

## Using Tableau Prep to Join Data Streams

In this exercise using Tableau Prep Builder, you will:

**E**xtract data files from Excel.

**T**ransform 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.

**L**oad the transformed data into Excel.

There are five streams of data in the Excel file, *Stuffed Animals.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	<b>Cuddly Stuffed Animals</b>					
2	<b>Sales</b>					
3	<b>Sales Order #</b>	<b>Salesperson #</b>	<b>Region</b>	<b>State</b>	<b>Product #</b>	<b>Units Sold</b>
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	<b>Cuddly Stuffed Animals</b>					
2	<b>Sales</b>					
3	<b>Sales Order #</b>	<b>Salesperson #</b>	<b>Region</b>	<b>State</b>	<b>Product #</b>	<b>Units Sold</b>
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.

	B	C	D	E	F	G
10						
11						
12	Cuddly Stuffed Animals					
13	Product Information					
14						
15	Product ID	Product	Unit Selling Price	Unit Manufacturing Cost	Profit Margin by Product	
16	101	Stuffed Lamb	20.00	8.75	11.25	
17						
18	102	Stuffed Giraffe	22.00	9.75	12.25	
19						
20	103	Stuffed Elephant	24.00	10.75	13.25	
21						
22	104	Stuffed Unicorn	21.00	8.50	12.50	
23						
24	105	Stuffed Horse	23.00	10.75	12.25	
25						
26	106	Stuffed Pig	25.00	12.50	12.50	
27						
28						

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

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	<b>Cuddly Stuffed Animals</b>					
13	<b>Salesperson Information</b>					
15	<b>Code</b>	<b>Salesperson</b>		<b>Title</b>		
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						

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				
Region				
Product ID	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	Units Sold	Sales	Shipping Cost
Ulysses Grant	57,924	\$ 1,317,384.00	\$ 117,878.40
Thomas Jefferson	141,156	\$ 3,210,924.00	\$ 287,129.76
James Polk	43,200	\$ 985,560.00	\$ 87,710.64
James Monroe	121,236	\$ 2,748,540.00	\$ 246,882.36
James Madison	59,136	\$ 1,349,352.00	\$ 120,158.04
Grover Cleveland	76,680	\$ 1,733,088.00	\$ 156,216.00
Total	499,332	\$ 11,344,848.00	\$ 1,015,975.20

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.

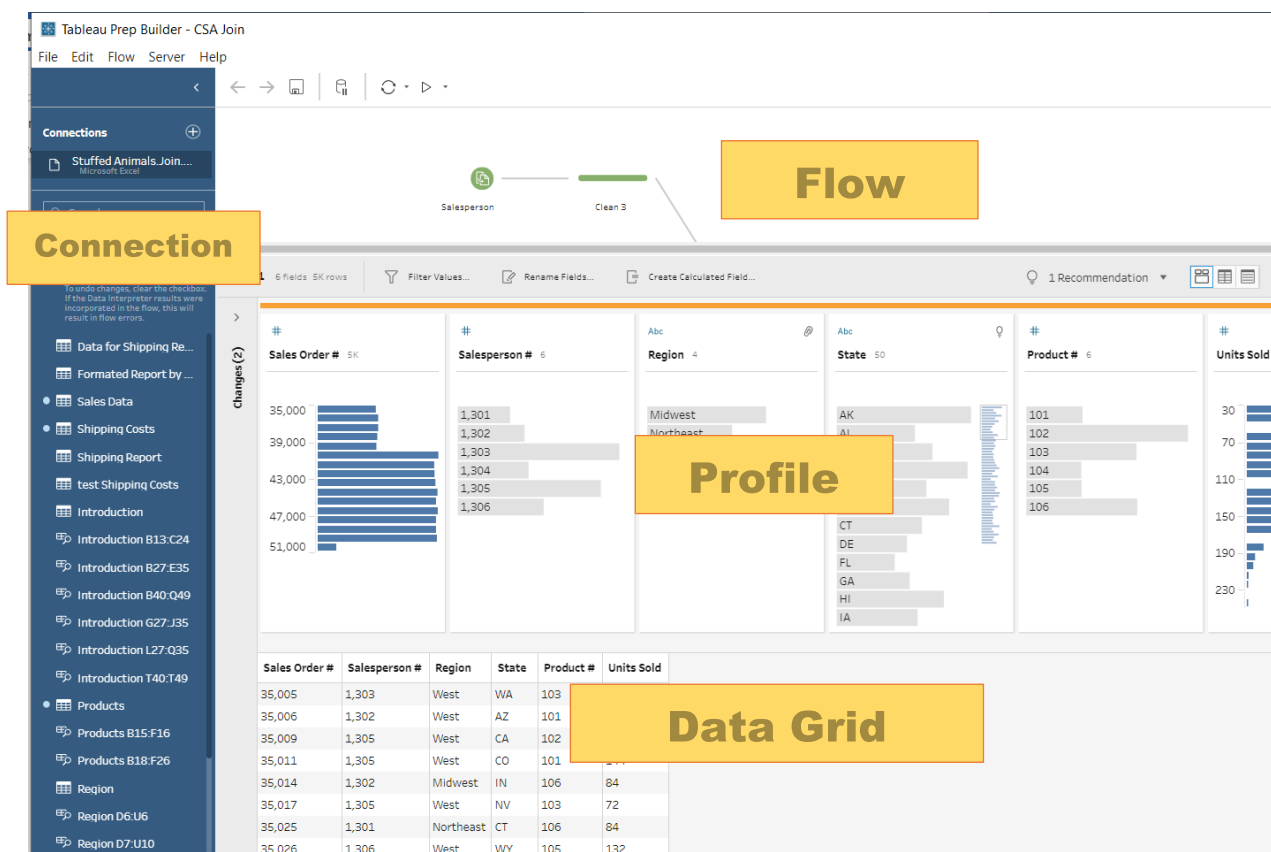
Shipping Costs by Product and Region					
	Midwest	Northeast	South	West	Grand Total
Stuffed Elephant	\$ 47,610.72	\$ 35,625.24	\$ 50,822.64	\$ 63,990.36	\$ 198,048.96
Stuffed Giraffe	\$ 76,744.92	\$ 56,958.72	\$ 80,686.08	\$ 91,095.84	\$ 305,485.56
Stuffed Horse	\$ 26,032.32	\$ 17,062.56	\$ 25,488.00	\$ 31,916.88	\$ 100,499.76
Stuffed Lamb	\$ 30,263.40	\$ 19,614.24	\$ 29,877.12	\$ 38,905.20	\$ 118,659.96
Stuffed Pig	\$ 47,898.24	\$ 30,566.40	\$ 47,361.60	\$ 59,648.40	\$ 185,474.64
Stuffed Unicorn	\$ 25,999.68	\$ 18,458.88	\$ 29,987.40	\$ 33,360.36	\$ 107,806.32
Grand Total	\$ 254,549.28	\$ 178,286.04	\$ 264,222.84	\$ 318,917.04	\$ 1,015,975.20

## Start Tableau Prep

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.

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.

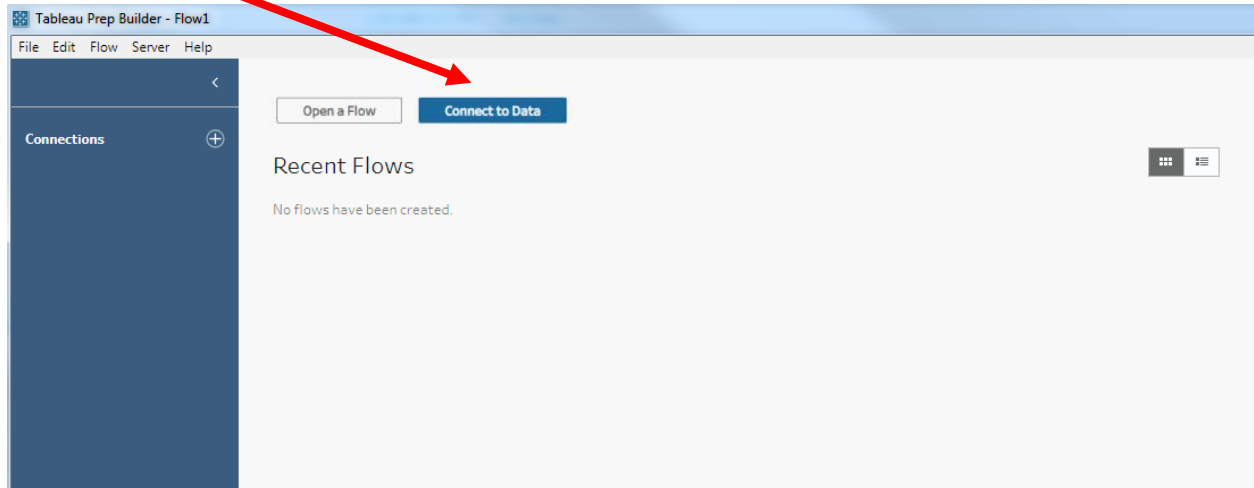


After you connect to your data and begin building your flow, you add steps in the Flow pane. These steps function as a lens into the structure of your data, as well as a summary of operations that is applied to your data. Each step represents a different category of operations that you define.

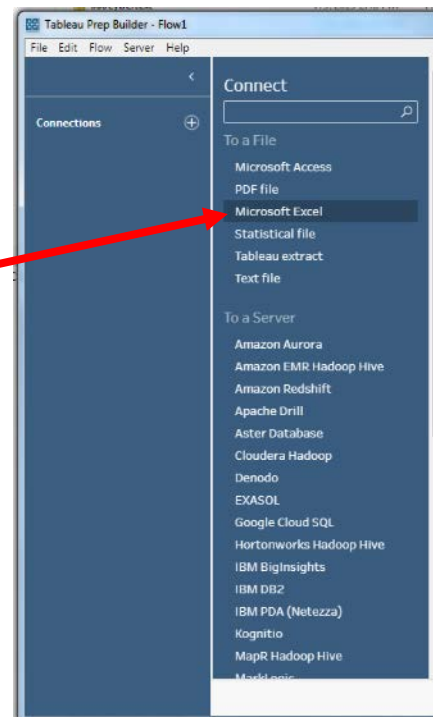
In the Configuration area, select the “Connect a File or Database” drop-down arrow.

**Step 1: Input Excel file**

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

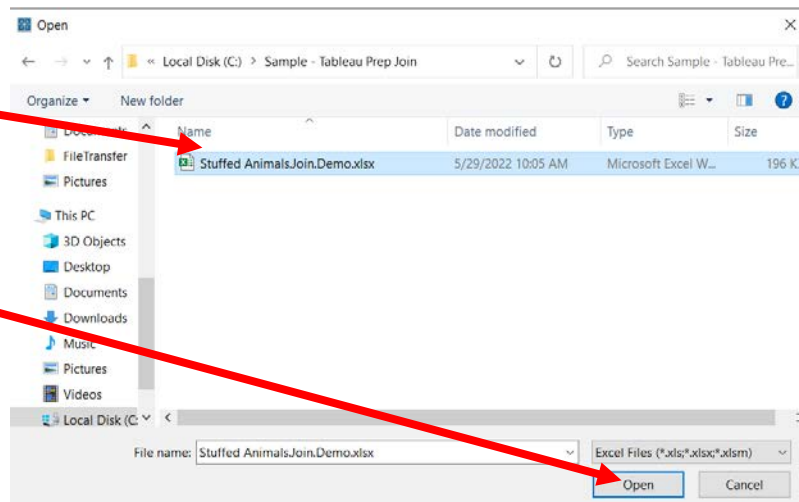


On the connections pane select Microsoft Excel.



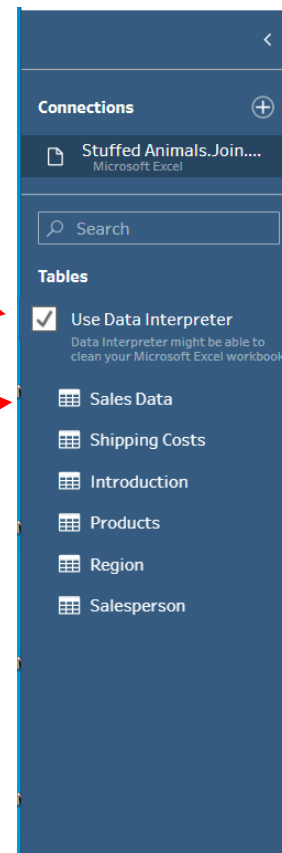
Select the file.

Select "Open".

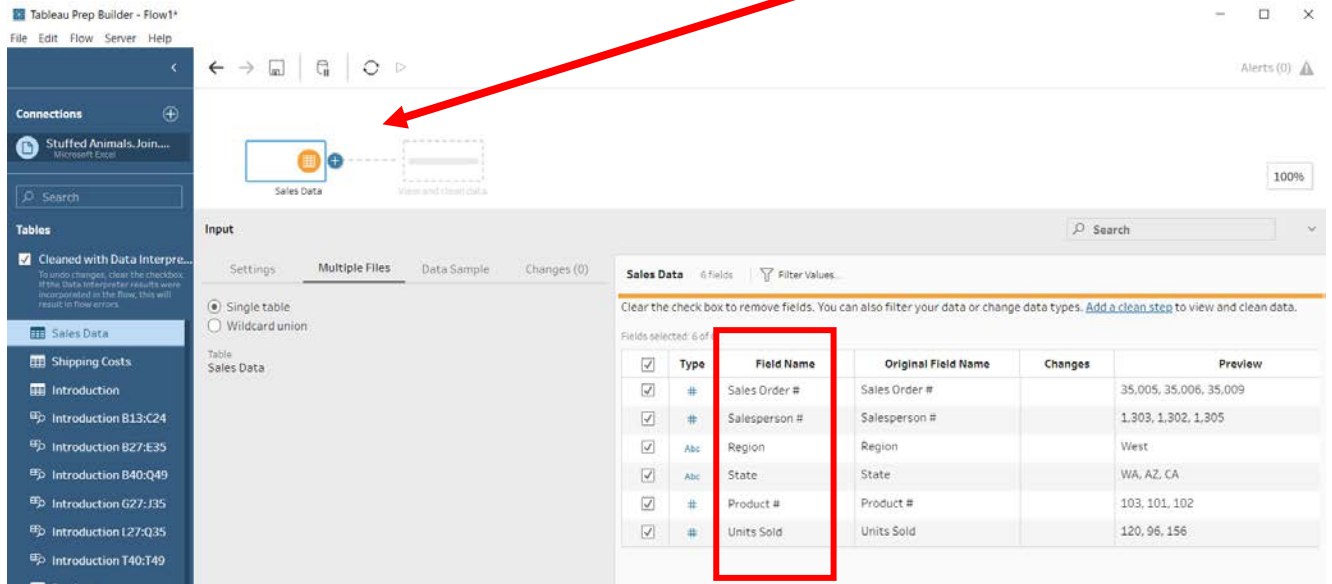


Remember that the field names are in the third row. Check "Use Data Interpreter" tick box to let Tableau Prep adjust for the input data.

