Using Tableau Prep Builder

Tableau Prep **E**xtracts data from one or more source files, **T**ransforms the data to conform with consistent standards and **L**oads the data into a repository. ETL software is used to automate the redundant manipulation of data. In this exercise, you will **E**xtract the **January Sales Data** file, **T**ransform the data into the required two reports: # of units, Sales and Profit Margin by Product and Salesperson. Then **L**oad the transformed data into Excel.

Notice that the column headings are in the ninth row, some rows have partial duplicate data and some rows are blank.

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3											
4											
5											
6											
7					Cuddly Stuffed Animals						
8					January Sales Data						
9	Sales Order #	Salesperson #	Salesperson	Title	Profit Margin	Region #	Region	State	Product	# of units	Sales
10	35003	1304	Martin van Bure	Sales Associate I	1,908.00		3	South	TN	Stuffed Horse	144	3,168.00
11										Stuffed Horse		3,168.00
12												
13	35006	1302	Grover Clevelan	Senior Sales Associate	936.00		1	Midwest	NE	Stuffed Panda	72	1,656.00
14										Stuffed Panda		1,656.00
15												
16	35009	1301	Thomas Jefferso	Sales Associate I	2,028.00		2	Northeast	PA	Stuffed Panda	156	3,588.00
17					2,028.00					Stuffed Panda		
18												
19	35011	1305	George Washing	Sales Associate I	1,560.00		1	Midwest	MN	Stuffed Panda	120	2,760.00
20					1,560.00					Stuffed Panda		
21												
22	35014	1303	Ulysses Grant	Senior Sales Associate	1,215.00		1	Midwest	MN	Stuffed Giraffe	108	2,268.00
23					1,215.00					Stuffed Giraffe		
24												
25	35017	1305	George Washing	Sales Associate I	708.00		2	Northeast	RI	Stuffed Elephant	48	1,200.00
26					708.00					Stuffed Elephant		
27												
28	35025	1303	Ulysses Grant	Senior Sales Associate	1,272.00		2	Northeast	ME	Stuffed Horse	96	2,112.00
29										Stuffed Horse	96	
30												
31	35026	1301	Thomas Jefferso	Sales Associate I	1,716.00		4	West	WY	Stuffed Penguin	156	3,120.00

| Introduction | **January Sales Data** | February-December Sales Data | Products | Salesperson | Region | Sheet3 | Sh ... ⊕ |

Sample Reports

Product	# of units	Sales	Profit Margin
Stuffed Frog	1,524	$ 33,528	$20,574
Stuffed Unicorr	408	8,160	4,692
Stuffed Panda	1,140	26,220	15,675
Stuffed Koala	624	14,976	8,112
Stuffed Elephai	4,044	84,924	49,539
Stuffed Bear	2,376	59,400	35,046
Total	10,116	$227,208	$133,638

Salesperson	# of units	Sales	Profit Margin
1405 - James Polk	1,812	$41,568	$24,438
1425 - John Tyler	2,172	48,336	28,110
1541 - Ulysses Grant	756	17,112	10,062
1576 - James Madison	1,776	39,216	23,160
1588 - George Washington	1,584	35,784	21,168
1776 - James Monroe	2,016	45,192	26,700
Total	10,116	$227,208	$133,638

Start Tableau Prep

The Tableau Prep Builder workspace consists of the Connections pane and three coordinated areas that help you interact with and explore your data:

- Flow pane: A visual representation of your operation steps as you prepare your data.
- Profile pane: A summary of each field in your data sample. See the shape of your data and quickly find outliers and nulls.
- Data grid pane: The row level detail for your data.

After you connect to your data and begin building your flow, you add steps in the Flow pane. These steps function as a lens into the structure of your data, as well as a summary of operations that is applied to your data. Each step represents a different category of operations that you define.

