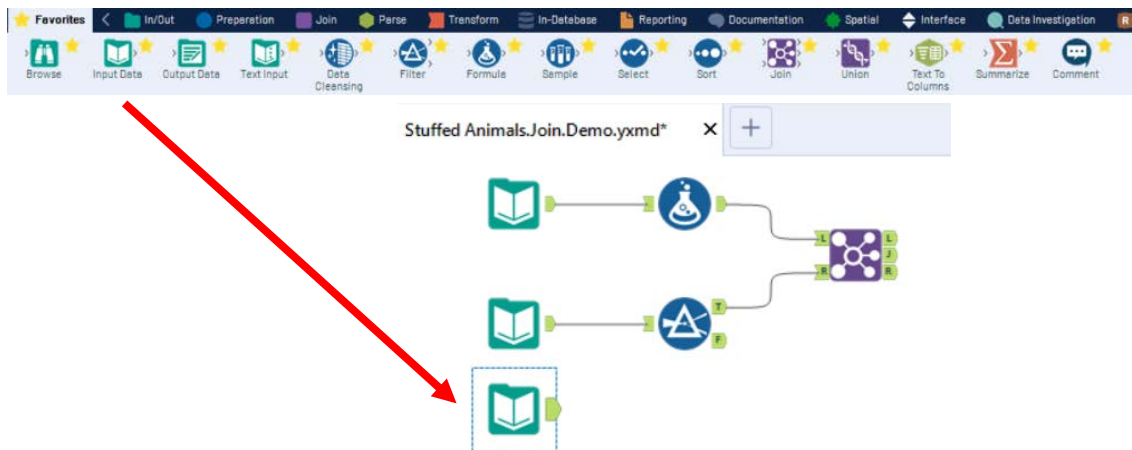
	C	D	E	F	G
10					
11					
12					
13					
14					
15	Code		Salesperson		Title
16	1301		James Polk		Sales Associate I
17	1302		Ulysses Grant		Senior Sales Associate
18	1303		Thomas Jefferson		Senior Sales Associate
19	1304		James Madison		Sales Associate I
20	1305		James Monroe		Sales Associate I
21	1306		Grover Cleveland		Senior Sales Associate
22					
23					

Tool 6: **Input Data Tool** – Input Excel file

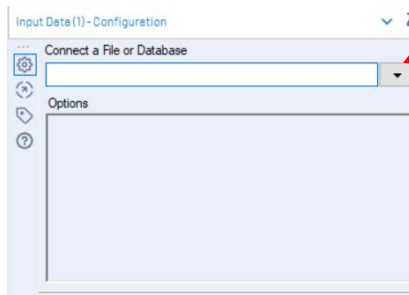
Drag the "Input Data Tool"



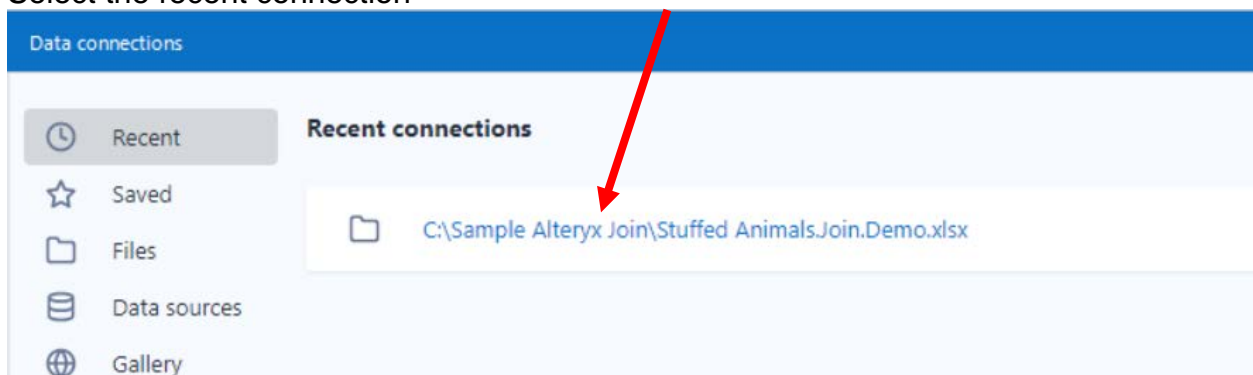
from the Tools Palette to the Canvas.



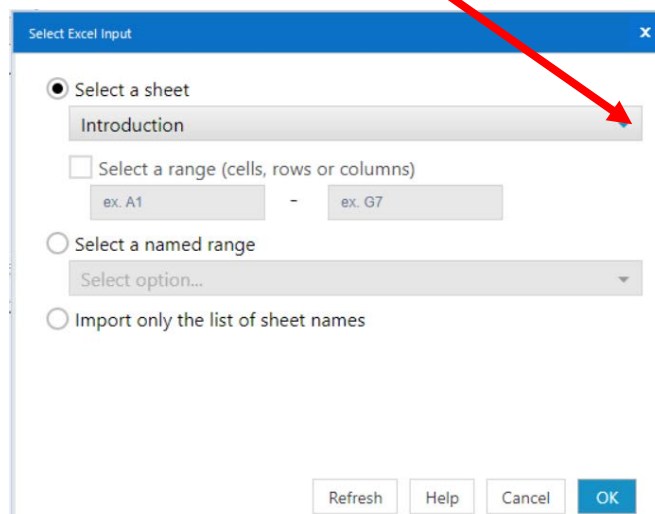
In the Configuration area, select the “Connect a File or Database” drop-down arrow.



Select the recent connection

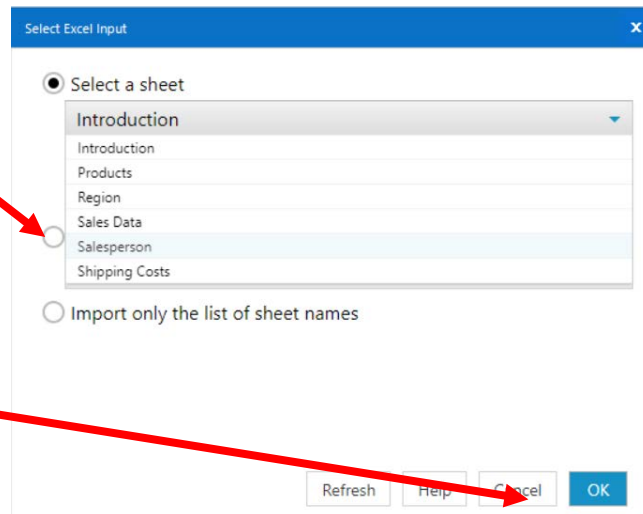


Select the drop-down arrow to the right of “Introduction”.



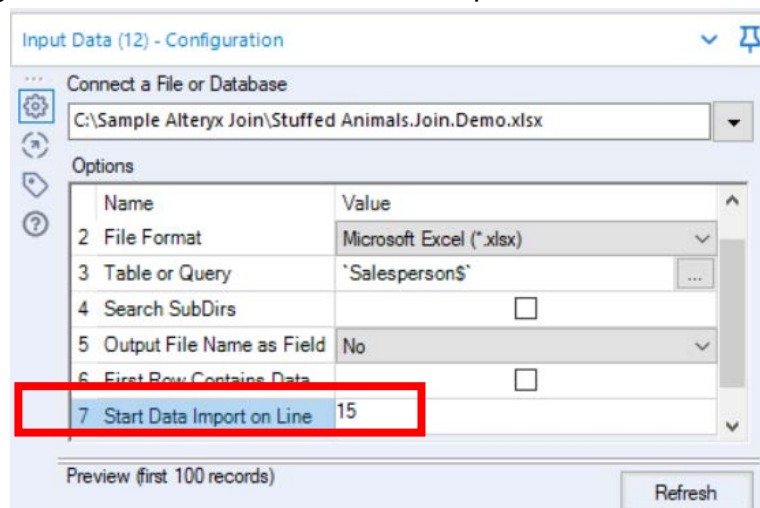
Select "Salesperson".

Select "OK".

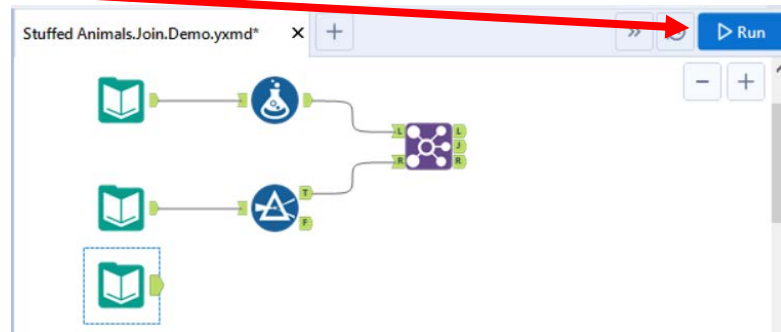


Remember that the field names are in the fifteenth row.

In the Configuration area, change the number in "7 Start Data Import on Line" to "15".



Select "Run".



The Results window will display the field headings and the data. Notice that five fields were extracted and that there are six records. The extra columns that contain [Null] must be removed. The last salesperson's "Code" is 1306.

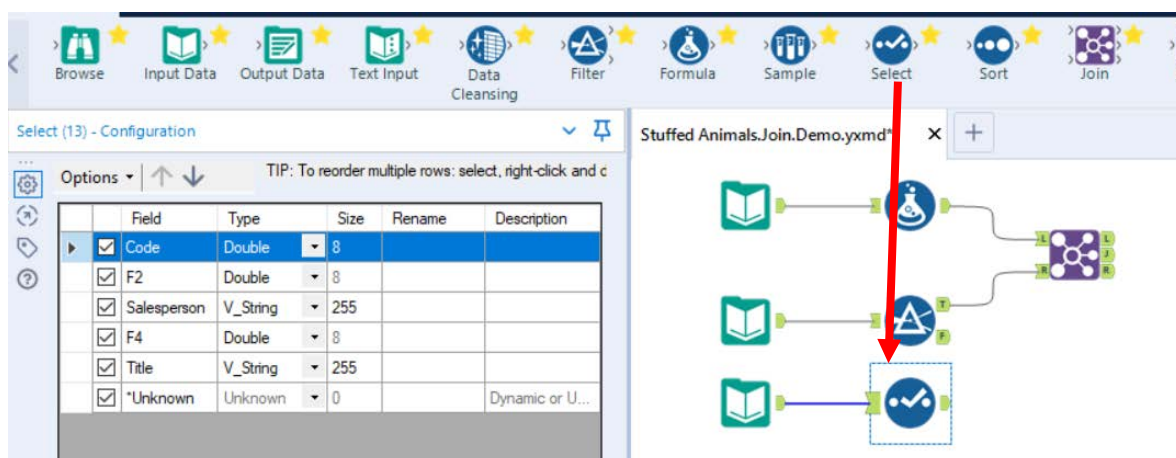
Results - Input Data (12) - Output

5 of 5 Fields 6 records displayed

Record	Code	F2	Salesperson	F4	Title
1	1301	[Null]	James Polk	[Null]	Sales Associate I
2	1302	[Null]	Ulysses Grant	[Null]	Senior Sales Associate
3	1303	[Null]	Thomas Jefferson	[Null]	Senior Sales Associate
4	1304	[Null]	James Madison	[Null]	Sales Associate I
5	1305	[Null]	James Monroe	[Null]	Sales Associate I
6	1306	[Null]	Grover Cleveland	[Null]	Senior Sales Associate

Tool 7: **Select Tool** – Remove fields that contain no data, F5 and Unknown

Drag the "**Select Tool**" from the Tools Palette to the output of the third "**Input Data Tool**" on the Canvas. The two will be connected automatically.



Results - Input Data (12) - Output

5 of 5 Fields | 6 records displayed

Record	Code	F2	Salesperson	F4	Title
1	1301	[Null]	James Polk	[Null]	Sales Associate I
2	1302	[Null]	Ulysses Grant	[Null]	Senior Sales Associate
3	1303	[Null]	Thomas Jefferson	[Null]	Senior Sales Associate
4	1304	[Null]	James Madison	[Null]	Sales Associate I
5	1305	[Null]	James Monroe	[Null]	Sales Associate I
6	1306	[Null]	Grover Cleveland	[Null]	Senior Sales Associate

Remove the extra fields, F2, F4 and Unknown, by unchecking them in the Configuration area.

Options | TIP: To reorder multiple rows: select, right-click and c

	Field	Type	Size	Rename	Description
<input checked="" type="checkbox"/>	Code	Double	8		
<input type="checkbox"/>	F2	Double	8		
<input checked="" type="checkbox"/>	Salesperson	V_String	255		
<input type="checkbox"/>	F4	Double	8		
<input checked="" type="checkbox"/>	Title	V_String	255		
<input type="checkbox"/>	*Unknown	Unknown	0		Dynamic or U...

Select "Run".

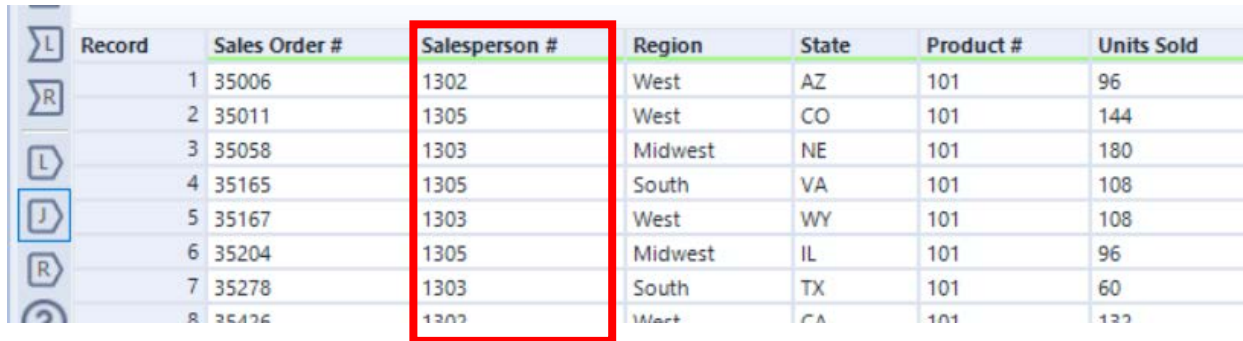
Results - Select (13) - Output

3 of 3 Fields | 6 records displayed

Record	Code	Salesperson	Title
1	1301	James Polk	Sales Associate I
2	1302	Ulysses Grant	Senior Sales Associate
3	1303	Thomas Jefferson	Senior Sales Associate
4	1304	James Madison	Sales Associate I
5	1305	James Monroe	Sales Associate I
6	1306	Grover Cleveland	Senior Sales Associate

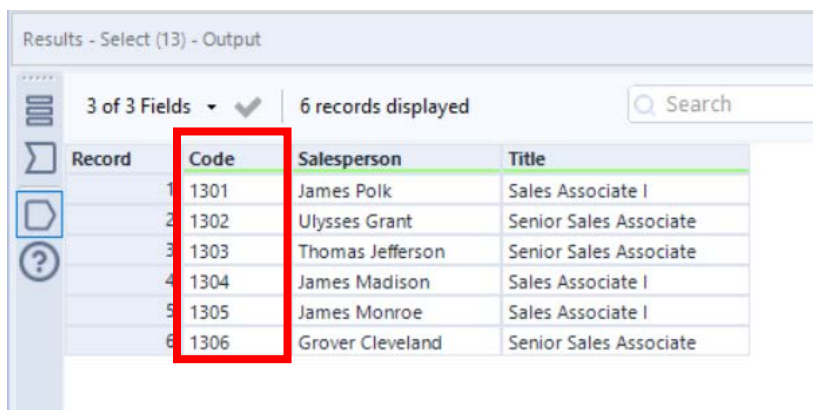
Tool 8: Join Tool – Combine Data From Two Streams On A Common Field

The first stream and second stream, “Sales Data” + “Products”.



