Part 1: The First Stream ----------1

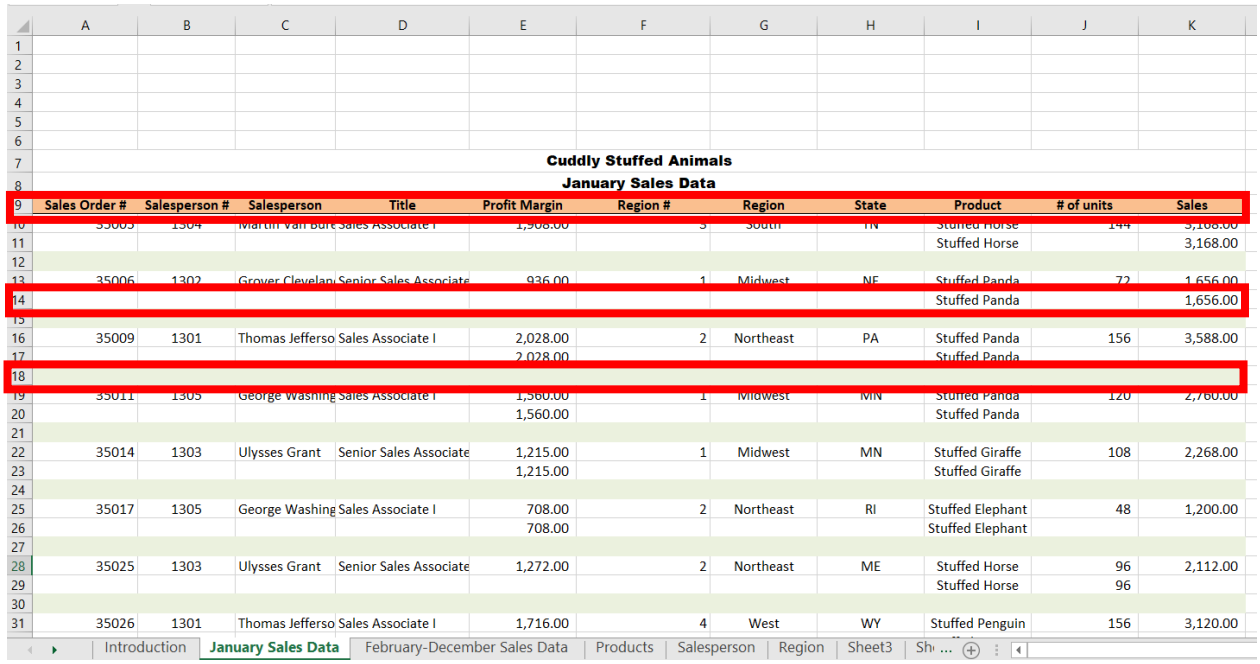
Part 2: Append the Second Stream----------25

Mark Friedman
University of Miami

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.



Cuddly Stuffed Animals
January Sales Data

Sales Order #	Salesperson #	Salesperson	Title	Profit Margin	Region #	Region	State	Product	# of units	Sales
								Stuffed Horse		3,168.00
35006	1302	Grover Clevelan	Senior Sales Associate	936.00	1	Midwest	NE	Stuffed Panda	72	1,656.00
								Stuffed Panda		1,656.00
35009	1301	Thomas Jefferso	Sales Associate I	2,028.00	2	Northeast	PA	Stuffed Panda	156	3,588.00
				2,028.00				Stuffed Panda		
35011	1305	George Washing	Sales Associate I	1,560.00	1	Midwest	MN	Stuffed Panda	120	2,760.00
				1,560.00				Stuffed Panda		
35014	1303	Ulysses Grant	Senior Sales Associate	1,215.00	1	Midwest	MN	Stuffed Giraffe	108	2,268.00
				1,215.00				Stuffed Giraffe		
35017	1305	George Washing	Sales Associate I	708.00	2	Northeast	RI	Stuffed Elephant	48	1,200.00
				708.00				Stuffed Elephant		
35025	1303	Ulysses Grant	Senior Sales Associate	1,272.00	2	Northeast	ME	Stuffed Horse	96	2,112.00
								Stuffed Horse	96	
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

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

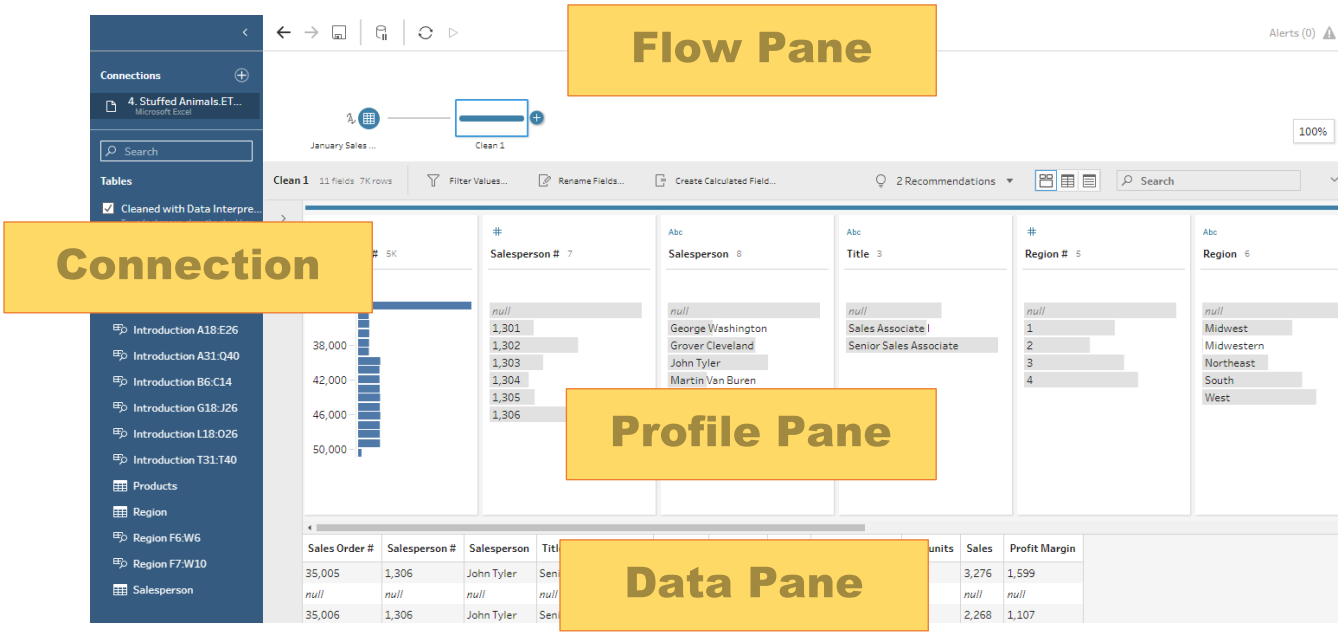
Salesperson	# of units	Sales	Profit Margin
1405 - James Polk	1,812	$41,568	$24,438
1425 - John Tyler	2,172	48,336	28,110
1541 - Ulysses Grant	756	17,112	10,062
1576 - James Madison	1,776	39,216	23,160
1588 - George Washington	1,584	35,784	21,168
1776 - James Monroe	2,016	45,192	26,700
Total	10,116	$227,208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

Mark Friedman
University of Miami

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.

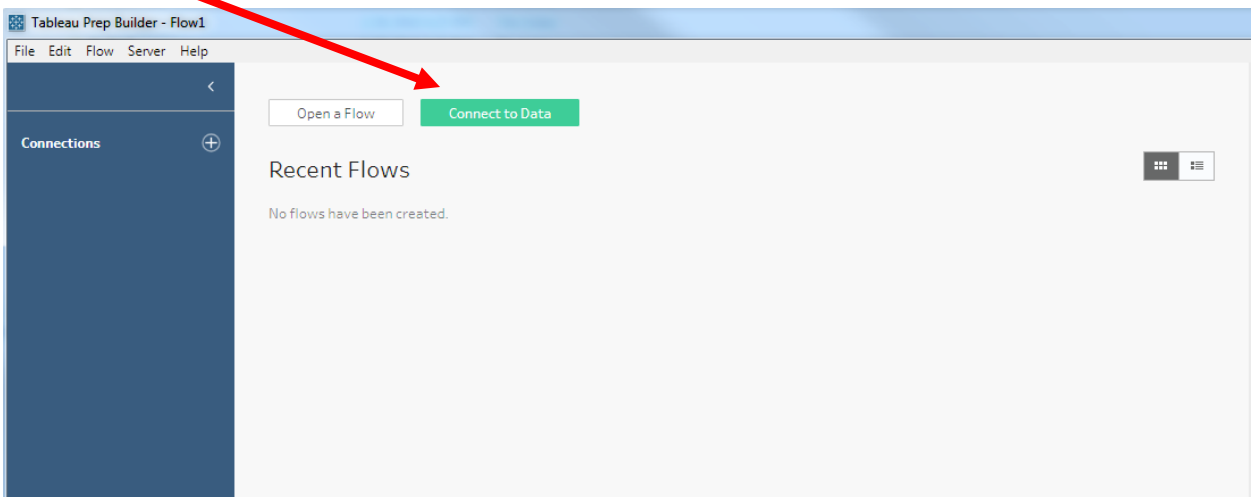


After you connect to your data and begin building your flow, you add steps in the Flow pane. These steps function as a lens into the structure of your data, as well as a summary of operations that is applied to your data. Each step represents a different category of operations that you define.

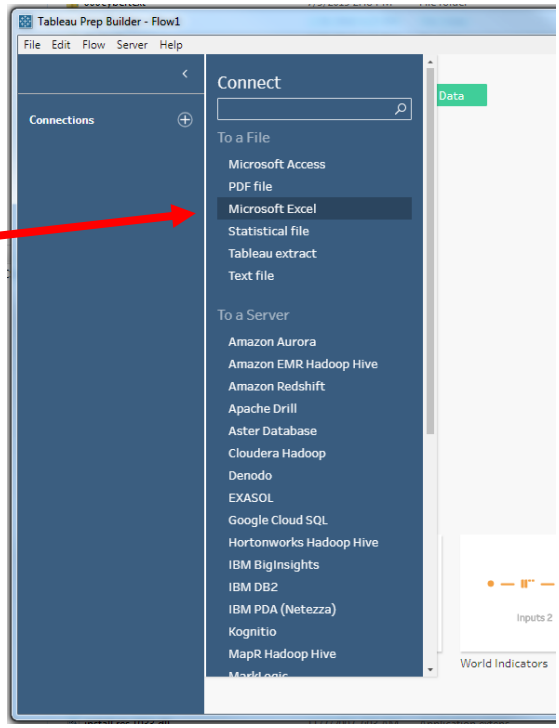
Mark Friedman
University of Miami

Step 1: Input Excel file

To open Excel file via Tableau Prep, press green button “Connect to Data” on the Flow pane.



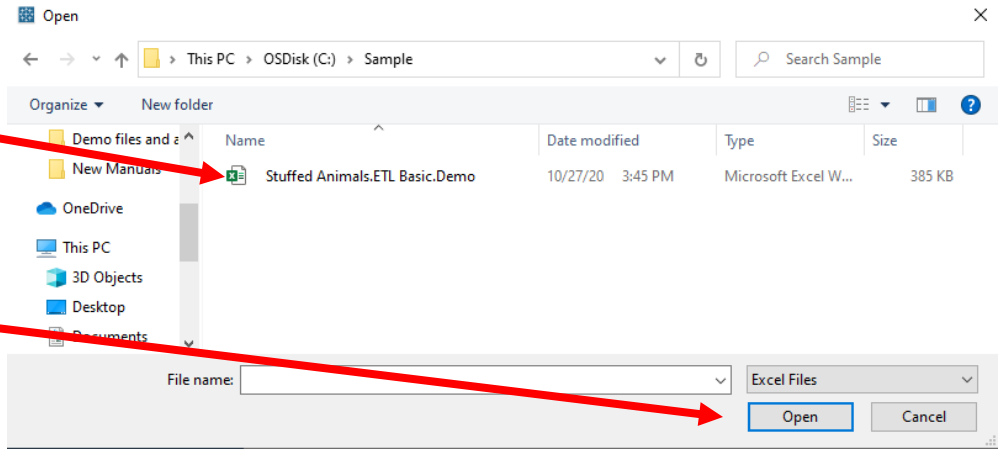
On the connections pane select Microsoft Excel.



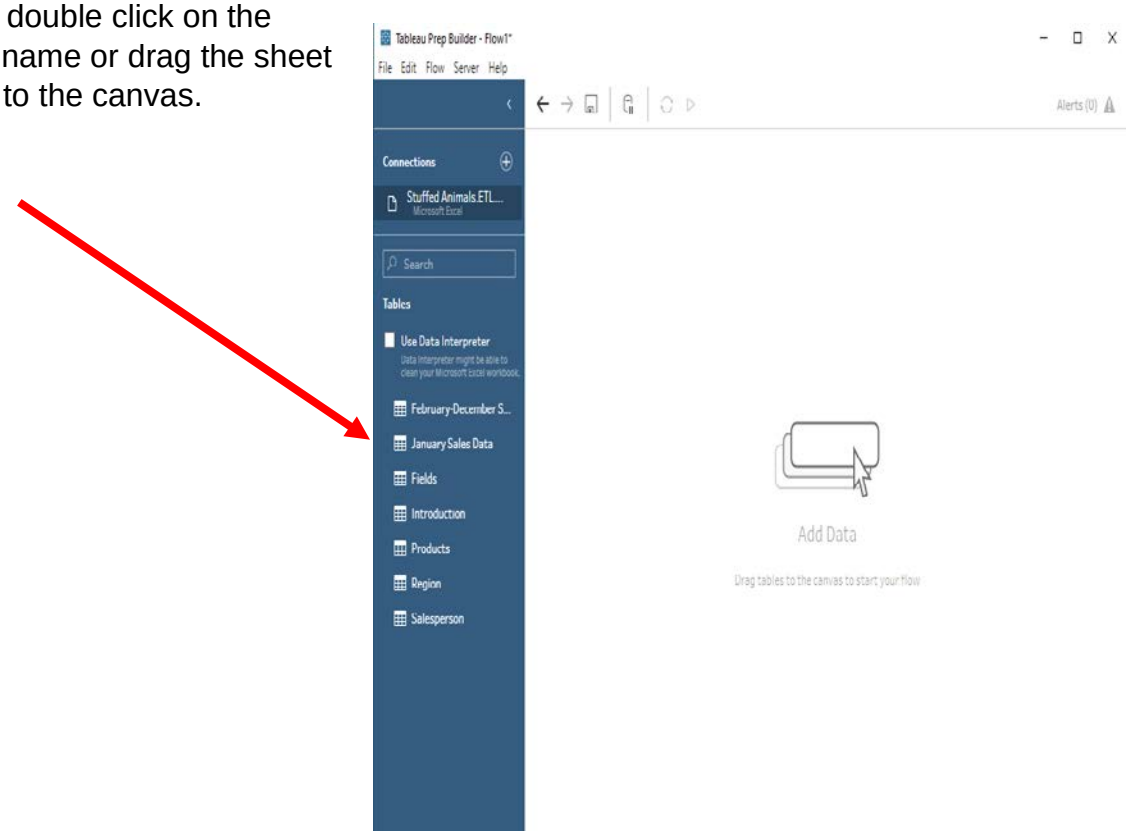
Mark Friedman
University of Miami

Select the file.