Step 1: Input Excel file

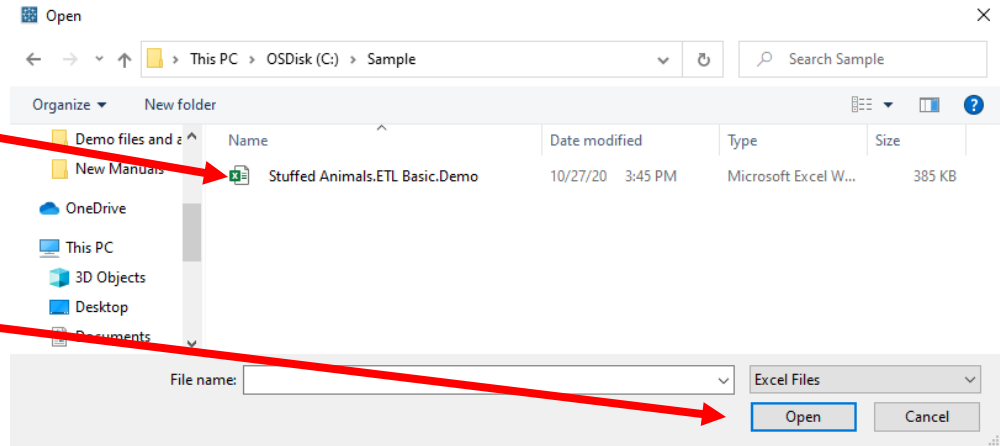To open Excel file via Tableau Prep, press green button "Connect to Data" on the Flow pane.
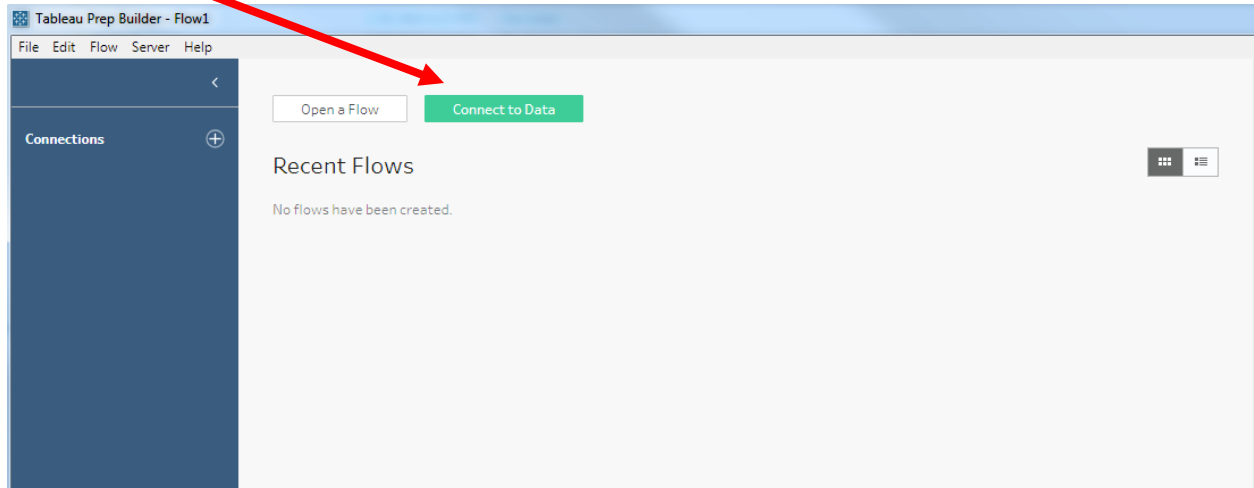
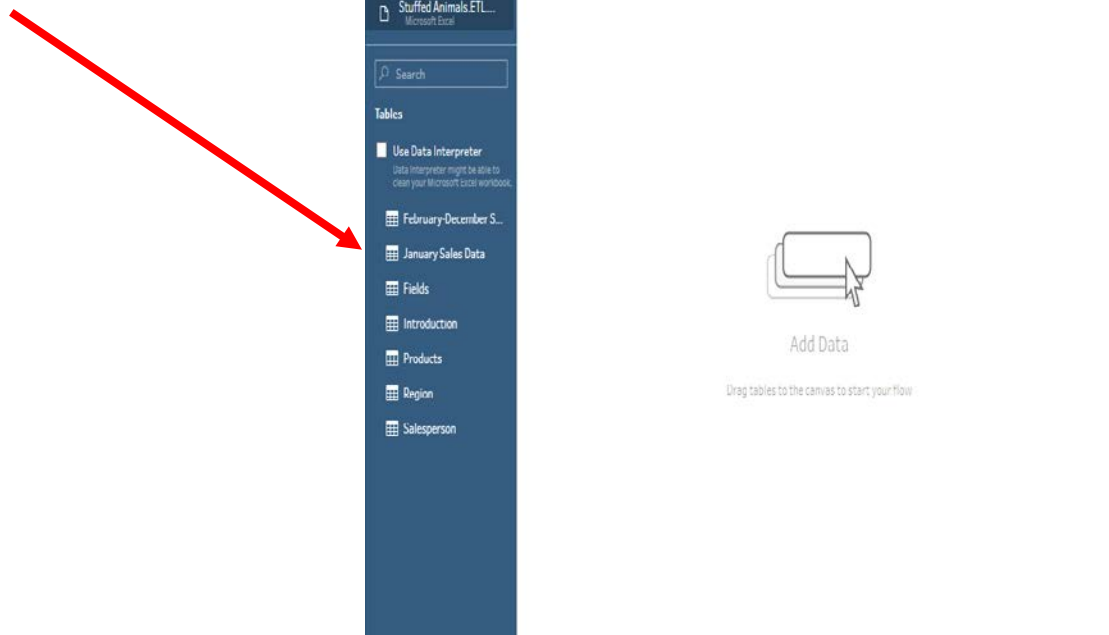On the connections pane select Microsoft Excel.
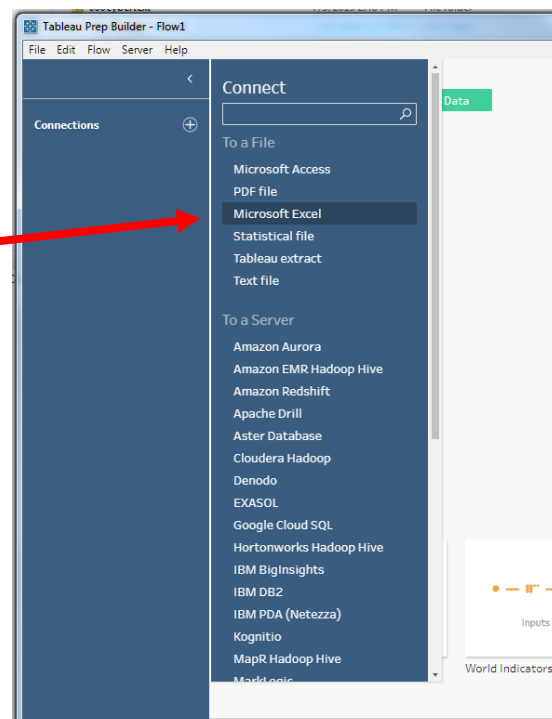
Select the file.

Select "Open".

After you open the file, select the "January Sales Data" sheet on the Connections pane.
Either double click on the sheet name or drag the sheet name to the canvas.

After selecting the sheet, "January Sales Data" will appear on the Flow pane.

Sample data will be visible on the Profile pane. Notice that the Field Names are not correct.

Step 2: Cleaning the data

The field names are in the ninth row.

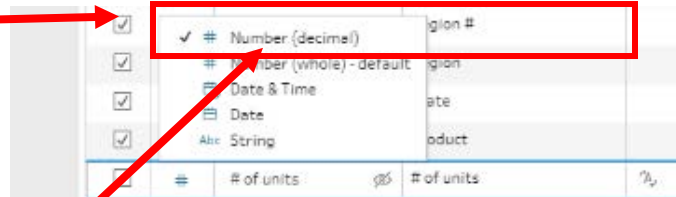Sometimes Tableau can find the headers by checking the box to the left of "Use Data Interpreter".

If it works, the field names will be dispalyed correctly.

If not, double click on the "Field Name" and change the names as shown below.

All number fields that will be added, subtracted, multiplied, divided must be coded as "Number (decimal)".

Change the field "Type" by selecting the "Type" to be changed and selecting the desired "Type".

- "Profit Margin": "Number (decimal)"
- "# of Units": "Number (decimal)"
- "Sales": "Number (decimal)"

	Type	Field Name	Original Field Name	Changes	Preview
✓	#	Sales Order #	Sales Order #		35,005, null, 35,006
✓	#	Salesperson #	Salesperson #		1,304, null, 1,302
✓	Abc	Salesperson	Salesperson		Martin Van Buren, null, Grover Cleveland
✓	Abc	Title	Title		Sales Associate I, null, Senior Sales Associate
✓	#	Profit Margin	Profit Margin	⌁	1,908, null, 936
✓	#	Region #	Region #		3, null, 1
✓	Abc	Region	Region		South, null, Midwest
✓	Abc	State	State		TN, null, NE
✓	Abc	Product	Product		Stuffed Horse, Stuffed Panda
✓	#	# of units	# of units	⌁	144, null, 72
✓	#	Sales	Sales	⌁	3,168, 1,656