Record	Sales Order #	Salesperson #	Region	State	Product #	Units Sold
1	35006	1302	West	AZ	101	96
2	35011	1305	West	CO	101	144
3	35058	1303	Midwest	NE	101	180
4	35165	1305	South	VA	101	108
5	35167	1303	West	WY	101	108
6	35204	1305	Midwest	IL	101	96
7	35278	1303	South	TX	101	60
8	35426	1302	West	CA	101	122

The Third stream, “Salesperson”.

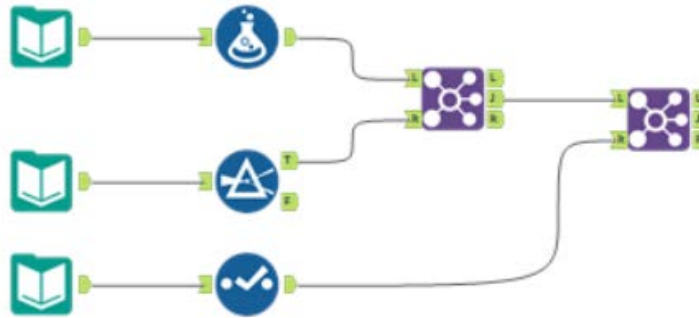


Record	Code	Salesperson	Title
1	1301	James Polk	Sales Associate I
2	1302	Ulysses Grant	Senior Sales Associate
3	1303	Thomas Jefferson	Senior Sales Associate
4	1304	James Madison	Sales Associate I
5	1305	James Monroe	Sales Associate I
6	1306	Grover Cleveland	Senior Sales Associate

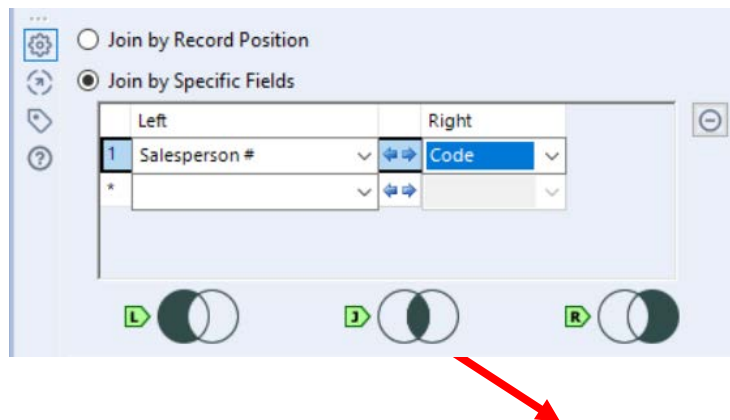
The field “Salesperson #” in the joined first and second stream can be joined, linked, to the “Code” field in the “Salesperson” stream.

Drag the **Join Tool** from the Tools Palette to the right of the first **Join Tool** associated with the first and second data streams. The two will be connected automatically to the L,(Left),inflow of the Join Tool. The outflow from the tool on the left is the inflow to the one on the right.

Connect the output of the **Select Tool** from the third data stream to the R,(Right) inflow of the **Join Tool**.



With the **Join Tool** selected, in the Configuration area select “Salesperson #” on the left and ”Code” on the right.



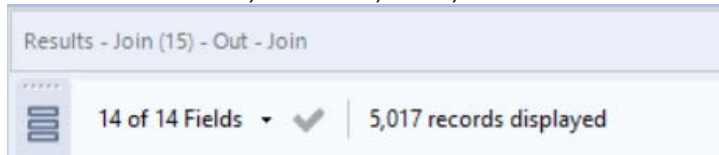
The two fields must be the same type and the same size. Remember that Join should not use fields defined as “Double” or “Float”. Change the field type of the “Salesperson #” and “Code” from Double to Integer to Int 16.

	Input	Field	Type	Size	R
<input checked="" type="checkbox"/>	Left	Sales Order #	Double	8	
<input checked="" type="checkbox"/>	Left	Salesperson #	Double	8	
<input checked="" type="checkbox"/>	Left	Region	Bool		
<input checked="" type="checkbox"/>	Left	State	Byte		
<input checked="" type="checkbox"/>	Left	Product #	Int16		
<input checked="" type="checkbox"/>	Left	Units Sold	Int32		
<input checked="" type="checkbox"/>	Left	Product ID	Int64		
<input checked="" type="checkbox"/>	Left	Product	FixedDecimal		
<input checked="" type="checkbox"/>	Left	Unit Selling Price	Float		
<input checked="" type="checkbox"/>	Left	Unit Manufacturing Cost	Double		
<input checked="" type="checkbox"/>	Left	Profit Margin by Product	Double	8	
<input checked="" type="checkbox"/>	Right	Code	Double	8	
<input checked="" type="checkbox"/>	Right	Salesperson	Double: Forced		
<input checked="" type="checkbox"/>	Right	Title	String		
<input checked="" type="checkbox"/>	Right		V_String	255	
<input checked="" type="checkbox"/>	Right		V_String	255	

	Input	Field	Type	Size	F
<input checked="" type="checkbox"/>	Left	Sales Order #	Double	8	
<input checked="" type="checkbox"/>	Left	Salesperson #	Int16	2	
<input checked="" type="checkbox"/>	Left	Region	V_String	255	
<input checked="" type="checkbox"/>	Left	State	V_String	255	
<input checked="" type="checkbox"/>	Left	Product #	Int16	2	
<input checked="" type="checkbox"/>	Left	Units Sold	Double	8	
<input checked="" type="checkbox"/>	Left	Product ID	Int16	2	
<input checked="" type="checkbox"/>	Left	Product	V_String	255	
<input checked="" type="checkbox"/>	Left	Unit Selling Price	Double	8	
<input checked="" type="checkbox"/>	Left	Unit Manufacturing Cost	Double	8	
<input checked="" type="checkbox"/>	Left	Profit Margin by Product	Double	8	
<input checked="" type="checkbox"/>	Right	Code	Int16	2	

Select Run to apply.

In the result area, select J, Join, and review the output.



There are now 14 fields and all 5,017 rows we joined.

Results - Join (15) - Out - Join

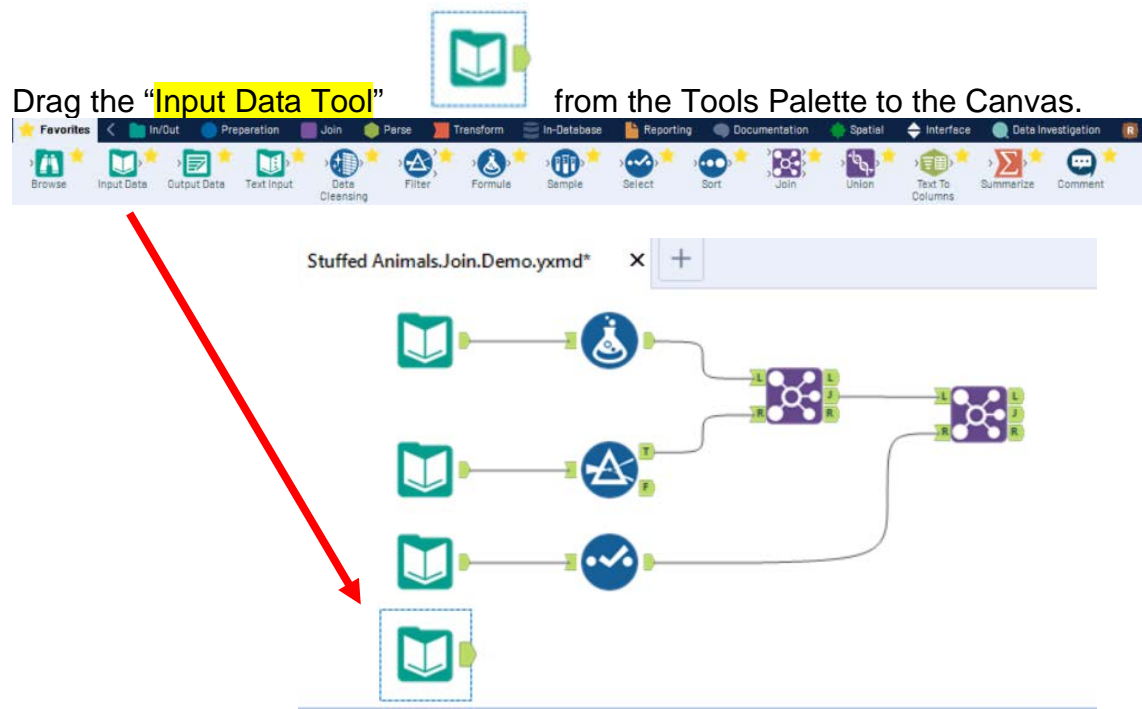
14 of 14 Fields ✓ Cell Viewer 5,017 records displayed

Record	Product ID	Product	Unit Selling Price	Unit Manufacturing Cost	Profit Margin by Product	Code	Salesperson	Title
1	101	Stuffed Lamb	20	8.75	11.25	1301	James Polk	Sales Associate I
2	101	Stuffed Lamb	20	8.75	11.25	1301	James Polk	Sales Associate I
3	101	Stuffed Lamb	20	8.75	11.25	1301	James Polk	Sales Associate I
4	101	Stuffed Lamb	20	8.75	11.25	1301	James Polk	Sales Associate I
5	101	Stuffed Lamb	20	8.75	11.25	1301	James Polk	Sales Associate I
6	101	Stuffed Lamb	20	8.75	11.25	1301	James Polk	Sales Associate I
7	101	Stuffed Lamb	20	8.75	11.25	1301	James Polk	Sales Associate I
8	101	Stuffed Lamb	20	8.75	11.25	1301	James Polk	Sales Associate I
9	101	Stuffed Lamb	20	8.75	11.25	1301	James Polk	Sales Associate I
10	101	Stuffed Lamb	20	8.75	11.25	1301	James Polk	Sales Associate I
11	101	Stuffed Lamb	20	8.75	11.25	1301	James Polk	Sales Associate I

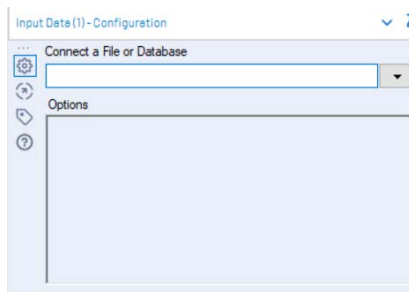
Requirement 4: Extract the data from the "Shipping Costs" sheet. Review the data, create required fields and join to the previously combined streams. Verify that the number of rows has not changed.

	C	D	E	F	G	H	I	J
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								

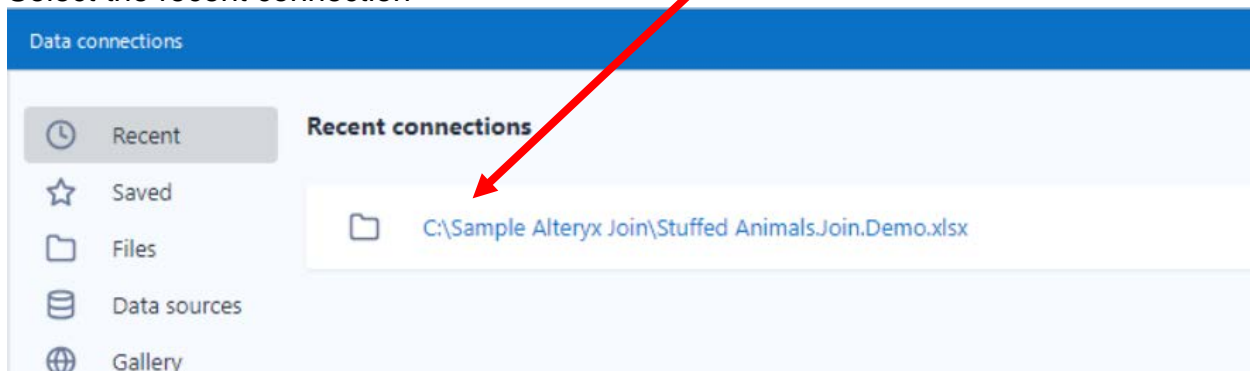
The shipping cost per unit is a function of the region in which the sales was made and the product that was shipped. This information is stored in an Excel sheet titled "Shipping Costs". Notice that the column headings are in the thirteenth row.

Tool 9: Input Data Tool – Input Excel File

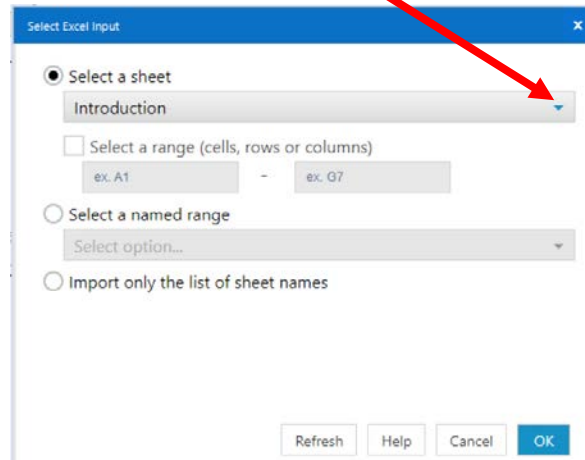
In the Configuration area, select the “Connect a File or Database” drop-down arrow.



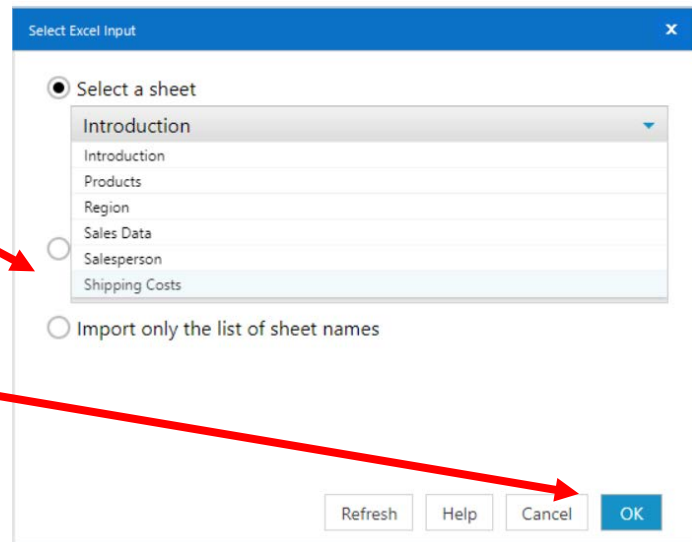
Select the recent connection



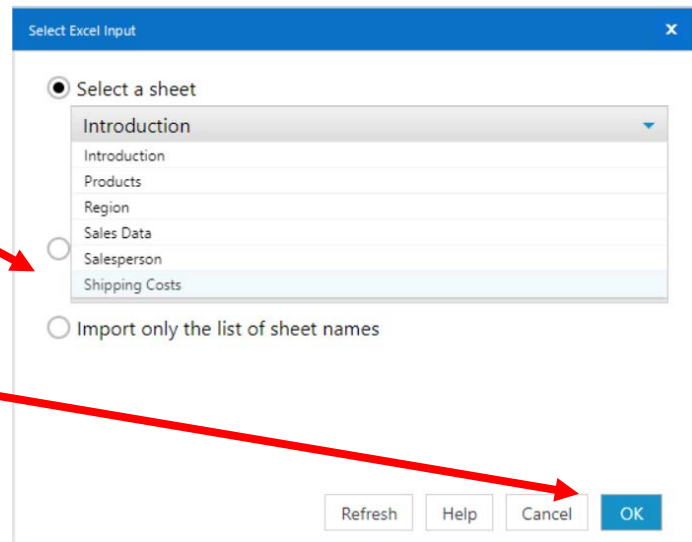
Select the drop-down arrow to the right of “Introduction”.



Select “Shipping Costs”.