Select the "Sales Data" sheet on the Connections pane by either double click on the sheet name or drag the sheet name to the canvas.



The step, “Sales Data”, will appear on the Flow pane.



The screenshot shows the Tableau Prep Builder interface. On the left, the 'Connections' pane lists 'Stuffed Animals Join...' and 'Sales Data'. The 'Tables' pane shows 'Sales Data' selected. The 'Flow' pane shows a flow with 'Sales Data' and 'View and clean data' steps. A red arrow points to the 'Sales Data' step. The 'Input' pane for 'Sales Data' is open, showing a table with 6 fields. The 'Field Name' column is highlighted with a red box.

	Type	Field Name	Original Field Name	Changes	Preview
<input checked="" type="checkbox"/>	#	Sales Order #	Sales Order #		35,005, 35,006, 35,009
<input checked="" type="checkbox"/>	#	Salesperson #	Salesperson #		1,303, 1,302, 1,305
<input checked="" type="checkbox"/>	Abc	Region	Region		West
<input checked="" type="checkbox"/>	Abc	State	State		WA, AZ, CA
<input checked="" type="checkbox"/>	#	Product #	Product #		103, 101, 102
<input checked="" type="checkbox"/>	#	Units Sold	Units Sold		120, 96, 156

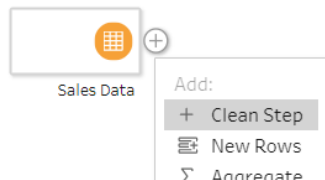
The sample data will be visible on the Profile pane.

The “Use Data Interpreter” worked, and the “Field Name” column is correct.

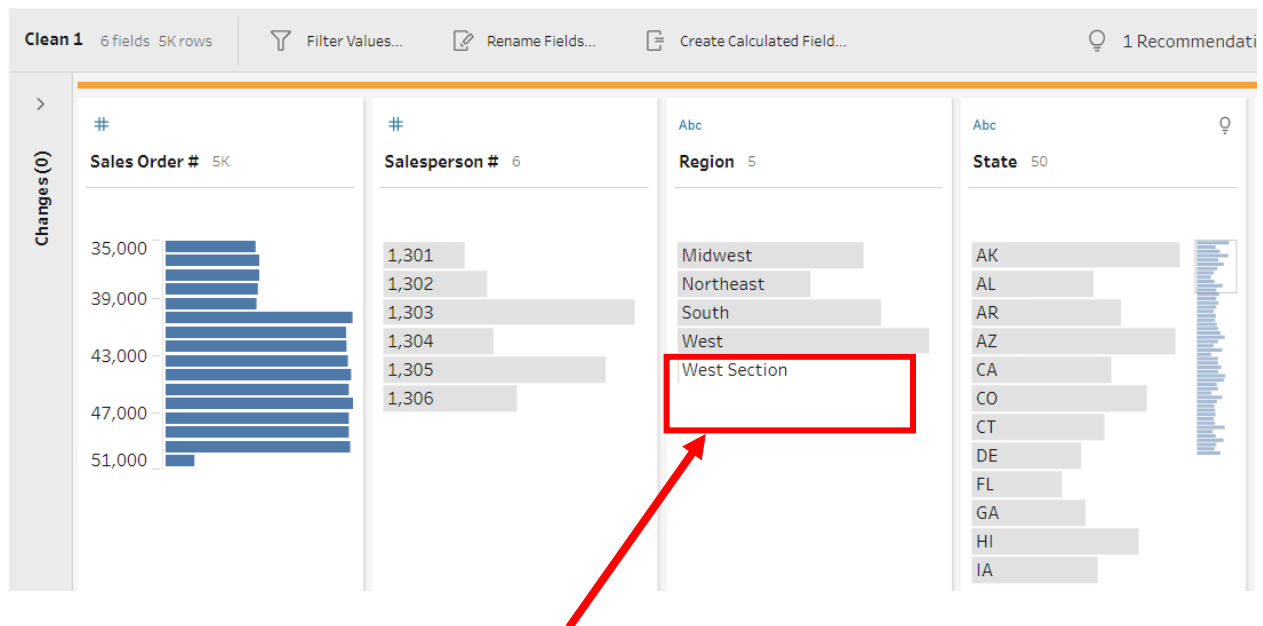


**Step 2: Clean the data**

Review the data that we just added. Click the plus sign and add “Clean Step”.



On the profile pane review the data.



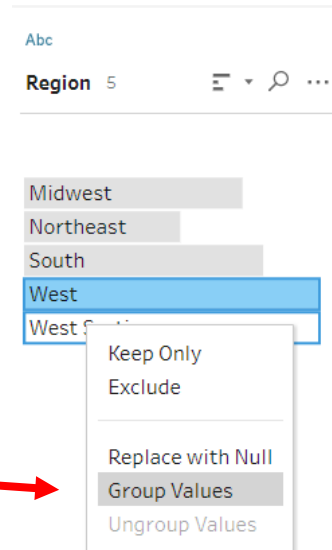
The “Region” column can only contain one of the four regions. Browse the data in the “Region” field to verify that all the data is correct. Make any changes that are required.

In the “Region” column two regions looks similar, but not the same: “West” and “West Section”. In one or more records “West” was incorrectly entered as “West Section”. We need to correct this manual input error.

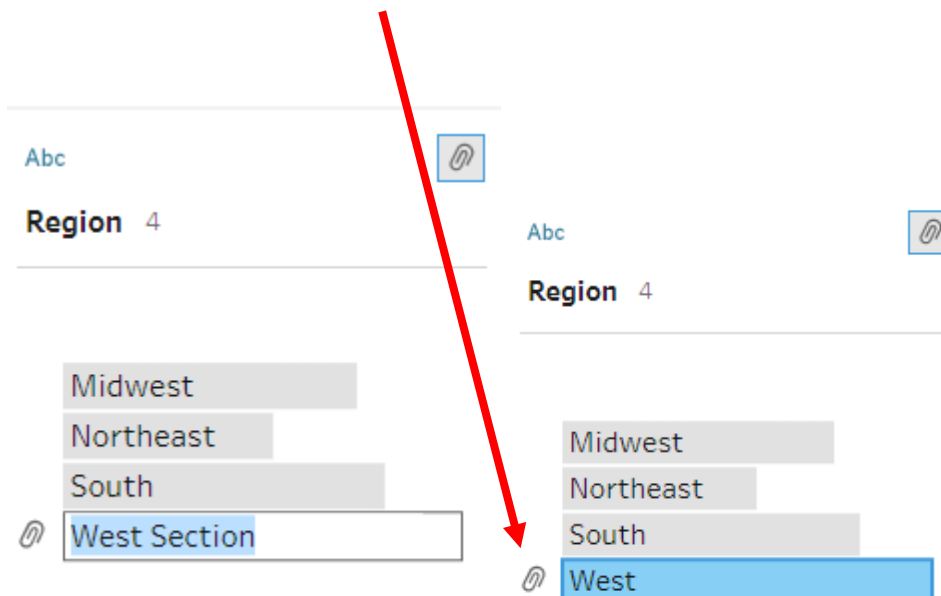
Code	Region
1	Midwest
2	Northeast
3	South
4	West

Group both values and give the group a common name. Select the two regions by first clicking on “West” then hold down the control key click on the second field “Western Section” and then release the control key (Ctrl+).

While hovering over the selected items right click and select “Group”.



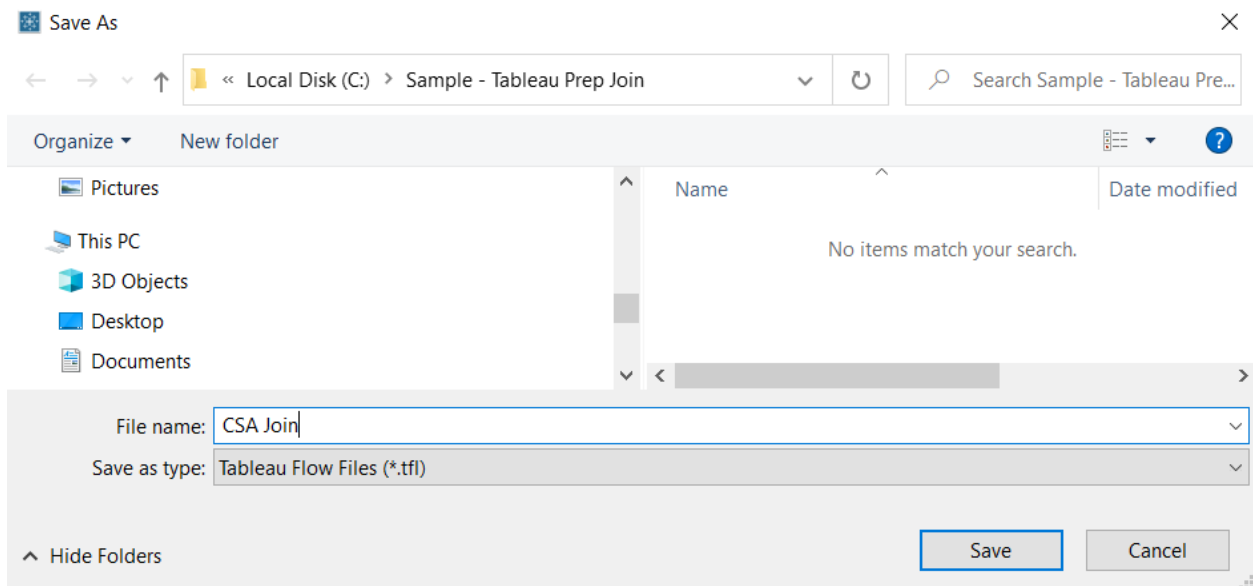
If necessary, name the group “West”.



Save the workflow regularly.

Select “File > Save As...”

Enter the “File name:” and Save.

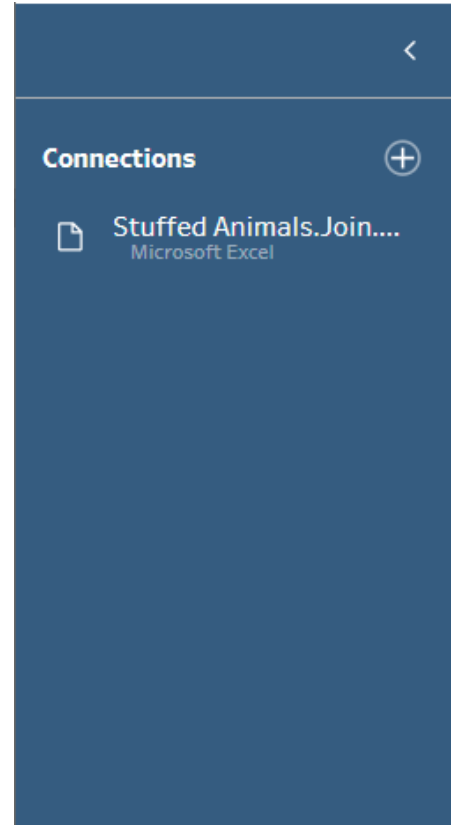


**Step 3:** Extract the data from the "Products" sheet. Review the data, join to Stream 1 and verify the number of rows.

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.

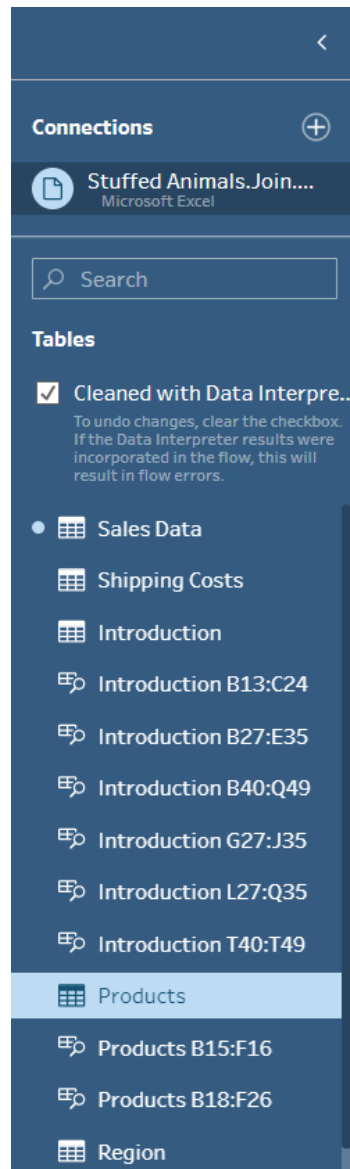
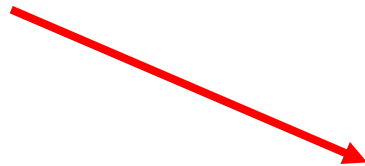
	B	C	D	E	F	G
10						
11						
12						
13						
14						
15	Product ID	Product	Unit Selling Price	Unit Manufacturing Cost	Profit Margin by Product	
16	101	Stuffed Lamb	20.00	8.75	11.25	
17						
18	102	Stuffed Giraffe	22.00	9.75	12.25	
19						
20	103	Stuffed Elephant	24.00	10.75	13.25	
21						
22	104	Stuffed Unicorn	21.00	8.50	12.50	
23						
24	105	Stuffed Horse	23.00	10.75	12.25	
25						
26	106	Stuffed Pig	25.00	12.50	12.50	
27						
28						

Select "Stuffed Animals.Join.Demo.xlsx".