In the "Flow Pane" select the "+" on the right of the "January Sales" input.
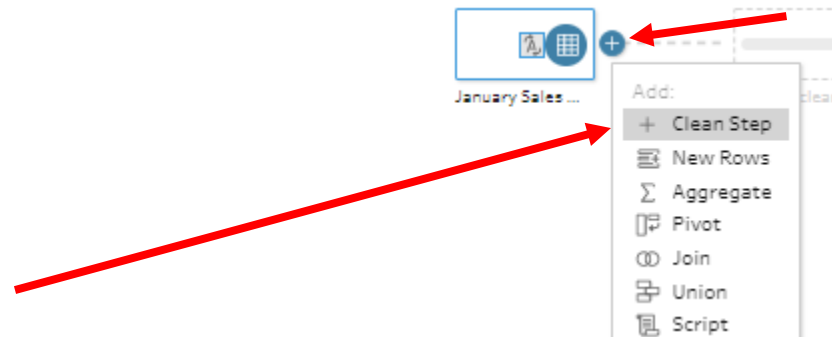
Select "+ Clean Step".
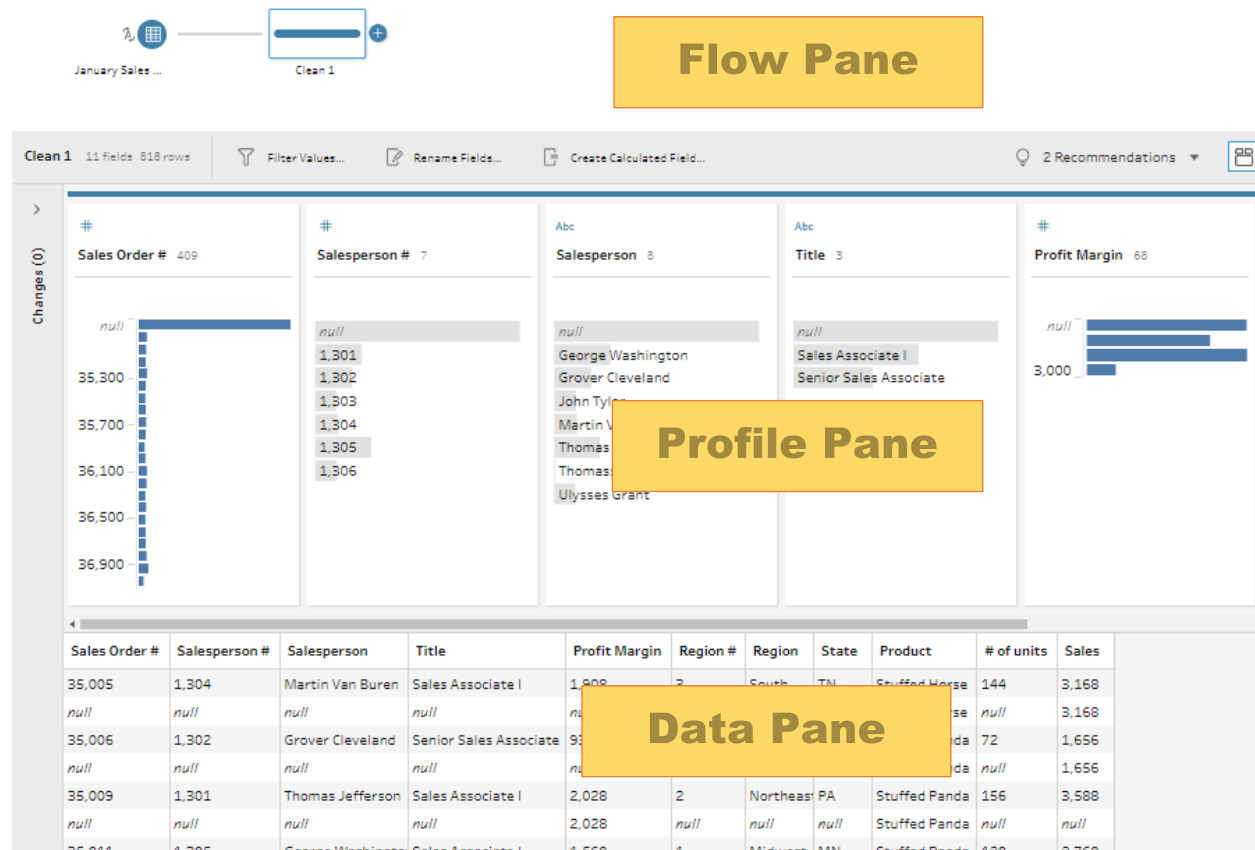
If a field is displayed as a graph, hover over the field in the Profile Pane. Click on the three dots that appear in the upper right corner and select "Detail".

Exclude all rows that do not contain data

A quick review of the "Data Pane" shows that the blank rows were deleted and that some rows that do not contain a value for "Salesperson #".
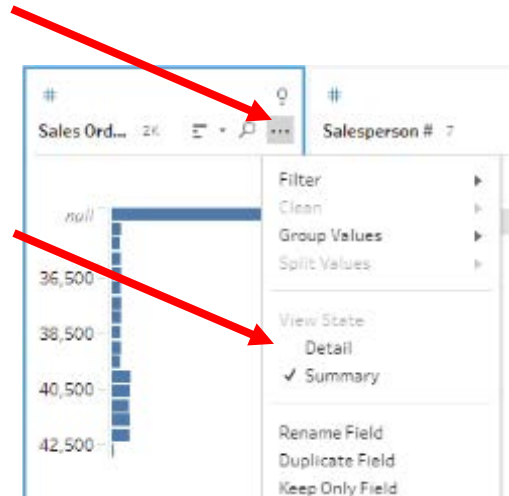
Sales Order #	Salesperson #	Salesperson	Title	Profit Margin	Region #	Region	State	Product	# of units	Sales
35,005	1,304	Martin Van Buren	Sales Associate I	1,908	3	South	TN	Stuffed Horse	144	3,168
null	*null*	*null*	*null*	*null*	*null*	*null*	*null*	Stuffed Horse	*null*	3,168
35,006	1,302	Grover Cleveland	Senior Sales Associate	936	1	Midwest	NE	Stuffed Panda	72	1,656
null	*null*	*null*	*null*	*null*	*null*	*null*	*null*	Stuffed Panda	*null*	1,656
35,009	1,301	Thomas Jefferson	Sales Associate I	2,028	2	Northeast	PA	Stuffed Panda	156	3,588
null	*null*	*null*	*null*	2,028	*null*	*null*	*null*	Stuffed Panda	*null*	*null*
35,011	1,305	George Washington	Sales Associate I	1,560	1	Midwest	MN	Stuffed Panda	120	2,760
null	*null*	*null*	*null*	1,560	*null*	*null*	*null*	Stuffed Panda	*null*	*null*
35,014	1,303	Ulysses Grant	Senior Sales Associate	1,215	1	Midwest	MN	Stuffed Giraffe	108	2,268
null	*null*	*null*	*null*	1,215	*null*	*null*	*null*	Stuffed Giraffe	*null*	*null*
35,017	1,305	George Washington	Sales Associate I	708	2	Northeast	RI	Stuffed Elephant	48	1,200
null	*null*	*null*	*null*	708	*null*	*null*	*null*	Stuffed Elephant	*null*	*null*
35,025	1,303	Ulysses Grant	Senior Sales Associate	1,272	2	Northeast	ME	Stuffed Horse	96	2,112
null	*null*	*null*	*null*	*null*	*null*	*null*	*null*	Stuffed Horse	96	*null*
35,026	1,301	Thomas Jefferson	Sales Associate I	1,716	4	West	WY	Stuffed Penguin	156	3,120
null	*null*	*null*	*null*	*null*	*null*	*null*	*null*	Stuffed Penguin	*null*	3,120

Modify the working dataset, so it includes only those rows that are not coded as *"null"*.