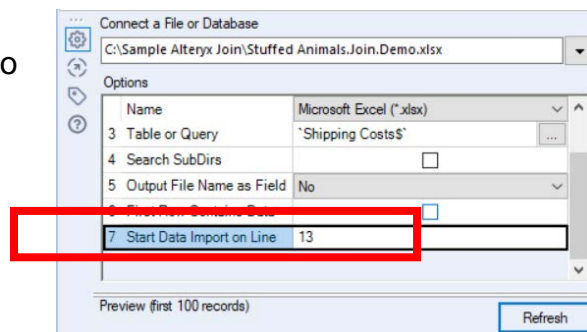


Select “OK”.



Remember that the field names are in the thirteenth row.

In the Configuration area, change the number in “7 Start Data Import on Line” to “13”.



Select "Run".

Results - Input Data (16) - Output

10 of 10 Fields ✓ Cell Viewer 11 records displayed

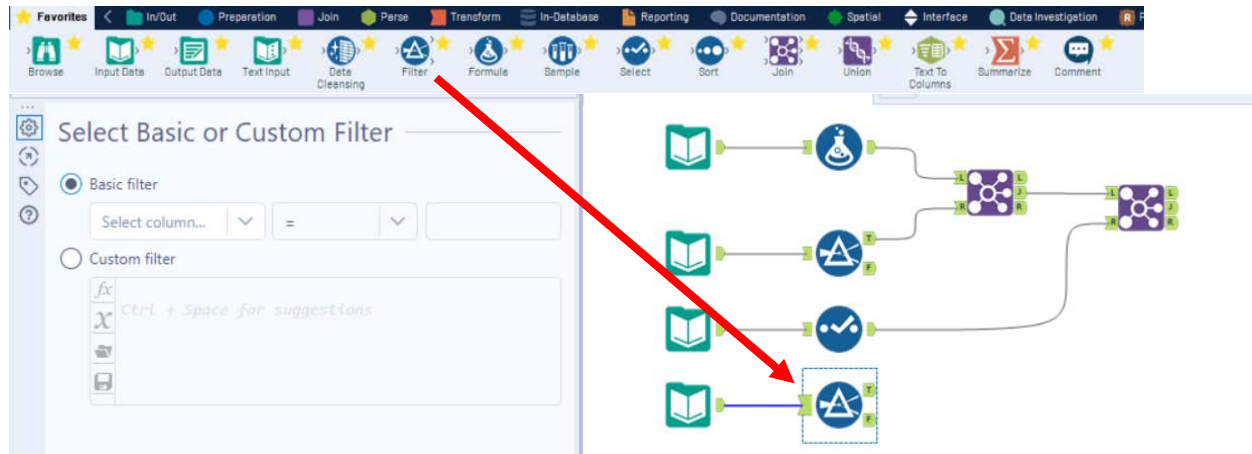
Record	Product ID	Midwest	Northeast	South	West	F6	F7	F8	F9	F10
1	101	2.15	2.12	2.08	2.02	[Null]	[Null]	[Null]	[Null]	[Null]
2	102	2.11	2.08	2.04	1.98	[Null]	[Null]	[Null]	[Null]	[Null]
3	103	2.14	2.11	2.0699999999999998	2.01	[Null]	[Null]	[Null]	[Null]	[Null]
4	104	2.12	2.09	2.0499999999999998	1.99	[Null]	[Null]	[Null]	[Null]	[Null]
5	105	2.07	2.04	2	1.94	[Null]	[Null]	[Null]	[Null]	[Null]
6	106	2.02	1.99	1.95	1.89	[Null]	[Null]	[Null]	[Null]	[Null]
7	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]
8	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]
9	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]
10	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]
11	[Null]	[Null]	[Null]	*	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]

The Results window will display the field headings and the data. Notice that ten fields were extracted and that there are eleven records. The extra rows and columns that contain [Null] must be removed. The last "Product ID" 106.

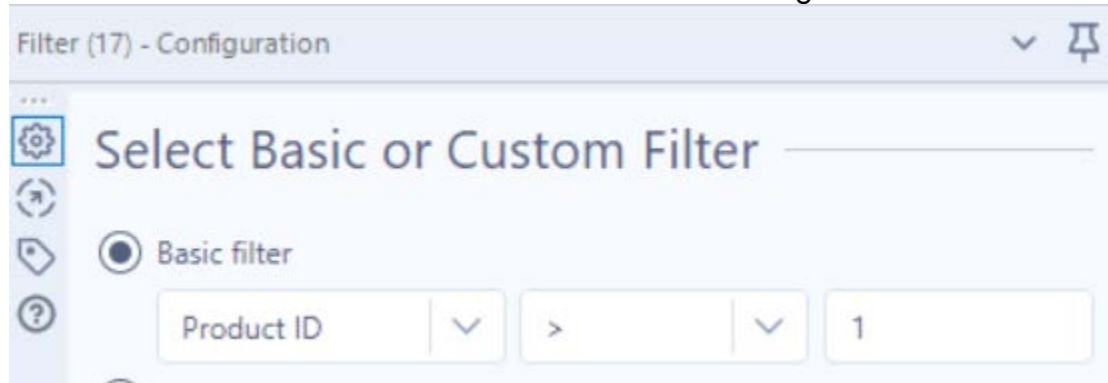
There is something strange about some of the numbers. Numbers that are to be used in arithmetic formulas must be stored as fixed decimals.

Tool 10: Filter Tool – Exclude All Rows That Do Not Contain Data

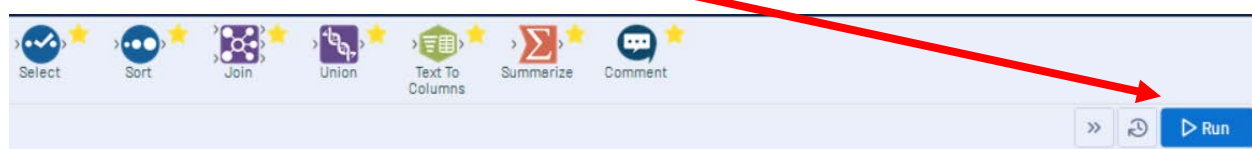
Drag the **Filter Tool** from the Tools Palette to the right of the **Input Data Tool** on the Canvas. The two will be connected automatically. The outflow from the tool on the left is the inflow to the one on the right.



With the **Filter Tool** selected, the “Basic” filter in the Configuration area can be modified to filter the field “Product ID” for those records greater than “1”.



To apply the latest change, select “Run”.



Results - Filter (17) - Out - True

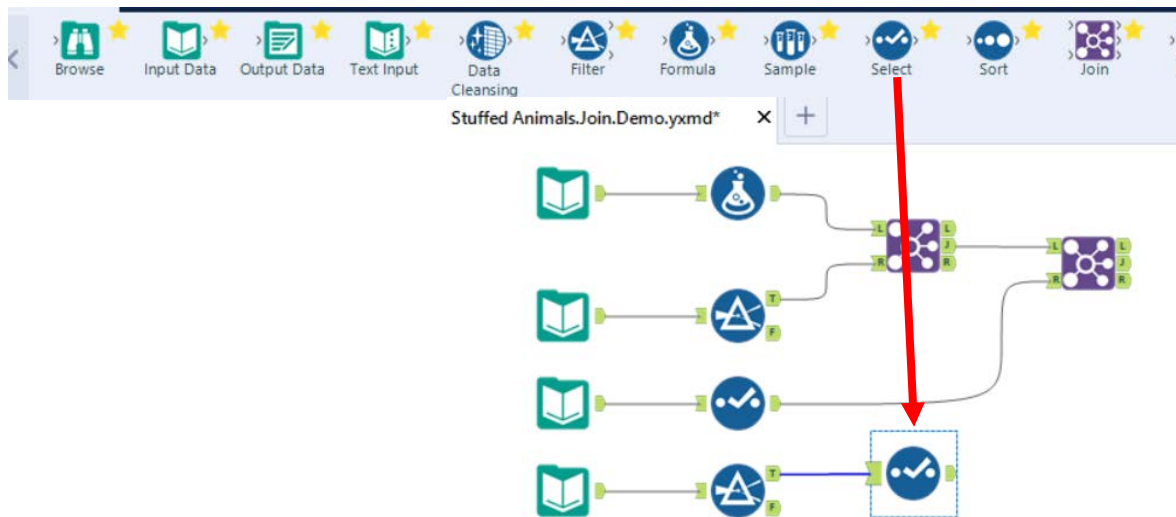
10 of 10 Fields ✓ Cell Viewer 6 records displayed ↑ ↓ Search Data Metadata Actions α

Record	Product ID	Midwest	Northeast	South	West	F6	F7	F8	F9	F10
1	101	2.15	2.12	2.08	2.02	[Null]	[Null]	[Null]	[Null]	[Null]
2	102	2.11	2.08	2.04	1.98	[Null]	[Null]	[Null]	[Null]	[Null]
3	103	2.14	2.11	2.0699999999999998	2.01	[Null]	[Null]	[Null]	[Null]	[Null]
4	104	2.12	2.09	2.0499999999999998	1.99	[Null]	[Null]	[Null]	[Null]	[Null]
5	105	2.07	2.04	2	1.94	[Null]	[Null]	[Null]	[Null]	[Null]
6	106	2.02	1.99	1.95	1.89	[Null]	[Null]	[Null]	[Null]	[Null]

The records that meet the filtered conditions, “Product ID> 1”, are displayed as T (True). There are six records displayed and the last one is 106 for “Product ID”.

Tool 11: Select Tool – Remove fields that contain no data, and change field types

Drag the “**Select Tool**”  from the Tools Palette to the True output of the “**Filter Tool**” on the Canvas. The two will be connected automatically.



Remove the extra fields, F6:F10 F4 and Unknown, by unchecking them in the Configuration area.

Select (18) - Configuration

Options | TIP: To reorder multiple rows: select, right-click and

	Field	Type	Size	Rename	Description
<input checked="" type="checkbox"/>	Product ID	Double	8		
<input checked="" type="checkbox"/>	Midwest	Double	8		
<input checked="" type="checkbox"/>	Northeast	Double	8		
<input checked="" type="checkbox"/>	South	V_String	255		
<input checked="" type="checkbox"/>	West	Double	8		
<input type="checkbox"/>	F6	Double	8		
<input type="checkbox"/>	F7	Double	8		
<input type="checkbox"/>	F8	Double	8		
<input type="checkbox"/>	F9	Double	8		
<input type="checkbox"/>	F10	Double	8		
<input type="checkbox"/>	*Unknown	Unknown	0		Dynamic or U...

Change "Product ID" to the Type Int16.
Change all the regions to FixedDecimal, 19.2.

Options ▾ | ↑ ↓ | TIP: To reorder multiple rows: select

	Field	Type	Size	Rename
<input checked="" type="checkbox"/>	Product ID	Int16	2	
<input checked="" type="checkbox"/>	Midwest	FixedDecimal	19.2	
<input checked="" type="checkbox"/>	Northeast	FixedDecimal	19.2	
<input checked="" type="checkbox"/>	South	FixedDecimal	19.2	
<input checked="" type="checkbox"/>	West	FixedDecimal	19.2	

Select Run.

Results - Select (18) - Output

5 of 5 Fields ▾ ✓ | Cell Viewer ▾ 6 records displayed | ↑ ↓ | Search

Record	Product ID	Midwest	Northeast	South	West
1	101	2.15	2.12	2.08	2.02
2	102	2.11	2.08	2.04	1.98
3	103	2.14	2.11	2.07	2.01
4	104	2.12	2.09	2.05	1.99
5	105	2.07	2.04	2.00	1.94
6	106	2.02	1.99	1.95	1.89

The “Shipping Costs”, stream must be transformed before combining it with the already combined streams of data. The “Shipping Costs” stream of data must be arranged into a data set with each shipping cost on a separate row.

The fourth stream before transformation:

Results - Select (18) - Output

5 of 5 Fields | Cell Viewer | 6 records displayed

Record	Product ID	Midwest	Northeast	South	West
1	101	2.15	2.12	2.08	2.02
2	102	2.11	2.08	2.04	1.98
3	103	2.14	2.11	2.07	2.01
4	104	2.12	2.09	2.05	1.99
5	105	2.07	2.04	2.00	1.94
6	106	2.02	1.99	1.95	1.89

The fourth stream after transformation into a data set with each shipping cost on a separate row:

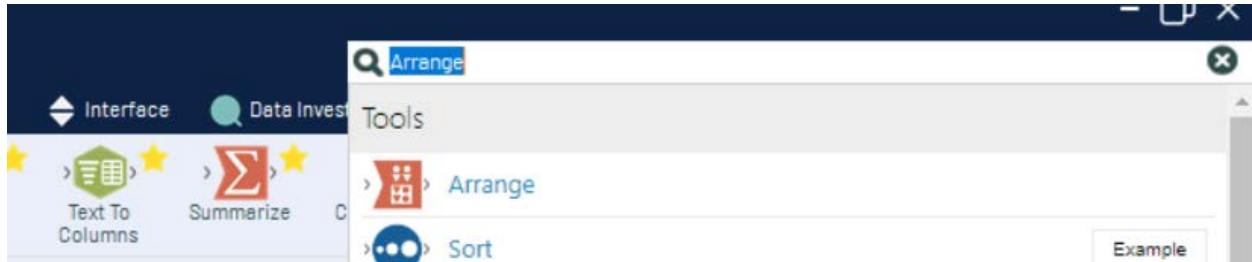
Results - Arrange (22) - Output

3 of 3 Fields | Cell Viewer | 24 records displayed

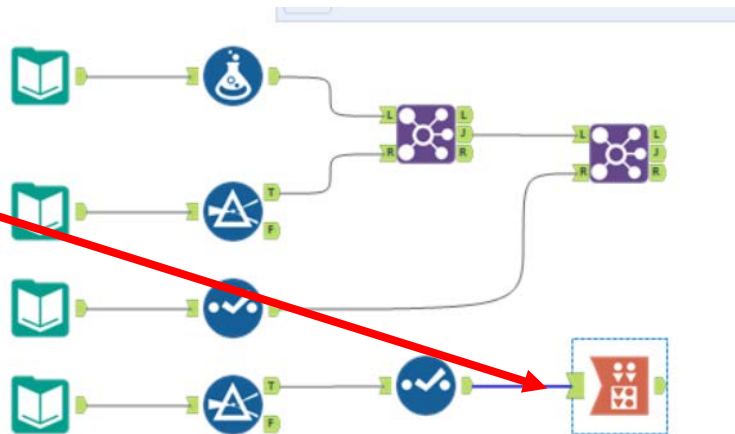
Record	Product ID	Region	Shipping Costs
1	101	Midwest	2.15
2	101	Northeast	2.12
3	101	South	2.08
4	101	West	2.02
5	102	Midwest	2.11
6	102	Northeast	2.08
7	102	South	2.04
8	102	West	1.98
9	103	Midwest	2.14
10	103	Northeast	2.11
11	103	South	2.07
12	103	West	2.01
13	104	Midwest	2.12
14	104	Northeast	2.09
15	104	South	2.05
16	104	West	1.99
17	105	Midwest	2.07
18	105	Northeast	2.04
19	105	South	2.00
20	105	West	1.94
21	106	Midwest	2.02
22	106	Northeast	1.99
23	106	South	1.95
24	106	West	1.89

Tool 12: Arrange Tool – Each Record Is Turned Into Multiple Records

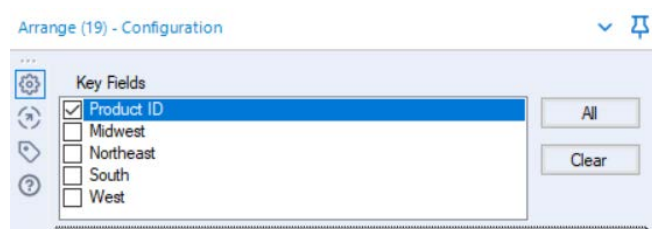
In the “Search for Tools Box” type “Arrange”.



Drag the “**Arrange Tool**” from the Tools Palette to the right of the “**Select Tool**” from the fourth data streams, on the Canvas. The two will be connected automatically. The outflow from the tool on the left is the inflow to the one on the right.