The “Cleaned with Data Interpreter” should still be selected.

Select the “Products” sheet on the Connections pane by either double click on the sheet name or drag the sheet name to the canvas.



**Step 4: Clean the data**

Review the data that we just added. Click the plus sign and add step “Clean 2”.

The screenshot shows the Tableau Prep interface. At the top, a workflow diagram shows 'Sales Data' connected to 'Clean 1', which is then connected to a plus sign. Below this, a 'Products' data source is connected to a plus sign, which is connected to a 'View and clean data' step. A dropdown menu is open, showing options: '+ Clean Step', 'Aggregate', and 'Pivot'. Below the menu, a 'Clean 2' step is added to the workflow, connected to the 'Products' data source.

The main view shows the 'Clean 2' step with 5 fields and 6 rows. The fields are: Product ID, Product, Unit Selling Price, Unit Manufacturing Cost, and Profit Margin by Product. The data is displayed in a table format with 6 rows and 5 columns. The 'Product ID' column is highlighted with a red box.

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

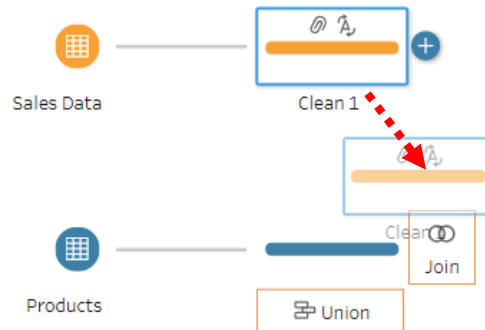
Everything looks correct. Step “Clean 2”, from the second stream has 6 records and contains the field “Product ID”. Step “Clean 1”, from the first steam has 5,017 records and contains the field “Product #”.

The screenshot shows the 'Clean 1' step with 6 fields and 5K rows. The fields are: Sales Order #, Salesperson #, Region, State, and Product #. The data is displayed in a table format with 5K rows and 5 columns. The 'Product #' column is highlighted with a red box.

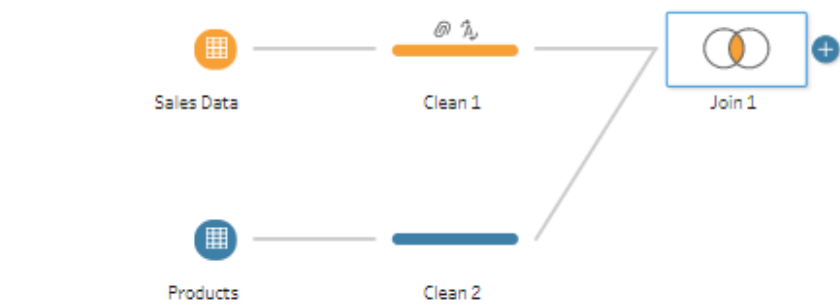
Sales Order #	Salesperson #	Region	State	Product #
35,005	1,301	Midwest	AK	101
35,006	1,302	Northeast	AL	102
35,009	1,303	South	AR	103
35,011	1,304	West	AZ	104
35,014	1,305		CA	105
35,017	1,306		CO	106
35,025			CT	

**Step 5: Combine Data from first Two Streams on a Common Field**

Now join the two the streams. Drag the “Clean 1” step from the first stream to the “Clean 2” step in the second stream and a Join box will appear. Drop the “Clean 1” step into the Join box.



As a result, step “Join 1”, a combination of the two streams will appear. Select the “Join 1” step.



The diagram illustrates the Tableau Prep workflow. It starts with two data sources: 'Sales Data' and 'Products'. 'Sales Data' is cleaned into 'Clean 1', and 'Products' is cleaned into 'Clean 2'. These two cleaned datasets are then joined into 'Join 1'.

The interface for 'Join 1' shows the following details:

- Settings:**
  - Applied Join Clauses:** Clean 2 (Product ID) = Clean 1 (Product #)
  - Join Type:** inner
  - Summary of Join Results:**
    - Clean 2: 6 (Included)
    - Clean 1: 5,017 (Included)
    - Join Result: 5,017
- Join Clauses:**
  - Clean 2: ↑ Product ID
  - Clean 1: ↑ Product #

The 'Summary of Join Results' section includes a bar chart showing the distribution of values. The 'Included' bar for Clean 2 is 6, and the 'Included' bar for Clean 1 is 5,017. The 'Join Result' bar is 5,017. The 'Mismatched values' bar is empty.

Tableau automatically joined based upon a match of “Product ID” in Clean 2 and “Product #” in Clean 1.

The Join result is 5,017 rows, everything was matched, there are no errors.



Sales Order #	Salesperson #	Region	State	Product #	Units Sold	Product ID	Product	Unit Selling Price	Unit Manufacturing Cost	Profit Margin by Product
35,005	1,303	West	WA	103	120	103	Stuffed Elephant	24	10.75	13.25
35,006	1,302	West	AZ	101	96	101	Stuffed Lamb	20	8.75	11.25
35,009	1,305	West	CA	102	156	102	Stuffed Giraffe	22	9.75	12.25
35,011	1,305	West	CO	101	144	101	Stuffed Lamb	20	8.75	11.25
35,014	1,302	Midwest	IN	106	84	106	Stuffed Pig	25	12.5	12.5

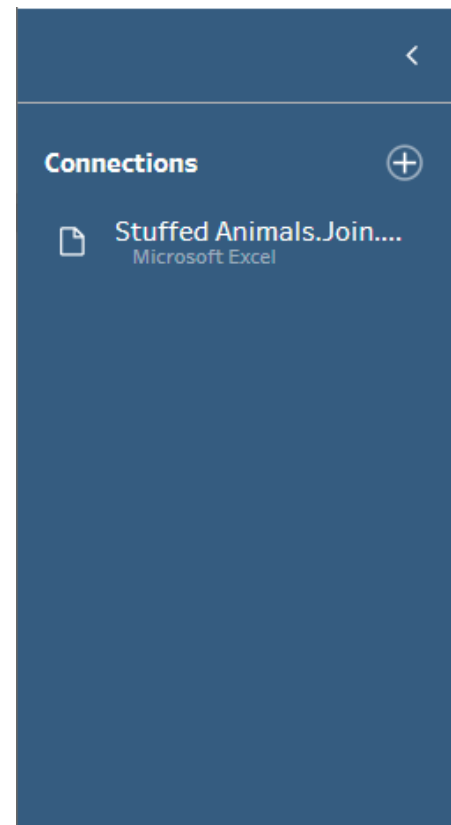
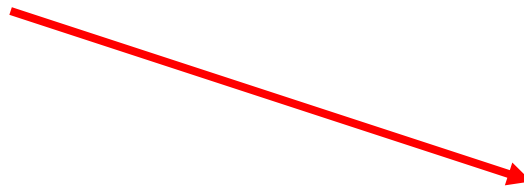
“Join 1” contains detailed product information.

**Step 6:** Extract the data from the "Salesperson" sheet. Review the data, join to the first two streams, and verify the number of rows.

	C	D	E	F	G	H
10						
11						
12			<b>Cuddly Stuffed Animals</b>			
13			<b>Salesperson Information</b>			
14						
	Code		Salesperson		Title	
15						
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						

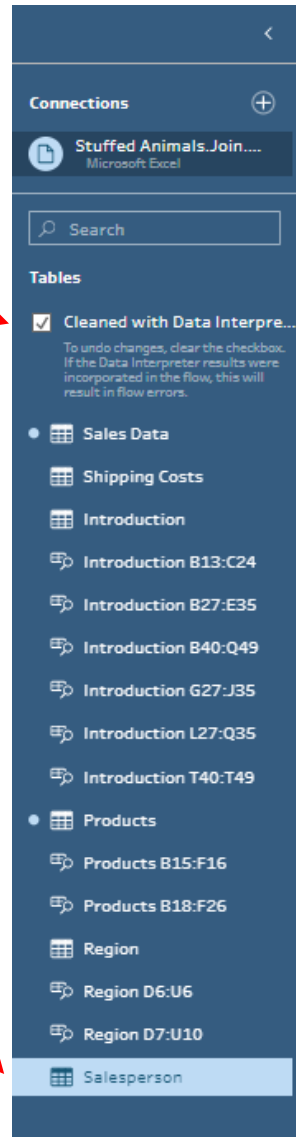
Data regarding the salesperson's name and title are stored in an Excel sheet titled "Salesperson". Notice that the column headings are in the fifteen row.

Select "Stuffed Animals.Join.Demo.xlsx" from the "Connections" panel.



The “Cleaned with Data Interpreter” should still be selected.

Select the “Salesperson” sheet on the Connections pane by either double click on the sheet name or drag the sheet name to the canvas.



**Step 7: Clean the data**

Review the data that we just added.



Click the “View and clean data” tool and add step “Clean 3”.



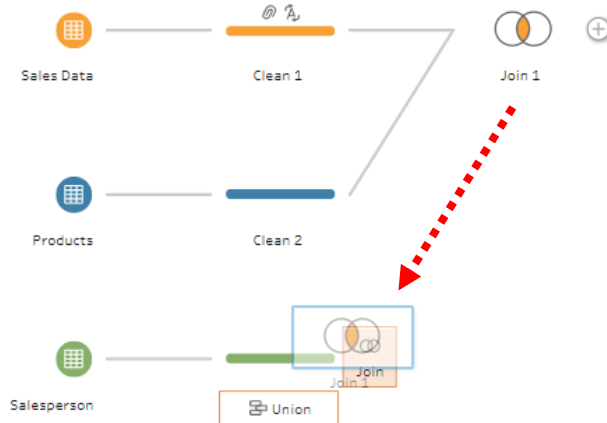
Clean 3 3 fields 6 rows		
Filter Values... Rename Fields... Create Calculated Field...		
#	Abc	Abc
Code 6	Salesperson 6	Title 2
1,301	Grover Cleveland	Sales Associate I
1,302	James Madison	Senior Sales Associate
1,303	James Monroe	
1,304	James Polk	
1,305	Thomas Jefferson	
1,306	Ulysses Grant	

Everything looks correct. Join 1 with 5,017 records contains the field “Salesperson #” which contains the same data as the field “Code” in Clean 3.

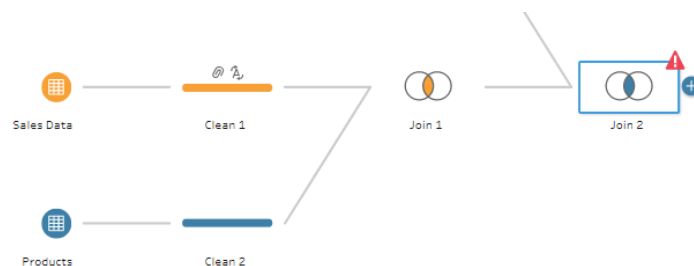
Join Results	
#	#
Sales Order # 5K	Salesperson # 6
35,005	1,301
35,006	1,302
35,009	1,303
35,011	1,304
35,014	1,305
35,017	1,306
35,025	

## Step 8: Combine Data from first Two Streams on a Common Field

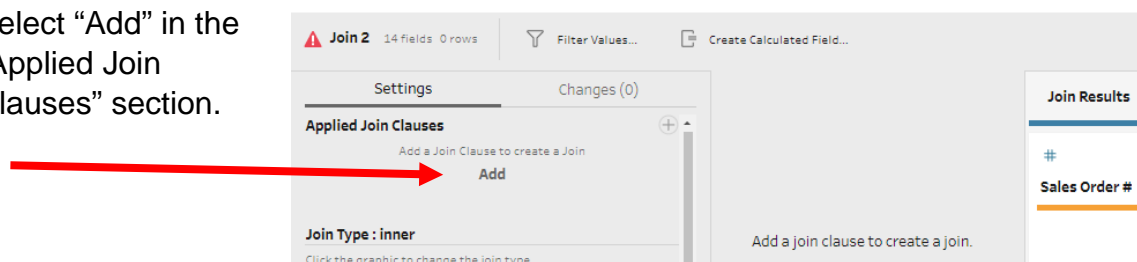
Now join the two the streams. Drag the “Join 1” step from the first and second stream to the “Clean 3” step from the third stream and a Join box will appear. Drop the “Join 1” step into the Join box.



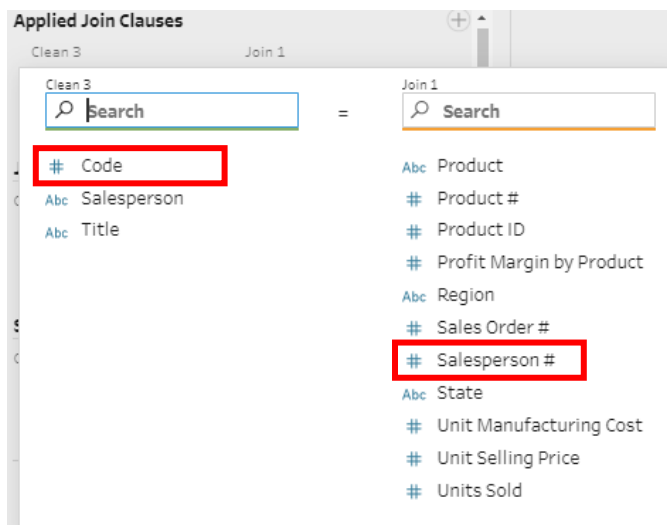
As a result, step “Join 2”, a combination of the two streams will appear. Tableau did not determine what fields should be used to join the two streams. Select the “Join 2” step.