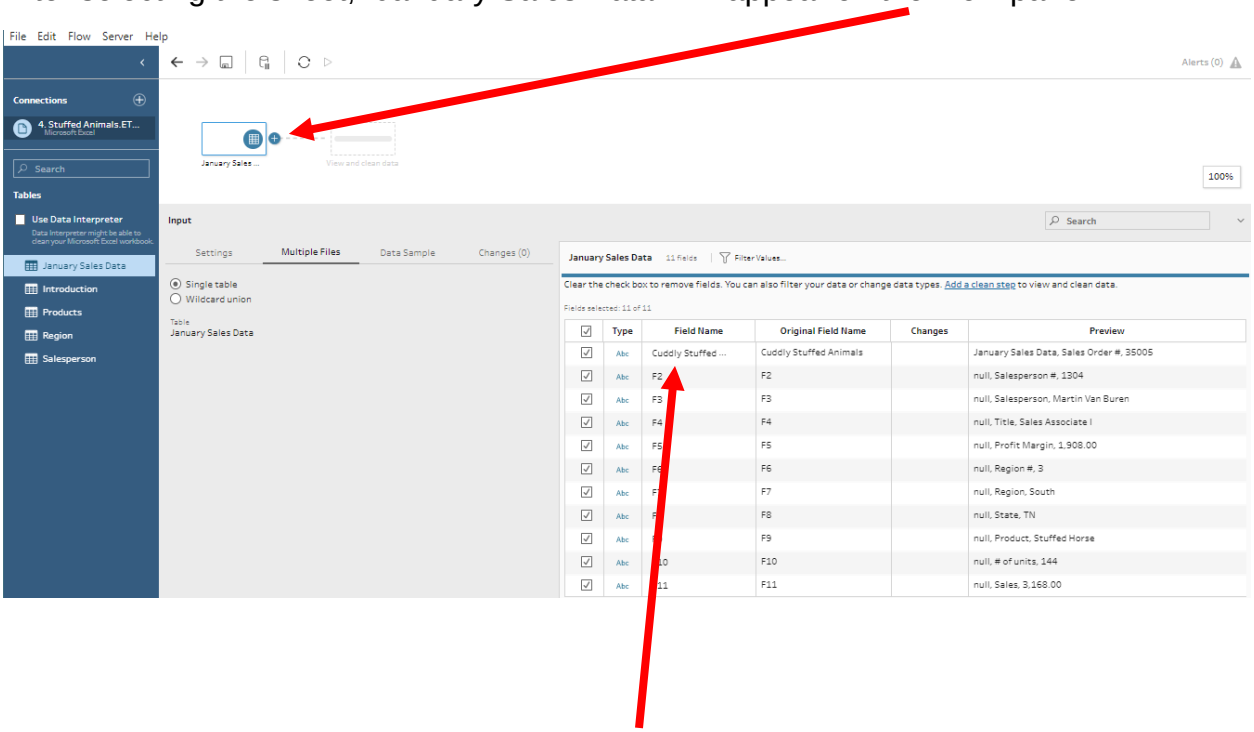
Select “Open”.



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.



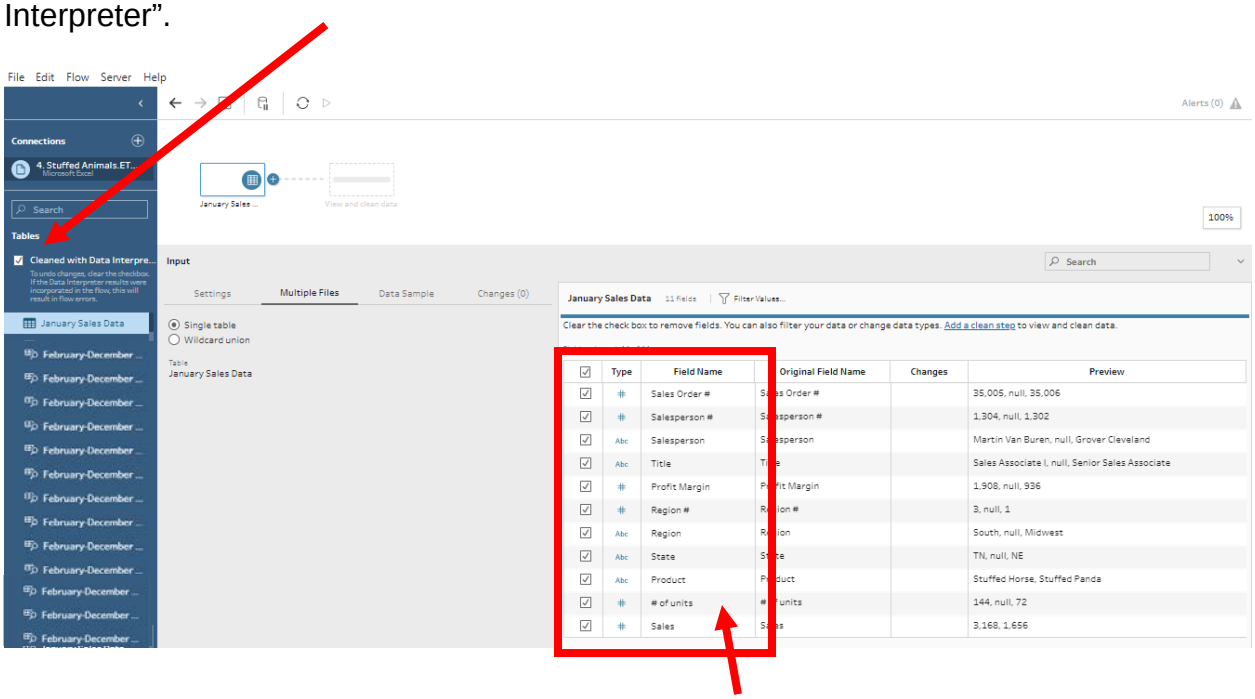
Sample data will be visible on the Profile pane. Notice that the Field Names are not correct.

Mark Friedman
University of Miami

Step 2: Cleaning the data

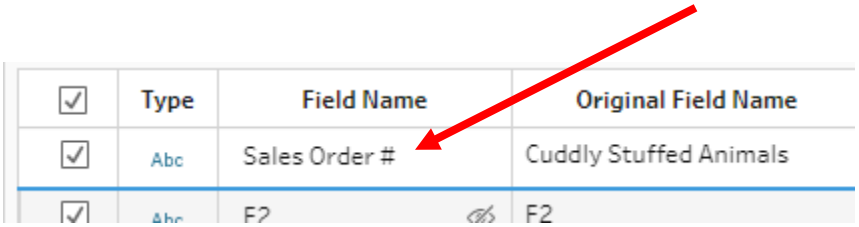
The field names are in the ninth row.

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



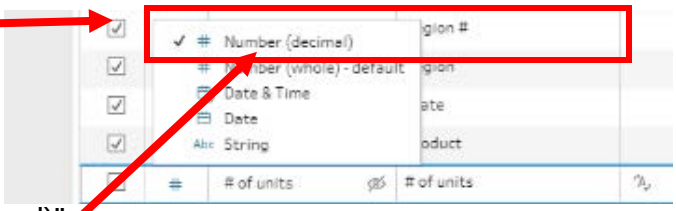
If it works, the field names will be dispalyed correctly.

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

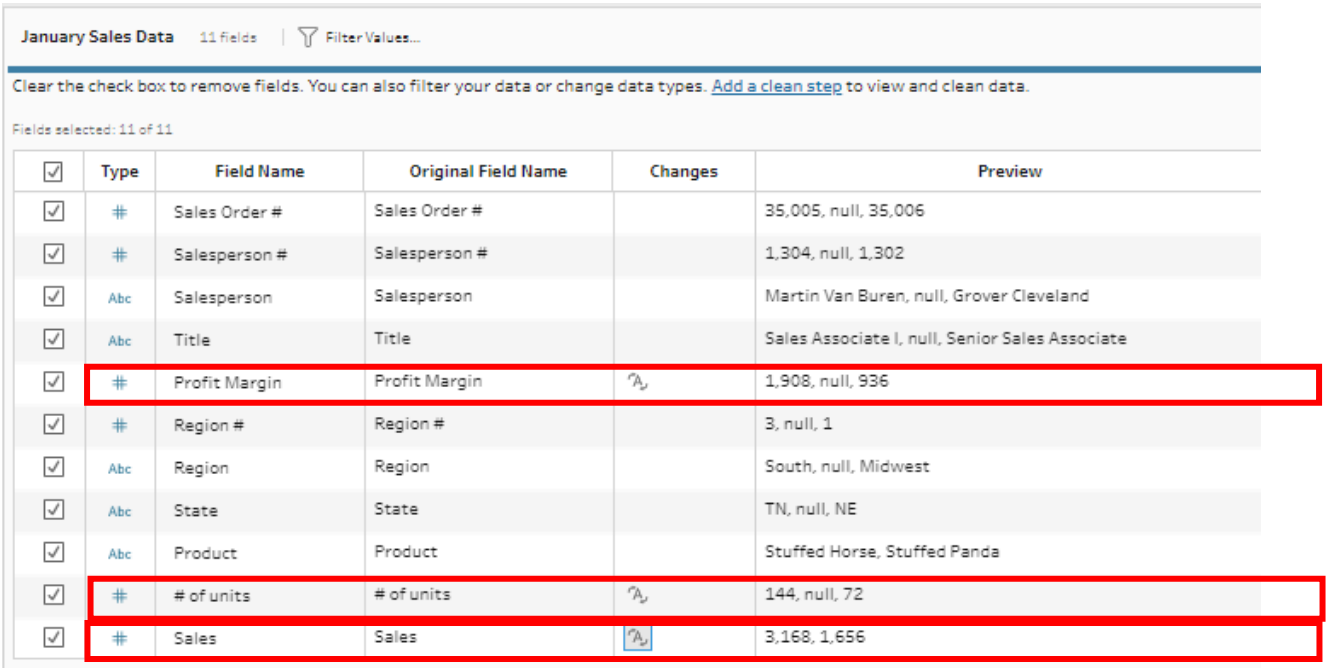


All number fields that will be added, subtracted, multiplied, divided must be coded as "Number (decimal)".

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

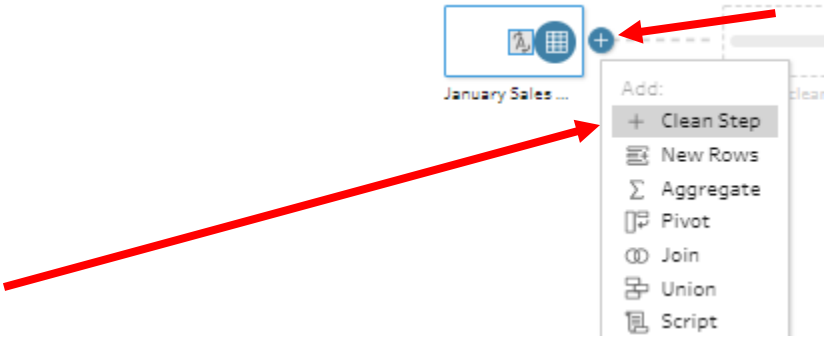


- "Profit Margin": "Number (decimal)"
- "# of Units": "Number (decimal)"
- "Sales": "Number (decimal)"

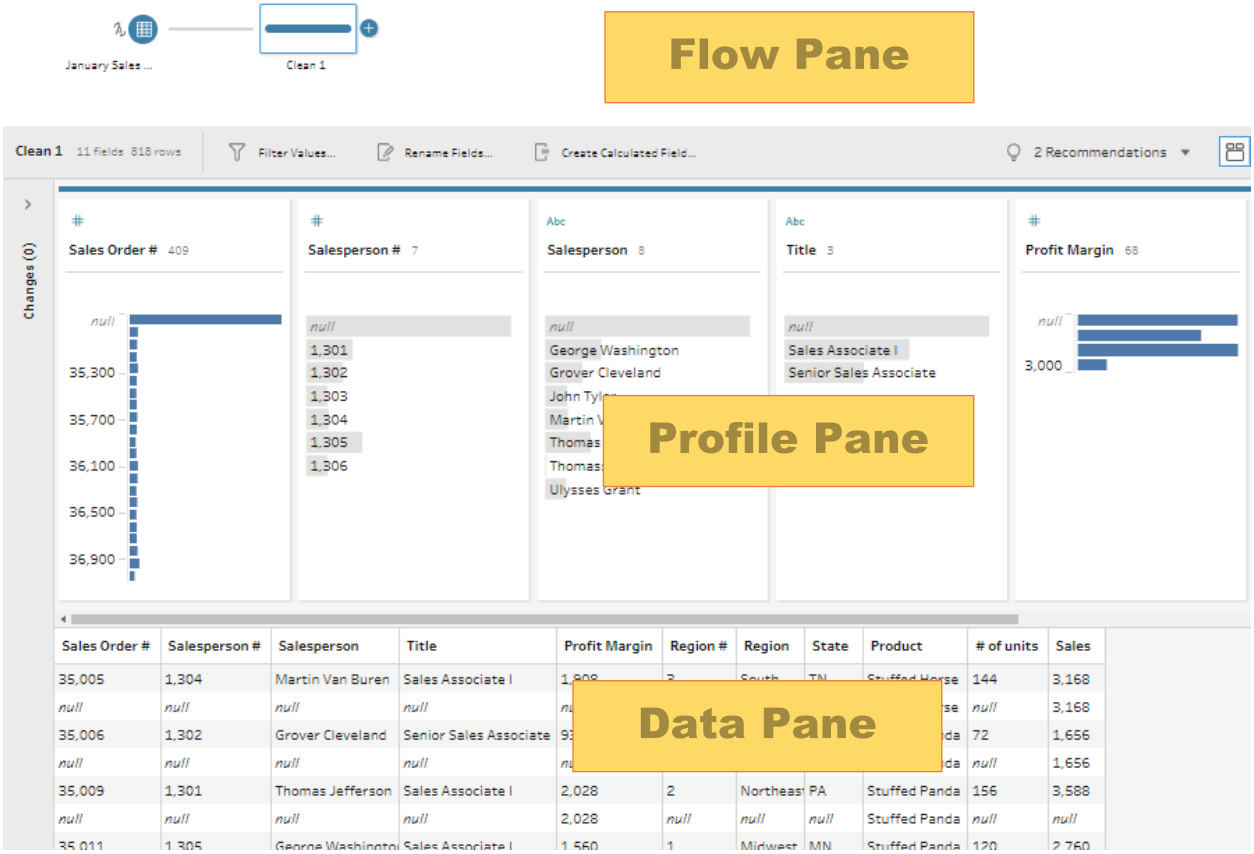


Type	Field Name	Original Field Name	Changes	Preview
#	Sales Order #	Sales Order #		35,005, null, 35,006
#	Salesperson #	Salesperson #		1,304, null, 1,302
Abc	Salesperson	Salesperson		Martin Van Buren, null, Grover Cleveland
Abc	Title	Title		Sales Associate I, null, Senior Sales Associate
#	Profit Margin	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.