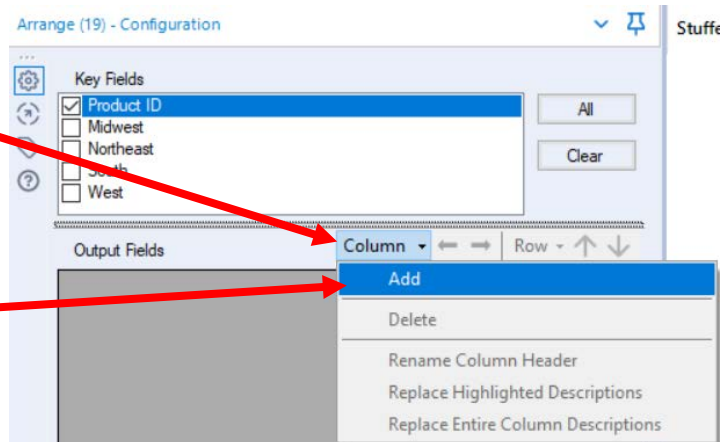


With the “**Arrange Tool**” selected, in the Configuration Area, under “Key Fields”, check Product ID.



Select column.

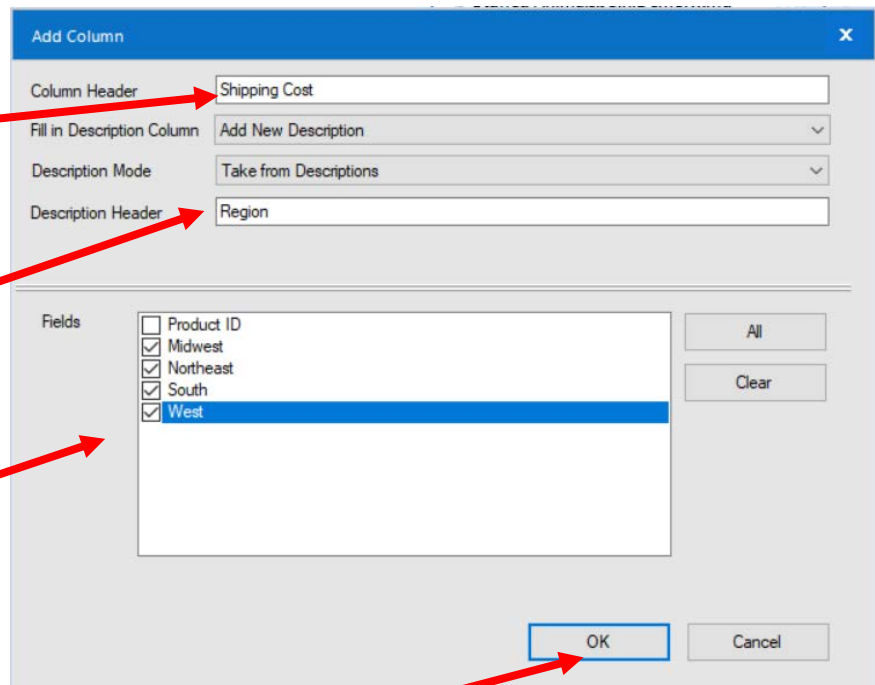
Select Add.



In the "Column Header" type: Shipping Cost

In the "Description Header" type: Region.

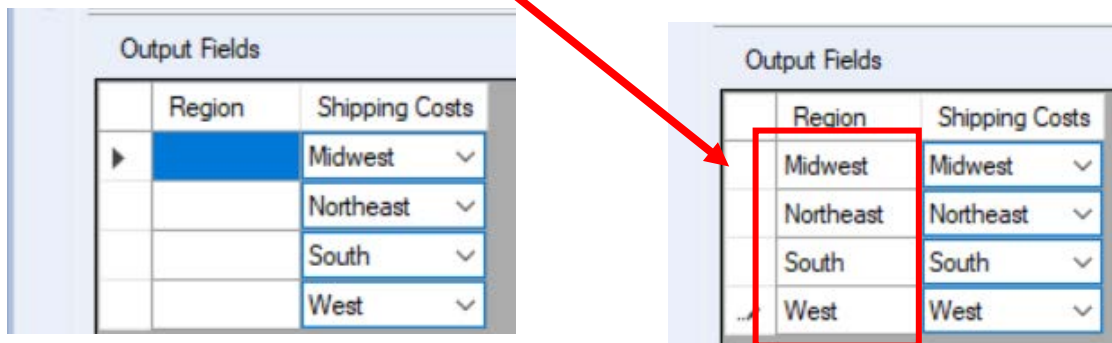
In the "Fields" check: Midwest, Northeast, South and West.



Select "OK".

Select Run to apply.

In the “Output Fields Description” type, Midwest, Northeast, South and West.



Select Run to apply.

Results - Arrange (22) - Output

3 of 3 Fields | Cell Viewer | 24 records displayed

Record	Product ID	Region	Shipping Costs
1	101	Midwest	2.15
2	101	Northeast	2.12
3	101	South	2.08
4	101	West	2.02
5	102	Midwest	2.11
6	102	Northeast	2.08
7	102	South	2.04
8	102	West	1.98
9	103	Midwest	2.14
10	103	Northeast	2.11
11	103	South	2.07
12	103	West	2.01
13	104	Midwest	2.12
14	104	Northeast	2.09
15	104	South	2.05
16	104	West	1.99
17	105	Midwest	2.07
18	105	Northeast	2.04
19	105	South	2.00
20	105	West	1.94
21	106	Midwest	2.02
22	106	Northeast	1.99
23	106	South	1.95
24	106	West	1.89

Compare to original data.


Results - Select (18) - Output

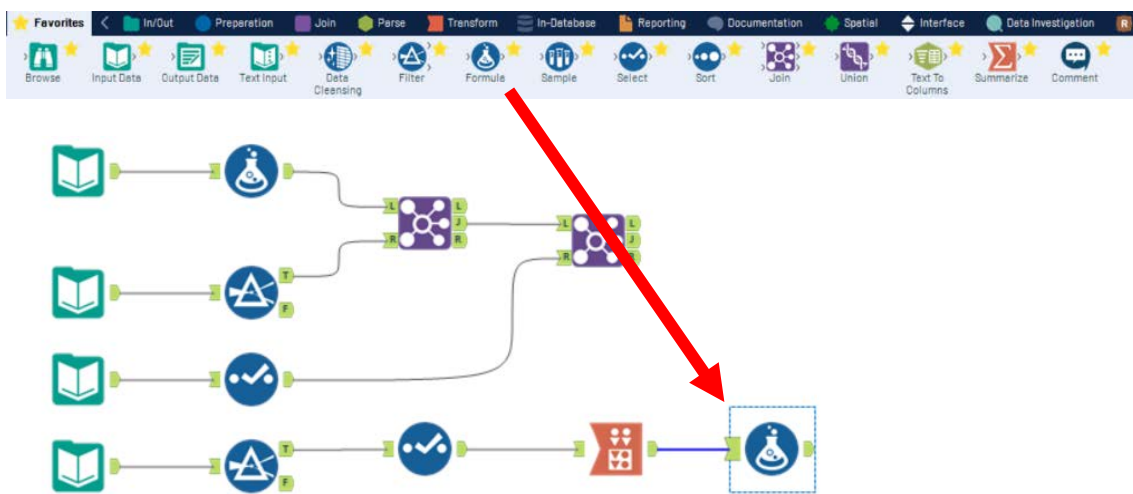
5 of 5 Fields | Cell Viewer | 6 records displayed

Record	Product ID	Midwest	Northeast	South	West
1	101	2.15	2.12	2.08	2.02
2	102	2.11	2.08	2.04	1.98
3	103	2.14	2.11	2.07	2.01
4	104	2.12	2.09	2.05	1.99
5	105	2.07	2.04	2.00	1.94
6	106	2.02	1.99	1.95	1.89

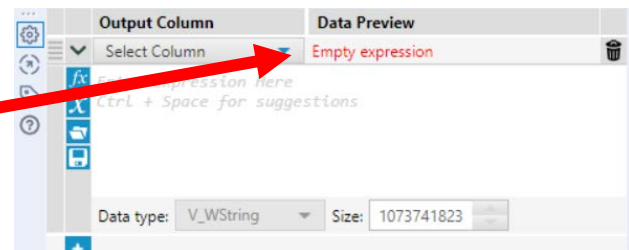
To combine two streams in a one-to-many relationship there must be one field that is unique for each row of the data. That unique field is referred to as the primary key. The primary field will be created by combining the “Product ID” field with a “.” and the “Region” field. The primary key for the first record will be “1301.East”.

Tool 13: **Formula Tool** - Create A Primary Key Field

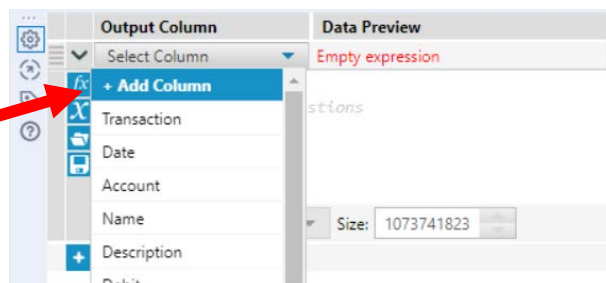
Drag a “**Formula Tool**”  from the Tools Palette to the output of the “**Arrange Tool**” on the fourth stream. The two will be connected automatically. The outflow from the tool on the left is the inflow to the one on the right.



In the output Configuration area, click on the drop-down arrow of “Select Column”.



Select “Add Column”.



Type "Product ID and Region".

Type the following formula
ToString([Product ID])+". "+[Region]

Formula (25) - Configuration

Output Column	Data Preview
Product ID and Region	101.Midwest

ToString([Product ID])+". "+[Region]

Data type: V_WString Size: 1073741823

Select "Run".

Results - Formula (25) - Output

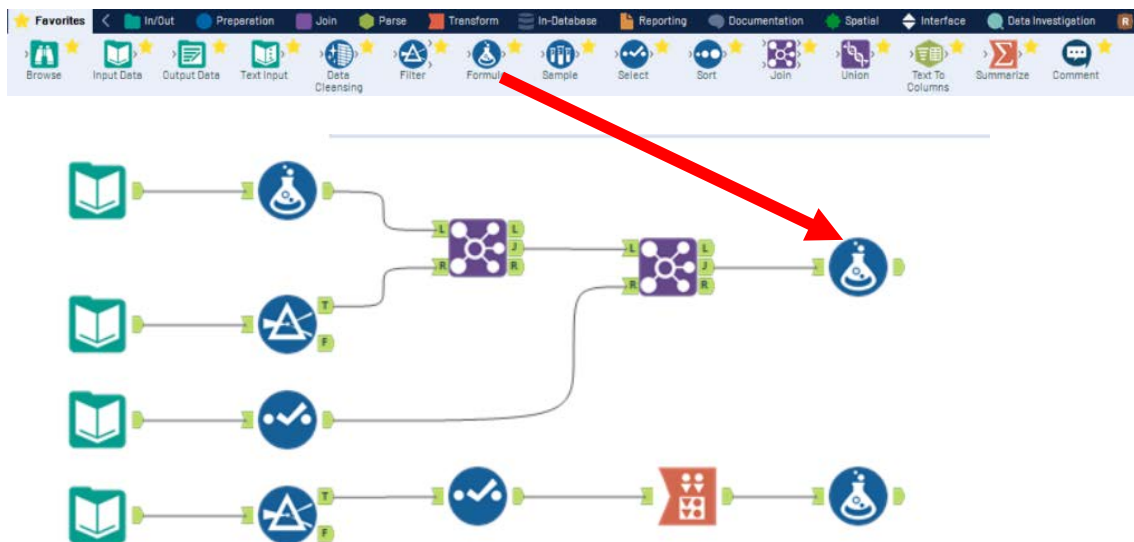
4 of 4 Fields | Cell Viewer | 24 records displayed

Record	Product ID	Region	Shipping Costs	Product ID and Region
1	101	Midwest	2.15	101.Midwest
2	101	Northeast	2.12	101.Northeast
3	101	South	2.08	101.South
4	101	West	2.02	101.West
5	102	Midwest	2.11	102.Midwest
6	102	Northeast	2.08	102.Northeast
7	102	South	2.04	102.South
8	102	West	1.98	102.West
9	103	Midwest	2.14	103.Midwest
10	103	Northeast	2.11	103.Northeast
11	103	South	2.07	103.South
12	103	West	2.01	103.West
13	104	Midwest	2.12	104.Midwest
14	104	Northeast	2.09	104.Northeast
15	104	South	2.05	104.South
16	104	West	1.99	104.West
17	105	Midwest	2.07	105.Midwest
18	105	Northeast	2.04	105.Northeast
19	105	South	2.00	105.South
20	105	West	1.94	105.West
21	106	Midwest	2.02	106.Midwest
22	106	Northeast	1.99	106.Northeast
23	106	South	1.95	106.South
24	106	West	1.89	106.West

Now add the “SPID and Region” field to the combined first, second and third stream of data. The field is referred to as a foreign key. It will permit us to link to the data in the third stream.

Tool 14: **Formula Tool** - Create A Foreign Key Field

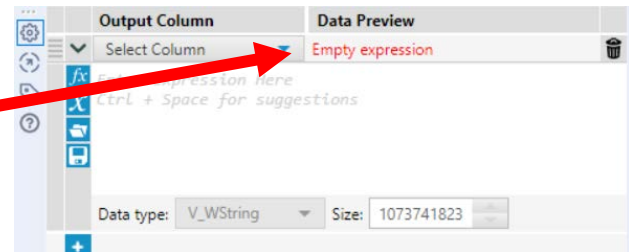
Drag a “**Formula Tool**” from the Tools Palette to the J, Join, output of the second “**Join Tool**”. The two will be connected automatically. The outflow from the tool on the left is the inflow to the one on the right.



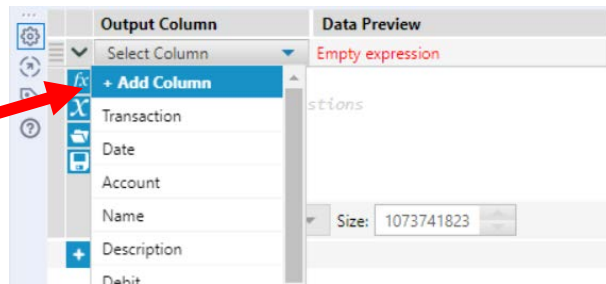
This is the layout of the combined stream 1, stream 2 and stream 3.

14 of 14 Fields Cell Viewer 5,017 records displayed						
Record	Sales Order #	Salesperson #	Region	State	Product #	Units Sold
1	36658	1301	Midwest	KS	101	48
2	36710	1301	Midwest	IN	101	72
3	37329	1301	Northeast	RI	101	144
4	38318	1301	West	NV	101	96
5	39854	1301	Midwest	MI	101	132
6	40243	1301	Midwest	NE	101	120

In the output Configuration area, click on the drop-down arrow of “Select Column”.

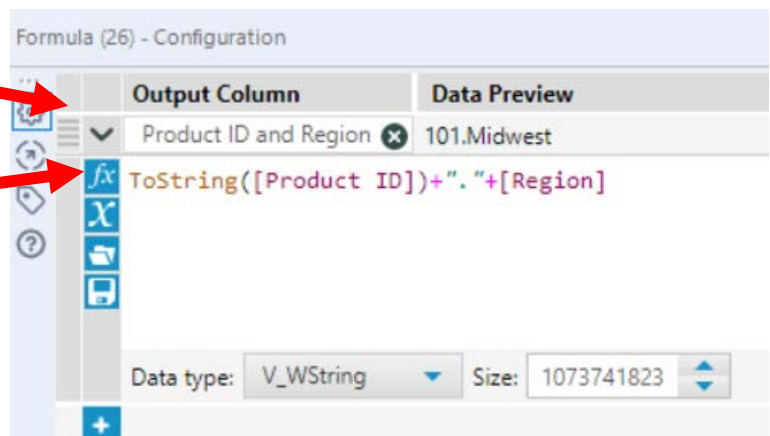


Select “Add Column”.



