MIS0855: Data Science
In-Class Exercise for Wed, Apr 17 – Simple Predictive Analytics Using Tableau

Objective: Analyze a data set to make inferences about future outcomes

Learning Outcomes:

- Forecast future sales based on order transaction data
- Perform association analysis to determine which products are purchased together
- Interpret the meaning of the results from these analyses

In this exercise, you'll once again be working with a data set of orders for an imaginary company, Vandelay Industries.

The data set contains 102,531 line items for 60,011 orders placed between January 1, 2009 and December 31, 2013.

Part 1: Download the data file

1) Download "VandelayOrdersAll.xlsx" and save it to your computer.

2) Open the data file in Excel. Take a quick look through the data and the Data Dictionary tab.

Part 2: Forecast future sales in Tableau

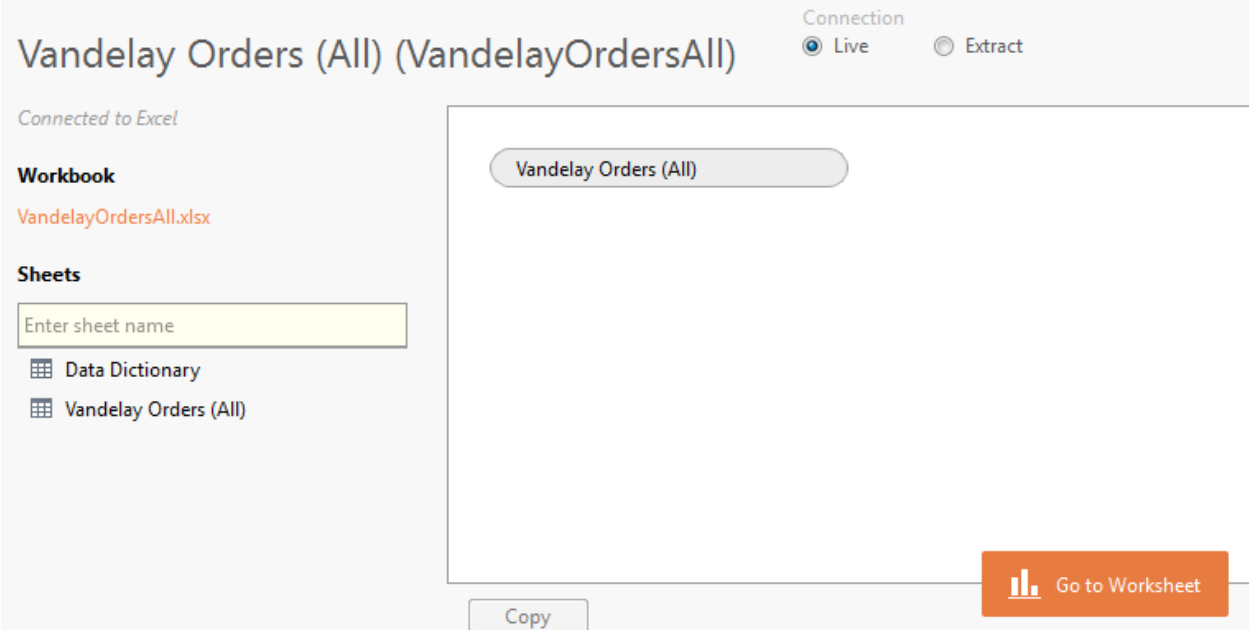
The first thing we'll do is use Tableau to predict "future" sales based on daily sales from 2009 through 2013. Tableau has a forecasting feature built in, so it's easy to do.

1) Start Tableau and click "Connect to data."

2) Click "Microsoft Excel."

3) Open "VandelayOrdersAll.xslx."

4) Drag the “Vandelay Orders (All)” sheet to the whitespace. Wait for the data to show up and click “Go to Worksheet.”