Right click on the "null" value in the "Salesperson #" field and select "Exclude".
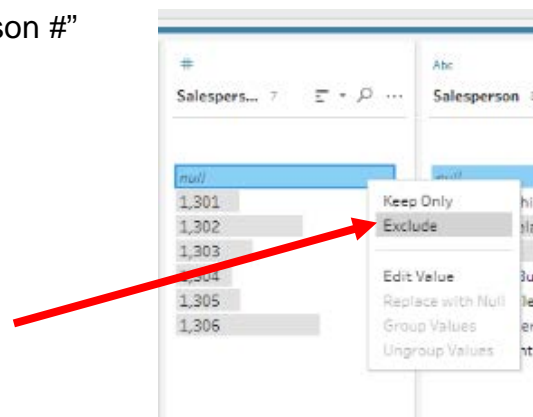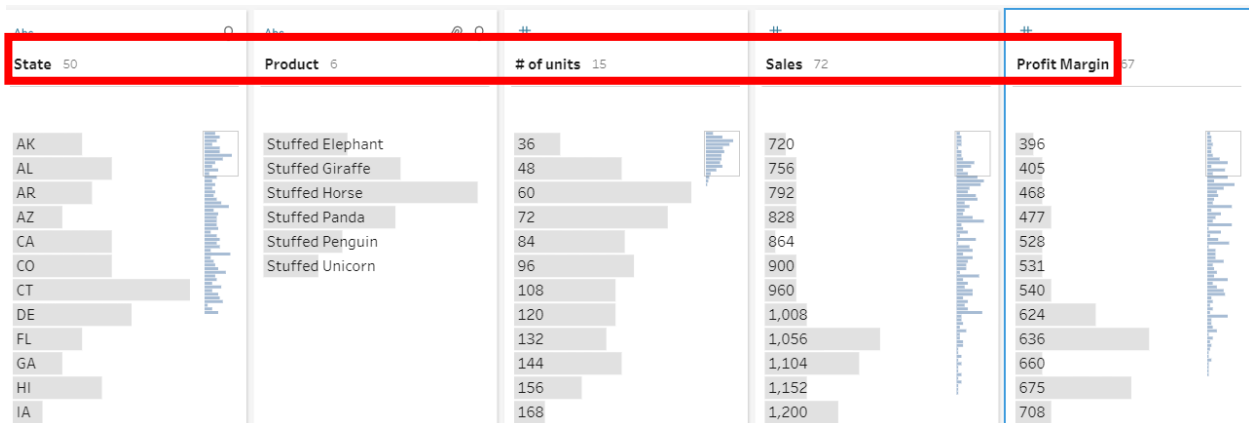
Reorder the fields
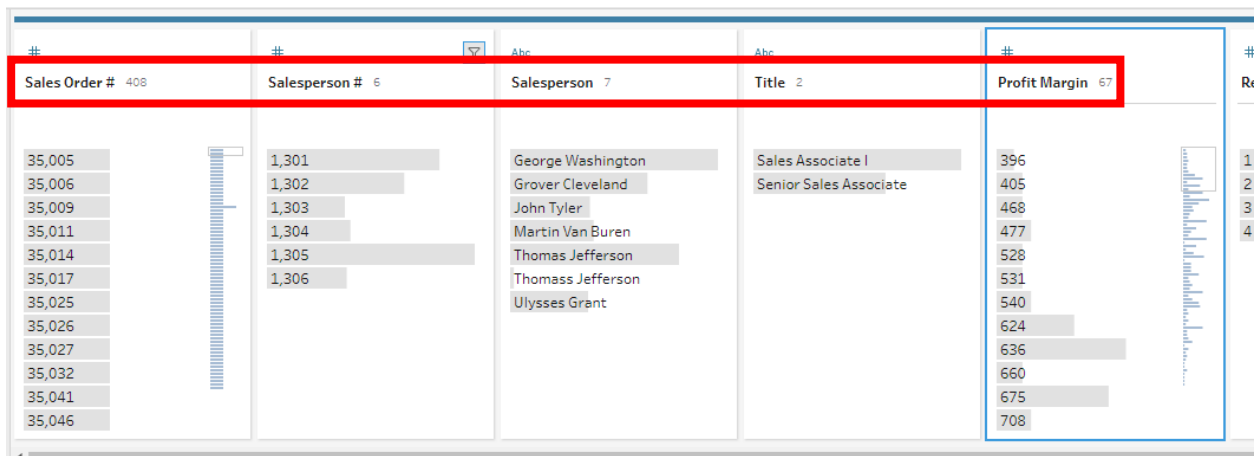
In the "Data Pane" select the field "Profit Margin" and drag it to the last field position.

Now the fields are in a different order.

Clean the Data in the Fields

The Cuddly Stuffed Animals company provided additional information in the Excel file.

Product ID	Product
101	Stuffed Penguin
102	Stuffed Horse
103	Stuffed Unicorn
104	Stuffed Giraffe
105	Stuffed Panda
106	Stuffed Elephant

Code	Salesperson
1301	Thomas Jefferson
1302	Grover Cleveland
1303	Ulysses Grant
1304	Martin Van Buren
1305	George Washington
1306	John Tyler

Code	Region
1	Midwest
2	Northeast
3	South
4	West

Sometime data is entered into a system incorrectly. The "Salesperson" field has entries for "Thomass Jefferson" that must be corrected to "Thomas Jefferson".