Select “Add” in the “Applied Join Clauses” section.

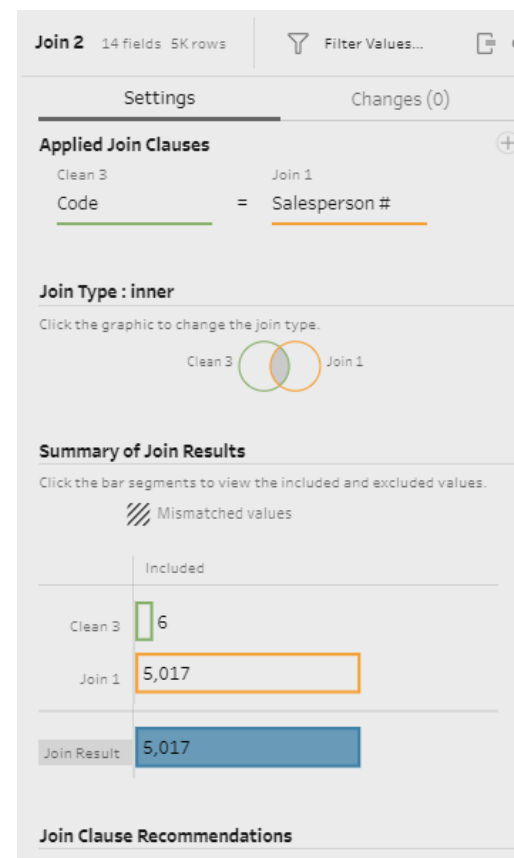


Select the fields to join on, “Code” from Clean 3 and “Salesperson #” from “Join 1”.



The default “Join Type” is “inner” which does not have to be changed.

The join was successful as 5,017 rows were combined.



**Step 9:** Extract the data from the "Shipping Costs" sheet. Review the data, join to Join 2 and verify the number of rows.

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

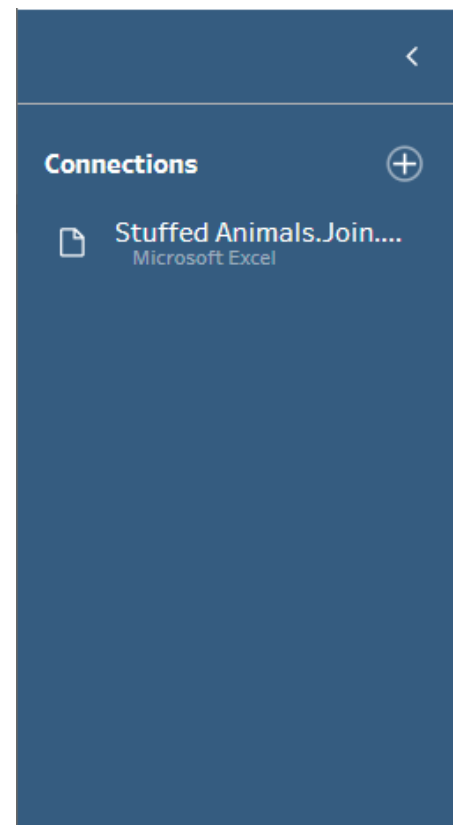
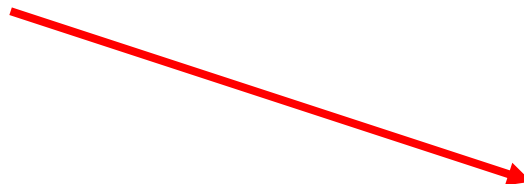
Cuddly Stuffed Animals Shipping Cost Per Unit				
Product ID	Region			
	Midwest	Northeast	South	West
101	2.15	2.12	2.08	2.02
102	2.11	2.08	2.04	1.98
103	2.14	2.11	2.07	2.01
104	2.12	2.09	2.05	1.99
105	2.07	2.04	2.00	1.94
106	2.02	1.99	1.95	1.89

...	Salesperson	Products	Region	Shipping Costs
-----	-------------	----------	--------	----------------

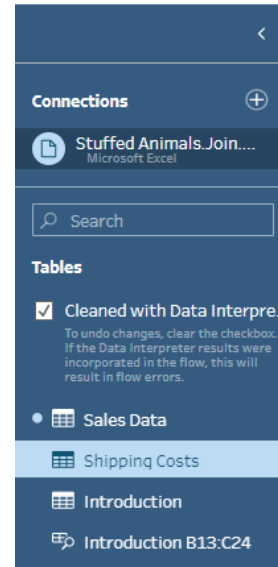
Notice that the column headings are in the thirteenth row.

Select "Stuffed Animals.Join.Demo.xlsx" from the "Connections" panel.



The “Cleaned with Data Interpreter” should still be selected.

Select the “Shipping Costs” sheet on the Connections pane by either double click on the sheet name or drag the sheet name to the canvas.



Input

Settings Multiple Files Data Sample Changes (0)

Single table  
Wildcard union

Table  
Shipping Costs

**Shipping Costs** 5 fields | Filter Values...

Clear the check box to remove fields. You can also filter your data or change data types. [Add a clean step](#) to view and

Fields selected: 5 of 5

<input checked="" type="checkbox"/>	Type	Field Name	Original Field Name	Changes
<input checked="" type="checkbox"/>	#	Product ID	Product ID	101, 102, 103
<input checked="" type="checkbox"/>	#	Region Midwest	Region Midwest	2.15, 2.11, 2.14

Change the field name from “Region Midwest” to “Midwest”.

**Shipping Costs** 5 fields | Filter Values...

Clear the check box to remove fields. You can also filter your data or change data types. [Add a clean step](#) to view and

Fields selected: 5 of 5

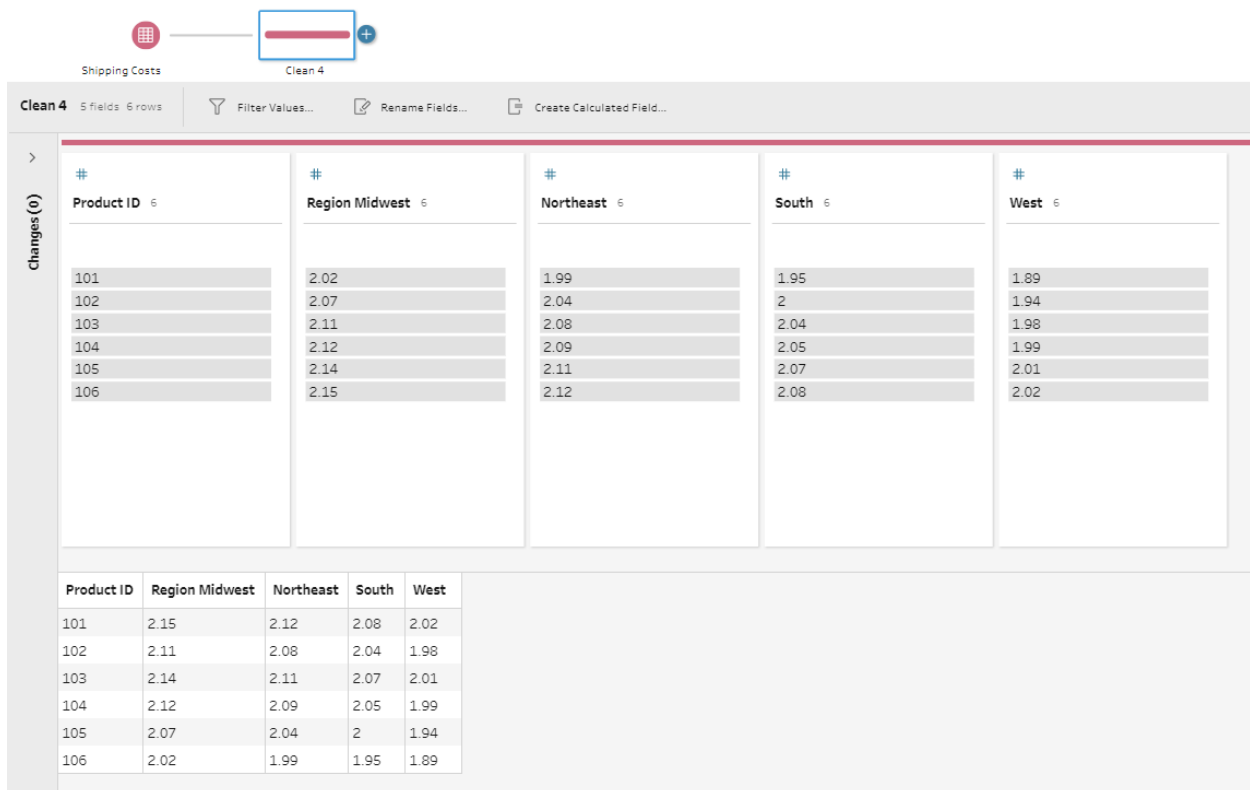
<input checked="" type="checkbox"/>	Type	Field Name	Original Field Name	Changes
<input checked="" type="checkbox"/>	#	Product ID	Product ID	101, 102, 103
<input checked="" type="checkbox"/>	#	Midwest	Region Midwest	2.15, 2.11, 2.14



**Step 10: Clean the data**

Review the data that we just added.

Click the “View and clean data” tool and add step “Clean 4”.



**Clean 4** 5 fields 6 rows

Filter Values... Rename Fields... Create Calculated Field...

#	Product ID	Region Midwest	Northeast	South	West
101		2.02	1.99	1.95	1.89
102		2.07	2.04	2	1.94
103		2.11	2.08	2.04	1.98
104		2.12	2.09	2.05	1.99
105		2.14	2.11	2.07	2.01
106		2.15	2.12	2.08	2.02

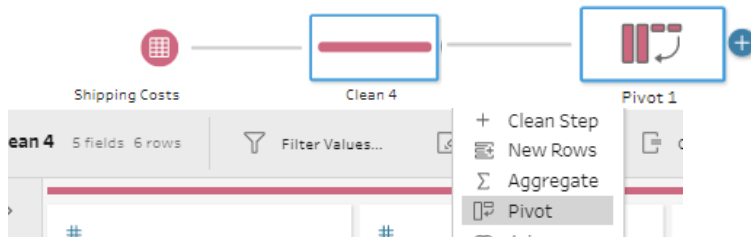
  

Product ID	Region Midwest	Northeast	South	West
101	2.15	2.12	2.08	2.02
102	2.11	2.08	2.04	1.98
103	2.14	2.11	2.07	2.01
104	2.12	2.09	2.05	1.99
105	2.07	2.04	2	1.94
106	2.02	1.99	1.95	1.89

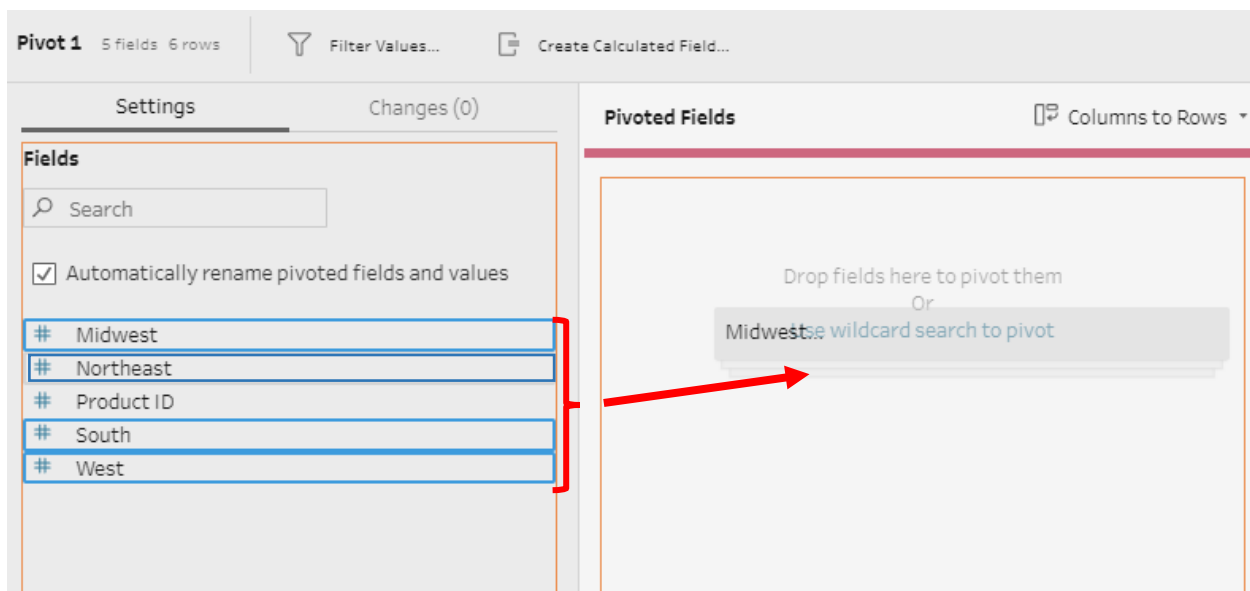
Data in a matrix format cannot be joined to the previous inputs in the current format. Each record must be converted into multiple records.