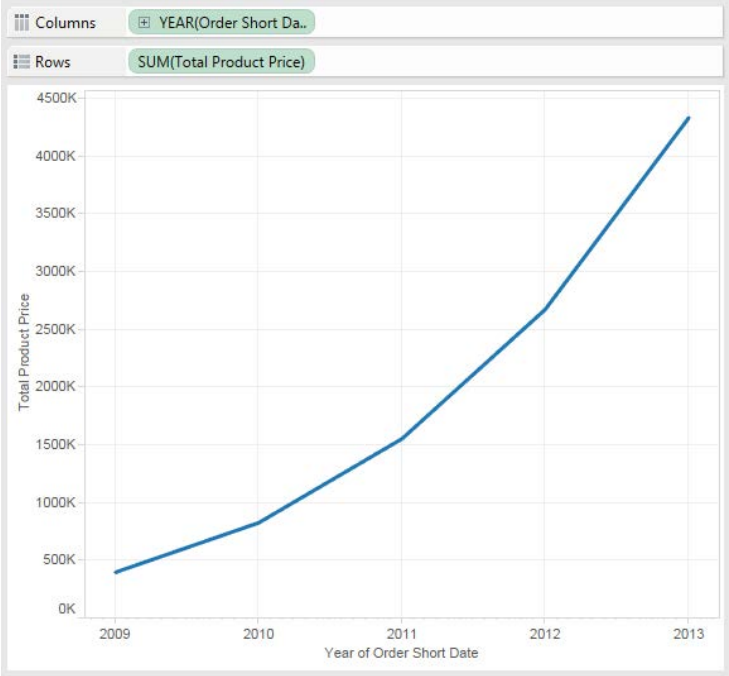


5) Drag the “Order Short Date” dimension to the Columns shelf and “Total Product Price” to the Rows shelf.

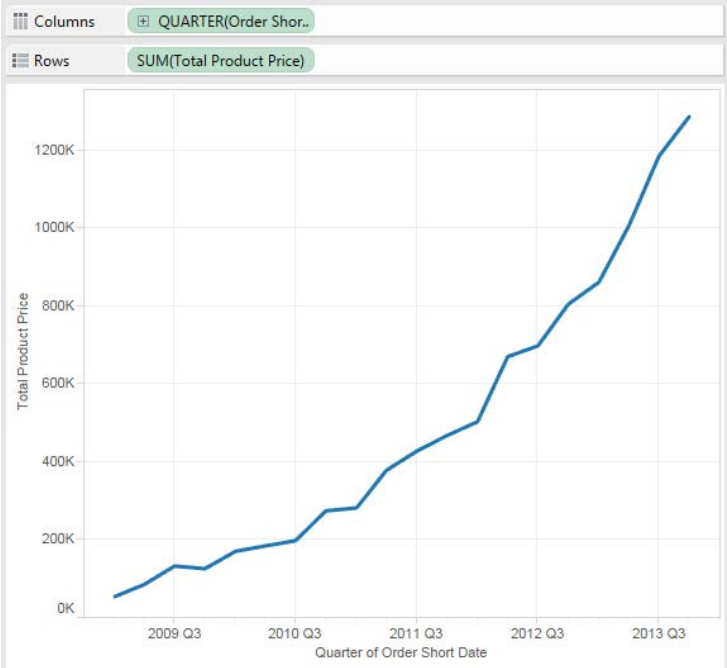
6) Click the line graph under the “Show Me” area.

7) You'll see a line graph of the year-to-year aggregate sales.