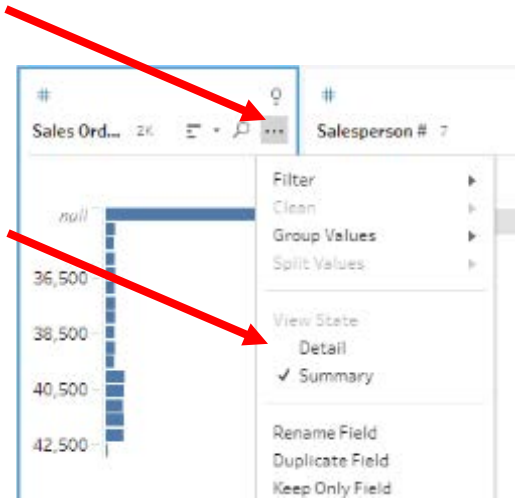


Select “+ Clean Step”.



Mark Friedman
University of Miami

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”.



Mark Friedman
University of Miami

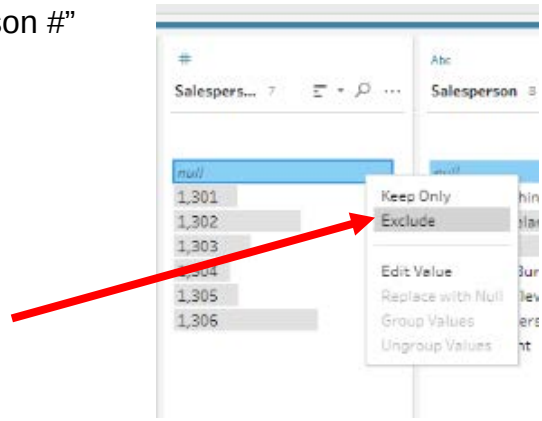
Exclude all rows that do not contain data

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 #”.

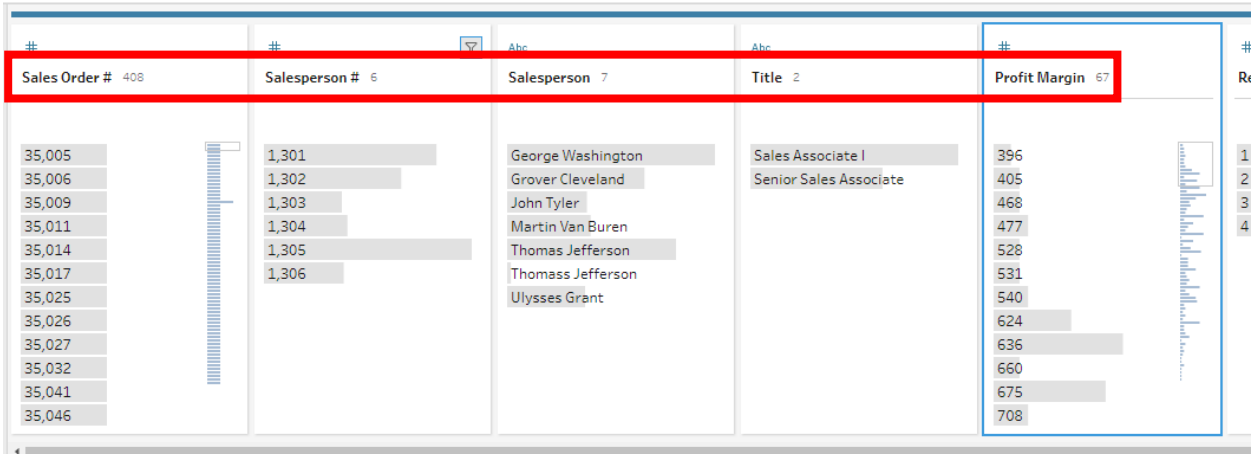
Sales Order #	Salesperson #	Salesperson	Title	Profit Margin	Region #	Region	State	Product	# of units	Sales
35,005	1,304	Martin Van Buren	Sales Associate I	1,908	3	South	TN	Stuffed Horse	144	3,168
null	*null*	*null*	*null*	*null*	*null*	*null*	*null*	Stuffed Horse	*null*	3,168
35,006	1,302	Grover Cleveland	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	Northeas	PA	Stuffed Panda	156	3,588
null	*null*	*null*	*null*	2,028	*null*	*null*	*null*	Stuffed Panda	*null*	*null*
35,011	1,305	George Washingto	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 Washingto	Sales Associate I	708	2	Northeas	RI	Stuffed Elepha	48	1,200
null	*null*	*null*	*null*	708	*null*	*null*	*null*	Stuffed Elepha	*null*	*null*
35,025	1,303	Ulysses Grant	Senior Sales Associate	1,272	2	Northeas	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 Penguii	156	3,120
null	*null*	*null*	*null*	*null*	*null*	*null*	*null*	Stuffed Penguii	*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”.

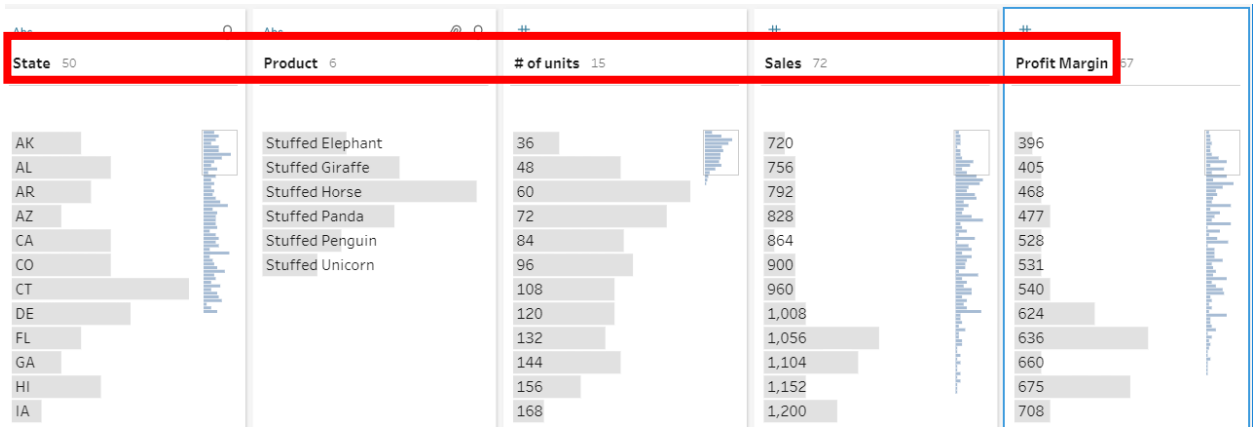


Reorder the fields



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.

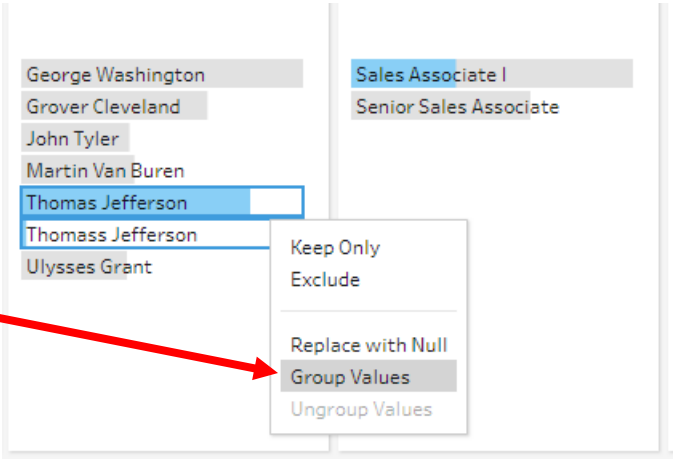
Product ID	Product
101	Stuffed Penguin
102	Stuffed Horse
103	Stuffed Unicorn
104	Stuffed Giraffe
105	Stuffed Panda
106	Stuffed Elephant

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

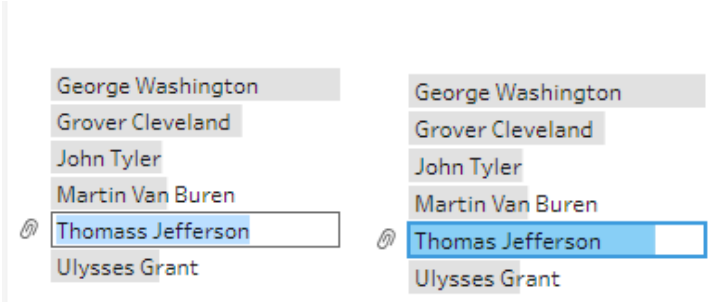
Code	Region
1	Midwest
2	Northeast
3	South
4	West

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”.



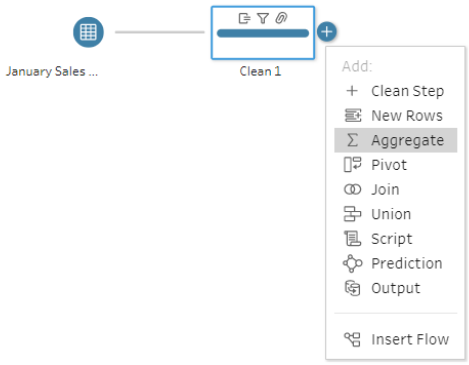
If the system groups with the incorrect name, type the correct name and hit the enter key.



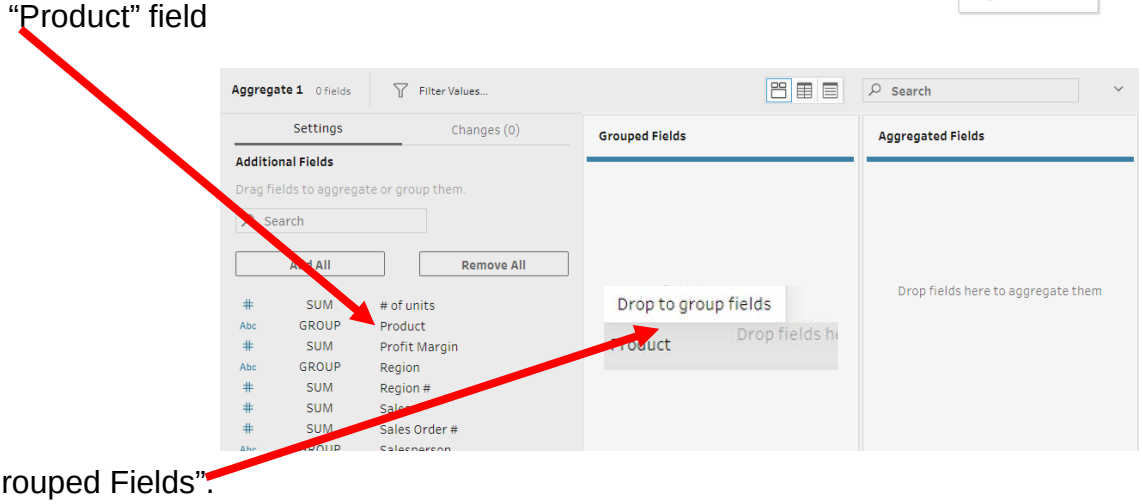
Check and correct the “Product” and “Region” Fields.

Step 3: Aggregate Balances

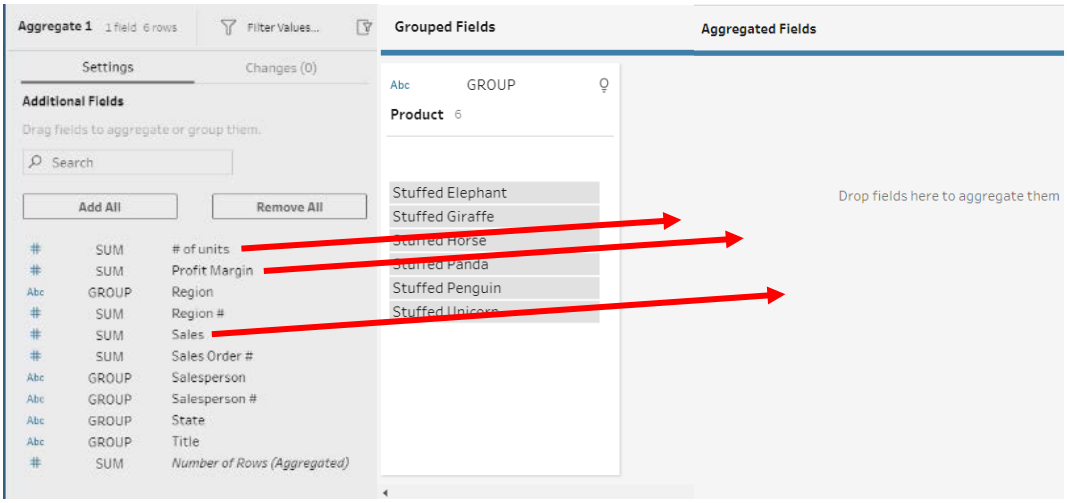
To accumulate the values by product, Click on the "+" to the right of "Clean 1" and select "Aggregate".