Select both, right click and select "Group Values".
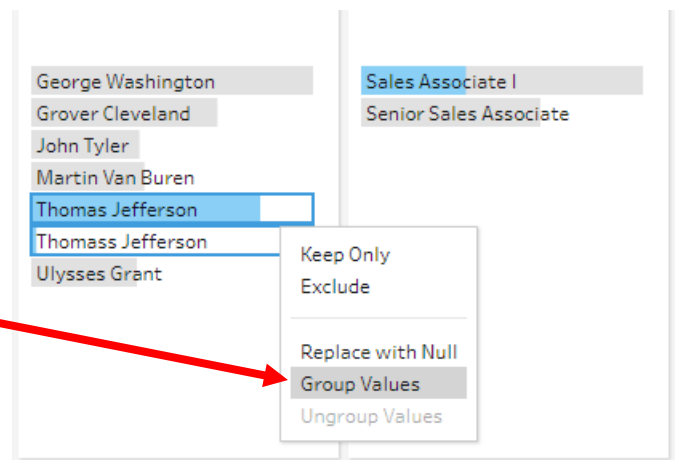
If the system groups with the incorrect name, type the correct name and hit the enter key.
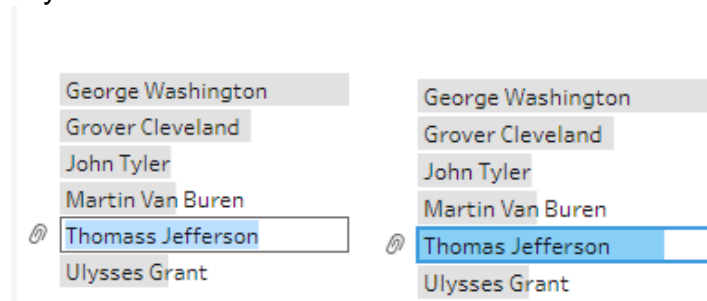
Check and correct the "Product" and "Region" Fields.

Step 3: Aggregate Balances

To accumulate the values by product, Click on the "+" to the right of "Clean 1" and select "Aggregate".

Drag "Product" field

to "Grouped Fields".
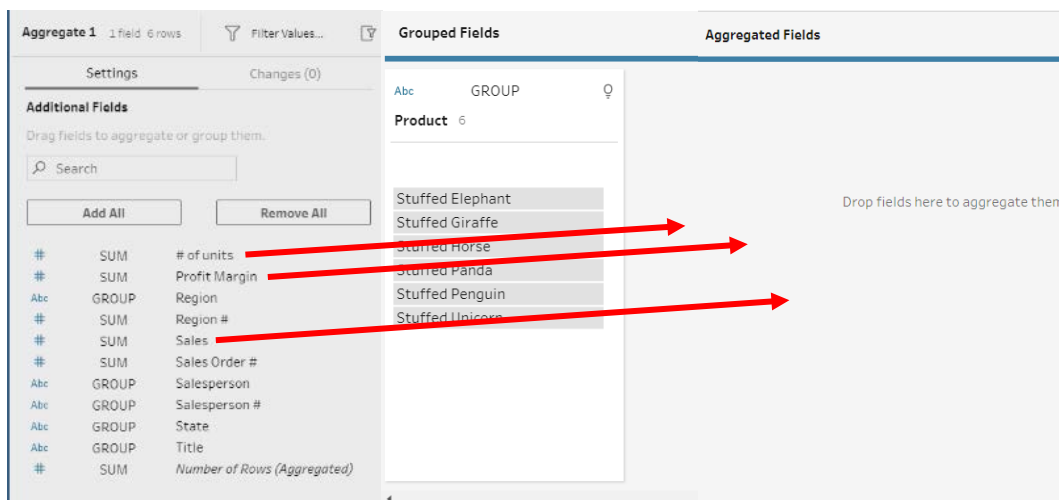
Drag "# of units", "Sales" and Profit Margin" to "Aggregated Fields".
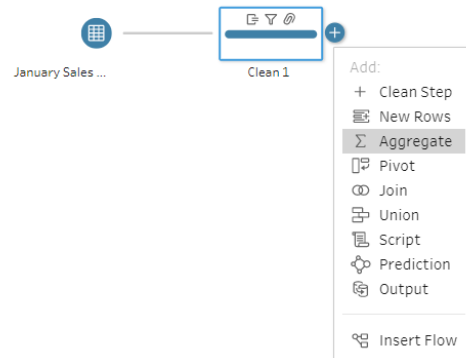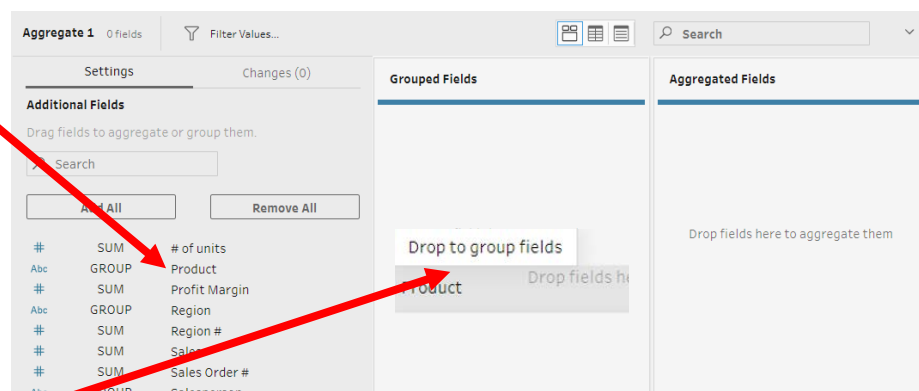
Grouped Fields

Abc GROUP		# SUM	# SUM	# SUM
Product 6		# of units 6	Sales 6	Profit Margin 6
Stuffed Elephant		3,060	73,440	40,545
Stuffed Giraffe		4,392	100,320	55,176
Stuffed Horse		5,016	109,800	64,782
Stuffed Panda		7,488	169,092	90,585
Stuffed Penguin		8,052	172,224	97,344
Stuffed Unicorn		12,120	266,640	160,590

Aggregated Fields

Product	# of units	Sales	Profit Margin
Stuffed Unicorn	3,060	73,440	40,545
Stuffed Horse	12,120	266,640	160,590
Stuffed Panda	7,488	172,224	97,344
Stuffed Giraffe	8,052	169,092	90,585
Stuffed Elephant	4,392	109,800	64,782
Stuffed Penguin	5,016	100,320	55,176

Step 4: Output the cleaned data that will be used for the Report by Product

Click on the plus sign of the "Aggregate 1"
tile and select "Output".
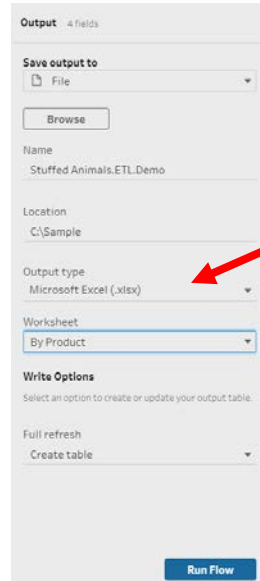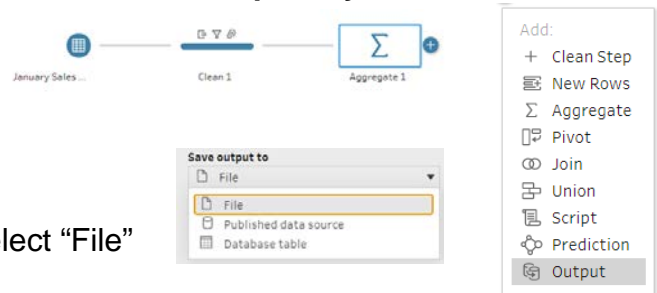
1.   "Save output to", Select "File"