Type “Product ID and Region”.

Type the following formula
ToString([Product ID])+". "+[Region]



Select “Run” to apply.

Data Metadata Actions 000	
Title	Product ID and Region
Sales Associate I	101.Midwest
Sales Associate I	101.Midwest
Sales Associate I	101.Northeast
Sales Associate I	101.West
Sales Associate I	101.Midwest
Sales Associate I	101.Midwest
Sales Associate I	101.South
Sales Associate I	101.Midwest
Sales Associate I	101.West
Sales Associate I	101.Midwest

Tool 15: Join Tool - Combine First, Second and Third Stream Of Data To The Fourth Stream Of Data

The combined first, second and third stream of data:

Results - Formula (26) - Output

15 of 15 Fields | 5,017 records displayed

Record	Profit Margin by Product	Code	Salesperson	Title	Product ID and Region
1	11.25	1301	James Polk	Sales Associate I	101.Midwest
2	11.25	1301	James Polk	Sales Associate I	101.Midwest
3	11.25	1301	James Polk	Sales Associate I	101.Northeast
4	11.25	1301	James Polk	Sales Associate I	101.West
5	11.25	1301	James Polk	Sales Associate I	101.Midwest
6	11.25	1301	James Polk	Sales Associate I	101.Midwest
7	11.25	1301	James Polk	Sales Associate I	101.South

The fourth stream of data:

4 of 4 Fields | 24 records displayed

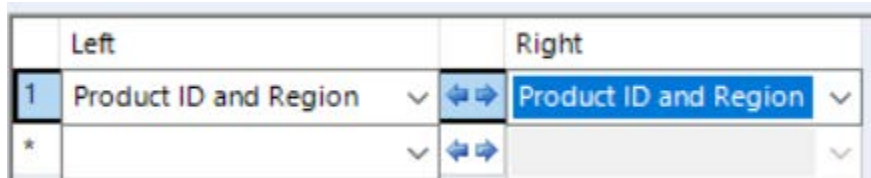
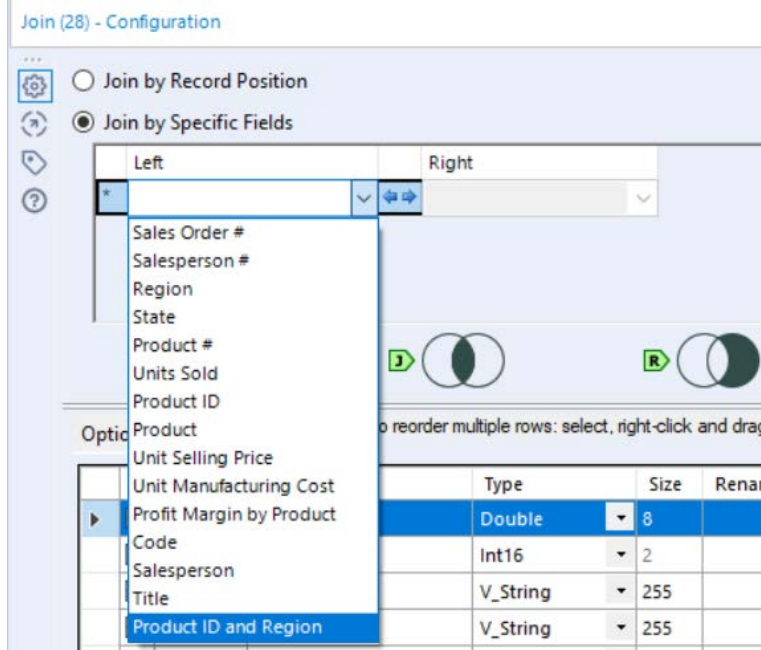
Record	Product ID	Region	Shipping Costs	Product ID and Region
1	101	Midwest	2.15	101.Midwest
2	101	Northeast	2.12	101.Northeast
3	101	South	2.08	101.South
4	101	West	2.02	101.West
5	102	Midwest	2.11	102.Midwest
6	102	Northeast	2.08	102.Northeast

Drag from the right of the “Formula Tool” in the first three data streams to the L (Left) inflow of the “Join Tool”.

Drag from the output of the “Formula Tool” in the fourth data stream to the R (Right) inflow of the “Join Tool”.



With the “**Join Tool**” selected, in the Configuration area, from the drop-down choose “Product ID and Region” on the left. and if it does not self-populate, “Product ID and Region” on the right. The two fields must be the same type.



Select Run to apply. If necessary, select the J of the Join output and select Run.

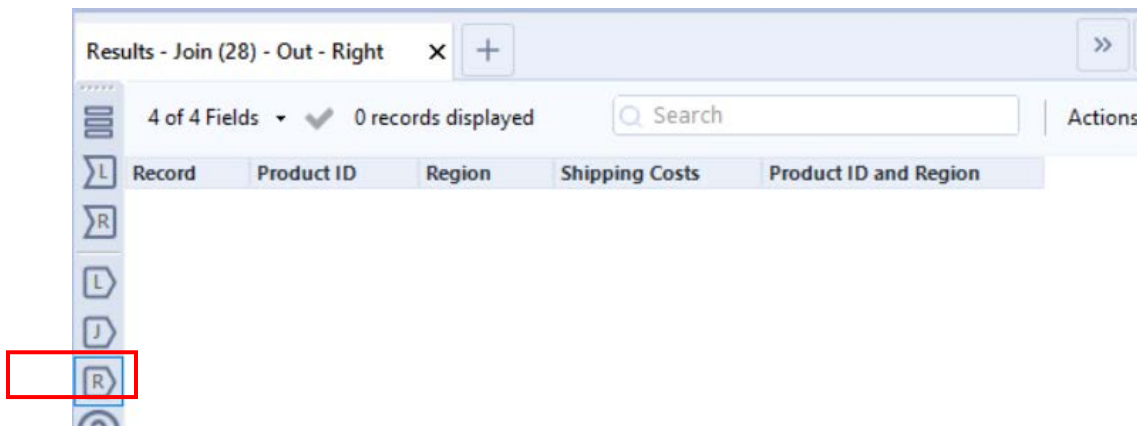
In the result area select J, Join, and review the output.

Results - Join (28) - Out - Join

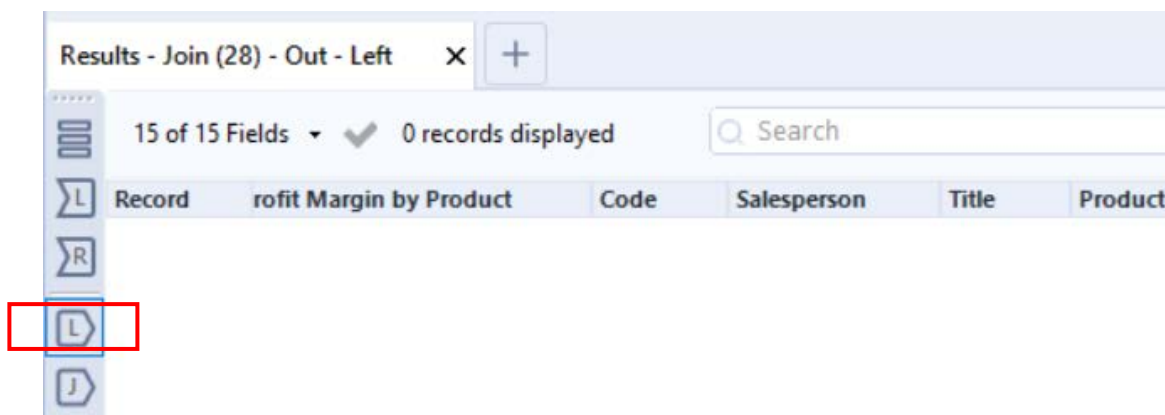
* 4,715 of 5,017 records displayed (partial results)

Record	Right_Product ID	Right_Region	Shipping Costs	Right_Product ID and Region
1	101	Midwest	2.15	101.Midwest
2	101	Midwest	2.15	101.Midwest
3	101	Midwest	2.15	101.Midwest
4	101	Midwest	2.15	101.Midwest
5	101	Midwest	2.15	101.Midwest
6	101	Midwest	2.15	101.Midwest
7	101	Midwest	2.15	101.Midwest

In the result area select R, Right, stream and verify that there are no shipping costs that were not used in any sales invoice.



In the result area select L, Left, stream and verify that there are no sales invoices without a shipping cost. If the region was spelled incorrectly it would not have been joined and would appear as a Left stream error.



Results - Join (28) - Out - Join

19 of 19 Fields Cell Viewer *4,715 of 5,017 records displayed(partial results)

Record	Sales Order #	Salesperson #	Region	State	Product #	Units Sold	Product ID
1	36658	1301	Midwest	Memory limit reached. Use a Browse tool to view full results.			
2	36710	1301	Midwest	IN	101	72	101
3	39854	1301	Midwest	MI	101	132	101
4	40243	1301	Midwest	NE	101	120	101

The 5,017 records have been joined but not displayed because of a memory limit.

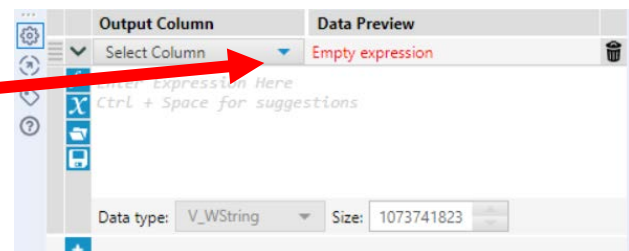
Requirement 5: Create fields to calculate sales in dollars and the shipping cost for each invoice.

Tool 16: Formula Tool: Calculate The Sales in Dollars for Every Sales Order

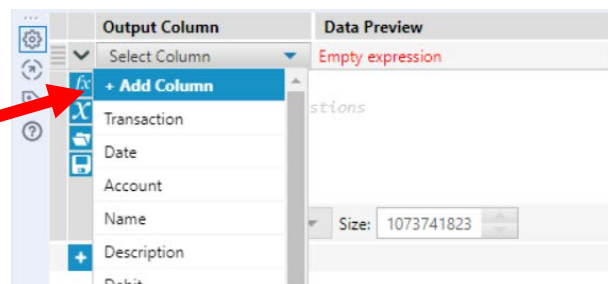
Drag a “**Formula Tool**” from the Tools Palette to the output of the Join output of the last “**Join Tool**”. The two will be connected automatically. The outflow from the tool on the left is the inflow to the one on the right.



In the output Configuration area, click on the drop-down arrow of “Select Column”.



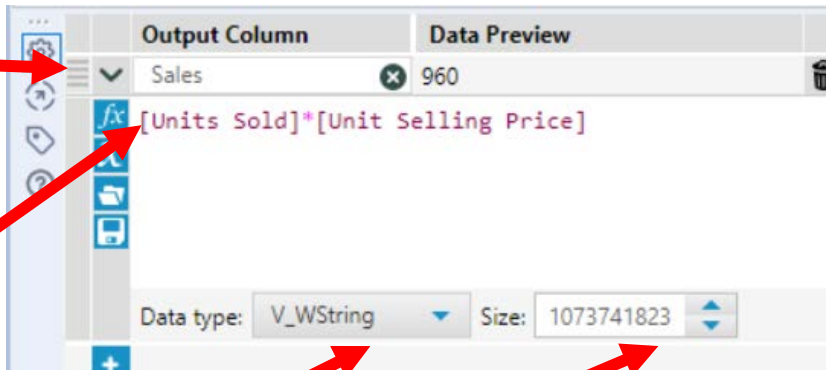
Select “Add Column”.



Type "Sales"

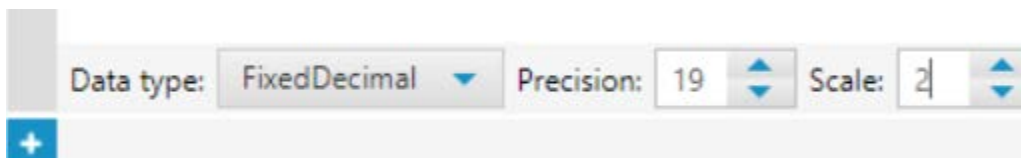
Type the following formula:

[Sales]*[Unit Selling Price]



Change the "Data type:" to FixedDecimal.

Change the Scale (number of decimal places) to two.



Select "Run" to apply.

Check the math:

Record 1: 48 x 20 = 960.00

Record 2: 72 x 20 = 1,440.00

Record	Units Sold	Product ID	Product	Unit Selling Price
1	48	101	Stuffed Lamb	20
2	72	101	Stuffed Lamb	20

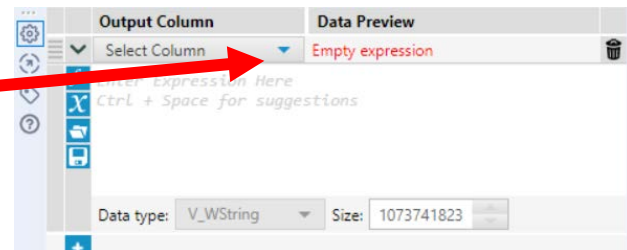
Record	and Region	Sales
1		960.00
2		1440.00

Tool 17: Formula Tool: Calculate The Shipping Cost in Dollars for Every Sales Order

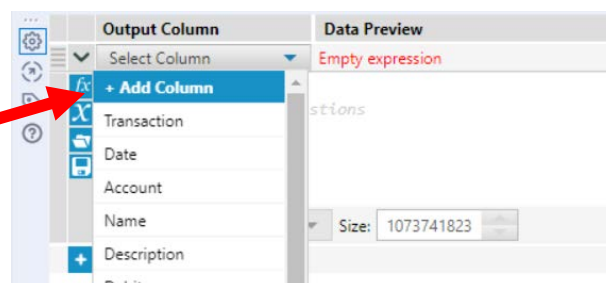
Drag a “**Formula Tool**” from the Tools Palette to the output of the “**Formula Tool**” that was previously created. The two will be connected automatically. The outflow from the tool on the left is the inflow to the one on the right.



In the output Configuration area, click on the drop-down arrow of “Select Column”.



Select “Add Column”.



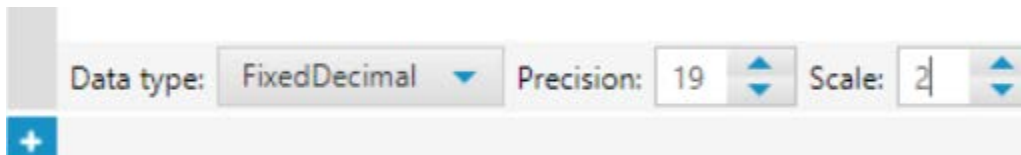
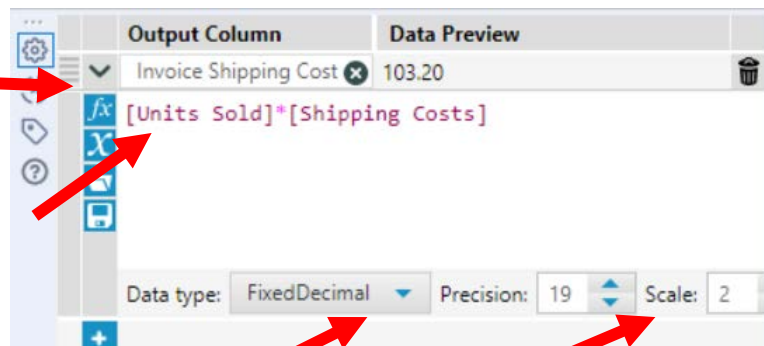
Type "Invoice Shipping Cost"

Type the following formula:

[Units Sold]*[Shipping Cost]

Change the "Data type:" to FixedDecimal.

Change the Scale (number of decimal places) to two.



Select "Run" to apply.