**Step 11: Pivot the data.**

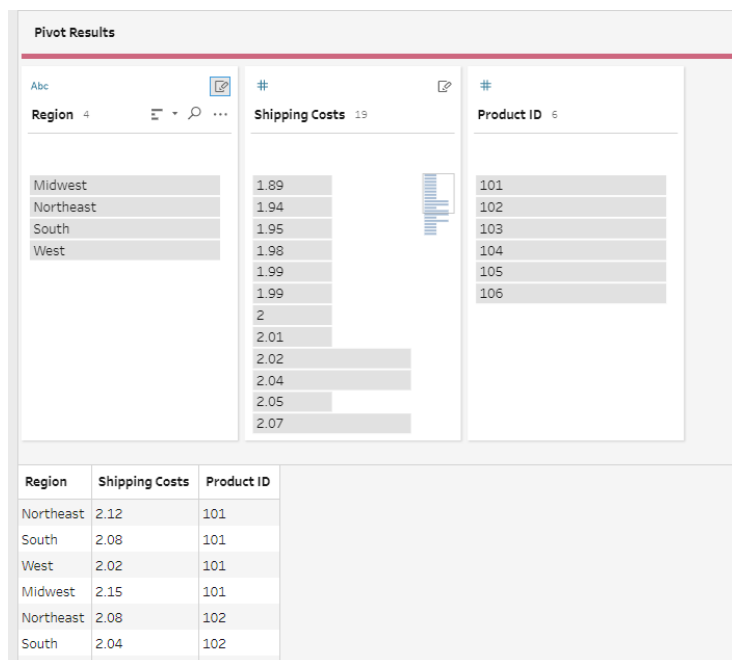
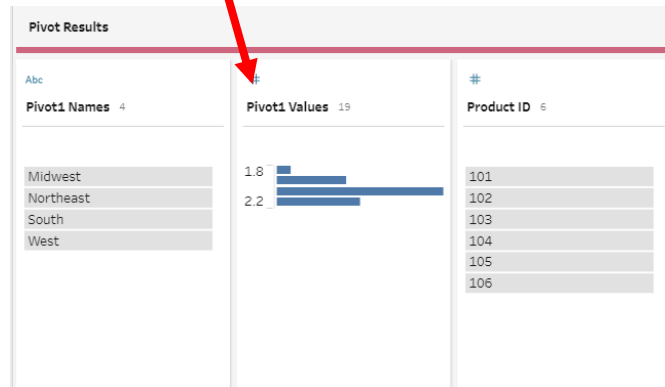
Add a “Pivot” step to the “Clean 4” step.



Select all four region columns and drag them to Pivoted Fields Area.



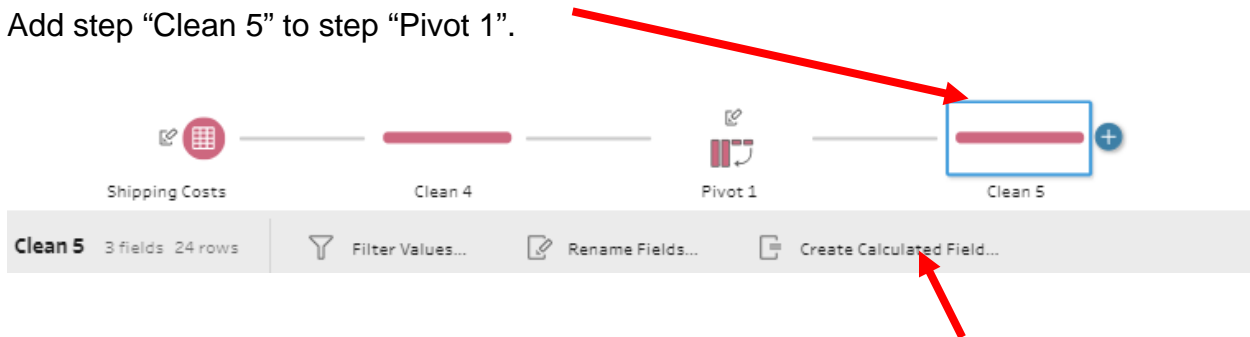
Select the “Pivot Results” section, double click on “Pivot1 Values”, and rename the column “Shipping Costs “. Then rename “Pivot1 Names” with “Region”.



To combine two streams of data there must be a unique key. That unique field is referred to as the primary key in the one side of the join. 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 “101.Northeast”. Since a numeric field “Product ID” is being combined with a string field, “Region” the numeric field must be converted to a string field.

**Step 12: Create a Primary Key Field**

Add step "Clean 5" to step "Pivot 1".



To create a Primary Key field, select "Create Calculated Field".

The 'Add Field' dialog box is shown. The 'Field Name' field contains the text 'Calculation1'.

Change the "Field Name" to "Product ID and Region".

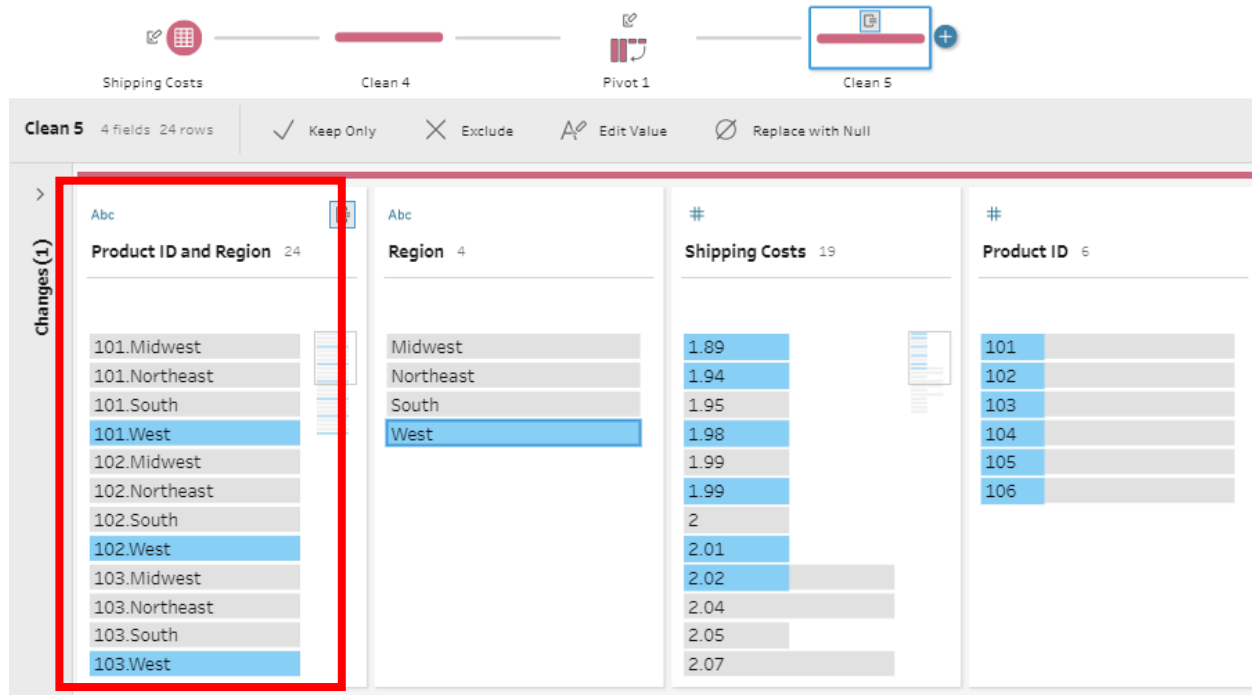
The 'Add Field' dialog box is shown. The 'Field Name' field contains the text 'Product ID and Region'. The formula field contains the text `str([Product ID]) + "." + [Region]`. A red arrow points to the 'Field Name' field.

Type: **STR([Product ID]) + "." + [Region]**

then click Save.

The 'Apply' and 'Save' buttons are shown. A red arrow points to the 'Save' button.

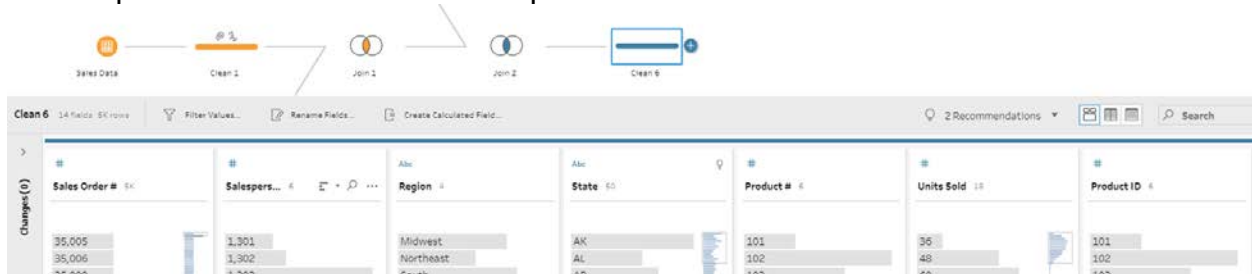
The “SPID and Region” field has been populated.



### Step 13: Create a Foreign Key Field

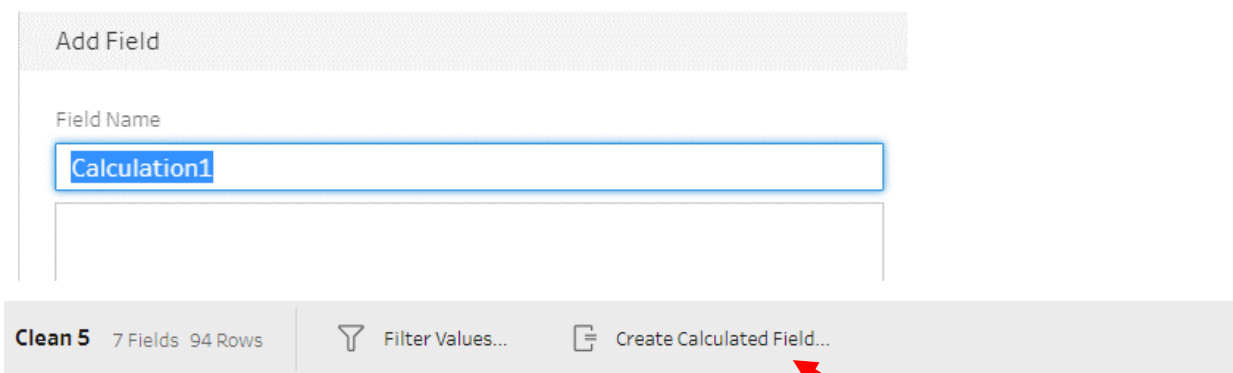
The combined first, second and third stream of data, now combined in step Join 2, also needs a field that combines the “Product ID” and the “Region” in the same exact format that was used in the fourth stream. This new field is referred to as a foreign key and will be used to link to the data in the third stream.

Add step “Clean 6” to the “Join 2” step.



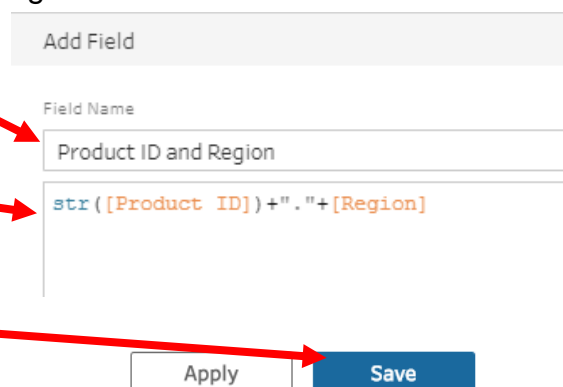
Since step “Join 2” contains both “Product #” and “Product ID” either can be used in creating the new field.

To create a Foreign Key field, select “Create Calculated Field”.



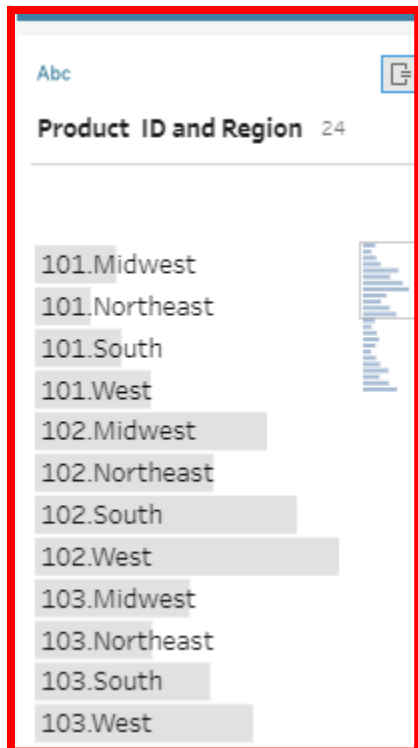
Change the “Field Name” to “Product ID and Region”.

Type: **STR([Product ID]) + "." + [Region]**



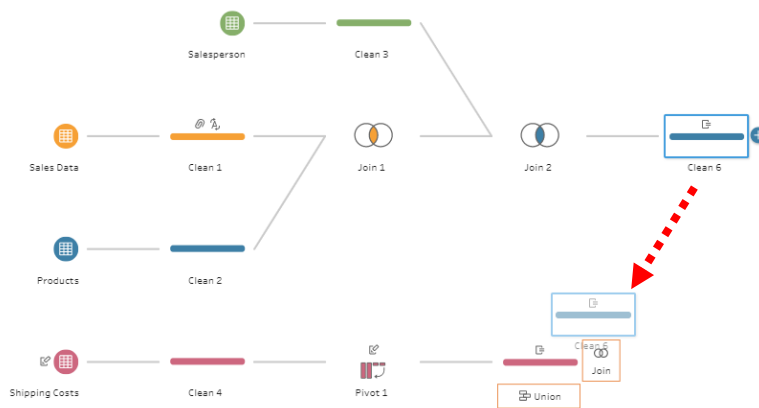
then click Save.

The “SPID and Region” field has been populated with 24 different values for “Product ID and Region”.



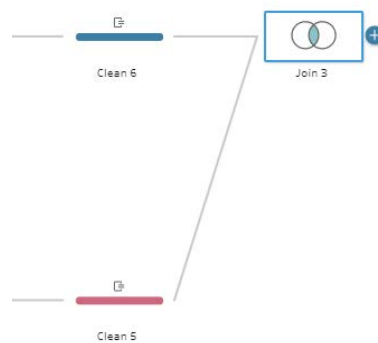
## Step 14: Combine First Three Streams of Data to the Fourth Stream of Data

Now that we have a primary key and a foreign key, we can join the fourth stream of data with the combined first three streams. Drag the “Clean 6” step from the combined first, second and third stream to the “Clean 5” step in the fourth stream. Drop the “Clean 6” step into the Join box.



Step “Join 3” will appear. If the Join is named differently, double click on the name and change it.

Select the “Join 3” step.