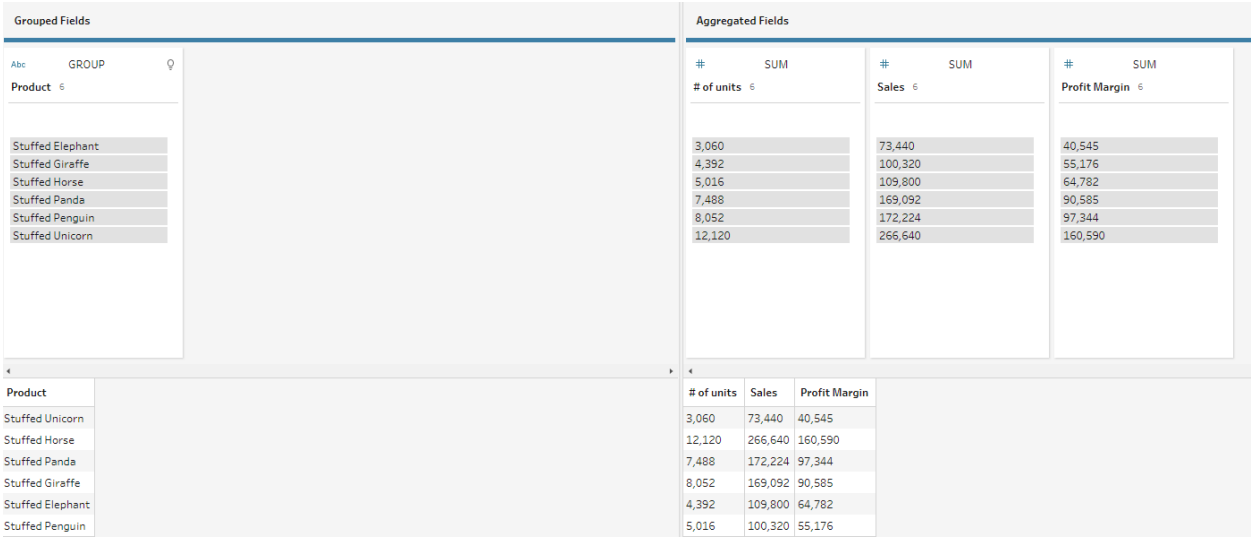
Drag "Product" field



to "Grouped Fields".



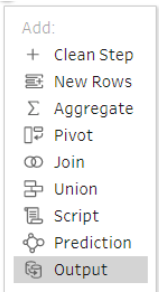
Drag "# of units", "Sales" and Profit Margin" to "Aggregated Fields".

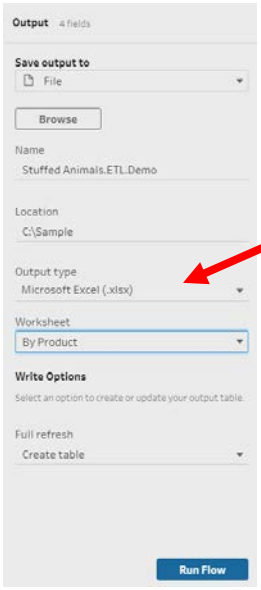


Product	# of units	Sales	Profit Margin
Stuffed Unicorn	3,060	73,440	40,545
Stuffed Horse	12,120	266,640	160,590
Stuffed Panda	7,488	172,224	97,344
Stuffed Giraffe	8,052	169,092	90,585
Stuffed Elephant	4,392	109,800	64,782
Stuffed Penguin	5,016	100,320	55,176

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

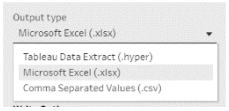
Click on the plus sign of the “Aggregate 1” tile and select “Output”.





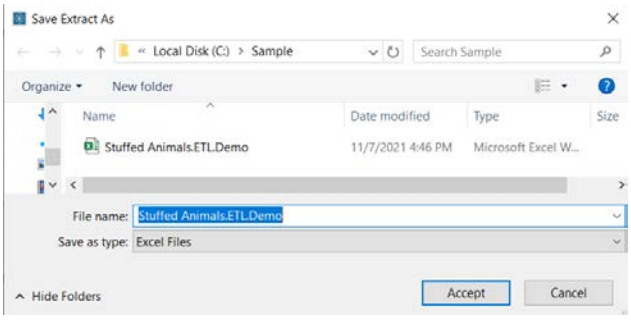
1. “Save output to”, Select “File”

2. “Output type”, Select “Microsoft Excel (.xlsx)”

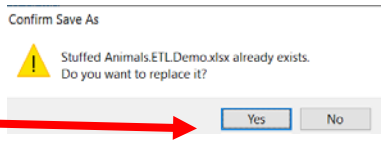


3. Select “Browse”

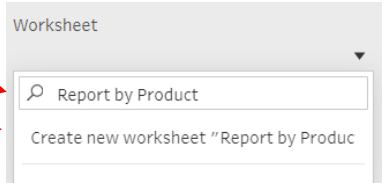
3.1 Select the original data file and select “Accept”.



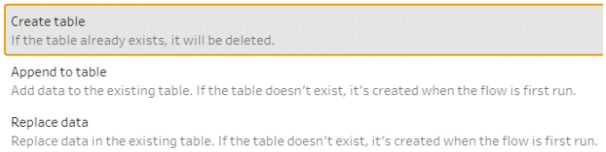
3.2 Confirm that you are replacing the original file.



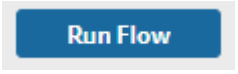
4. “Worksheet”, enter “Report by Product” and confirm by selecting “Create new worksheet “Report by Product”.



5. “Write Options” select “Create table” .

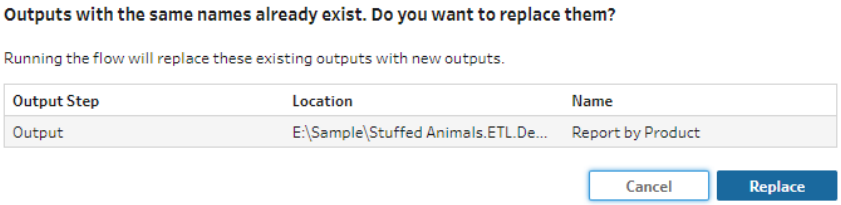


6. Select “Run Flow”.

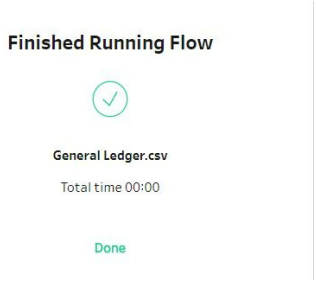


Mark Friedman
University of Miami

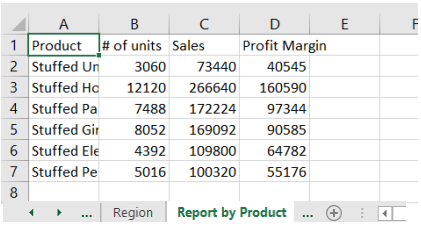
If the program asks to confirm select Replace



Select “Done”.



Open the original data file in Excel.

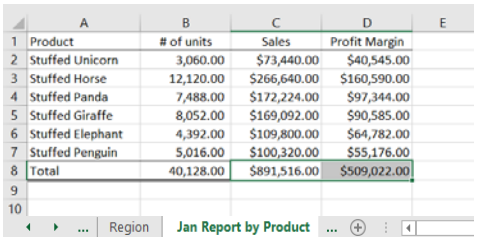


	A	B	C	D	E
1	Product	# of units	Sales	Profit Margin	
2	Stuffed Un	3060	73440	40545	
3	Stuffed Ho	12120	266640	160590	
4	Stuffed Pa	7488	172224	97344	
5	Stuffed Gir	8052	169092	90585	
6	Stuffed Ele	4392	109800	64782	
7	Stuffed Pe	5016	100320	55176	
8					

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.

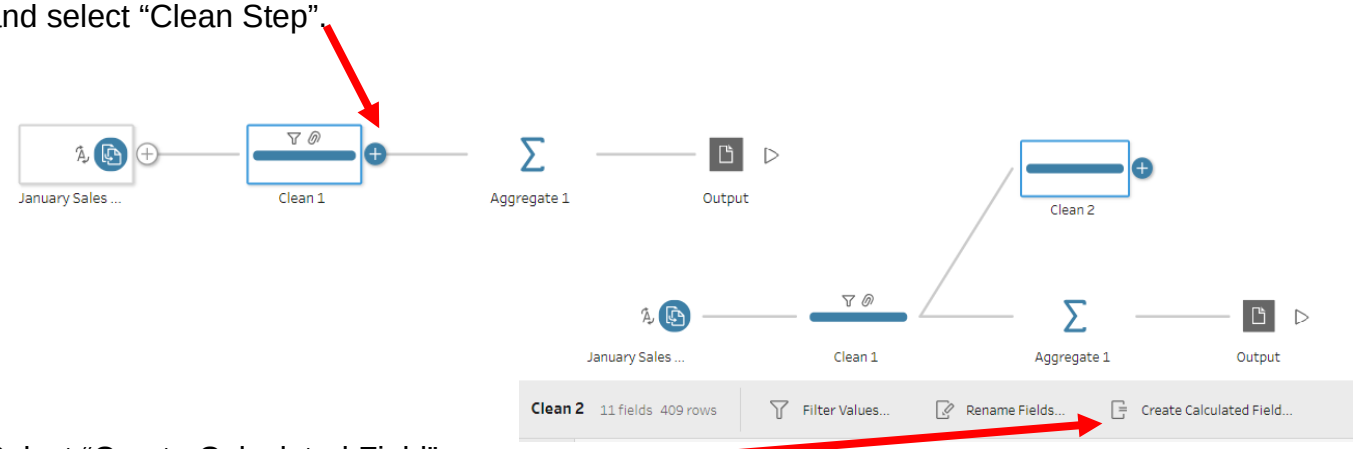
Save, close the Excel file, and return to Tableau.

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".

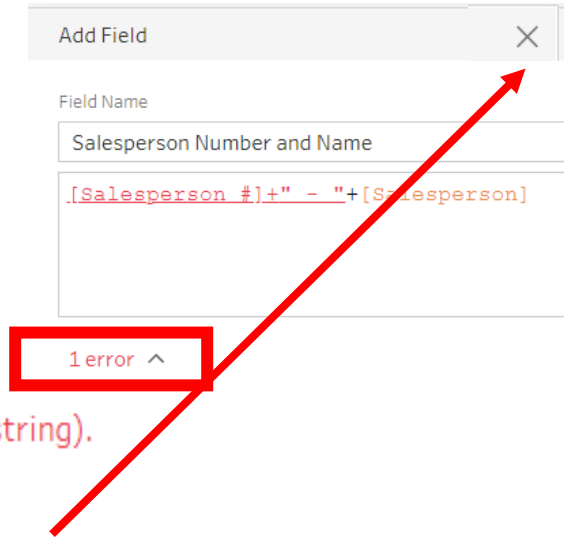


Select "Create Calculated Field".

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

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

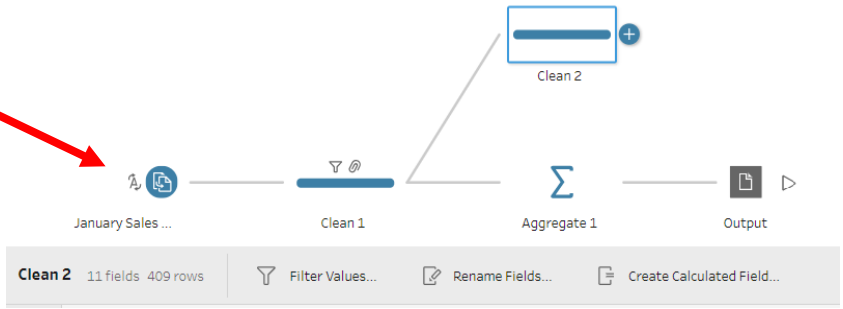
There is an error because "Salesperson #" is a number and "Salesperson" is a string.



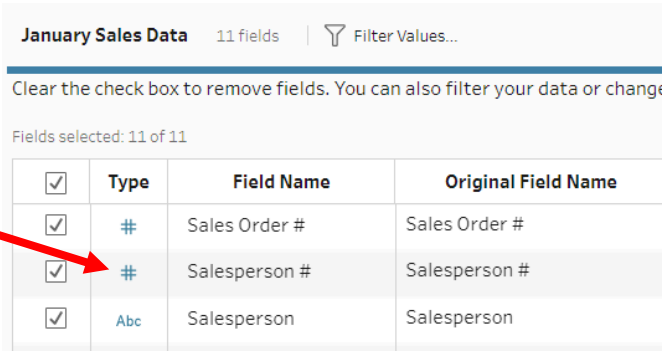
Select the "X" to close the window.

Change the field “Salesperson #” from a numeric to a string.

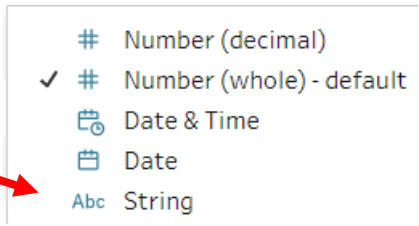
Select the first step, “January Sales”.



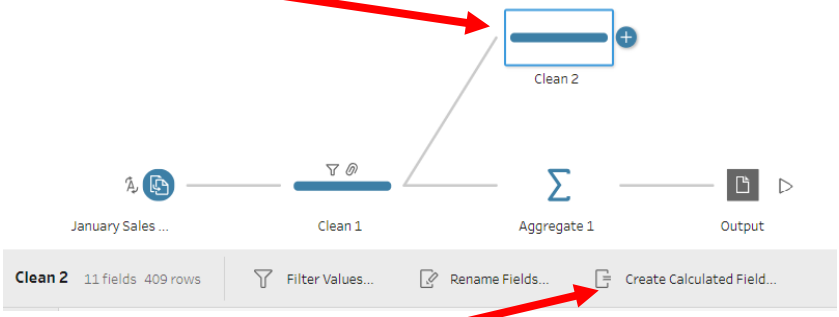
Select the “Type” field for “Salesperson #”.



Select “String”



Select the step Clean 2.