Check the math:

Record 1: $48 \times 2.15 = 103.20$

Record 2: $72 \times 2.15 = 154.80$

Record	Units Sold
1	48
2	72

	Shipping Costs
1	2.15
2	2.15

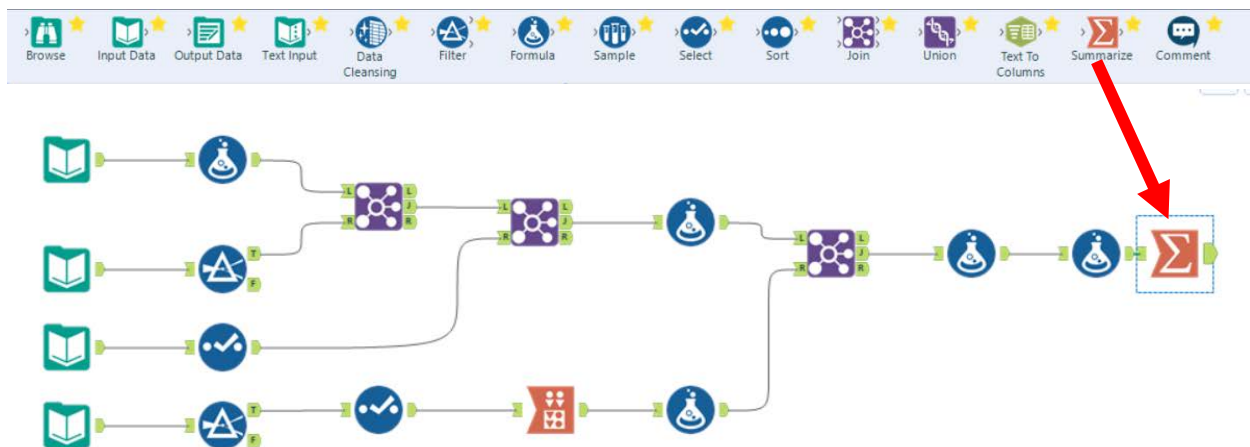
Record	Invoice Shipping Cost
1	103.20
2	154.80

Requirement 6: Create a report that show the number of units sold and sales by product.

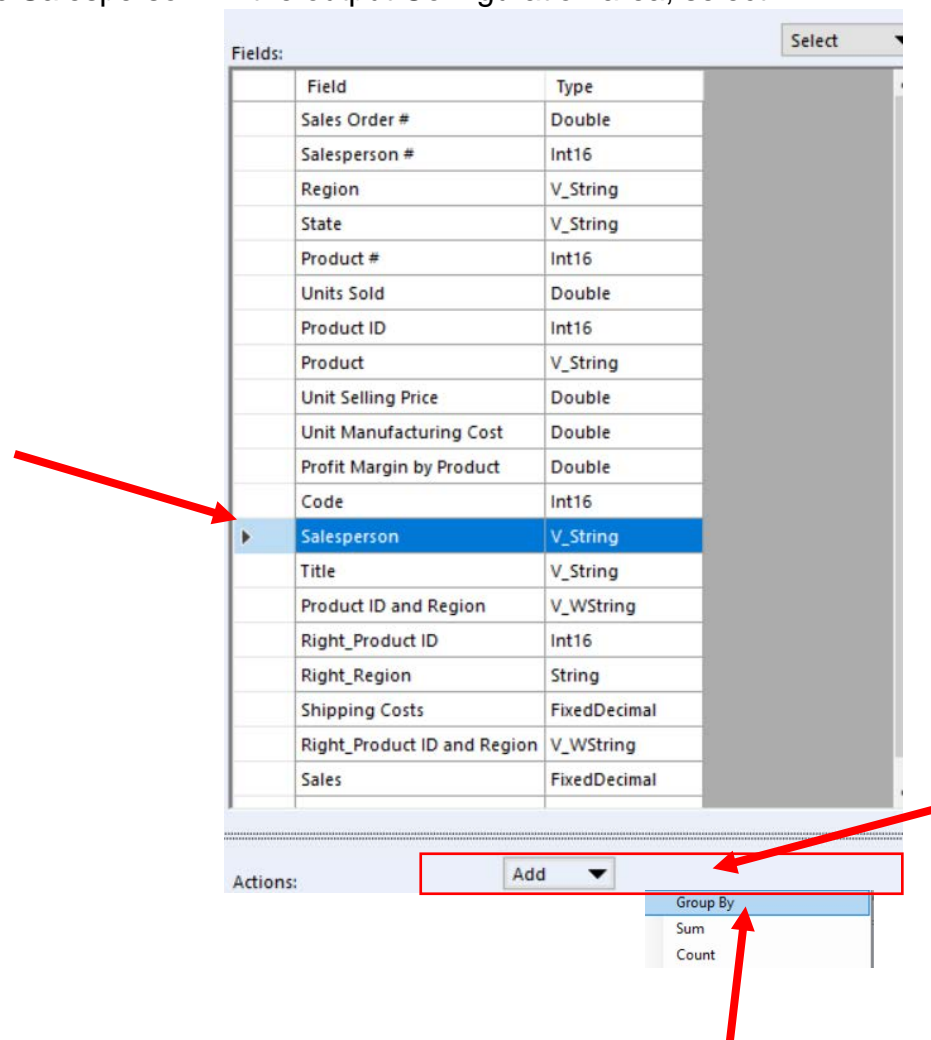
Tool 18: Select Tool Summary: We now want to create a report by salesperson.

By Salesperson			
Salesperson	# of units	Sales	Shipping Costs
Ulysses Grant	36,636	\$818,052.00	\$77,266.44
George Washington	58,284	\$1,299,192.00	\$123,242.28
James Monroe	97,152	\$2,162,436.00	\$205,622.64
John Tyler	78,924	\$1,751,352.00	\$166,596.00
James Madison	38,388	\$854,412.00	\$81,105.24
James Polk	77,484	\$1,723,956.00	\$163,717.08
Total	386,868	\$8,609,400.00	\$817,549.68

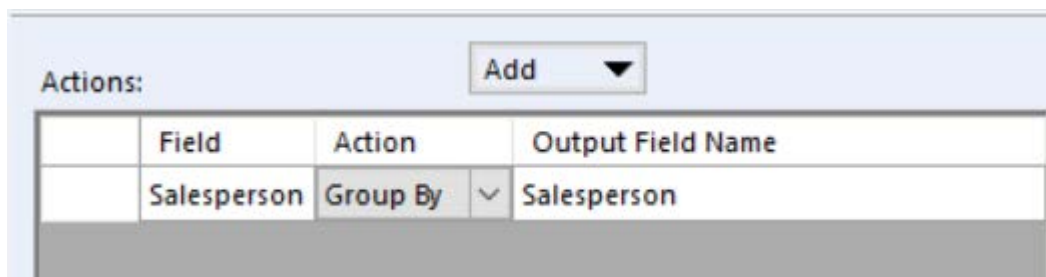
Drag a “**Summary Tool**” to the right of the “**Formula tool**” on the Canvas. The two will be connected automatically. The outflow from the tool on the left is the inflow to the one on the right.



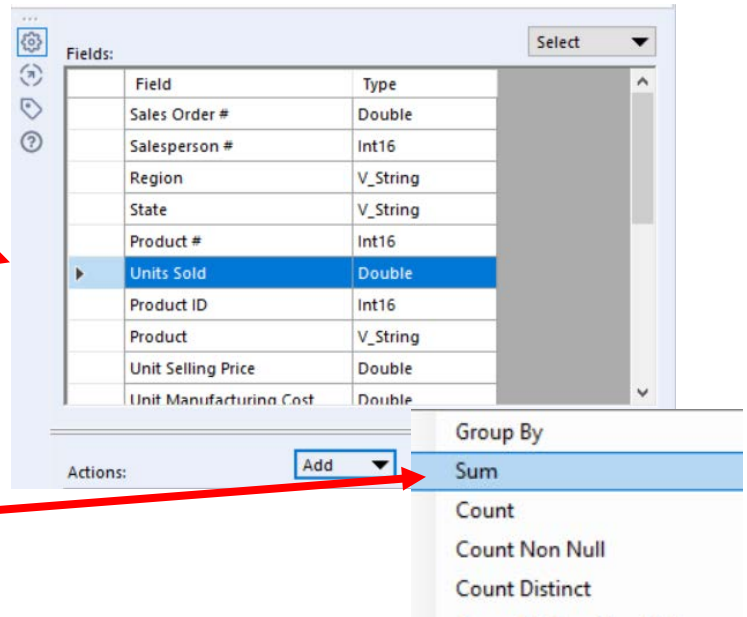
The first column will be Salesperson. In the output Configuration area, select “Salesperson”.



Click on the drop-down arrow on “Add” in “Actions” and select “Group By”.

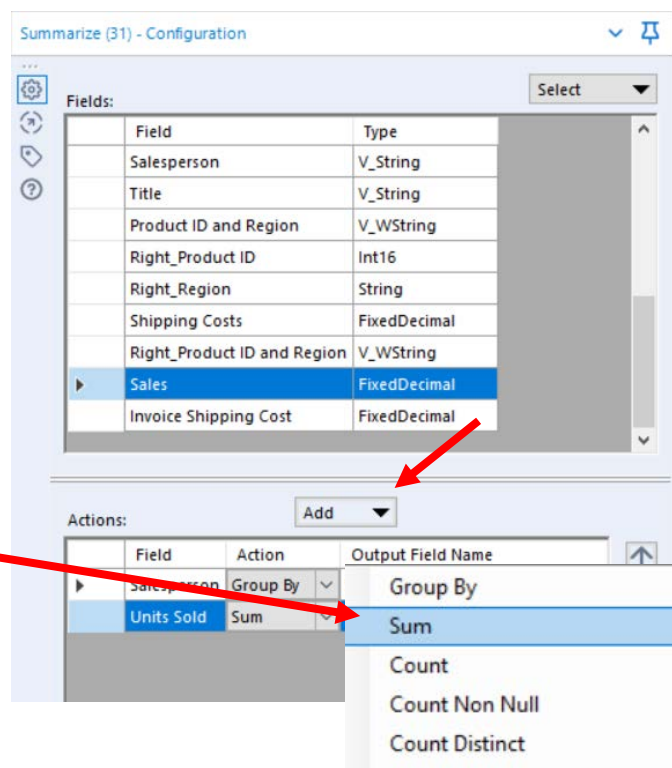


The second column will be “Units Sold”.



Click on the drop-down arrow on "Add" in the "Actions" and select "Sum".

The third column will be "Sales".



Click on the drop-down arrow on "Add" in the "Actions" and select "Sum".

The fourth column will be “Invoice Shipping Cost”.

Click on the drop-down arrow on “Add” in the “Actions” and select “Sum”.

Select “Run”.

The report is created.

Fields:

Field	Type
Salesperson	V_String
Title	V_String
Product ID and Region	V_WString
Right_Product ID	Int16
Right_Region	String
Shipping Costs	FixedDecimal
Right_Product ID and Region	V_WString
Sales	FixedDecimal
Invoice Shipping Cost	FixedDecimal

Actions:

Field	Action	Output Field Name
Salesperson	Group By	Salesperson
Units Sold	Sum	Sum_Units Sold
Sales	Sum	Sum_Sales
Invoice Shippin...	Sum	Sum_Invoice Ship...

Results - Summarize (31) - Output

Record	Salesperson	Sum_Units Sold	Sum_Sales	Sum_Invoice Shipping Cost
1	Grover Cleveland	76680	1733088.00	156216.00
2	James Madison	59136	1349352.00	120158.04
3	James Monroe	121236	2748540.00	246882.36
4	James Polk	43200	985560.00	87710.64
5	Thomas Jefferson	141156	3210924.00	287129.76
6	Ulysses Grant	57924	1317384.00	117878.40

Change the “Output Field Names” and select “Run”

Actions:

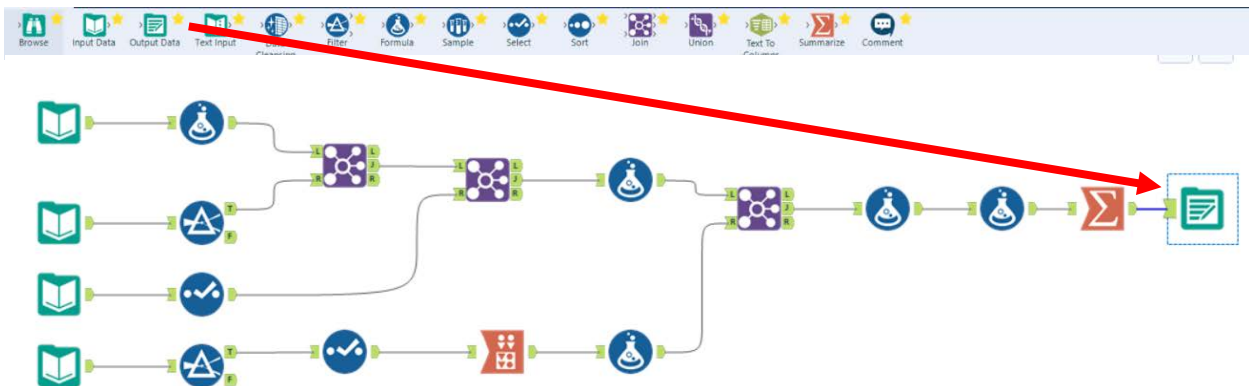
Field	Action	Output Field Name
Salesperson	Group By	Salesperson
Units Sold	Sum	Units
Sales	Sum	Sales
Invoice Shippin...	Sum	Shipping

Results - Summarize (31) - Output

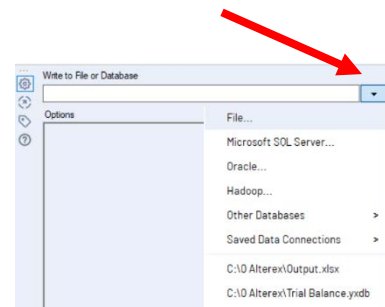
Record	Salesperson	Units	Sales	Shipping
1	Grover Cleveland	76680	1733088.00	156216.00
2	James Madison	59136	1349352.00	120158.04
3	James Monroe	121236	2748540.00	246882.36
4	James Polk	43200	985560.00	87710.64
5	Thomas Jefferson	141156	3210924.00	287129.76
6	Ulysses Grant	57924	1317384.00	117878.40

Tool 19: Select Tool Output Data Tool: We now want to export the cleaned and assembled Salesperson Report to Excel.

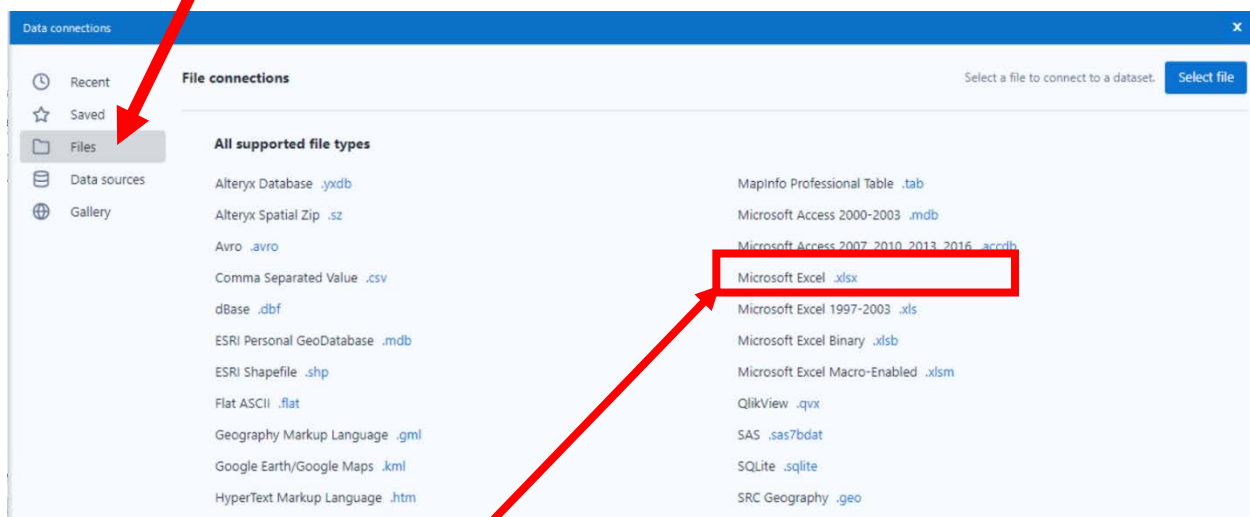
Drag an “**Output Data Tool**” to the right of the “**Summary Tool**” on the Canvas. The two will be connected automatically. The outflow from the tool on the left is the inflow to the one on the right.