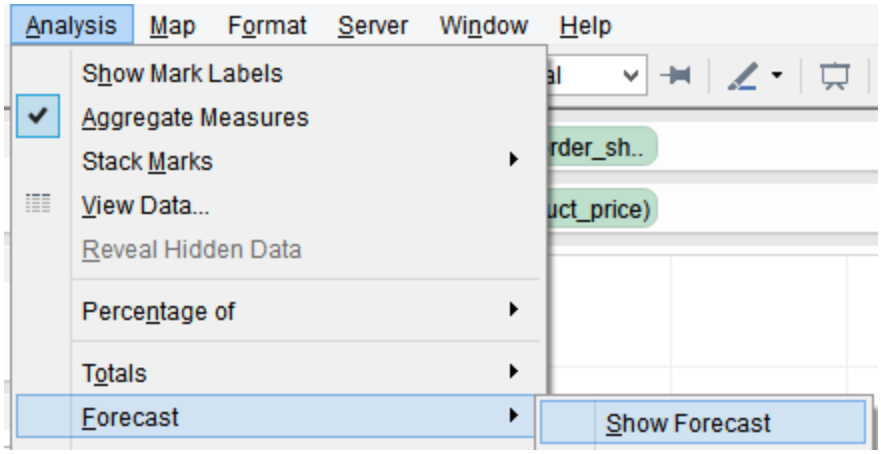
   Notice Order Short Date appears as YEAR(Order Short Date). Tableau automatically presents dates as hierarchies so you can drill down to Quarter (or Month or Day.)



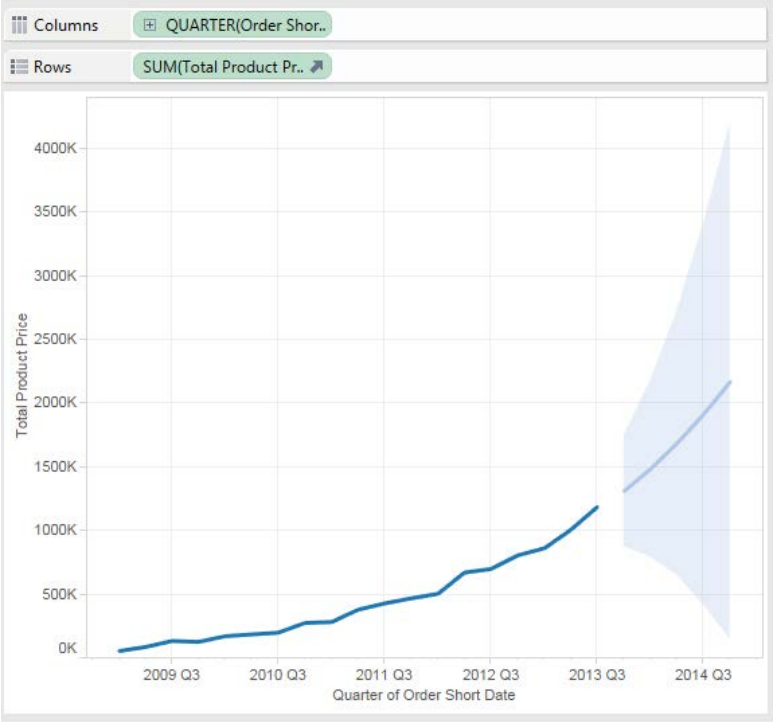
8) Click on the plus sign next to YEAR to drill-down to quarters. You'll see this:



9) Now we can run a forecast by selecting Analysis menu and then Forecast/Show Forecast.



You'll see this:

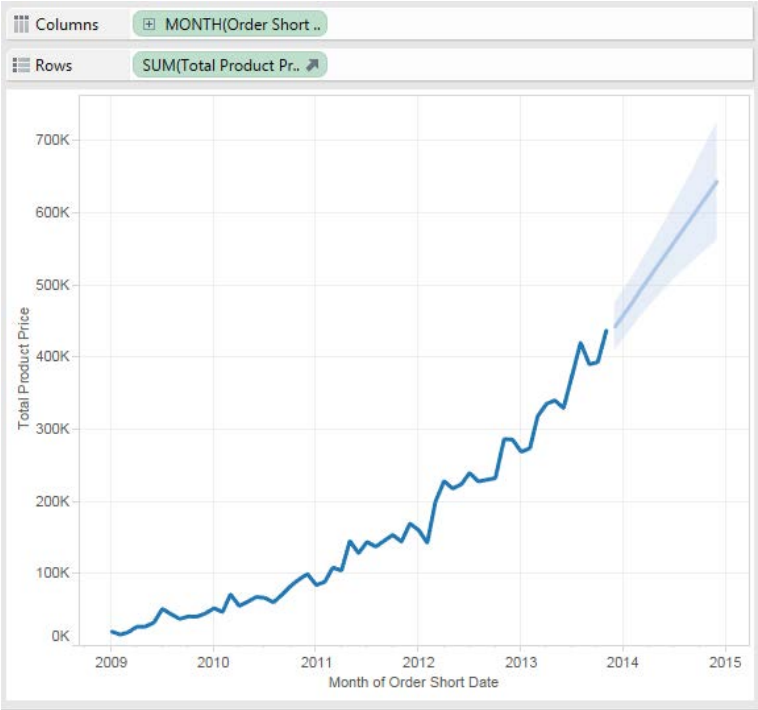


There is a gap because Tableau doesn't count the last data period in its analysis. In this case, our last data period is the fourth quarter of 2013.

Let's talk about some other aspects of this chart.

- The solid line to the right of the gap are the forecasted values – the prediction of future sales.
- The shaded area is the 95% prediction interval. This means that the actual values will fall somewhere in the shaded range 95% of the time. Note that the solid line is right down the middle of the prediction interval.
- The prediction interval is pretty wide – this means it is difficult for Tableau to be confident about its prediction using quarterly data. There's just not enough of it to make a good prediction.

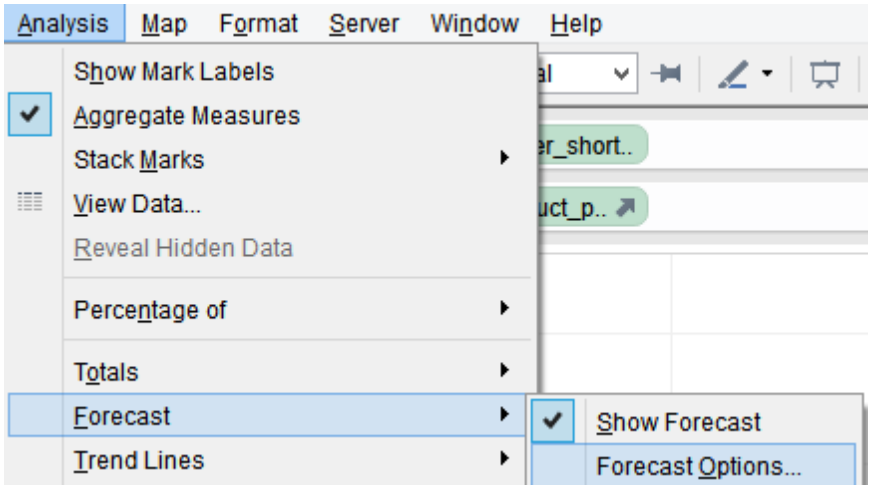
10) Click on the plus sign next to QUARTER to drill down to MONTH. You'll see this:



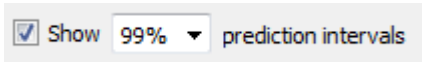
Notice hat the prediction interval is **much** narrower. The main reason for this is Tableau has much more data to work with (60 months instead of 20 quarters).

<u>The more data points you use, the better your predictions become.</u>

11) Let's change the confidence level of the prediction interval. Go to Analysis menu and select Forecast/Forecast Options.



12) Change the prediction interval to 99%. Then click "OK."



13) You'll see the prediction interval get slightly wider, since now you're asking Tableau to present a range of values that will contain the actual value 99% of the time (instead of 95%).

To see why this is true, think about a game where you throw crumpled-up paper into a wastebasket. Say you successfully get the paper into the wastebasket 95% of the time. If you want to make sure you get it into the wastebasket 99% of the time, one option is to buy a larger wastebasket!

A larger prediction interval is like a larger wastebasket.

14) Save your Tableau workbook and close it.

Part 3: Perform an association analysis in Tableau
(Adapted from kb.tableausoftware.com/articles/knowledgebase/market-basket-analysis)

An association analysis is discovering which events occur at the same time. In this case, we're looking for which products are purchased together (within the same order).