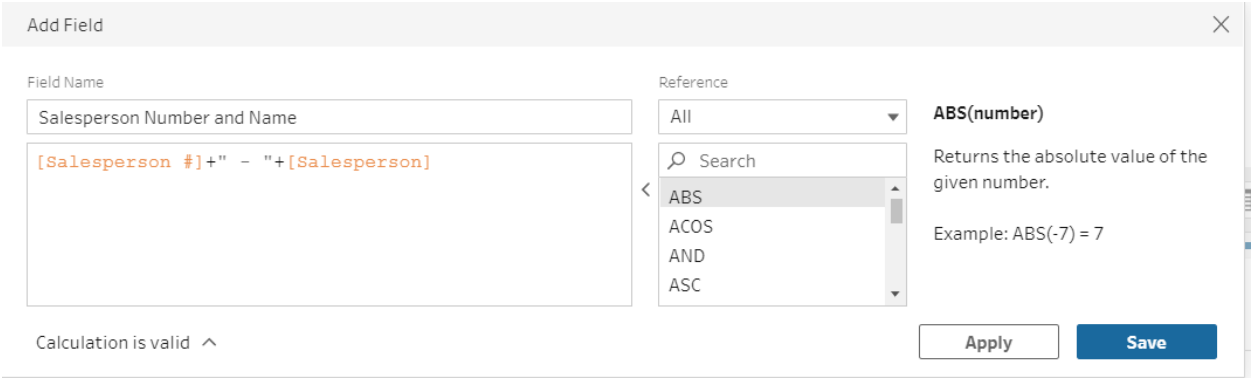


Select “Create Calculated Field”.

Enter “Salesperson Number and Name” as the “Field Name”.

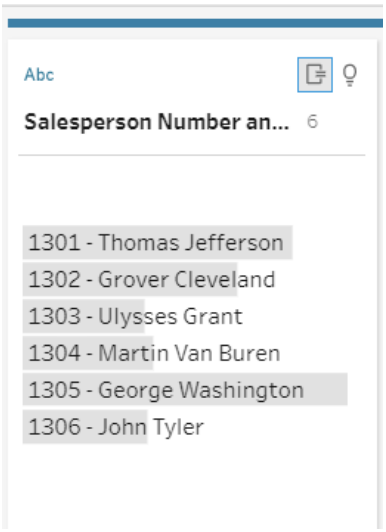
Enter the formula: [Salesperson #]+" - "+[Salesperson].

.



Select “Save”.

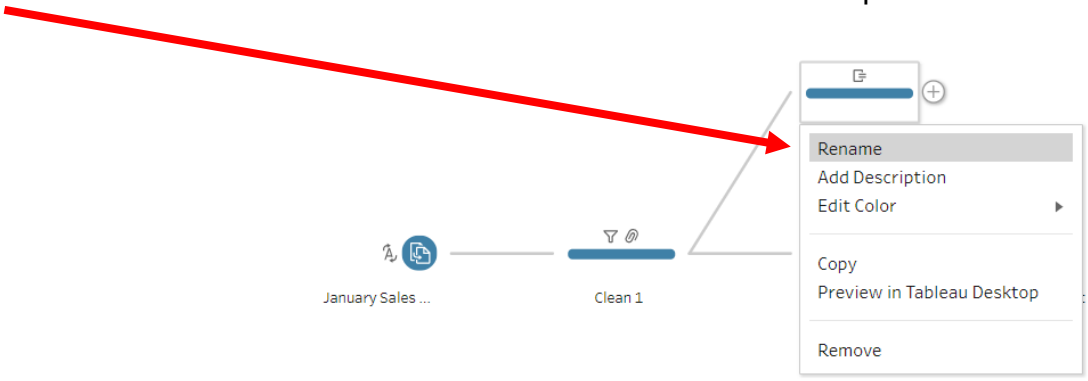
A new field has been created.



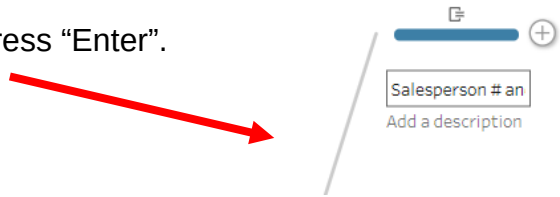
Mark Friedman
University of Miami

Step 6: Aggregate balances and export “Report by Salesperson”

Right Click on the “Clean 2” in the “Flow Pane” and select “Rename Step”.



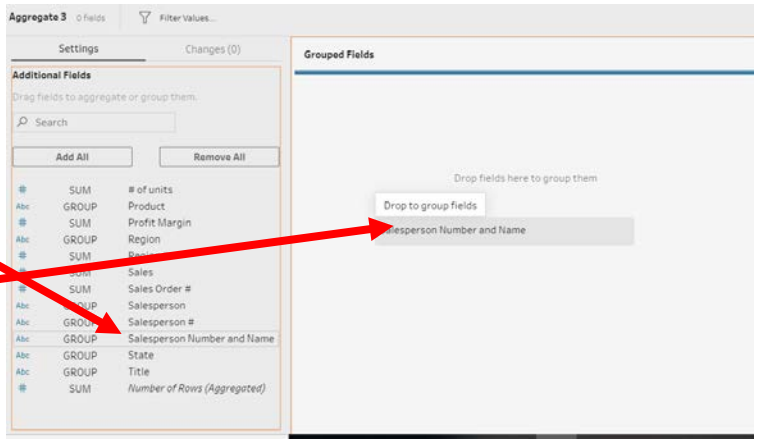
Type “Salesperson # and Name” and press “Enter”.



Click on the “+” to the right of “Salesperson # and Name”.

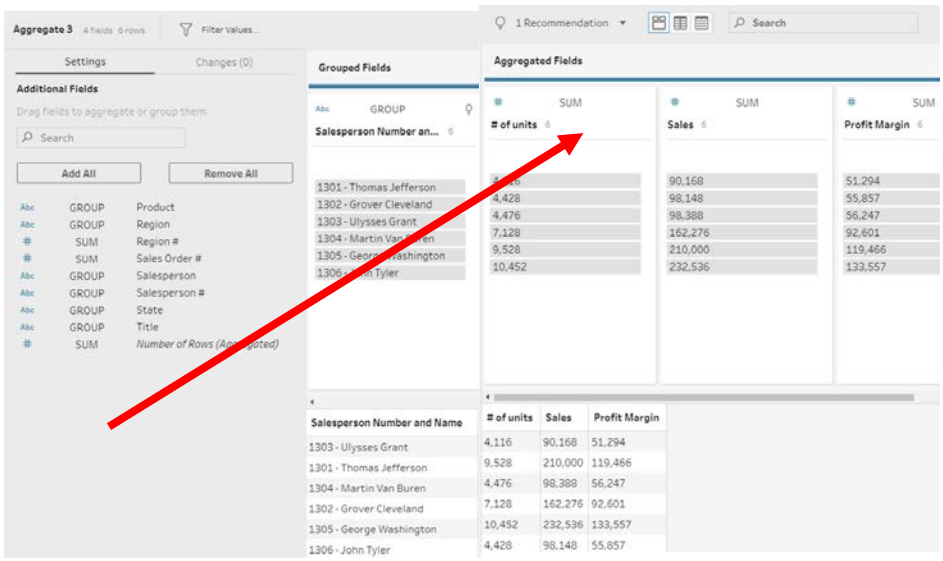
Select “Add Aggregate”.

Drag “Salesperson # and Name” field



to “Grouped Fields”

Drag “# of Units”, “Sales” and “Profit Margin” field

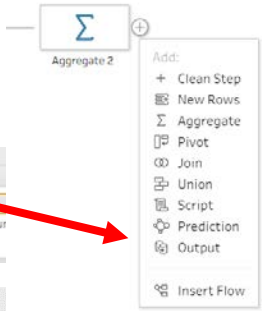


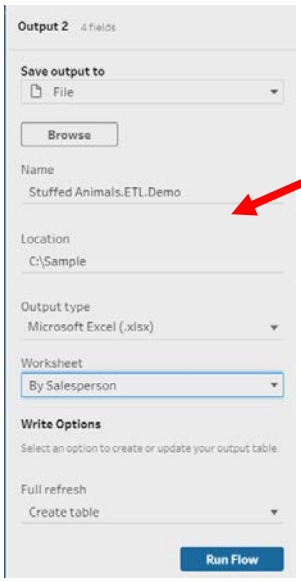
to “Aggregated Fields”.

Mark Friedman
University of Miami

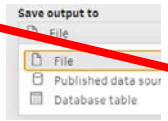
Step 7: Output the Report by Salesperson

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

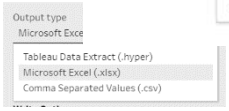




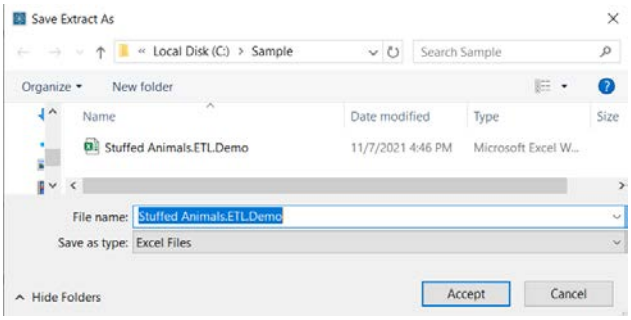
1. “Save output to”, Select “File”



2. “Output type”, Select “Microsoft Excel (.xlsx)”

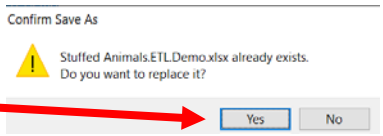


3. Select “Browse”

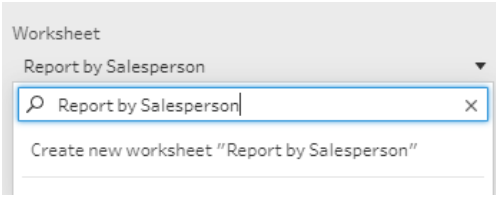


3.1 Select the original data file and select “Accept”.

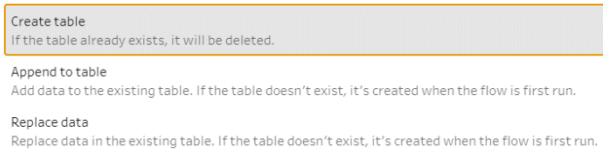
3.2 Confirm that you are replacing the original file.

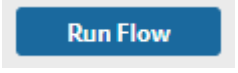


4. “Worksheet”, enter “Report By Salesperson” and confirm by selecting “Create new worksheet “Report by Salesperson”.

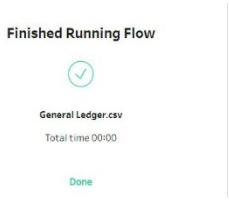


5. “Write Options” select “Create table” .

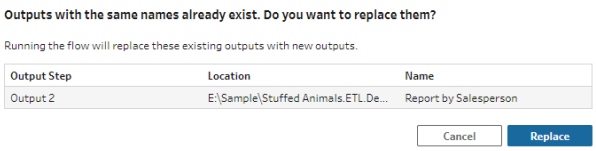


6. Select “Run  Flow”.

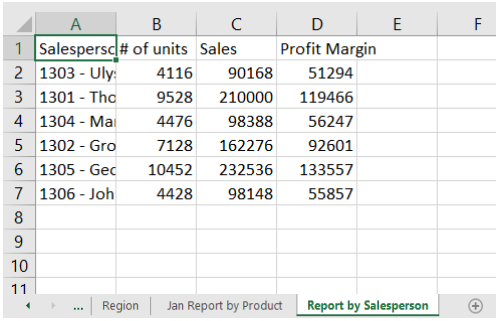
7. Select “Done”.



8. Select Replace.



Open the original data file in Excel.



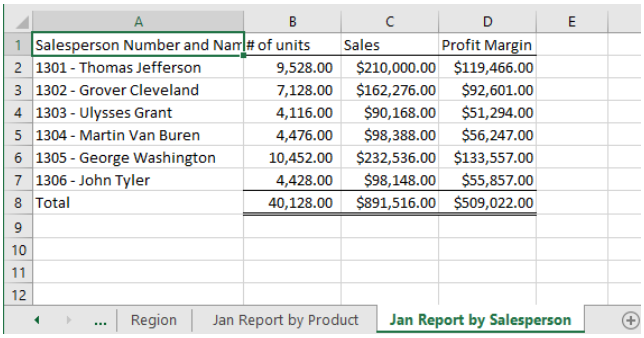
	A	B	C	D	E	F
1	Salespersc	# of units	Sales	Profit Margin		
2	1303 - Uly:	4116	90168	51294		
3	1301 - Tho	9528	210000	119466		
4	1304 - Mai	4476	98388	56247		
5	1302 - Gro	7128	162276	92601		
6	1305 - Gec	10452	232536	133557		
7	1306 - Joh	4428	98148	55857		
8						
9						
10						
11						

... | 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.

Double click on the tab “By Salesperson”, replace it with “Jan by Salesperson” and hit the Enter key.



	A	B	C	D	E
1	Salesperson Number and Nan	# 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					

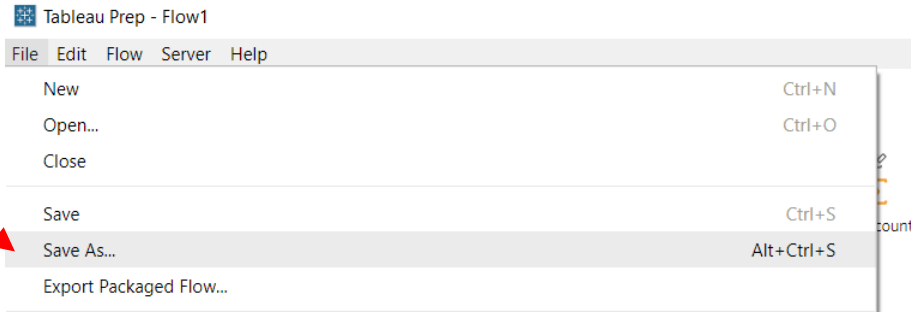
... | Region | Jan Report by Product | Jan Report by Salesperson

Save, close the Excel file and return to Tableau.

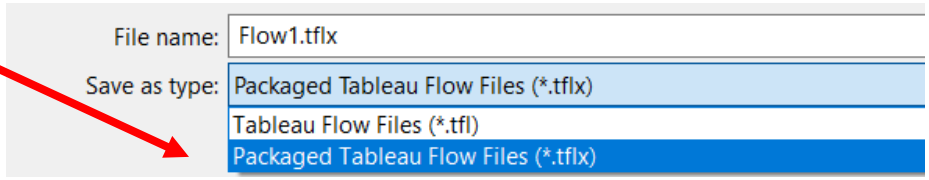
Save the Tableau flow and data

Select “File”.

Select “Save As”.



Change the type to “Packaged Tableau Flow Files (*.tflx)”.



Select “Save”.