In the output Configuration area, select the drop-down arrow.

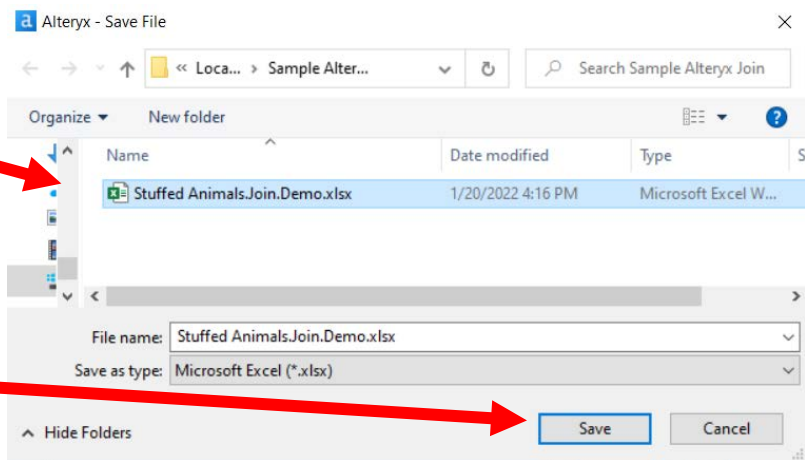


Select “File”.



Select “Microsoft Excel .xlsx”

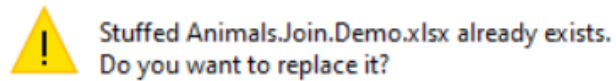
Select the required file name.



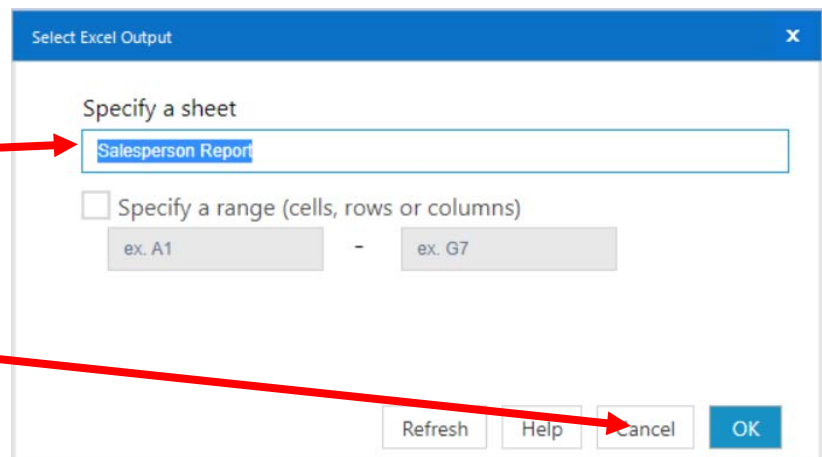
Select "Save".

Confirm Save As

Click on "Yes" to confirm "Save".



Enter the sheet name, "Salesperson Report".



Select "OK".

Select "Run".

Open the file in Excel.

	A	B	C	D	E
1	Salesperson	Units	Sales	Shipping	
2	Grover Cleveland	76680	#####	#####	
3	James Madison	59136	#####	#####	
4	James Monroe	121236	#####	#####	
5	James Polk	43200	#####	87710.64	
6	Thomas Jefferson	141156	#####	#####	
7	Ulysses Grant	57924	#####	#####	
8					
9					
10					
11					

Ready Scroll Lock

Change the columns' widths and formats.

	A	B	C	D	E
1	Salesperson	Units	Sales	Shipping	
2	Grover Cleveland	76,680.00	1,733,088.00	156,216.00	
3	James Madison	59,136.00	1,349,352.00	120,158.04	
4	James Monroe	121,236.00	2,748,540.00	246,882.36	
5	James Polk	43,200.00	985,560.00	87,710.64	
6	Thomas Jefferson	141,156.00	3,210,924.00	287,129.76	
7	Ulysses Grant	57,924.00	1,317,384.00	117,878.40	
8	Total	499,332.00	11,344,848.00	1,015,975.20	
9					

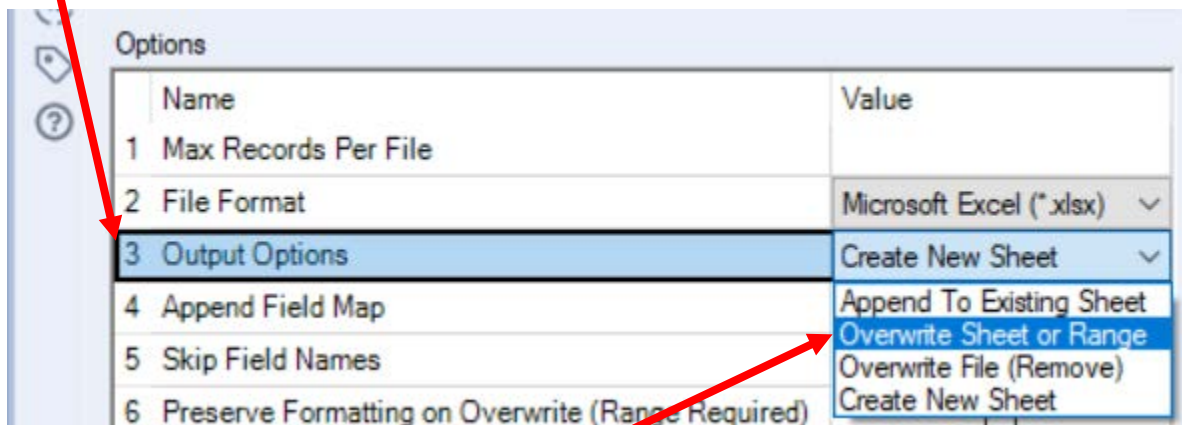
Ready Scroll Lock

“Save” and “Close” the Excel file.
Return to Alteryx and select “Run” for the second time.

But there is an error! The sheet already exists.



In the output Configuration area, select "Output Options" number 3.



Select "Overwrite Sheet or Range".

Select "Run".

Open the Excel file.

Open the file in Excel.

	A	B	C	D	E
1	Salespersc	Units	Sales	Shipping	
2	Grover Cle	76680	#####	#####	
3	James Ma	59136	#####	#####	
4	James Mo	121236	#####	#####	
5	James Polk	43200	#####	87710.64	
6	Thomas Je	141156	#####	#####	
7	Ulysses Gr	57924	#####	#####	
8					
9					
10					
11					

Salesperson Report

Reformat the Excel sheet and rename the formatted sheet "Formatted Salesman Report". When the Alteryx flow runs it will create a new unformatted sheet "Salesperson Report".

	A	B	C	D	E
1	Salesperson	Units	Sales	Shipping	
2	Grover Cleveland	76,680.00	1,733,088.00	156,216.00	
3	James Madison	59,136.00	1,349,352.00	120,158.04	
4	James Monroe	121,236.00	2,748,540.00	246,882.36	
5	James Polk	43,200.00	985,560.00	87,710.64	
6	Thomas Jefferson	141,156.00	3,210,924.00	287,129.76	
7	Ulysses Grant	57,924.00	1,317,384.00	117,878.40	
8	Total	499,332.00	11,344,848.00	1,015,975.20	
9					
10					
11					

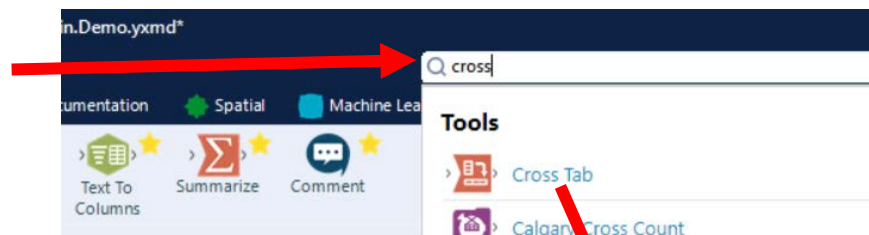
Formatted Salesperson Report

We now want to create a crosstab shipping cost report by product and region

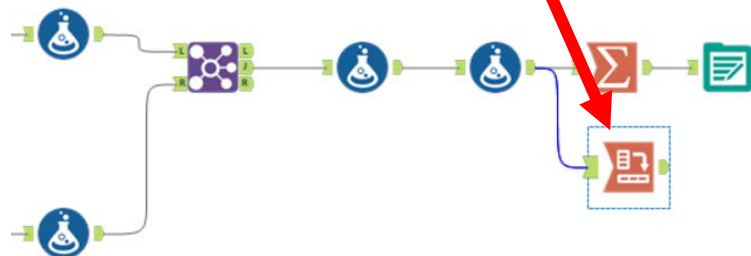
Shipping Costs by Product and Region					
	Midwest	Northeast	South	West	Grand Total
Stuffed Bear	\$ 26,904.00	\$ 43,048.44	\$ 22,066.20	\$ 41,647.80	\$ 133,666.44
Stuffed Elephant	\$ 54,810.00	\$ 72,060.84	\$ 48,486.60	\$ 86,971.80	\$ 262,329.24
Stuffed Frog	\$ 36,405.60	\$ 52,861.20	\$ 30,753.60	\$ 54,081.00	\$ 174,101.40
Stuffed Koala	\$ 14,094.00	\$ 20,699.28	\$ 14,572.80	\$ 25,364.40	\$ 74,730.48
Stuffed Panda	\$ 16,875.60	\$ 26,664.00	\$ 15,720.00	\$ 23,108.40	\$ 82,368.00
Stuffed Unicorn	\$ 13,648.20	\$ 30,604.32	\$ 18,144.00	\$ 27,957.60	\$ 90,354.12
Grand Total	\$ 162,737.40	\$ 245,938.08	\$ 149,743.20	\$ 259,131.00	\$ 817,549.68

Tool 20: **Cross Tab Tool**: Change Vertical Data to Multiple Horizontal Columns

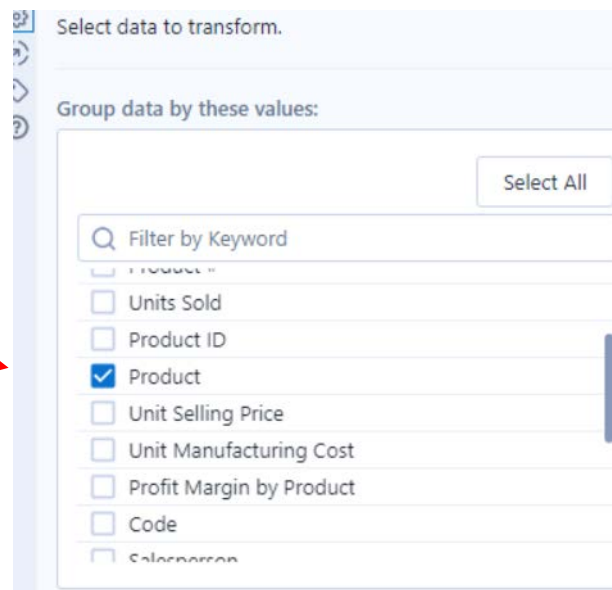
In the “Search for Tools” box type, “Cross Tab”.



Drag the “**Cross Tab Tool**” from the Tools Palette to the right of the “**Formula Tool**” and below the “**Summarize Tool**” on the Canvas. The two will be connected automatically. The outflow from the tool on the left is the inflow to the one on the right.



The leftmost column of the cross tab will be the “Product”. In the output Configuration area, “Group Data by these Values”, select “Product”.



The column headings of the cross tab will be the "Region". In the output Configuration area, "Change Column Headers" select "Region".

Change Column Headers

Region

Find

Sales Order #

Salesperson #

Region

In the output Configuration area, "Values for New Column" select "Invoice Shipping Cost".

Values for New Columns

Invoice Shipping Cost

In the output Configuration area, "Method for Aggregating Values" select: "Sum", "Total Column" and "Total Row".

Method for Aggregating Values

Select All

☒ Sum

☐ Average

☐ Count (Without Nulls)

☐ Count (With Nulls)

☐ Percent Row

☐ Percent Column

☒ Total Column

☒ Total Row

☐ First

☐ Last

Select "Run" to apply.

6 of 6 Fields | Cell Viewer | 7 records displayed | Search

Record	Product	Sum_Midwest	Sum_Northeast	Sum_South	Sum_West	Total
1	Stuffed Elephant	47610.72	35625.24	50822.64	63990.36	198048.96
2	Stuffed Giraffe	76744.92	56958.72	80686.08	91095.84	305485.56
3	Stuffed Horse	26032.32	17062.56	25488	31916.88	100499.76
4	Stuffed Lamb	30263.4	19614.24	29877.12	38905.2	118659.96
5	Stuffed Pig	47898.24	30566.4	47361.6	59648.4	185474.64
6	Stuffed Unicorn	25999.68	18458.88	29987.4	33360.36	107806.32
7	Total	254549.28	178286.04	264222.84	318917.04	1015975.2

Tool 21: Select Tool – Rename Fields

Drag the “**Select Tool**” from the Tools Palette to the output of the “**Cross Tab Tool**” on the Canvas. The two will be connected automatically.



Rename the column headers.

Change the type.

Options ▾ | ↑ ↓ | TIP: To reorder multiple rows: select, right-click and drag.

	Field	Type	Size	Rename	Description
<input checked="" type="checkbox"/>	Product	V_String	255		
<input checked="" type="checkbox"/>	Sum_Midwest	FixedDecimal	12.2	Midwest	
<input checked="" type="checkbox"/>	Sum_Northeast	FixedDecimal	12.2	Northeast	
<input checked="" type="checkbox"/>	Sum_South	FixedDecimal	12.2	South	
<input checked="" type="checkbox"/>	Sum_West	FixedDecimal	12.2	West	
<input checked="" type="checkbox"/>	Total	Double	8		
<input checked="" type="checkbox"/>	*Unknown	Unknown	0		Dynamic or Unknown Fields

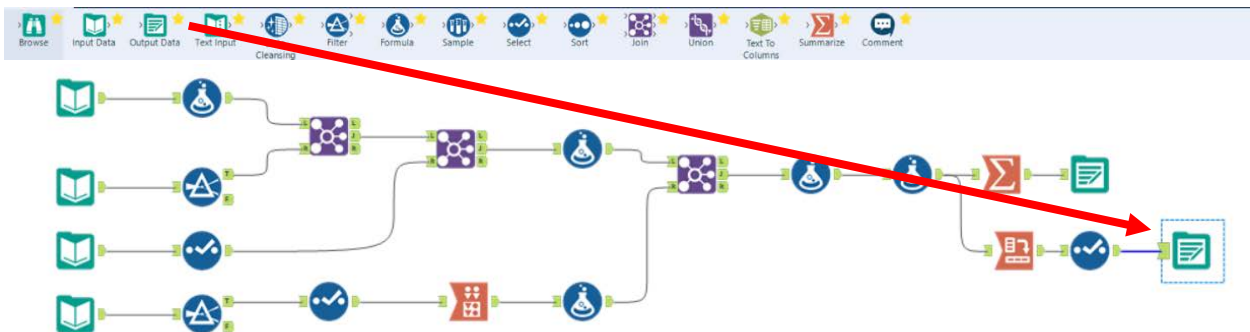
Select “Run” to apply.