Tableau doesn't have an "association analysis" function, but with some clever table joining, we can do a simple version of the type of analyses more sophisticated data mining programs do.

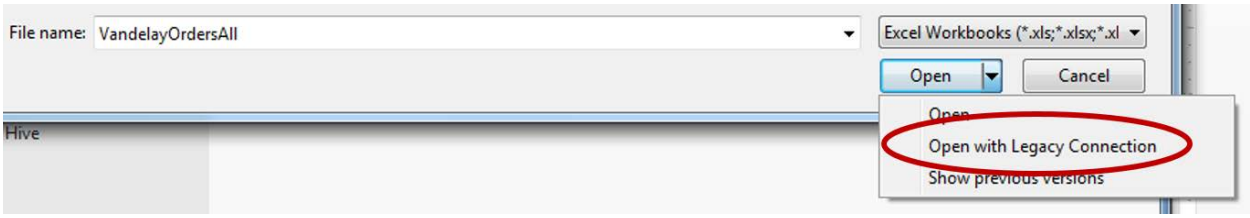
1) Open Tableau again. Make sure you're starting a new Tableau file.

2) Click "Connect to data."

3) Click "Microsoft Excel."

4) Click ONCE on VandelayOrdersAll.xlsx. <u>Just select the file – don't open it!</u>

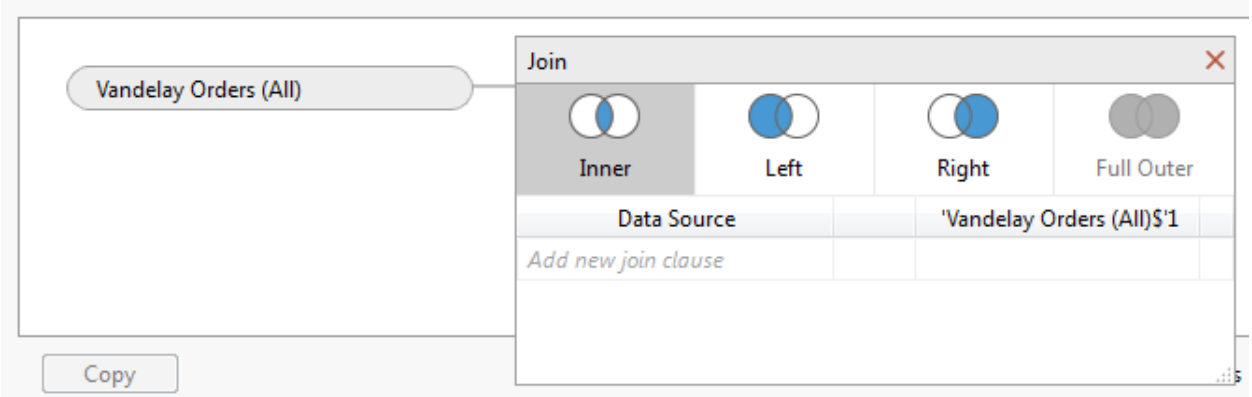
5) Click the "down arrow" next to Open and select "Open with Legacy Connection."



6) Drag the "Vandelay Orders (All)" sheet to the whitespace.

7) Again, drag the "Vandelay Orders (All)" sheet to the whitespace one more time time. It should look like this:

but the Join dialog may cover up the second "Vandelay Orders (All)" sheet.