**Join 3** 19 fields 20K rows Filter Values...

Settings Changes (0)

**Applied Join Clauses**

Clean 5  
Product ID = Product ID

**Join Type: inner**

Click the graphic to change the join type.

Clean 5 Clean 6

**Summary of Join Results**

Click the bar segments to view the included and excluded values.

Mismatched values

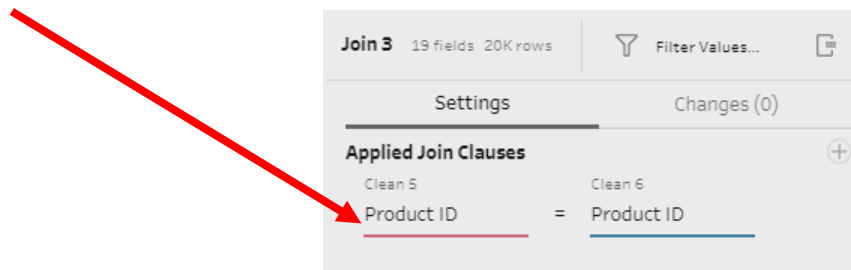
	Included
Clean 5	24
Clean 6	5,017
<b>Join Result</b>	<b>20,068</b>

**Join Clause Recommendations**

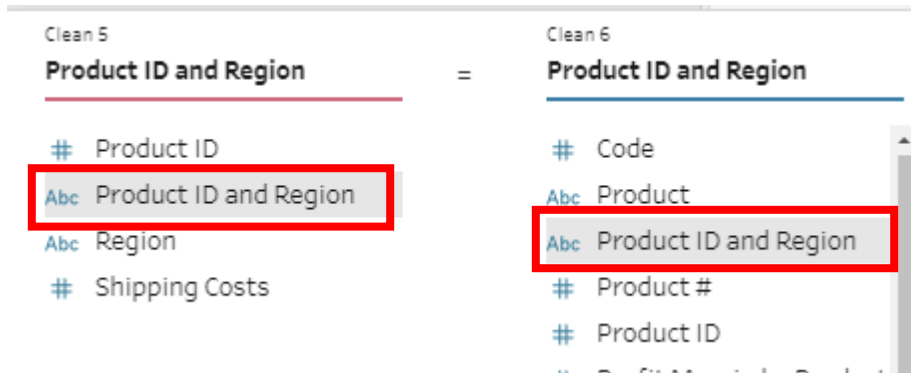
The join is incorrect because the result is **20,068** rows and there are supposed to be **5,017** rows.



Select "Product ID" in the "Applied Join Clauses" section.

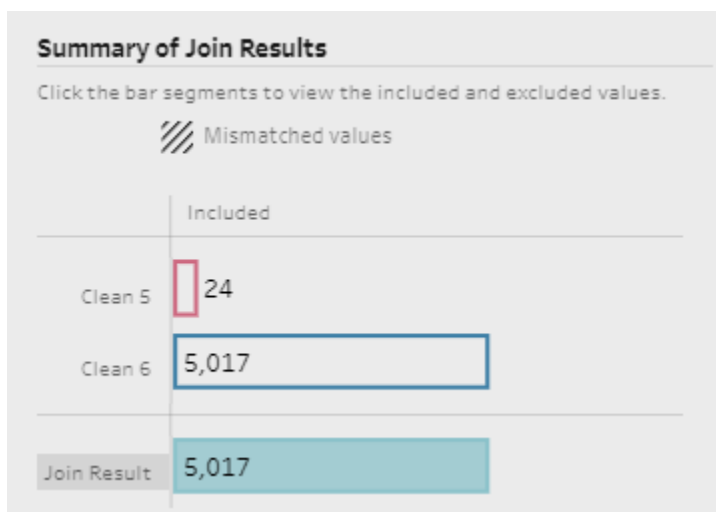


Change the fields from "Product ID" to "Product ID and Region" in both "Clean 5" and "Clean 6" steps.



The default "Join Type" is "inner" which does not have to be changed.

The join was successful as 5,017 rows were combined.



**Join 3** 19 fields 5K rows Filter Values... Create Calculated Field...


**Settings** Changes (0)

**Applied Join Clauses** +

Clean 5 = Clean 6

**Join Type: inner**

Click the graphic to change the join type.



**Summary of Join Results**

Click the bar segments to view the included and excluded values.

Mismatched values

	Included
Clean 5	24
Clean 6	5,017
<b>Join Result</b>	<b>5,017</b>

**Join Clause Recommendations**

Product ID = Product ID  
Region = Region  
Product ID = Product #

**Join Clauses** Show only mismatched values

**Clean 5** Product ID and Region

- 101.Midwest
- 101.Northeast
- 101.South
- 101.West
- 102.Midwest
- 102.Northeast
- 102.South
- 102.West
- 103.Midwest
- 103.Northeast
- 103.South
- 103.West
- 104.Midwest
- 104.Northeast
- 104.South
- 104.West
- 105.Midwest
- 105.Northeast
- 105.South
- 105.West
- 106.Midwest
- 106.Northeast

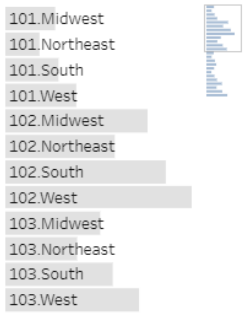
**Clean 6** Product ID and Region

- 101.Midwest
- 101.Northeast
- 101.South
- 101.West
- 102.Midwest
- 102.Northeast
- 102.South
- 102.West
- 103.Midwest
- 103.Northeast
- 103.South
- 103.West
- 104.Midwest
- 104.Northeast
- 104.South
- 104.West
- 105.Midwest
- 105.Northeast
- 105.South
- 105.West
- 106.Midwest
- 106.Northeast

**Join Results**

Abc

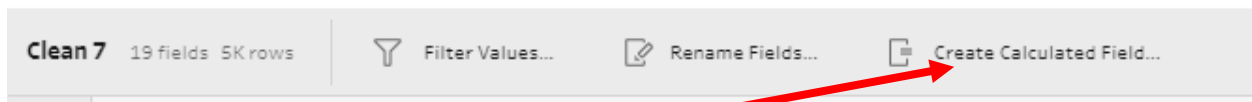
**Product ID and Region** 24



Product ID and Region	Product I
103.West	103.West
101.West	101.West
102.West	102.West
101.West	101.West
106.Midwest	106.Midw
103.West	103.West
106.Northeast	106.Nortl
105.West	105.West

**Step 15: Calculate the Sales and the Shipping Cost on Every Sales Order**

Add a step “Clean 7” after the step “Join 3” and calculate the sales for every sales order.



Click on “Create Calculated Field”

Enter “Sales” as the  
“Field Name”.

The screenshot shows the 'Add Field' dialog box. The 'Field Name' field contains the text 'Sales'. Below it, the formula field contains the text `Round([Units Sold]*[Unit Selling Price],2)`. A red arrow points to the 'Field Name' field, and another red arrow points to the formula field.

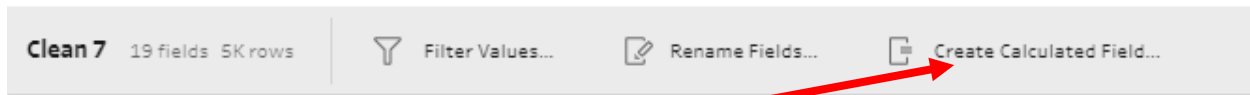
Type the formula  
`Round([Units Sold]*[Unit Selling Price],2)`

then click Save.

Save

Sales	Product ID and Region	Product ID and Region	Region	Shipping Costs	Sales Order #	Salesperson #	Region-1	State	Product #	Units Sold	Product ID-1	Product	Unit Selling Price
2,880	103.West	103.West	West	2.01	35,005	1,303	West	WA	103	120	103	Stuffed Elephant	24

$$120 \times 24 = 2,880.$$



Click on "Create Calculated Field".

Enter "Shipping by Sales Order" as the "Field Name".

Edit Field

Field Name

Shipping by Sales Order

`Round([Units Sold]*[Shipping Costs],2)`

Type the formula `Round([Units Sold]*[Unit Selling Price],2)`

then click Save.

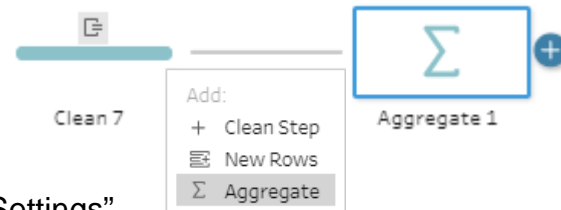


Shipping by Sales Order	Sales	Product ID and Region	Product ID and Region	Region	Shipping Costs	Sales Order #	Salesperson #	Region-1	State	Product #	Units Sold	Product ID-1
241.2	2,880	103.West	103.West	West	2.01	35,005	1,303	West	WA	103	120	103

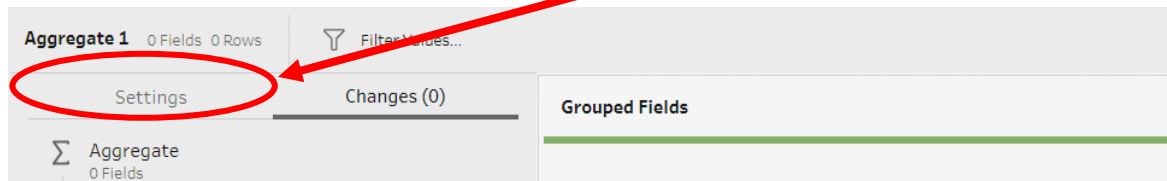
$$120 \times 2.01 = 241.2$$

**Step 16: Create a “Units Sold, Sales and Shipping Costs by Salesperson” Report**

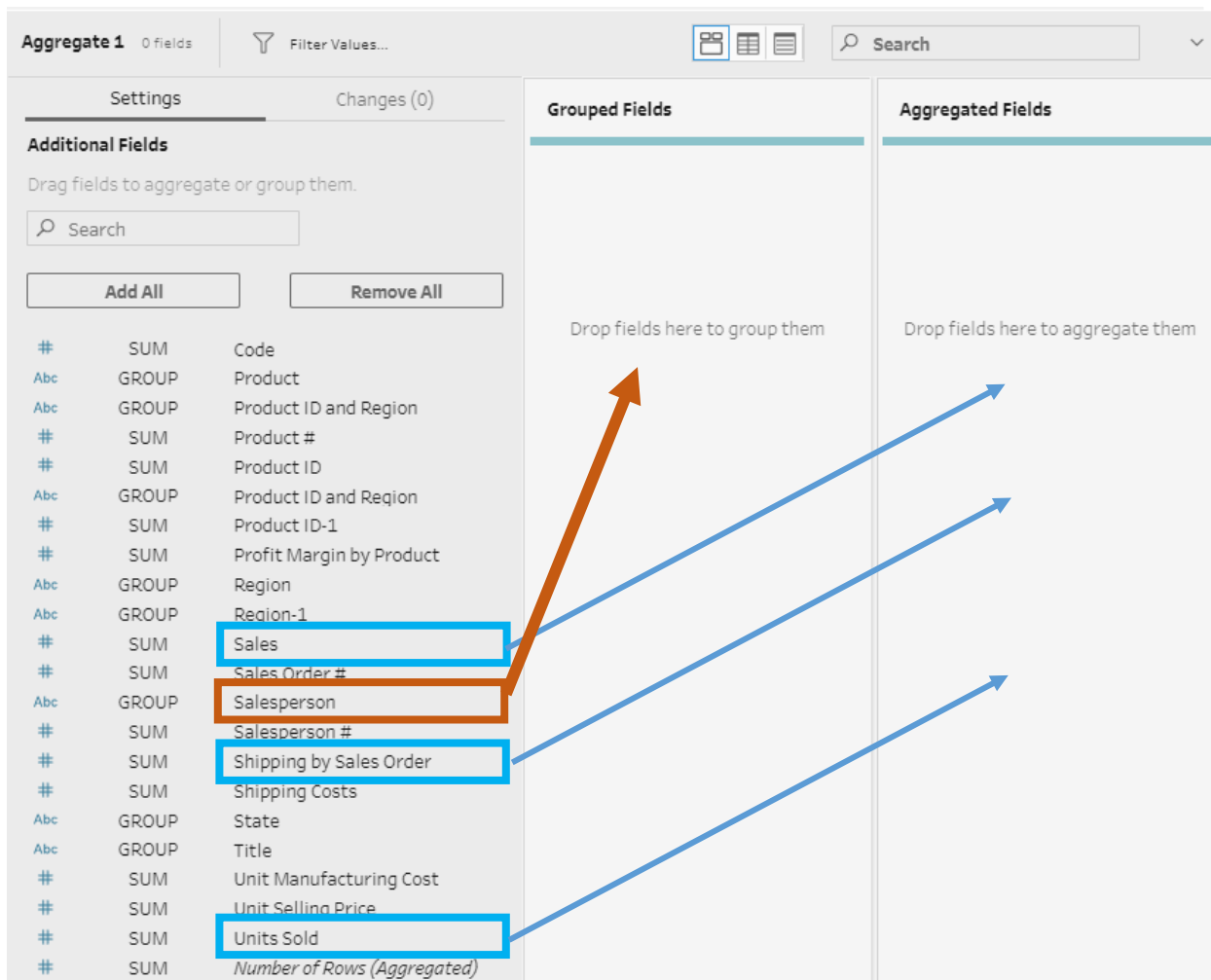
Add a step “Aggregate 1” after the step “Clean 7”.



Select the “Aggregate 1” step, then select “Settings”.



Drag the “Salesperson” field to the “Grouped Fields” area and “Sales”, “Shipping Costs by Sales Order” and “Units Sold” fields to the “Aggregated Fields” area.