6 of 6 Fields ▾ | Cell Viewer ▾ | 7 records displayed | ↑ ↓ | Search

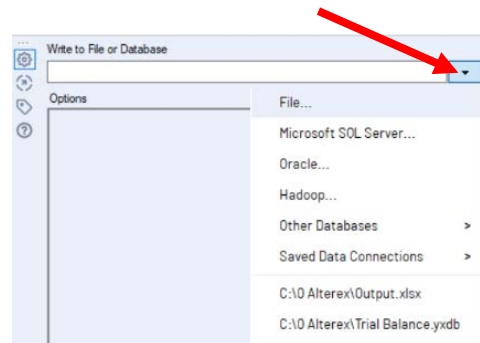
Record	Product	Midwest	Northeast	South	West	Total
1	Stuffed Elephant	47610.72	35625.24	50822.64	63990.36	198048.96
2	Stuffed Giraffe	76744.92	56958.72	80686.08	91095.84	305485.56
3	Stuffed Horse	26032.32	17062.56	25488	31916.88	100499.76
4	Stuffed Lamb	30263.4	19614.24	29877.12	38905.2	118659.96
5	Stuffed Pig	47898.24	30566.4	47361.6	59648.4	185474.64
6	Stuffed Unicorn	25999.68	18458.88	29987.4	33360.36	107806.32
7	Total	254549.28	178286.04	264222.84	318917.04	1015975.2

Tool 22: Select Tool Output Data Tool: We now want to export the cleaned and assembled Salesperson Report to Excel.

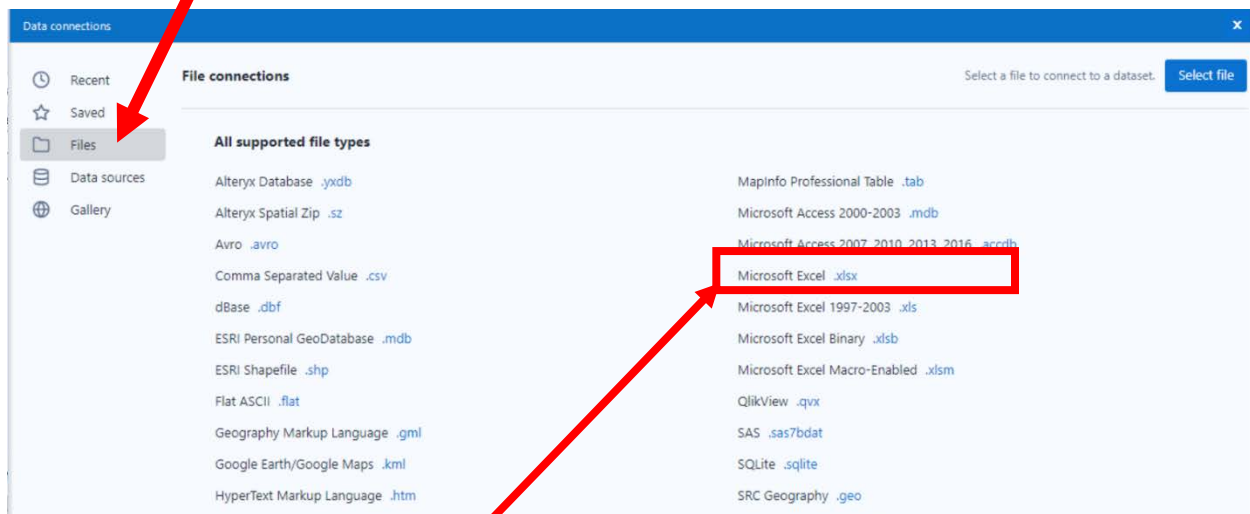
Drag an “**Output Data Tool**” to the right of the “**Select Tool**” on the Canvas. The two will be connected automatically. The outflow from the tool on the left is the inflow to the one on the right.



In the output Configuration area, select the drop-down arrow.



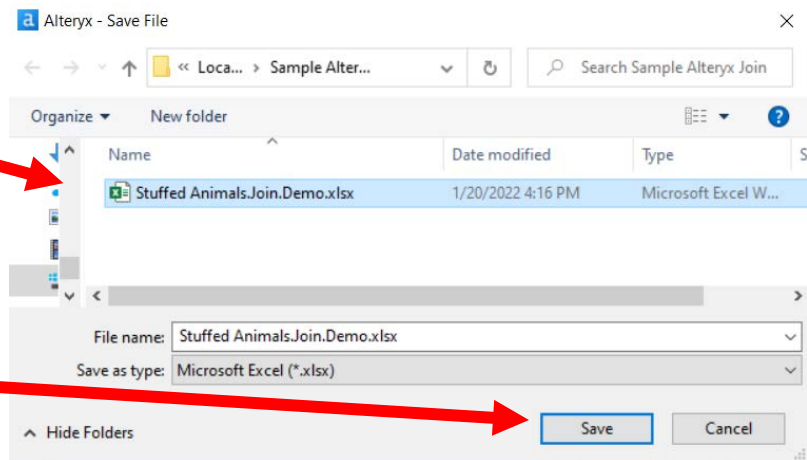
Select “File”.



Select “Microsoft Excel .xlsx”

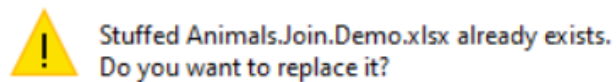
Select the required file name.

Select “Save”.



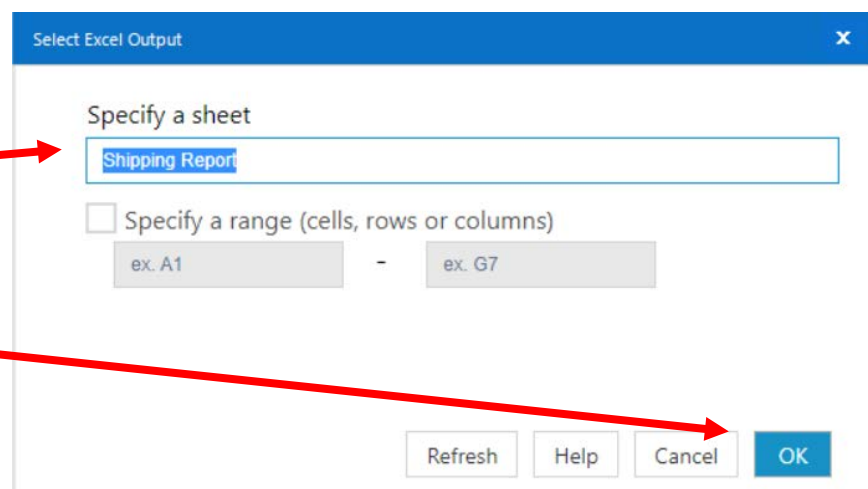
Confirm Save As

Click on “Yes” to confirm “Save”.



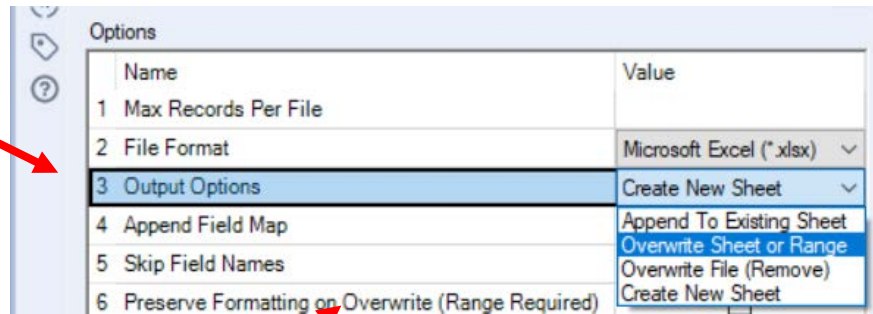
Enter the sheet name, “Shipping Report”.

Select “OK”.



In the output Configuration area, select “Output Options” number 3.

Select “Overwrite Sheet or Range”.



Select “Run”.

Open the file in Excel.

	A	B	C	D	E	F	G	H
1	Product	Midwest	Northeast	South	West	Total		
2	Stuffed Ele	47610.72	35625.24	50822.64	63990.36	198049		
3	Stuffed Gir	76744.92	56958.72	80686.08	91095.84	305485.6		
4	Stuffed Ho	26032.32	17062.56	25488	31916.88	100499.8		
5	Stuffed La	30263.4	19614.24	29877.12	38905.2	118660		
6	Stuffed Pig	47898.24	30566.4	47361.6	59648.4	185474.6		
7	Stuffed Un	25999.68	18458.88	29987.4	33360.36	107806.3		
8	Total	254549.3	178286	264222.8	318917	1015975		
9								
10								
11								

Navigation: Salesperson Report | Formatted Salesperson Report | **Shipping Report**

Reformat the Excel sheet and rename the formatted sheet “Formatted Shipping Report”. When the Alteryx flow runs it will create a new unformatted sheet “Shipping Report”.

	A	B	C	D	E	F
1	Product	Midwest	Northeast	South	West	Total
2	Stuffed Elephant	47,610.72	35,625.24	50,822.64	63,990.36	198,048.96
3	Stuffed Giraffe	76,744.92	56,958.72	80,686.08	91,095.84	305,485.56
4	Stuffed Horse	26,032.32	17,062.56	25,488.00	31,916.88	100,499.76
5	Stuffed Lamb	30,263.40	19,614.24	29,877.12	38,905.20	118,659.96
6	Stuffed Pig	47,898.24	30,566.40	47,361.60	59,648.40	185,474.64
7	Stuffed Unicorn	25,999.68	18,458.88	29,987.40	33,360.36	107,806.32
8	Total	254,549.28	178,286.04	264,222.84	318,917.04	1,015,975.20
9						
10						
11						

Navigation: Formatted Salesperson Report | **Formatted Shipping Report** (+)

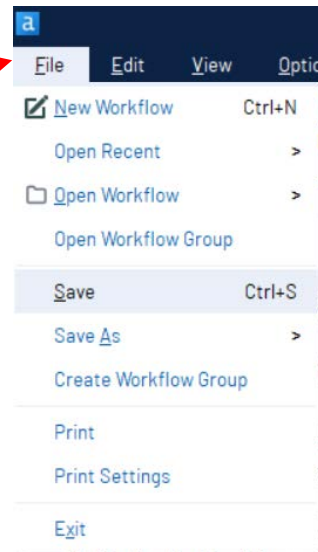
“Save” and “Close” the Excel file.

Return to Alteryx and select “Run” for the second time.

The workflow must be saved.

Select "File".

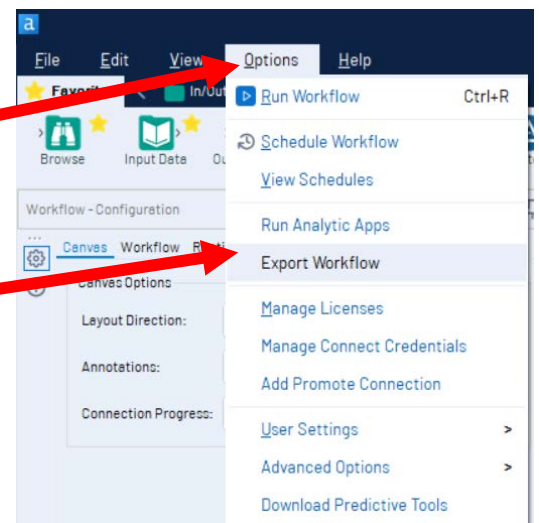
Select "Save".



Export the workflow and the related input files so that the project can be completed on another computer or uploaded to the web.

Select "Options".

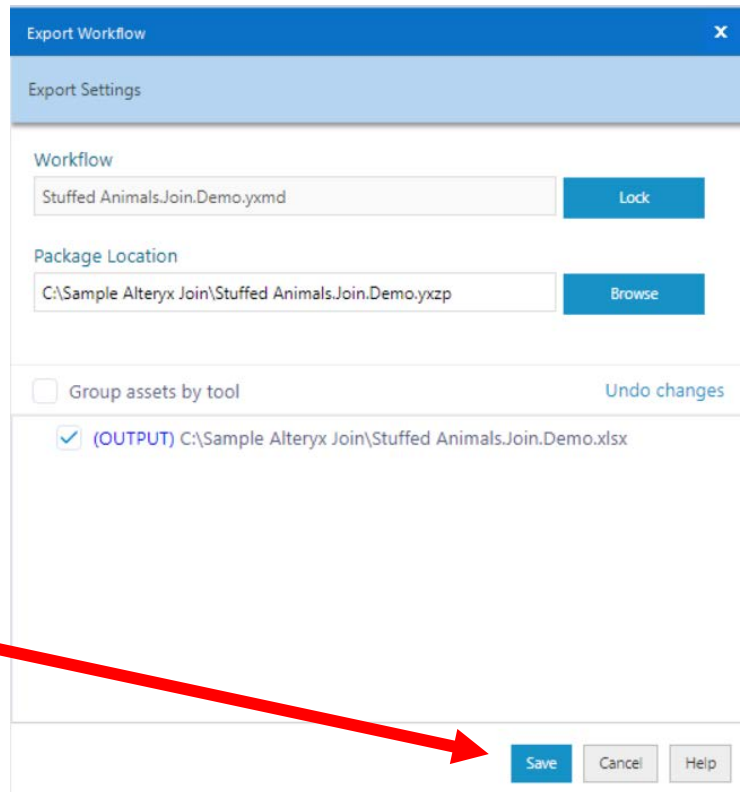
Select "Export Workflow".



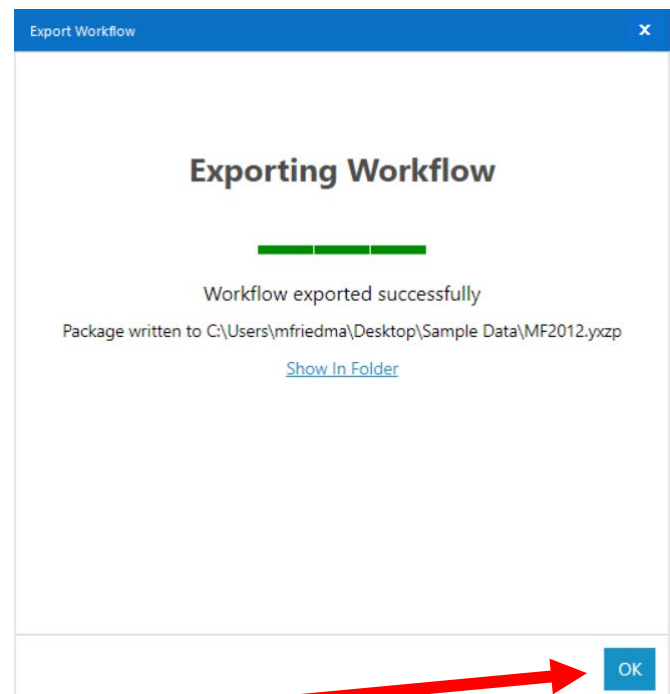
Verify the name of the "Workflow".

Verify the "Package Location".

Select "Save".



The "Export Workflow" dialog box is shown. It has a blue header bar with the title "Export Workflow" and a close button (X). Below the header is a section titled "Export Settings". Under this section, there are two input fields: "Workflow" and "Package Location". The "Workflow" field contains the text "Stuffed Animals.Join.Demo.yxml" and has a "Lock" button to its right. The "Package Location" field contains the text "C:\Sample Alteryx Join\Stuffed Animals.Join.Demo.yxzp" and has a "Browse" button to its right. Below these fields is a checkbox labeled "Group assets by tool" which is currently unchecked. To the right of this checkbox is a link that says "Undo changes". Below the checkbox is a list of output files, with the first one being "(OUTPUT) C:\Sample Alteryx Join\Stuffed Animals.Join.Demo.xlsx" and having a checkmark to its left. At the bottom of the dialog box are three buttons: "Save", "Cancel", and "Help".



The "Exporting Workflow" progress dialog box is shown. It has a blue header bar with the title "Export Workflow" and a close button (X). The main content area has a title "Exporting Workflow" in bold. Below the title is a green progress bar that is nearly full. Under the progress bar, the text "Workflow exported successfully" is displayed. Below that, the text "Package written to C:\Users\mfriedma\Desktop\Sample Data\MF2012.yxzp" is shown. Below this text is a link that says "Show In Folder". At the bottom right of the dialog box is an "OK" button.

Select "OK".

Exit Alteryx.