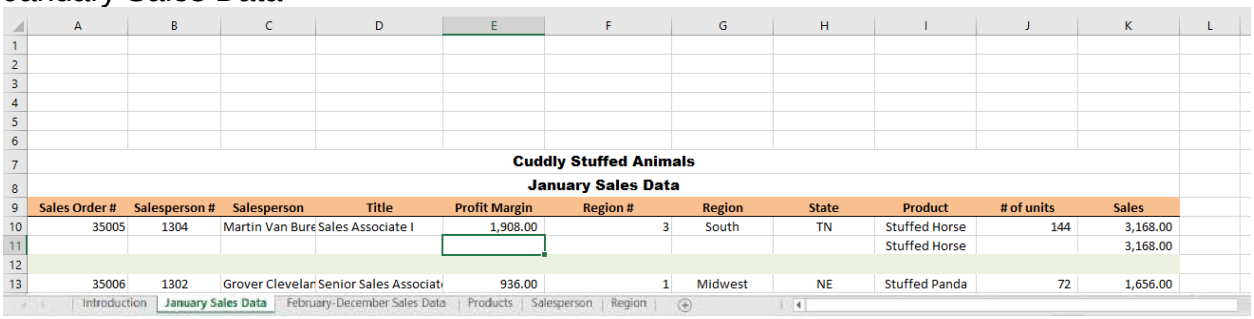
Save the file as a normal Tableau Prep file,” *.tfl” and close Tableau.

Mark Friedman
University of Miami

Part 2

The “February-December Sales Data” tab contains data that was created in a different system than the “January Sales Data”. You cannot append two data sets unless the field names and field types are identical. The fields however, may be ordered differently.

January Sales Data



	A	B	C	D	E	F	G	H	I	J	K	L
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	35005	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 Clevelar	Senior Sales Associat	936.00	1	Midwest	NE	Stuffed Panda	72	1,656.00	

Introduction | January Sales Data | February-December Sales Data | Products | Salesperson | Region

February-December Sales Data

	A	B	C	D	E	F	G	H	I	J	K
1											
2					Cuddly Stuffed Animals						
3					February-December Sales						
4	Order #	SalesPerson	Title	Region Name	Region Code	*State*	Article	# of pieces sold	Selling Price Per Unit	Profit Margin	
5	37060	1303 : Ulysses Grant	Senior Sales Associate	South	3	TX	Stuffed Horse	120	22.00	1,590.00	
6							Stuffed Horse			1,590.00	
7											
8											
9	37064	1301 : Thomas Jefferson	Sales Associate I	Northeast	2	CT	Stuffed Horse	144	22.00	1,908.00	
10							Stuffed Horse				

Introduction | January Sales Data | February-December Sales Data | Products | Salesperson | Region

Ready

As you review the new sheet note the differences:

- The column headings are in row 4.
- The column headings are different.
- The columns are in a different order.
- Column B contains both the Salesperson’s number and name.
- Selling Price Per Units has replaced Sales.

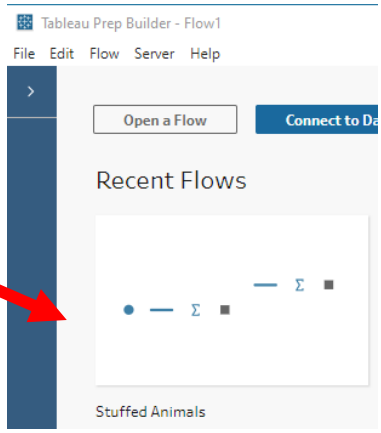
Sample Reports

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

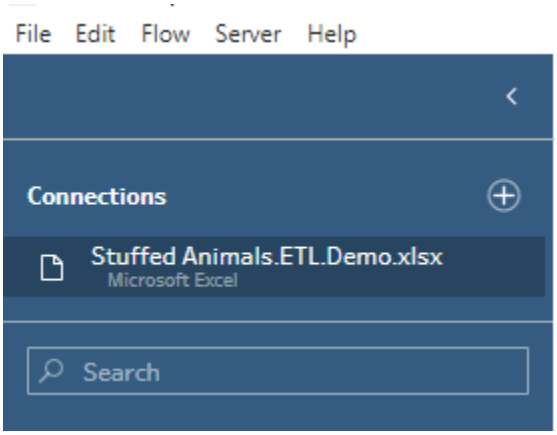
Open the flow that we completed in Part 1.

Select “File”, “Open”
Or Select from “Recent Flows”

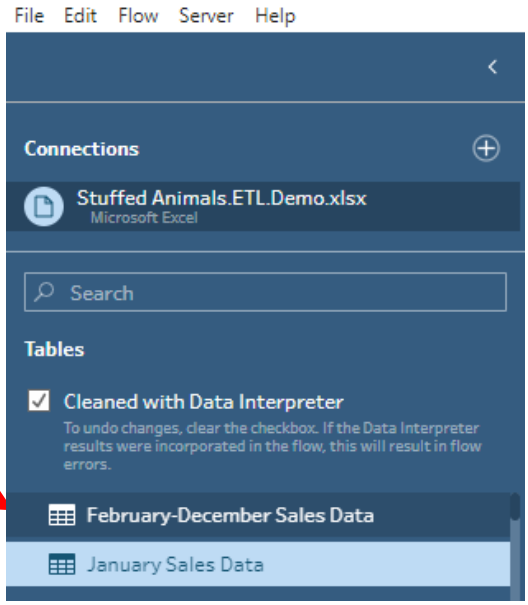


Step 8: Input Excel file
If the list of sheets in the file is not visible, click on the name of the file.

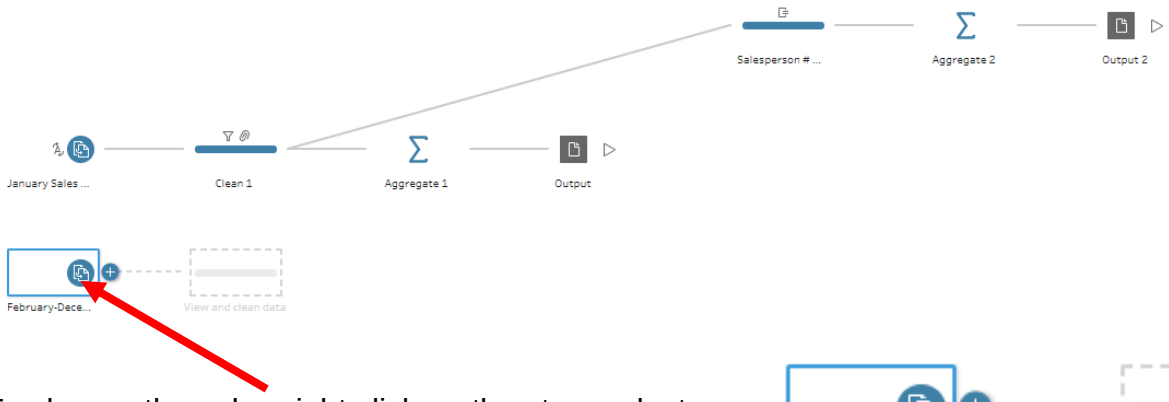
.



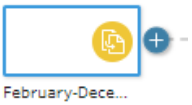
Double click on “February-December Sales Data”.

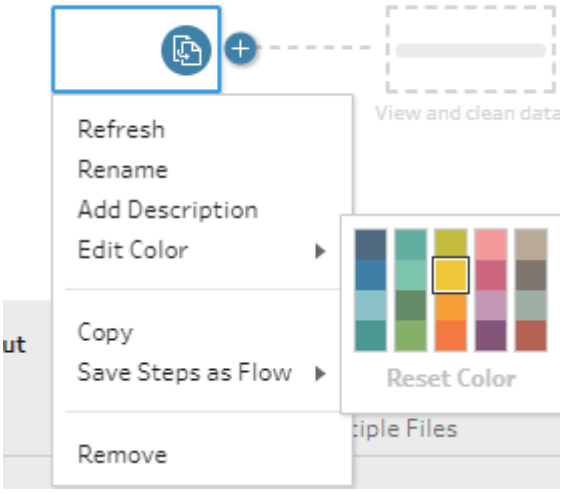


Mark Friedman
University of Miami



To change the color, right click on the step, select "edit Color" and select the "reset Color"





Step 9: Transforming and Cleaning the data

To append the sets of data the "Field Names" and "Types" of "February-December Sales Data" must be the same as those in the data set "January Sales Data". There are ten fields in the "February-December Sales Data" stream, and we will review them one by one making all necessary changes.

January Sales Data

#	Field Name	Sample	Type
1	Sales Order #	35005	Number (whole)
2	Salesperson #	1304	Number (whole)
3	Salesperson	Martin Van Buren	Abc
4	Title	Sales Associate I	Abc
5	Profit Margin	1,908.00	Number (decimal)
6	Region #	3	Number (whole)
7	Region	South	Abc
8	State	TN	Abc
9	Product	Stuffed Horse	Abc
10	# of units	144	Number (decimal)
11	Sales	3,168.00	Number (decimal)

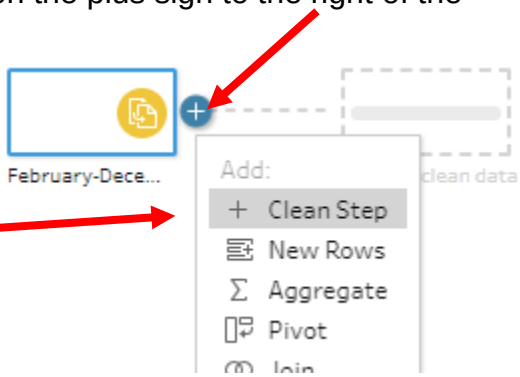
February-December Sales Data

#	Field Name	Sample	Type
1	Sales Order #	37060	Number (whole)
2	SalesPerson	1303 : Ulysses Grant	Abc
3	Title	Senior Sales Associate	Abc
4	Region	South	Abc
5	Region #	3	Number (whole)
6	State	TX	Abc
7	Article	Stuffed Horse	Abc
8	# of units	120	Number (whole)
9	Selling Price Per Unit	22.00	Number (whole)
10	Profit Margin	1,590.00	Number (whole)

1) The field "**Sales Order #"** has no changes

2) The field "**SalesPerson**" contains two fields, "Salesperson #" (1303) and "Salesperson" (Ulysses Grant") separated by, " : ", a space, a colon and another space. The characters that separate two fields are referred to as delimiters. Common delimiters are commas, dashes, and semicolons.

We will convert one field into two fields. Click on the plus sign to the right of the "February-December Sales Data"".

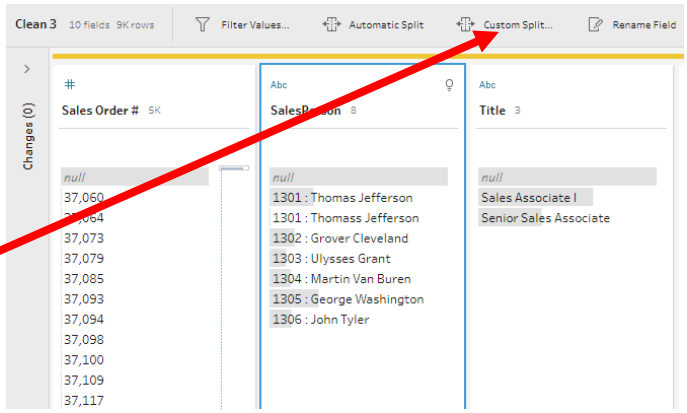


Select "Clean Step".

Select the "SalesPerson" column by clicking anywhere in the column, a blue border will appear around the column.

Mark Friedman
University of Miami

Select “Custom Split”.



Select the “Use the separator” box and type the “ : “, select “2” for the fields which will put the number in one field and the name in the second field. Select “Split”.

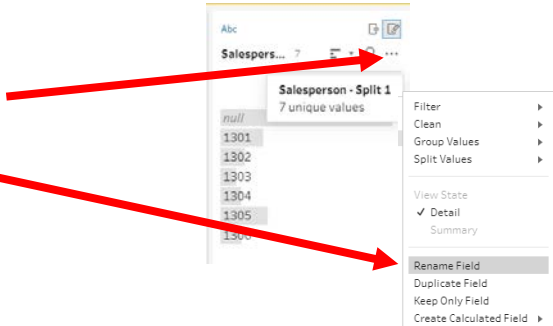
SalesPerson - Split 1	SalesPerson - Split 2	Order #	SalesPerson
1303	Ulysses Grant	37,060	1303 : Ulysses Grant
null	*null*	*null*	*null*
1301	Thomas Jefferson	37,064	1301 : Thomas Jefferson
null	*null*	*null*	*null*
1303	Ulysses Grant	37,073	1303 : Ulysses Grant

There are now three fields, “**SalesPerson - Split 1**” containing the “Salesperson #” “SalesPerson - Split 2” containing the “Salesperson” data and the original field in the second stream,”Salesperson”.

2.1) The field “**SalesPerson – Split 1**” has to be renamed “Salesperson #”. It is a string, so type must be changed to the type that was used in the first stream “Number (whole)”. There are “null” values that must be removed.

Select the field “SalesPerson – Split 1”.

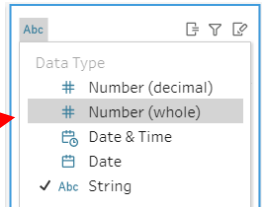
Select the three dots in the upper right.



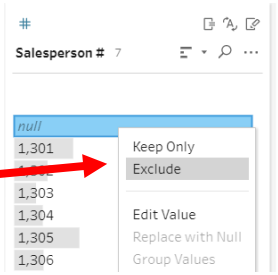
Select “Rename Field”.

Type “Salesperson #” and hit the Enter key.

The type of the field “Salesperson #” must be changed from Abc to Number (whole). Click the “Abc” in the upper left corner and select “Number (decimal)”.



To remove all the rows that do not contain good data right click on “null” and select “Exclude to remove all the extra rows.



2.2) The field “SalesPerson – Split 2” has to be renamed “Salesperson”. It is a string, so type does not have to be changed. Note that there are seven names when there should only be six names. Some names must have been entered incorrectly.

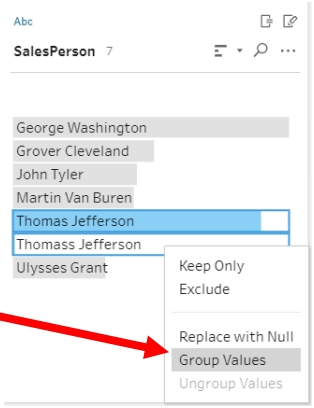
Select the field “SalesPerson – Split 2”.