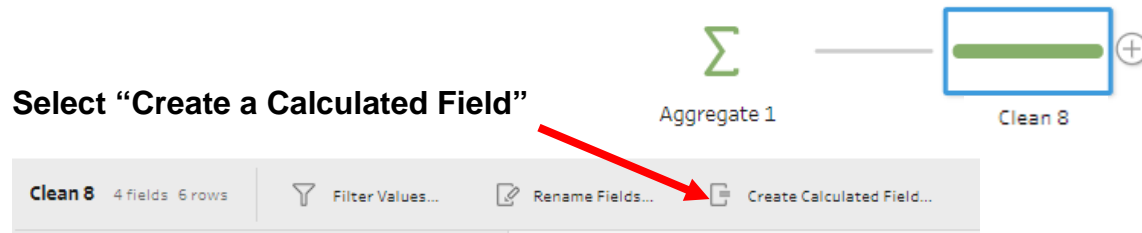


The decimal places in the “Shipping by Sales Order” field are not uniform.

Aggregated Fields					
#	SUM	#	SUM	#	SUM
Shipping by Sales Order 6		Sales 6		Units Sold 6	
87,710.639999999996		985,560		43,200	
117,878.399999999997		1,317,384		57,924	
120,158.04		1,349,352		59,136	
156,215.999999999968		1,733,088		76,680	
246,882.35999999989		2,748,540		121,236	
287,129.759999999995		3,210,924		141,156	

**Step 17: Round the Shipping Costs**

Add a step “Clean 8” after the step “Aggregate 1”.



Select “Create a Calculated Field”

Add Field

Field Name

Calculation1

Put the result back in the “Shipping Cost” field by naming the field “Shipping Cost”.

Edit Field

Field Name

Shipping by Sales Order

round([Shipping by Sales Order],2)

Reference

All

Search

ABS

ACOS

AND

ASC

DESC

ABS(number)

Returns the absolute value of the given number.

Example: ABS(-7) = 7

Calculation is valid ^

Apply Save

Type in the formula, Round([Shipping by Sales Order],2) and select “Save”.

The data in the “Shipping by Sales Order” field now has two decimal places.

Clean 8 4 fields 6 rows

Filter Values...

Changes (1)

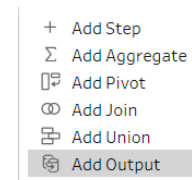
#	Shipping by Sales Order
1	87,710.64
2	117,878.4
3	120,158.04
4	156,216
5	246,882.36
6	287,129.76

**Step 18: Export Report to Excel**

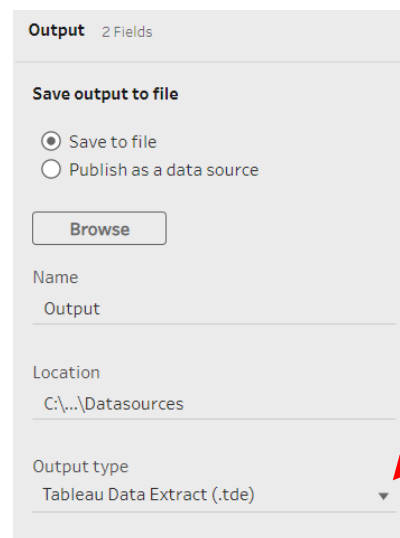
Add a step “Output” after the step “Clean 8”.



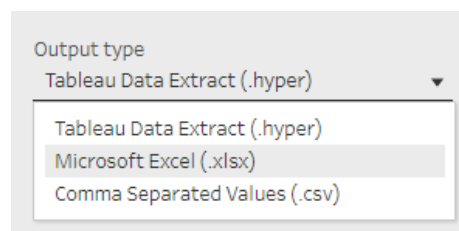
Select the step “Output”.



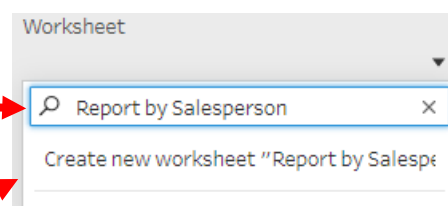
Select the “Output type” drop down.



Select the Microsoft Excel (.xlsx) file type.



Type the name of the worksheet, Report by Salesperson.

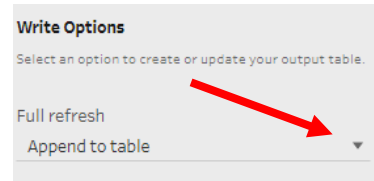


Select “Create new worksheet “Report by Salesperson”.

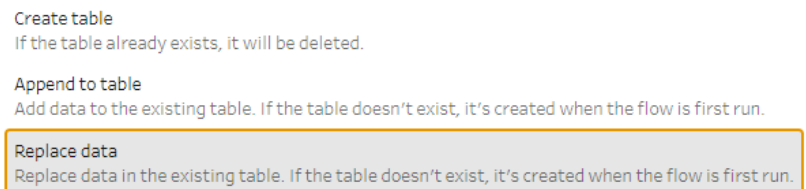




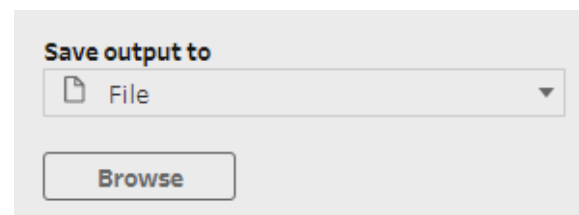
Select the “Full refresh” drop down.



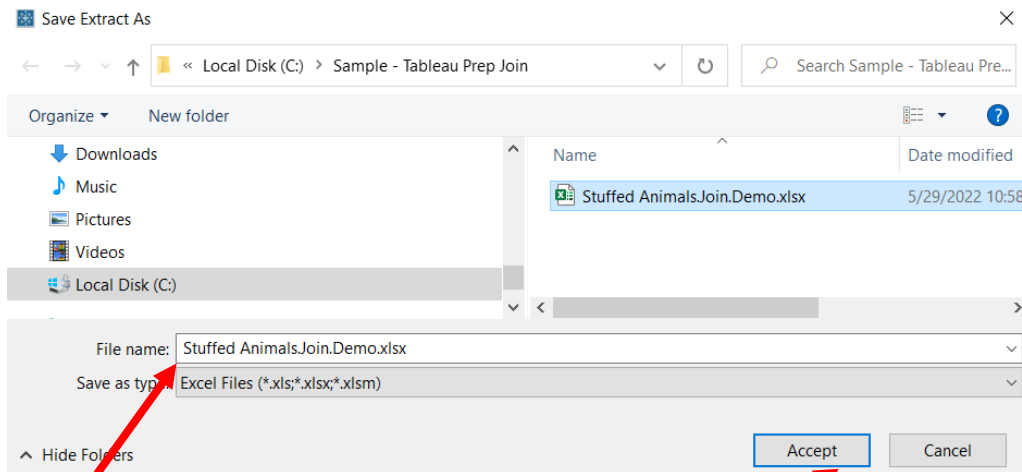
Select “Replace data”.



“Save output to” should be “File”.

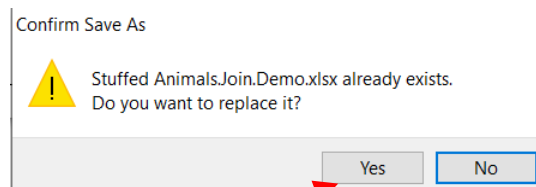


Select “Browse” button.



Select the location for the file, specify the file name and select “Accept”.

Confirm.



Output 4 fields

Save output to  
File

Name  
Stuffed Animals.Join.Demo

Location  
C:\Sample - Tableau Prep Join

Output type  
Microsoft Excel (.xlsx)

Worksheet  
Report by Salesperson

Write Options  
Select an option to create or update your output table.

Full refresh  
Replace data

Run Flow

Save to Stuffed Animals.Join.Demo.xlsx

From: Flow		To: Table	
Type	Field Name	Type	Field Name
#	Shipping by Sales Order	→	# Shipping by Sales Order
#	Sales	→	# Sales
Abc	Salesperson	→	Abc Salesperson
#	Units Sold	→	# Units Sold

Select "Run Flow".

### 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	C:\Sample - Tableau Prep Join\Stuff...	Report by Salesperson

Cancel Replace

If asked, select "Replace".

### Finished Running Flow

Select "Done".



Report by Salesperson.csv

Total time 00:00

Done

Open the *Stuffed Animals.Join.Demo.xlsx* file and select the Report by Salesperson sheet.

	A	B	C	D	E
1	Shipping by Sale Sales		Salesperso	Units Sold	
2	117878.4	1317384	Ulysses Gr	57924	
3	246882.36	2748540	James Moi	121236	
4	120158.04	1349352	James Mac	59136	
5	87710.64	985560	James Polk	43200	
6	287129.76	3210924	Thomas Je	141156	
7	156216	1733088	Grover Cle	76680	
8					
9					
10					

Report by Salesperson

Make any format changes and change the name of the tab so that it will not be automatically overwritten by Tableau.

	A	B	C	D	E
1	Salesperson	Units Sold	Sales	Shipping Costs	
2	Ulysses Grant	57,924	\$ 1,317,384.00	\$ 117,878.40	
3	James Monroe	121,236	\$ 2,748,540.00	\$ 246,882.36	
4	James Madison	59,136	\$ 1,349,352.00	\$ 120,158.04	
5	James Polk	43,200	\$ 985,560.00	\$ 87,710.64	
6	Thomas Jefferson	141,156	\$ 3,210,924.00	\$ 287,129.76	
7	Grover Cleveland	76,680	\$ 1,733,088.00	\$ 156,216.00	
8	Total	499,332	\$ 11,344,848.00	\$ 1,015,975.20	
9					
10					

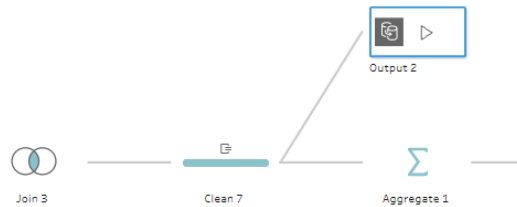
Formatted Report by Salesperson

Save and close the Excel file and return to Tableau.

**Step 19: Create a Shipping Cost report by Salesperson by Region.**

It is going to be much easier to create the “Commission by Salesperson by Region” report in Excel since all the data has already been transformed.

Add a step “Output 2” after the step “Clean 7”.



Select the step “Output 2”



Select the “Output type” drop down.

**Output** 2 Fields

**Save output to file**

☒ Save to file  
☐ Publish as a data source

**Browse**

Name  
Output

Location  
C:\...\Datasources

Output type  
Tableau Data Extract (.tde)

Select the Microsoft Excel (.xlsx) file type.

Output type  
Tableau Data Extract (.hyper)

Tableau Data Extract (.hyper)  
 Microsoft Excel (.xlsx)  
 Comma Separated Values (.csv)

Type the name of the worksheet, “Data for Shipping Report”.

Select “Create new worksheet “Data for Shipping Report”.

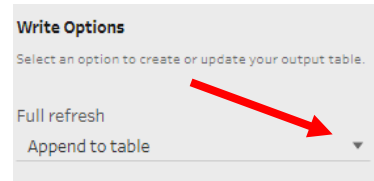
Output type  
Microsoft Excel (.xlsx)

Worksheet

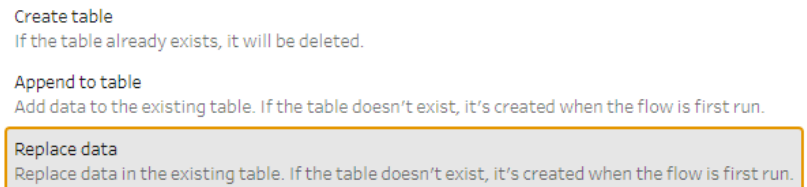
Data for Shipping Report

Create new worksheet “Data for Shipping Report”

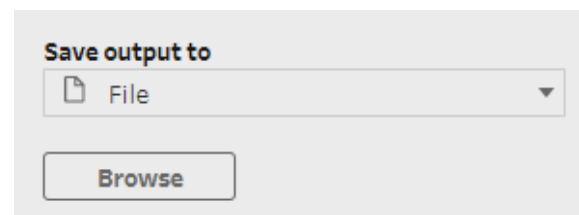
Select the “Full refresh” drop down.



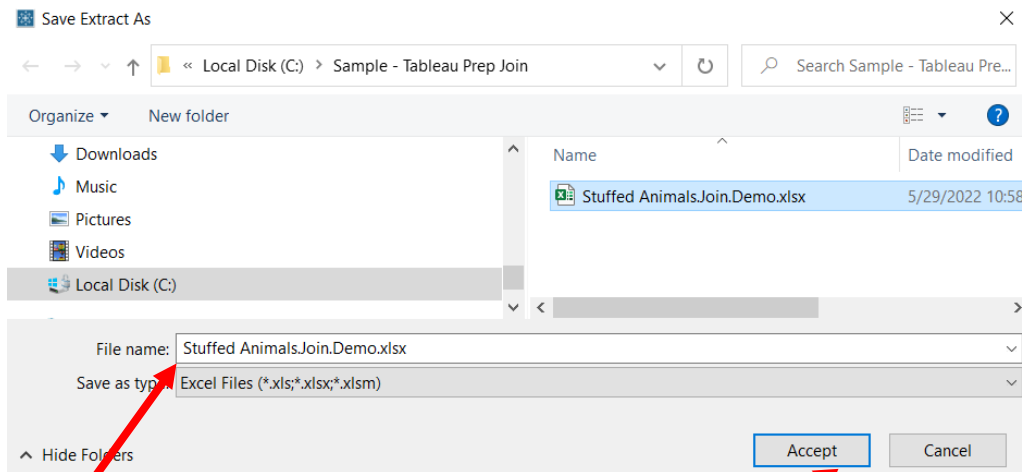
Select “Replace data”.



“Save output to” should be “File”.

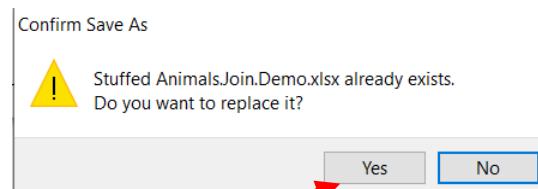


Select “Browse” button.