2.   "Output type", Select
       "Microsoft Excel (.xlsx)"

3.   Select "Browse"

3.1   Select the original
       data file and select
       "Accept".
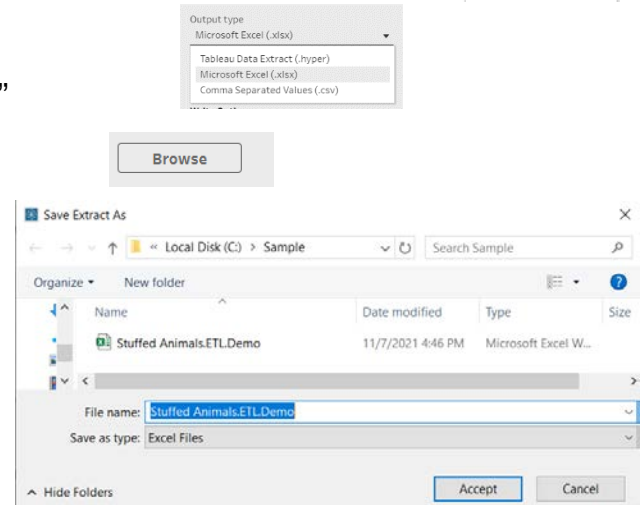
3.2 Confirm that you are replacing the original file.
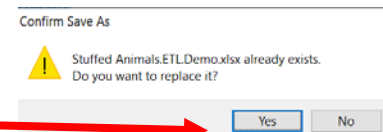
4.   "Worksheet", enter "Report by Product"
       and confirm by selecting "Create new worksheet
       "Report by Product".

5.   "Write Options" select "Create table" .

6.   Select "Run Flow".

If the program asks to confirm select Replace

Outputs with the same names already exist. Do you want to replace them?

Running the flow will replace these existing outputs with new outputs.

Output Step	Location	Name
Output	E:\Sample\Stuffed Animals.ETL.De...	Report by Product

Cancel   **Replace**
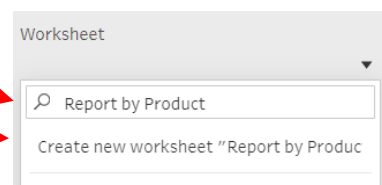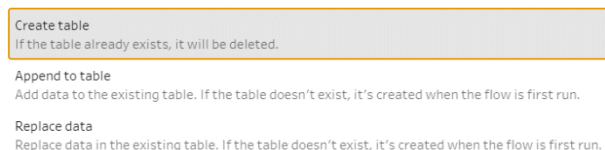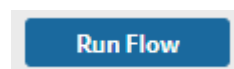
Select "Done".

Finished Running Flow

✓

General Ledger.csv

Total time 00:00

Done

Open the original data file in Excel.

	A	B	C	D	E	F
1	Product	# of units	Sales	Profit Margin		
2	Stuffed Un	3060	73440	40545		
3	Stuffed Ho	12120	266640	160590		
4	Stuffed Pa	7488	172224	97344		
5	Stuffed Gir	8052	169092	90585		
6	Stuffed Ele	4392	109800	64782		
7	Stuffed Pe	5016	100320	55176		
8						

◀ ▶ ... | Region | **Report by Product** ... ⊕ ⋮ ◀

`

Make format changes and to prevent the changes being inadvertently lost when other extracts are performed.

	A	B	C	D	E
1	Product	# of units	Sales	Profit Margin	
2	Stuffed Unicorn	3,060.00	$73,440.00	$40,545.00	
3	Stuffed Horse	12,120.00	$266,640.00	$160,590.00	
4	Stuffed Panda	7,488.00	$172,224.00	$97,344.00	
5	Stuffed Giraffe	8,052.00	$169,092.00	$90,585.00	
6	Stuffed Elephant	4,392.00	$109,800.00	$64,782.00	
7	Stuffed Penguin	5,016.00	$100,320.00	$55,176.00	
8	Total	40,128.00	$891,516.00	$509,022.00	
9					
10					

◀ ▶ ... | Region | **Jan Report by Product** ... ⊕ ⋮ ◀

To change the name of the tab, double click on the tab "Report by Product", replace it with "Jan Report by Product" and hit the Enter key.

Save, close the Excel file, and return to Tableau.

Step 5: Create a new field that combine the fields "Salesperson #" and Salesperson"

The "Salesperson" report will summarize the data by "Salesperson #" with the salesperson's name attached for clarity. We now want to combine the fields with " – " between the fields.

Salesperson	# of units	Sales	Profit Margin
1405 - James Polk	1,812	$41,568	$24,438
1425 - John Tyler	2,172	48,336	28,110
1541 - Ulysses Grant	756	17,112	10,062
1576 - James Madison	1,776	39,216	23,160
1588 - George Washington	1,584	35,784	21,168
1776 - James Monroe	2,016	45,192	26,700
Total	10,116	$227,208	$133,638

Click on the plus sign on the right of the "Clean 1" step and select "Clean Step".
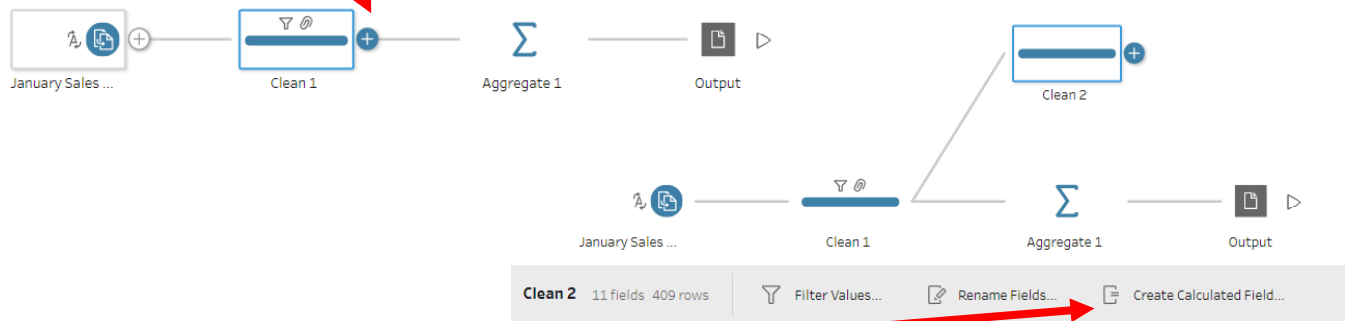
Select "Create Calculated Field".

Enter "Salesperson Number and Name" as the "Field Name".