Select the three dots in the upper right.

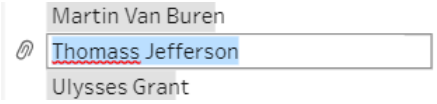
Select “Rename Field”.

Type “Salesperson” and hit the Enter key.

Select Both “Thomas Jefferson” and “Thomass Jefferson”, right click and select “Group Value”.



If it selects the wrong spelling, correct the spelling, and hit the Enter key.

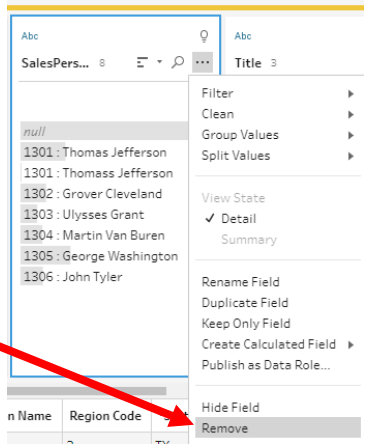


2.3) The field “SalesPerson” must be removed.

Select the field “SalesPerson”.

Select the three dots in the upper right.

Select “Remove”.



3) The field “**Title**” has no changes.

4) The field “**Region**” has no changes.

5) The field “**Region #**” has no changes.

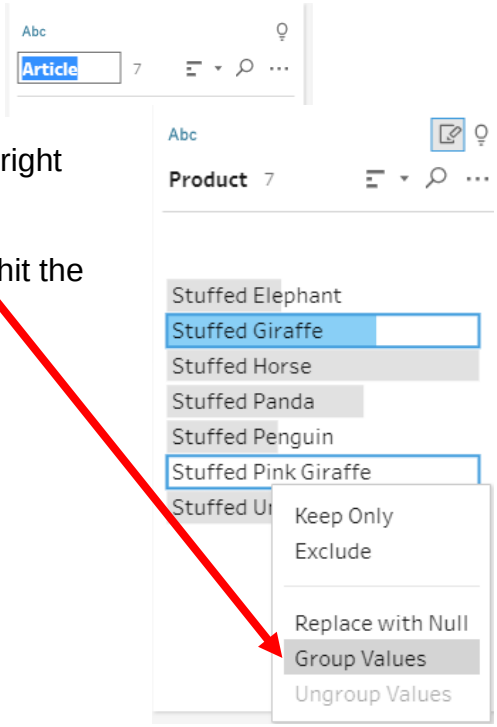
6) The field “**State**” has no changes.

7) The field “**Article**” has to be renamed “**Product”** and there are seven products listed while there should only be six. Some names must have been entered incorrectly.

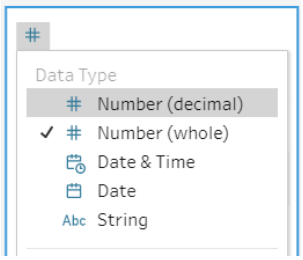
Double click on the old name replace with the new name and select the Enter key.

Select Both “Stuffed Giraffe” and “Stuffed Pink Giraffe”, right click and select “Group Value”.

If it selects the wrong spelling, correct the spelling, and hit the Enter key.



8) The field "**# of units**" is Number (whole) and the type must be changed to Number (decimal). Click the "#" in the upper left corner and select "Number (decimal)".

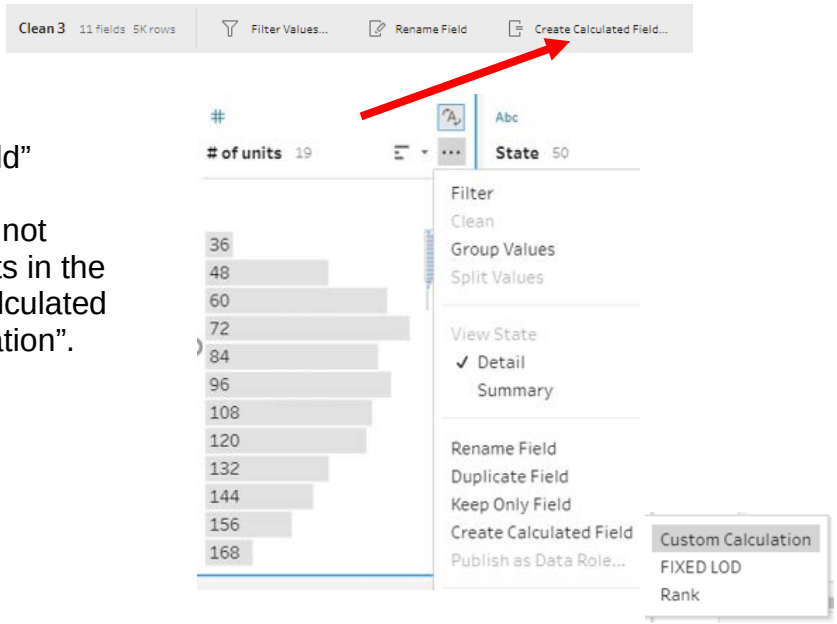


9) The field "**Selling Price Per Unit**" does not exist in the "January Sales Data" stream of data. Instead, there is a field "Sales", the product of the "# of units" and the "Selling Price Per Unit".

Select "# of units".

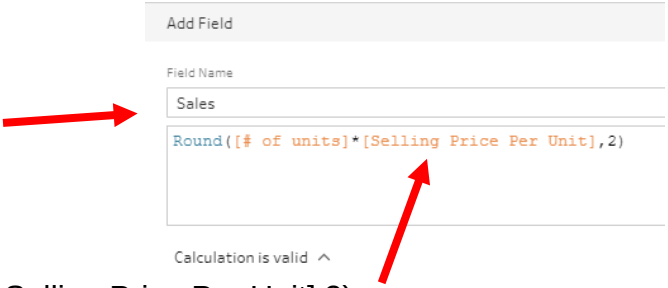
Select "Create Calculated Field"

If "Create Calculated Field" Is not displayed, select the three dots in the upper right, select "Create Calculated Field", select "Custom Calculation".



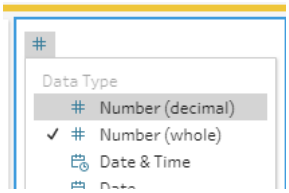
Type in the "Field Name", "Sales".

Type in the formula round([# of units]*[Selling Price Per Unit],2)



Select "Save".

The type of the field "Sales" must be changed from Number (whole) to Number (decimal). Click the "#" in the upper left courier and select "Number (decimal)".



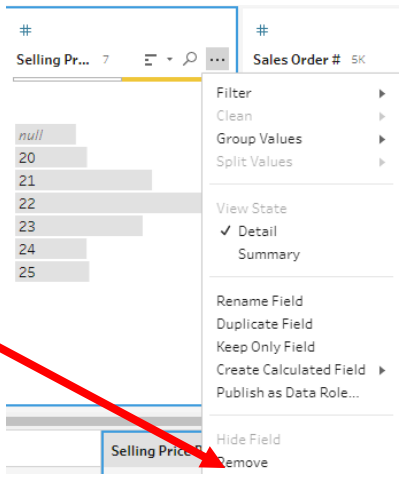
Mark Friedman
University of Miami

9.2) Remove “Selling Price Per Unit” which will not cause a problem with our calculation of “Sales”.

Select the field “Selling Price Per Unit”.

Select the three dots in the upper right.

Select “Remove”.



10) The field “Profit Margin” must be changed to Number (decimal). Click the “#” in the upper left corner and select “Number (decimal)”.

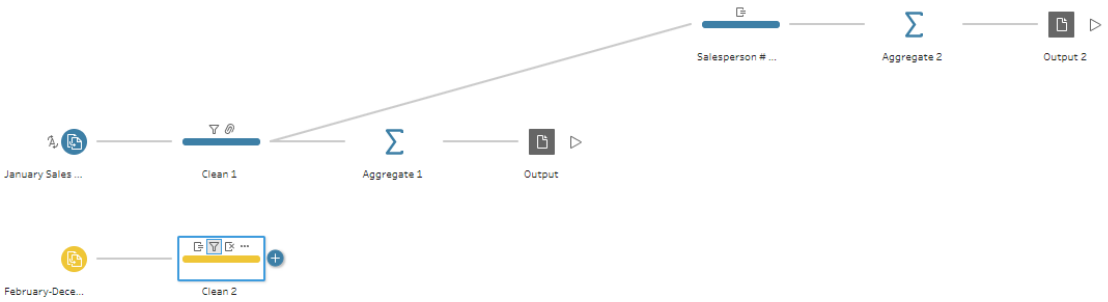
#	Field Name	Sample	Type
1	Sales Order #	35005	Number (whole)
2	Salesperson #	1304	Number (whole)
3	Salesperson	Martin Van Buren	Abc
4	Title	Sales Associate I	Abc
5	Profit Margin	1,908.00	Number (decimal)
6	Region #	3	Number (whole)
7	Region	South	Abc
8	State	TN	Abc
9	Product	Stuffed Horse	Abc
10	# of units	144	Number (decimal)
11	Sales	3,168.00	Number (decimal)

The “Field Names” and “Types” of “February-December Sales Data” are the same as those in the data set “January Sales Data” and we can combine, union, the two parts.

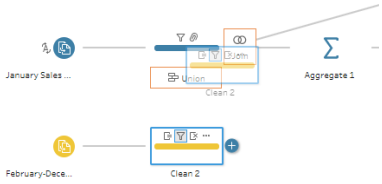
#	Field Name	Sample	Type
1	Sales Order #	35005	Number (whole)
2	Salesperson #	1304	Number (whole)
3	Salesperson	Martin Van Buren	Abc
4	Title	Sales Associate I	Abc
5	Profit Margin	1,908.00	Number (decimal)
6	Region #	3	Number (whole)
7	Region	South	Abc
8	State	TN	Abc
9	Product	Stuffed Horse	Abc
10	# of units	144	Number (decimal)
11	Sales	3,168.00	Number (decimal)

Step 10: Combine the two streams of data

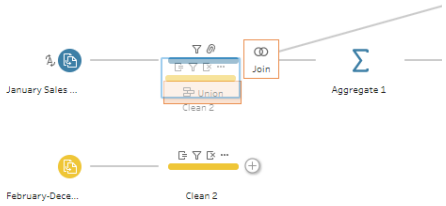
Rename the last “Cleaning” step “Clean 2”. As you drag the “Clean 2” step to the “Clean 1” step in the first stream of data, an option will be presented to “Join” or “Union”.



As you drag the “Clean 2” step to the “Clean 1” step in the first stream of data, an option will be presented to “Join” or “Union”.

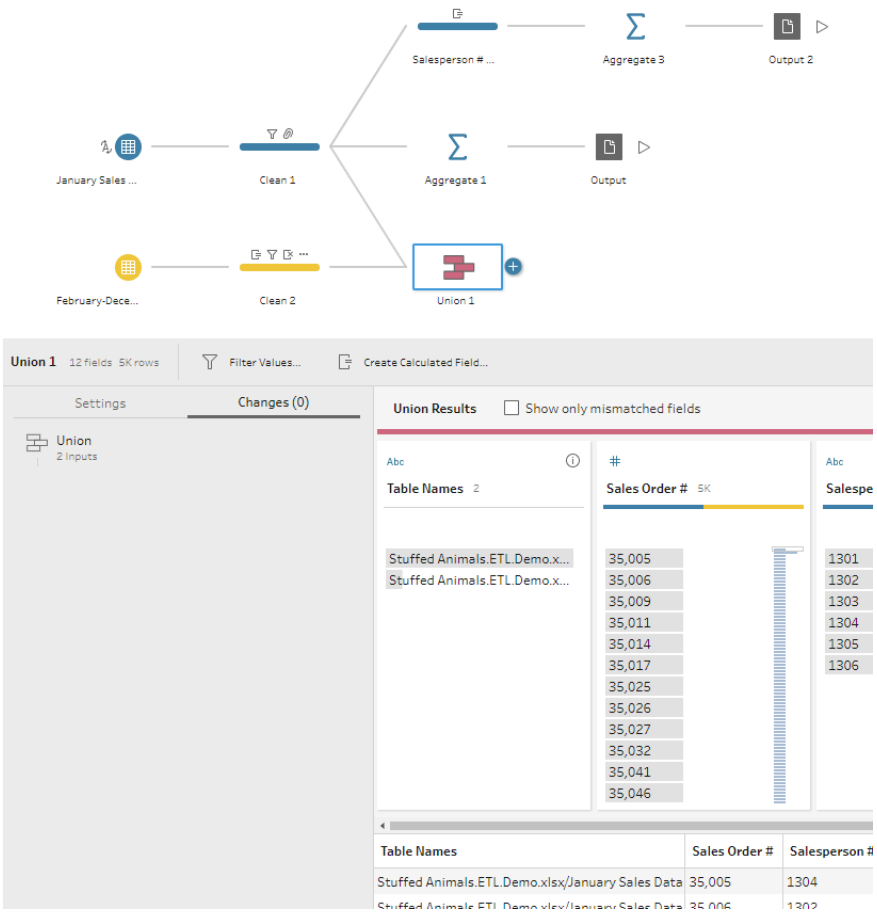


Drop the “Clean 2” step into the “Union” area to combine the two data streams.



.

The result will be a new tile “Union1”.

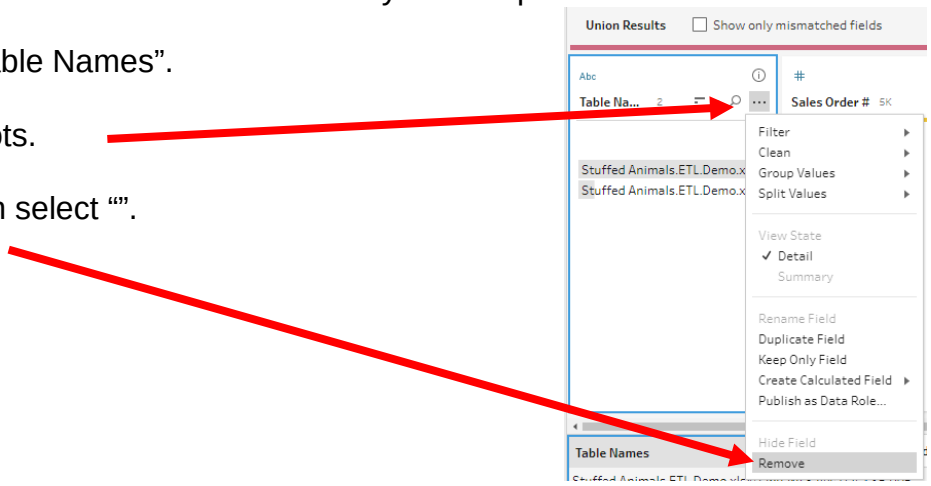


Remove the column “Table Names” created by Union operation.