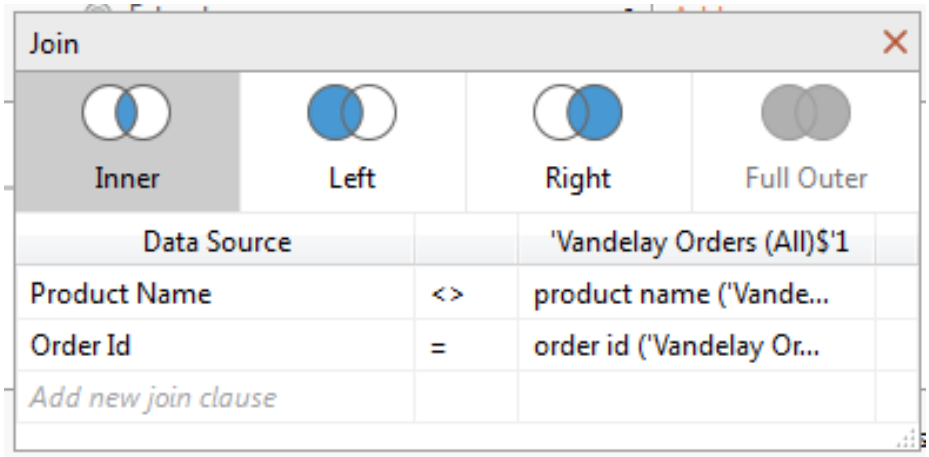
8) If you don't see the join dialog, click on the join area between the two sheets:

9) You'll create two joins:

Select Product Name from Data Source and 'Valenday Orders (All)$'1'
Select the "<>" symbol from the middle drop-down box.

Select Order ID from Data Source and 'Vandelay Orders (All)$'1'
Select the "=" symbol from the middle drop-down box.

It should look EXACTLY like this:



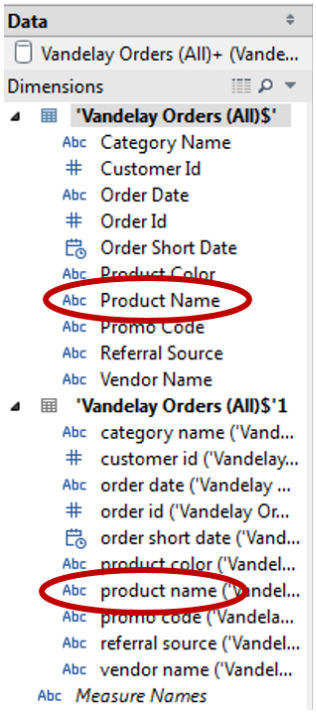
So what does this mean? It's called a self-join – you're connecting the table with itself.

You're asking Tableau to match up any combination of different products
(Product Name <> Product Name)
that are part of the same order
(Order Id = Order Id).

10) When you have this set up like the image above, click "Go to Worksheet."

11) Drag the “Product Name” dimension from ‘Vandelay Orders(All)$’ (from the first set of dimensions) to the Columns shelf.

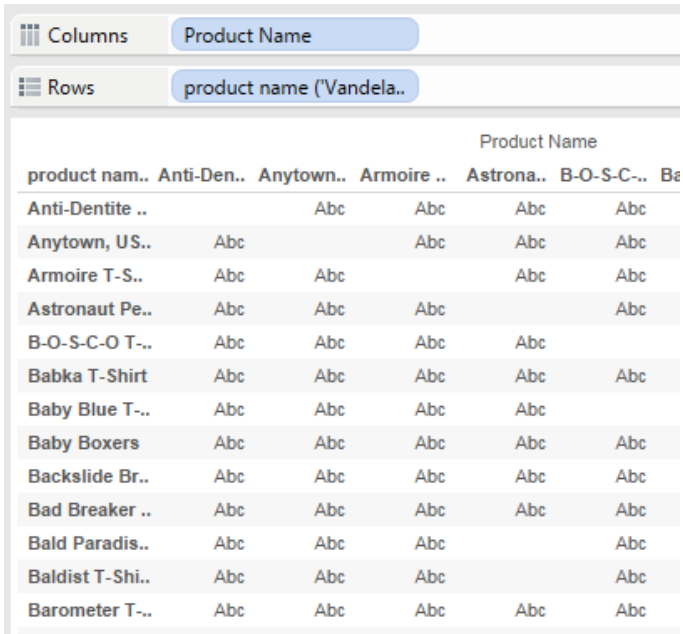
Then drag the “Product Name” dimension from ‘Vandelay Orders(All)$’1 (from the second set of dimensions) to the Rows shelf.