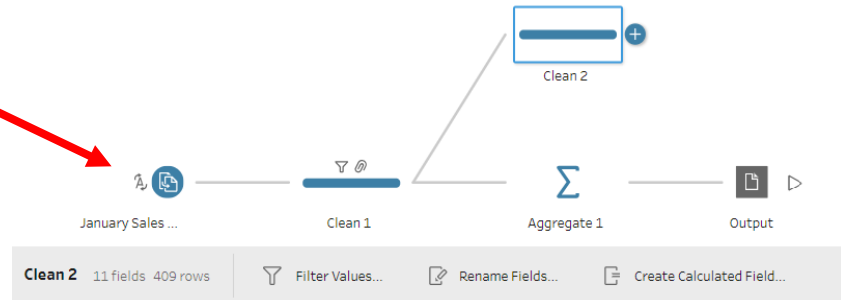Enter the formula: [Salesperson #]+" - "+[Salesperson].

There is an error because "Salesperson #" is a number and "Salesperson" is a string.

Function '+' can't accept parameters: (integer, string).

Select the "X" to close the window.

Change the field "Salesperson #" from a numeric to a string.

Select the first step, "January Sales".

January Sales ...   Clean 1   Clean 2   Aggregate 1   Output

Clean 2   11 fields 409 rows   Filter Values...   Rename Fields...   Create Calculated Field...

Select the "Type" field for "Salesperson #".

January Sales Data   11 fields   Filter Values...

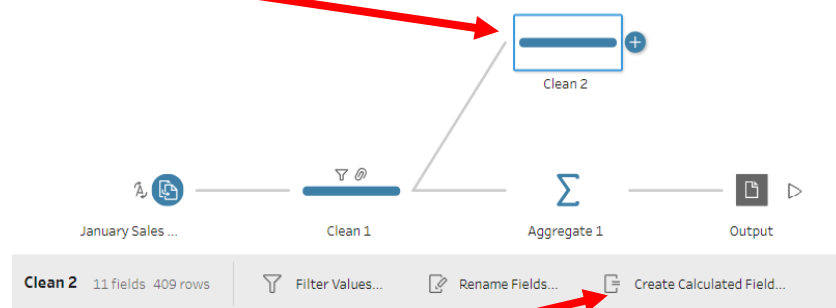Clear the check box to remove fields. You can also filter your data or change

Fields selected: 11 of 11

✓	Type	Field Name	Original Field Name
✓	#	Sales Order #	Sales Order #
✓	#	Salesperson #	Salesperson #
✓	Abc	Salesperson	Salesperson

Select "String"

	#	Number (decimal)
✓	#	Number (whole) - default
	🕒	Date & Time
	📅	Date
Abc		String

Select the step Clean 2.

Clean 2

January Sales ...   Clean 1   Aggregate 1   Output

Clean 2   11 fields 409 rows   Filter Values...   Rename Fields...   Create Calculated Field...

Select "Create Calculated Field".

Enter "Salesperson Number and Name" as the "Field Name".

Enter the formula: [Salesperson #]+" - "+[Salesperson].

.

Select "Save".

A new field has been created.

Step 6: Aggregate balances and export "Report by Salesperson"

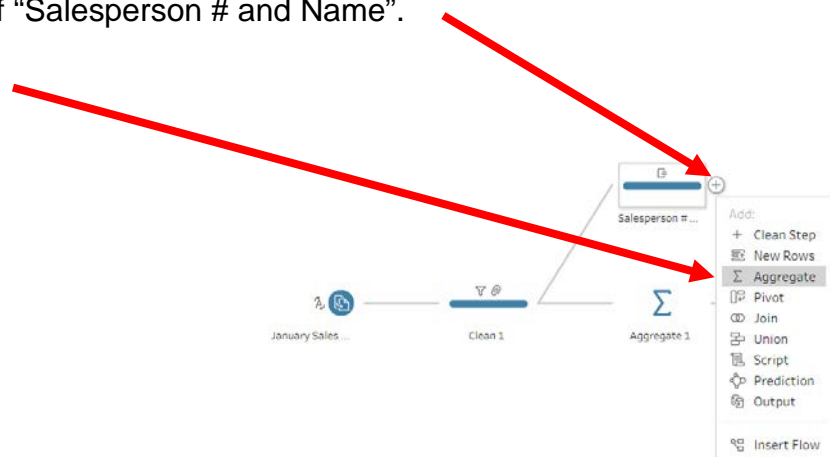Right Click on the "Clean 2" in the "Flow Pane" and select "Rename Step".
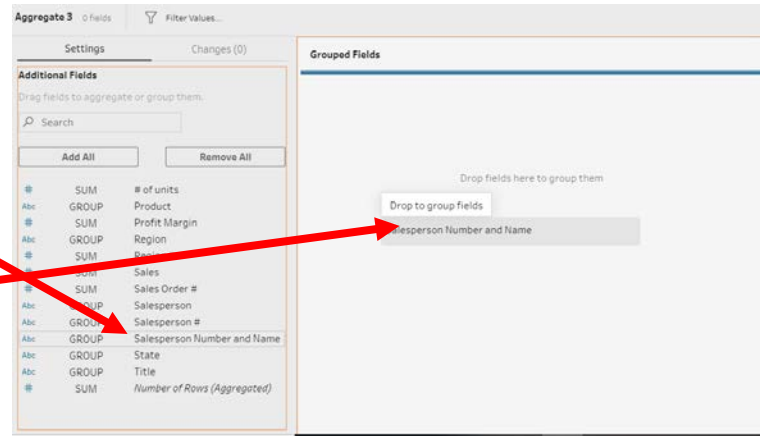
Type "Salesperson # and Name" and press "Enter".

Click on the "+" to the right of "Salesperson # and Name".
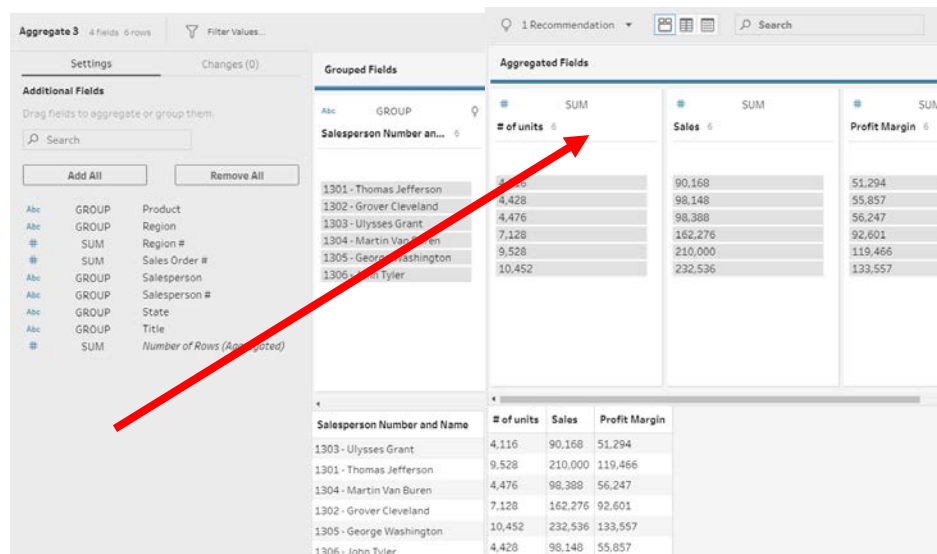
Select "Add Aggregate".