Select the field “Table Names”.

Select the three dots.

From the pull down select “”.



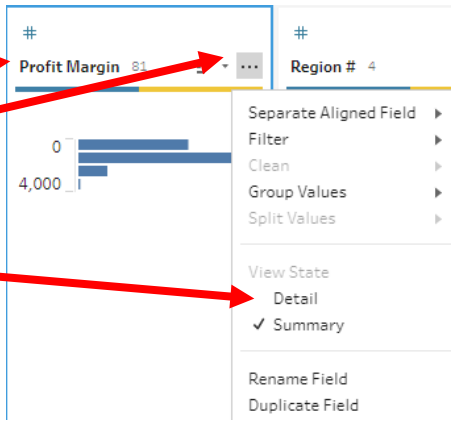
Mark Friedman
University of Miami

A field may display in a bar graph format.

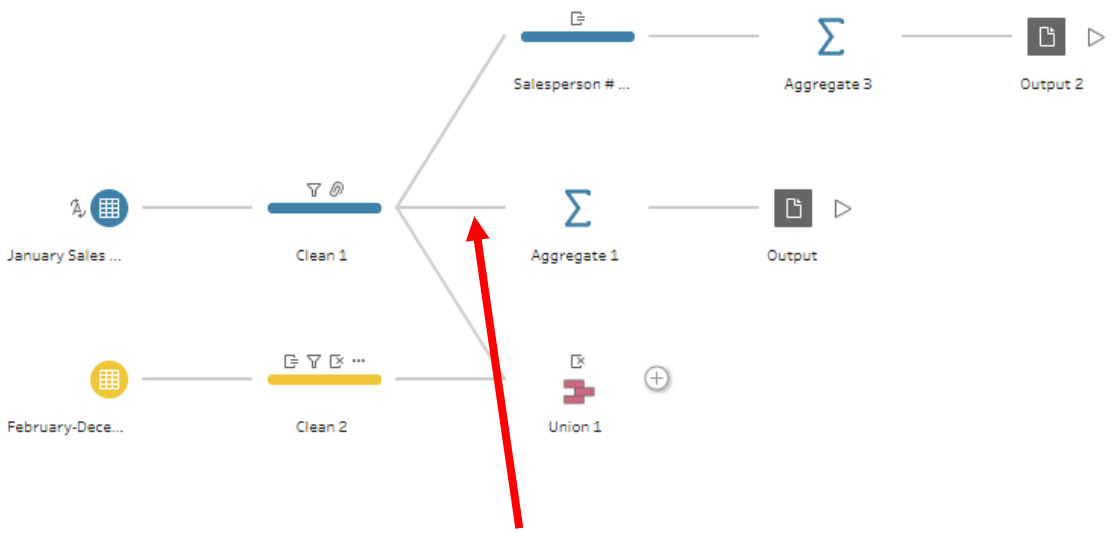
Select the field,

Select the three dots.

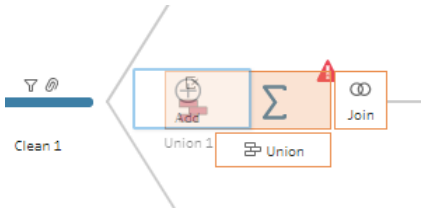
From the pull down select “Detail”.



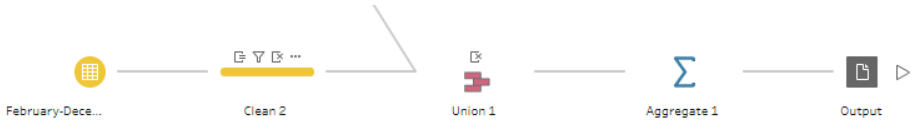
In Part 1 the "January Sales" were attached to two output branches. Instead of redoing the two output branches, just connect them to the "Union 1" tile.



Right click on the branch between the "Clean 1" and "Aggregate 1" tile and select "Remove".

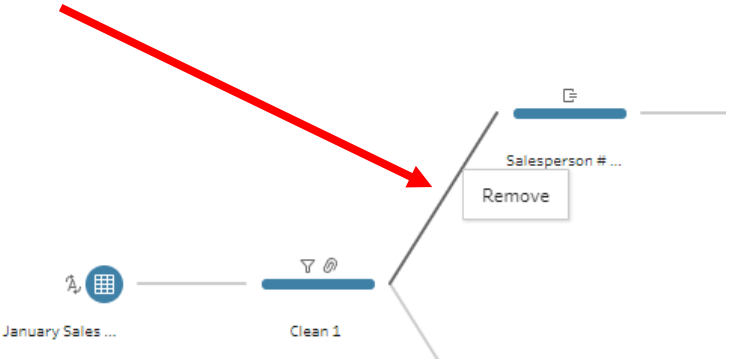


Drag the "Union 1" tile to the left of the "Aggregate 1" tile and drop it in the "Add" area.

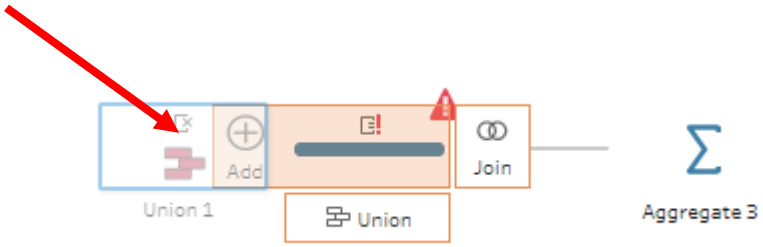


Step 8: Disconnect the output "Unadjusted TB" branch and connect it to the "Union".

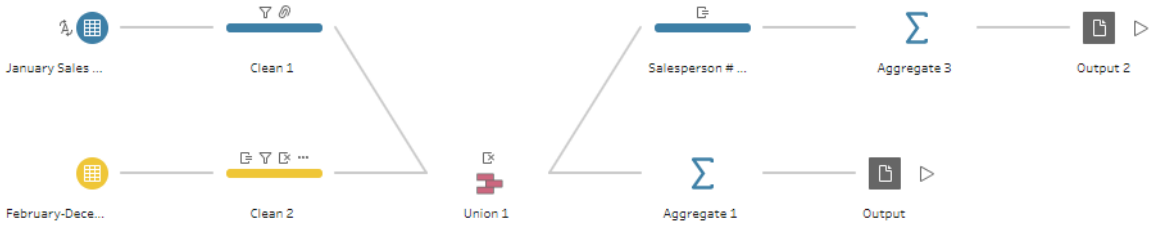
Right click on the branch between "Clean 1" and "Salesperson # and Name1" tile and select "Remove".



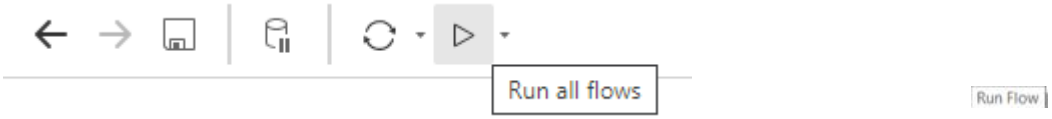
Drag the "Union 1" tile to the left of the "Salesperson # and Name1" tile and drop it in the "Add" area.



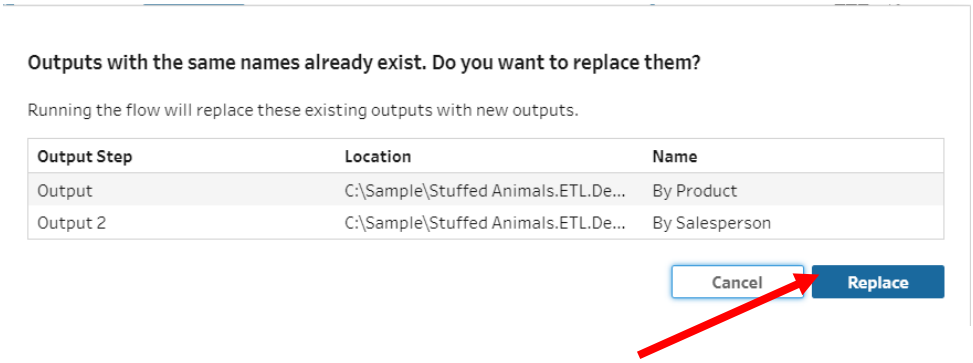
Both outputs will be attached to the "Union 1" tile.



Select the “Run All Flows” for each of the outputs and select “Done” when completed.

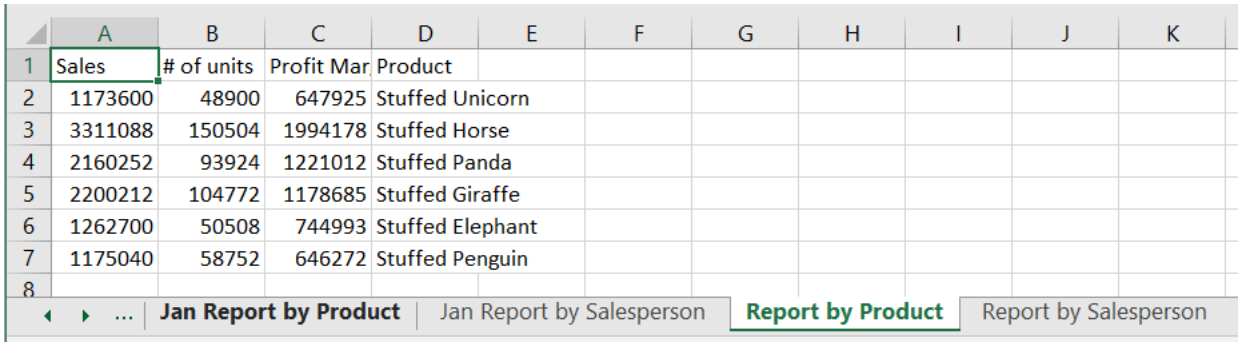


Replace the existing file.



Open the original data file in Excel.

To prevent the changes being inadvertently lost when other extracts are performed rename the Tabs. Double click on the tab “Report by Product” and replace it with “Jan-Dec Report by Product, Year” then hit the Enter key. Double click on the tab “Report by Salesperson” and replace it with “Jan-Dec Report by Salesperson, Year” then hit the Enter key.



	A	B	C	D	E	F	G	H	I	J	K
1	Sales	# of units	Profit Mar	Product							
2	1173600	48900	647925	Stuffed Unicorn							
3	3311088	150504	1994178	Stuffed Horse							
4	2160252	93924	1221012	Stuffed Panda							
5	2200212	104772	1178685	Stuffed Giraffe							
6	1262700	50508	744993	Stuffed Elephant							
7	1175040	58752	646272	Stuffed Penguin							
8											

... | Jan Report by Product | Jan Report by Salesperson | Report by Product | Report by Salesperson

Reformat the “Jan-Dec Report by Product” sheet.

	A	B	C	D
1	Product	# of units	Sales	Profit Margin
2	Stuffed Elephant	50,508	$1,262,700.00	$744,993.00
3	Stuffed Giraffe	104,772	$2,200,212.00	$1,178,685.00
4	Stuffed Horse	150,504	$3,311,088.00	$1,994,178.00
5	Stuffed Panda	93,924	$2,160,252.00	$1,221,012.00
6	Stuffed Penguin	58,752	$1,175,040.00	$646,272.00
7	Stuffed Unicorn	48,900	$1,173,600.00	$647,925.00
8		507,360	$11,282,892.00	$6,433,065.00
9				
10				
11				

... Jan-Dec Report by Product | Jan-Dec Report ...

Reformat the “Jan-Dec Report by Salesperson” sheet.

	A	B	C	D	E
1	Salesperson Number and Name	# of units	Sales	Profit Margin	
2	1301 - Thomas Jefferson	125,508.00	$2,793,660.00	$1,593,147.00	
3	1302 - Grover Cleveland	74,280.00	$1,650,192.00	$936,729.00	
4	1303 - Ulysses Grant	47,400.00	$1,050,024.00	$599,241.00	
5	1304 - Martin Van Buren	60,228.00	$1,329,636.00	$759,033.00	
6	1305 - George Washington	138,936.00	$3,099,840.00	$1,769,571.00	
7	1306 - John Tyler	61,008.00	$1,359,540.00	$775,344.00	
8	Total	507,360.00	$11,282,892.00	$6,433,065.00	
9					
10					
11					

... Jan-Dec Report by Salesperson

Compare your totals to the numbers given on the “Introduction” sheet,

6. The total annual sales for your project is $11,282,892.00 and the total annual profit margin is $6,433,065.00.

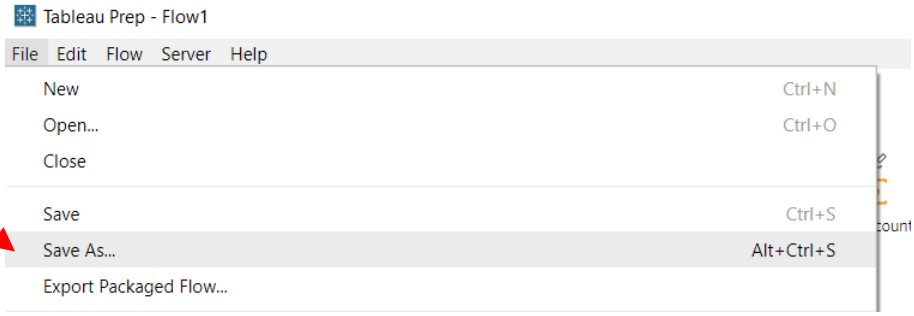
Save, close the Excel file, and return to Tableau.

Mark Friedman
University of Miami

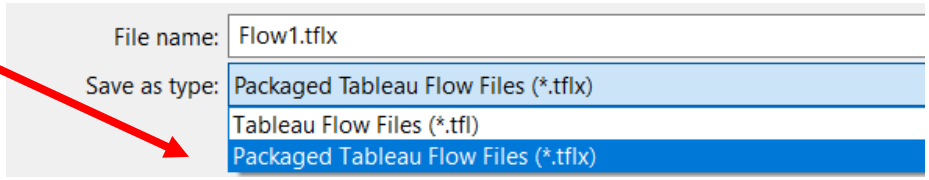
Save the Tableau flow and data

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.