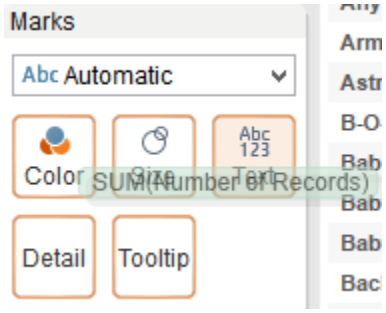
Columns shelf

Rows shelf

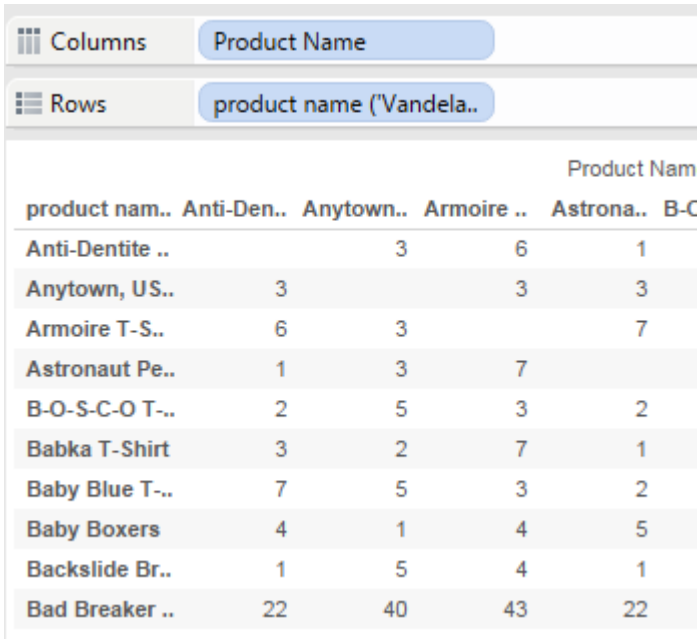
12) You’ll see something like this:



13) Under Measures, drag “Number of Records” to the Text icon under the Marks area.



14) You’ll now see this:



				Product Nam	
product nam..	Anti-Den..	Anytown..	Armoire ..	Astrona..	B-C
Anti-Dentite ..		3	6	1	
Anytown, US..	3		3	3	
Armoire T-S..	6	3		7	
Astronaut Pe..	1	3	7		
B-O-S-C-O T-..	2	5	3	2	
Babka T-Shirt	3	2	7	1	
Baby Blue T-..	7	5	3	2	
Baby Boxers	4	1	4	5	
Backslide Br..	1	5	4	1	
Bad Breaker ..	22	40	43	22	

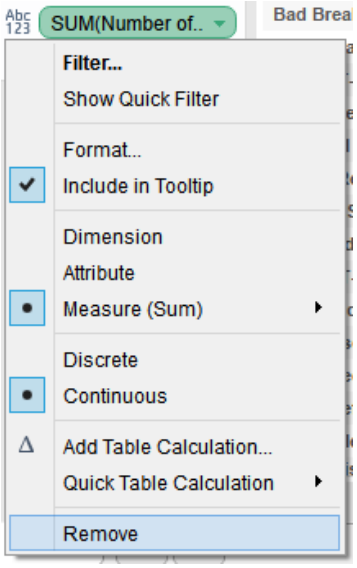
This shows how many orders contained both products. For example, look at the first row. We now know that “Anti-Dentite Jeans” and “Anytown, USA Sweatshirts” appeared together in the same order 3 times. (Hover your mouse over the product name to see the whole thing).

Here are a few more.

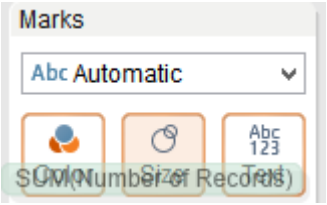
- “Bad Breaker Upper Socks” and “Armoire T-Shirts” appeared in the same order 43 times.
- “Baby Boxers” and “Astronaut Pen Boxers” appeared in the same order 5 times.
- “B-O-S-C-O T-Shirts” and “Anti-Dentite Jeans” appeared in the same order 2 times.