Select the location for the file, specify the file name and select “Accept”.

Confirm.



**Output 2** 21 fields

**Save output to**

File

Name  
Stuffed Animals.Join.Demo

Location  
C:\Sample - Tableau Prep Join

Output type  
Microsoft Excel (.xlsx)

Worksheet  
Data for Shipping Report

**Write Options**  
Select an option to create or update your output table.

Full refresh  
Replace data

**Run Flow**

**Save to Stuffed Animals.Join.Demo.xlsx**

From: Flow		To: Table		Status
Type	Field Name	Type	Field Name	Status
#	Shipping by Sales Order		No table field assigned	No match: Field is ignored.
#	Sales		No table field assigned	No match: Field is ignored.
Abc	Product ID and Region		No table field assigned	No match: Field is ignored.
Abc	Product ID and Region		No table field assigned	No match: Field is ignored.
Abc	Region		No table field assigned	No match: Field is ignored.
#	Shipping Costs		No table field assigned	No match: Field is ignored.
#	Sales Order #		No table field assigned	No match: Field is ignored.
#	Salesperson #		No table field assigned	No match: Field is ignored.
Abc	Region-1		No table field assigned	No match: Field is ignored.
Abc	State		No table field assigned	No match: Field is ignored.
#	Product #		No table field assigned	No match: Field is ignored.
#	Units Sold		No table field assigned	No match: Field is ignored.
#	Product ID-1		No table field assigned	No match: Field is ignored.
Abc	Product		No table field assigned	No match: Field is ignored.
#	Unit Selling Price		No table field assigned	No match: Field is ignored.

Select "Run Flow".

### 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	C:\Sample - Tableau Prep Join\Stuff...	Data for Shipping Report

Cancel Replace

If asked, select "Replace"

## Finished Running Flow



Select "Done".

**Report by Salesperson.csv**

Total time 00:00

Done

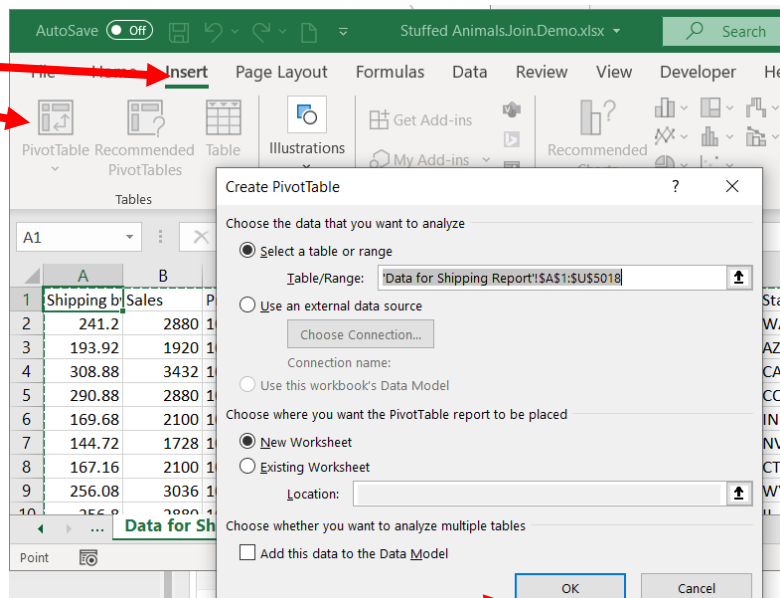
Open the *Stuffed Animals.Join.Demo.xlsx* file and select the Data for Shipping Report sheet.

Select a cell, perhaps A1, in the data range on the “Data by Shipping Report”.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Shipping b	Sales	Product ID	Product ID	Region	Shipping C	Sales Orde	Salesperso	Region-1	State	Product #	Units Sold	Product
2	241.2	2880	103.West	103.West	West	2.01	35005	1303	West	WA	103	120	1
3	193.92	1920	101.West	101.West	West	2.02	35006	1302	West	AZ	101	96	1
4	308.88	3432	102.West	102.West	West	1.98	35009	1305	West	CA	102	156	1
5	290.88	2880	101.West	101.West	West	2.02	35011	1305	West	CO	101	144	1
6	169.68	2100	106.Midw	106.Midw	Midwest	2.02	35014	1302	Midwest	IN	106	84	1
7	144.72	1728	103.West	103.West	West	2.01	35017	1305	West	NV	103	72	1
8	167.16	2100	106.North	106.North	Northeast	1.99	35025	1301	Northeast	CT	106	84	1
9	256.08	3036	105.West	105.West	West	1.94	35026	1306	West	WY	105	132	1
10	256.08	2880	103.Midw	103.Midw	Midwest	2.14	35027	1301	Midwest	IL	103	120	1

Select Insert.

Select Pivot table.



This will insert a Pivot Table in a new sheet. Pivot tables help summarize the data in various formats.

Select OK.



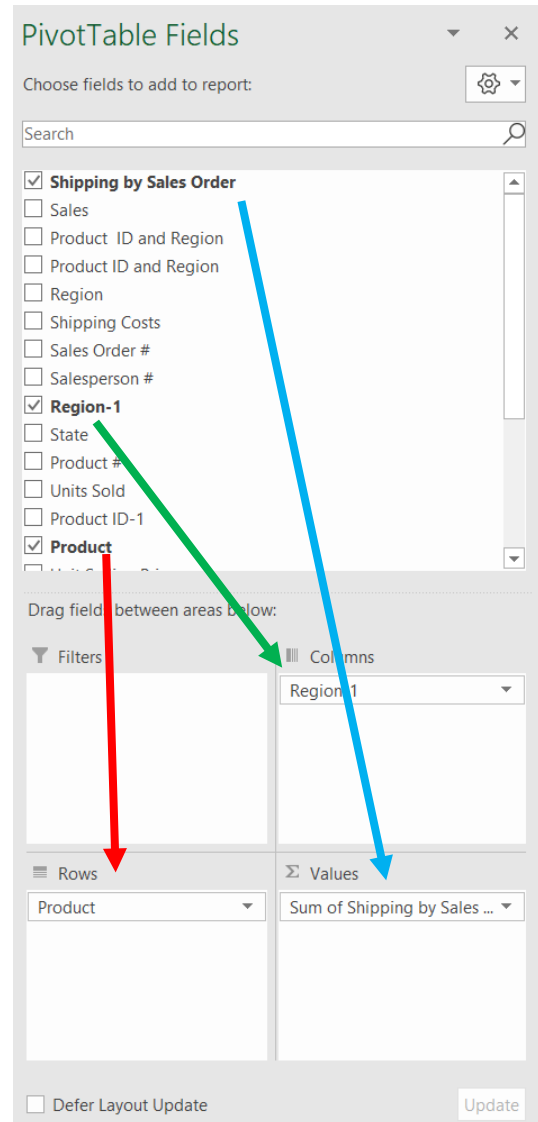
Once a Pivot table is inserted, the Pivot Table Fields are displayed.

Drag the fields

1. **Product to Rows**
2. **Region-1 to Columns** and
3. **Shipping by Sales Order to Values**

As indicated in the diagram.

If “Count of Shipping ...” appears instead of the “Sum of Shipping ...”, left click on “Count of Shipping ...”, select Value Field Setting > Sum > OK.



This summarizes the data in the following format:

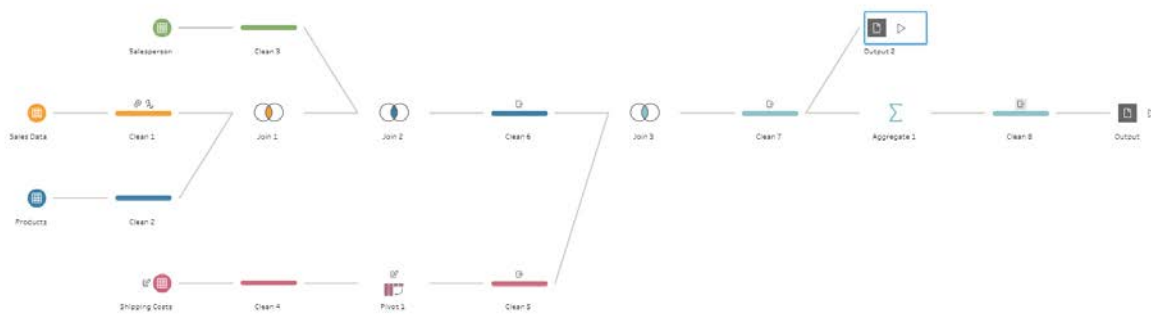
	A	B	C	D	E	F
1						
2						
3	Sum of Shipping by Sales Order	Column Labels				
4	Row Labels	Midwest	Northeast	South	West	Grand Total
5	Stuffed Elephant	47610.72	35625.24	50822.64	63990.36	198048.96
6	Stuffed Giraffe	76744.92	56958.72	80686.08	91095.84	305485.56
7	Stuffed Horse	26032.32	17062.56	25488	31916.88	100499.76
8	Stuffed Lamb	30263.4	19614.24	29877.12	38905.2	118659.96
9	Stuffed Pig	47898.24	30566.4	47361.6	59648.4	185474.64
10	Stuffed Unicorn	25999.68	18458.88	29987.4	33360.36	107806.32
11	Grand Total	254549.28	178286.04	264222.84	318917.04	1015975.2
12						
<div> <span>◀ ▶ ...</span> <span>Sheet1</span> <span>Data for Shipping Report</span> <span>+</span> </div>						

Right click on cell A3, number format > currency > 2.  
Rename “Sheet1” to “Shipping Report”

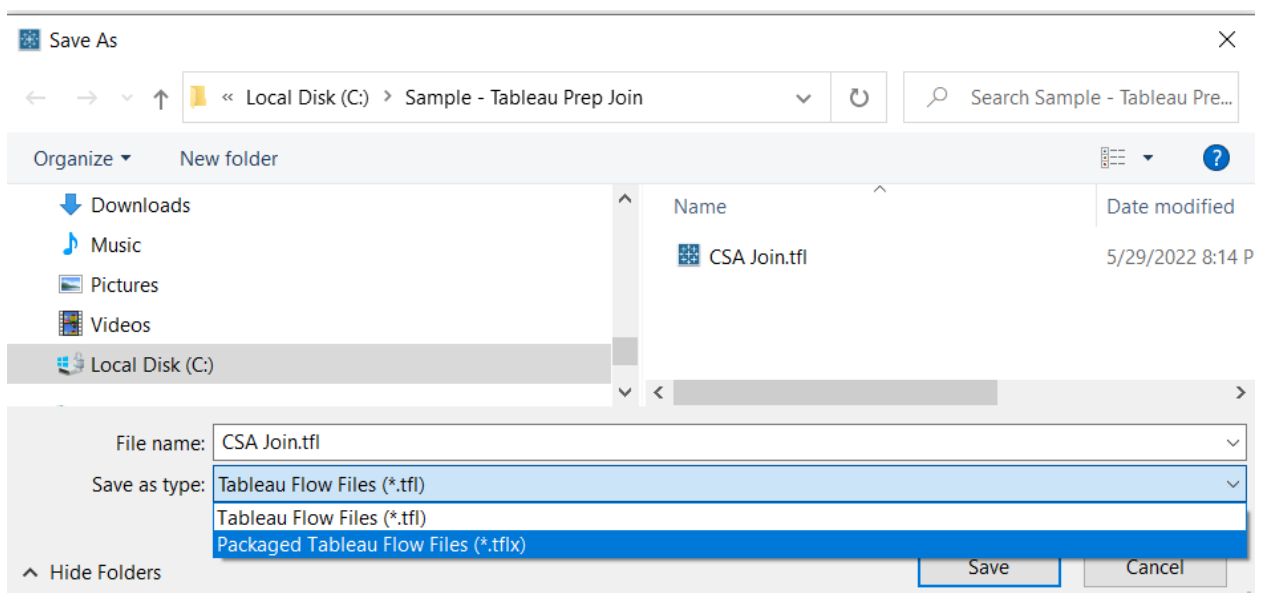
	A	B	C	D	E	F
1						
2						
3	Sum of Shipping by Sales Order	Column Labels				
4	Row Labels	Midwest	Northeast	South	West	Grand Total
5	Stuffed Elephant	\$47,610.72	\$35,625.24	\$50,822.64	\$63,990.36	\$198,048.96
6	Stuffed Giraffe	\$76,744.92	\$56,958.72	\$80,686.08	\$91,095.84	\$305,485.56
7	Stuffed Horse	\$26,032.32	\$17,062.56	\$25,488.00	\$31,916.88	\$100,499.76
8	Stuffed Lamb	\$30,263.40	\$19,614.24	\$29,877.12	\$38,905.20	\$118,659.96
9	Stuffed Pig	\$47,898.24	\$30,566.40	\$47,361.60	\$59,648.40	\$185,474.64
10	Stuffed Unicorn	\$25,999.68	\$18,458.88	\$29,987.40	\$33,360.36	\$107,806.32
11	Grand Total	\$254,549.28	\$178,286.04	\$264,222.84	\$318,917.04	\$1,015,975.20
12						
<div> <span>◀ ▶ ...</span> <span>Shipping Report</span> <span>Data for Shipping Report</span> <span>+</span> </div>						

Save and close the Excel file.

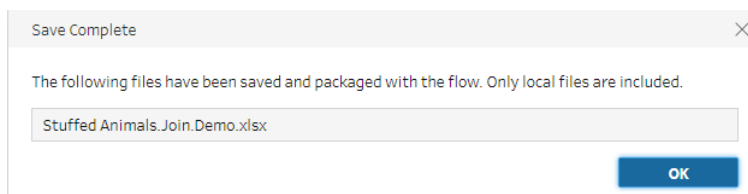
Return to the Tableau Prep flow file.



Select File > Save As..



Select the Packaged Tableau Flow Files (\*.tflx) which bundles the data file with the flow.



Select OK.

Close Tableau.