Drag "Salesperson # and Name" field

to "Grouped Fields"

Drag "# of Units", "Sales" and "Profit Margin" field

to "Aggregated Fields".

Step 7: Output the Report by Salesperson

Click on the plus sign of our Aggregate tile and select "Add Output".

1. "Save output to", Select "File"

2. "Output type", Select "Microsoft Excel (.xlsx)"

3. Select "Browse"

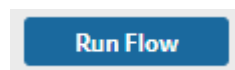3.1 Select the original data file and select "Accept".
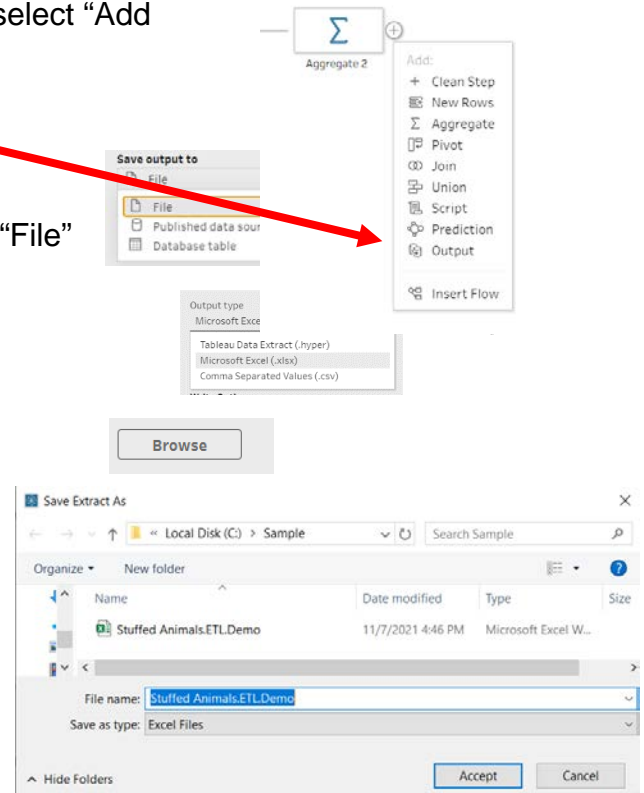
3.2 Confirm that you are replacing the original file.

4. "Worksheet", enter "Report By Salesperson" and confirm by selecting "Create new worksheet "Report by Salesperson".

5. "Write Options" select "Create table" .

6. Select "Run   Run Flow   Flow".

Finished Running Flow

✓

General Ledger.csv

Total time 00:00

Done

7. Select "Done".

Outputs with the same names already exist. Do you want to replace them?

Running the flow will replace these existing outputs with new outputs.

Output Step	Location	Name
Output 2	E:\Sample\Stuffed Animals.ETL.De...	Report by Salesperson

Cancel   **Replace**

8. Select Replace.

Open the original data file in Excel.

	A	B	C	D	E	F
1	Salespersc	# of units	Sales	Profit Margin		
2	1303 - Uly:	4116	90168	51294		
3	1301 - Tho	9528	210000	119466		
4	1304 - Mar	4476	98388	56247		
5	1302 - Gro	7128	162276	92601		
6	1305 - Gec	10452	232536	133557		
7	1306 - Joh	4428	98148	55857		
8						
9						
10						
11						

◄ ► ... | Region | Jan Report by Product | **Report by Salesperson** | ⊕

`

Make format changes and to prevent the changes inadvertently lost when other extracts are preformed change the name of the tab.

	A	B	C	D	E
1	Salesperson Number and Nam	# of units	Sales	Profit Margin	
2	1301 - Thomas Jefferson	9,528.00	$210,000.00	$119,466.00	
3	1302 - Grover Cleveland	7,128.00	$162,276.00	$92,601.00	
4	1303 - Ulysses Grant	4,116.00	$90,168.00	$51,294.00	
5	1304 - Martin Van Buren	4,476.00	$98,388.00	$56,247.00	
6	1305 - George Washington	10,452.00	$232,536.00	$133,557.00	
7	1306 - John Tyler	4,428.00	$98,148.00	$55,857.00	
8	Total	40,128.00	$891,516.00	$509,022.00	
9					
10					
11					
12					

◄ ► ... | Region | Jan Report by Product | **Jan Report by Salesperson** | ⊕

Double click on the tab "By Salesperson", replace it with "Jan by Salesperson" and hit the Enter key.
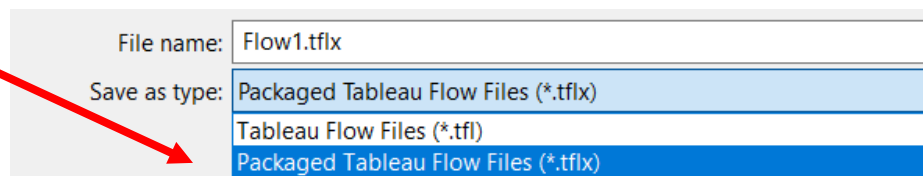
Save, close the Excel file and return to Tableau.

Save the Tableau flow and data

Select "File".

Select "Save As".
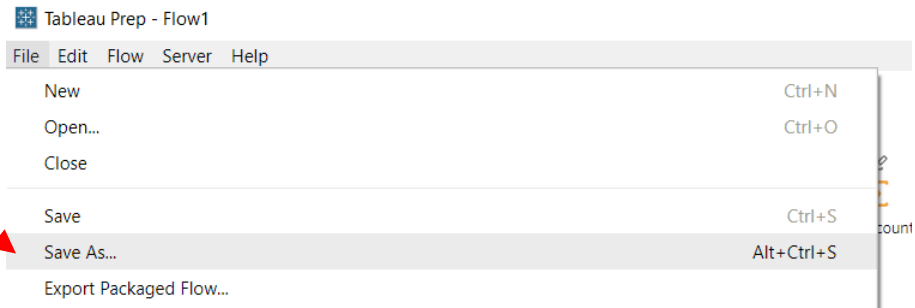
Change the type to "Packaged Tableau Flow Files (*.tflx)".

Select "Save".

Save the file as a normal Tableau Prep file," *.tfl" and close Tableau.