15) It's not difficult to understand, but it would be easier if we could generate an easy-to-read visual of this data. For that, we are creating a heat map.

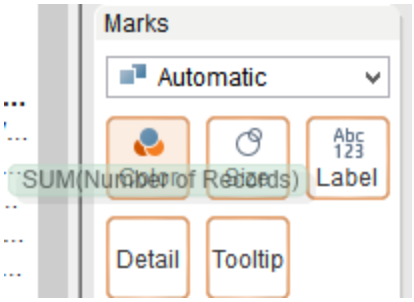
Review "SUM(Number of Records)" in the Marks area.



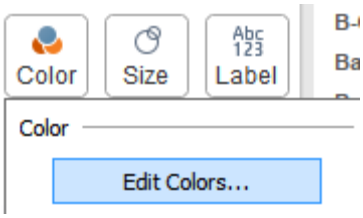
Drag "SUM(Number of Records)" in the Marks area to the Size icon in the Marks area.



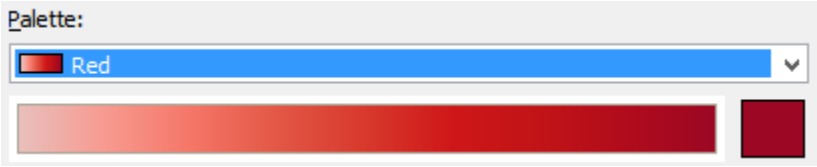
16) Go back to Measures and drag "Number of Records" again, but this time, to the Color icon in the Marks area.



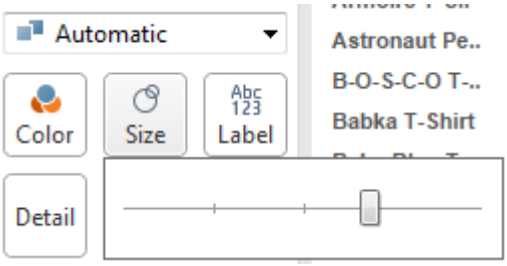
17) Click on the Color icon in the Marks area, then click “Edit Colors…”



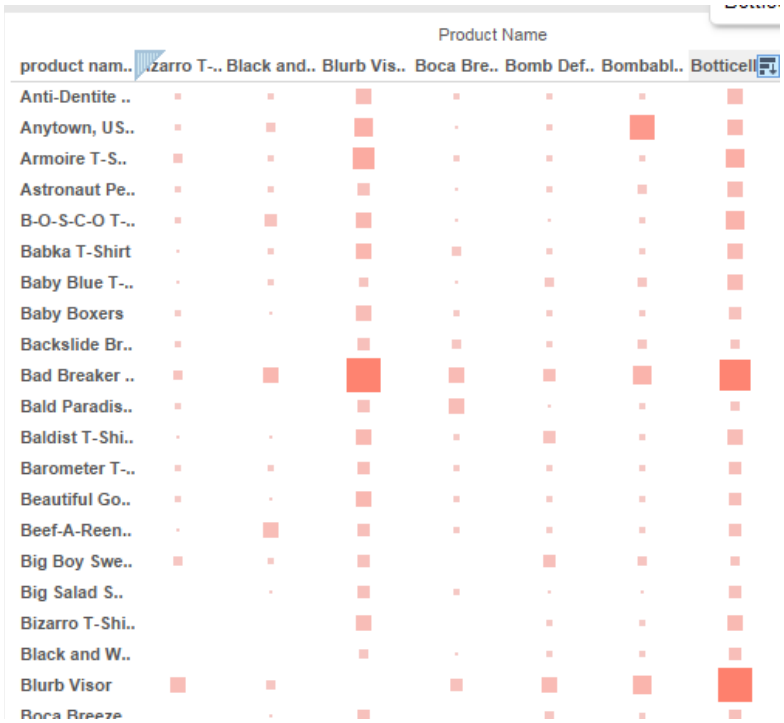
18) Choose “Area Red” for the Palette.



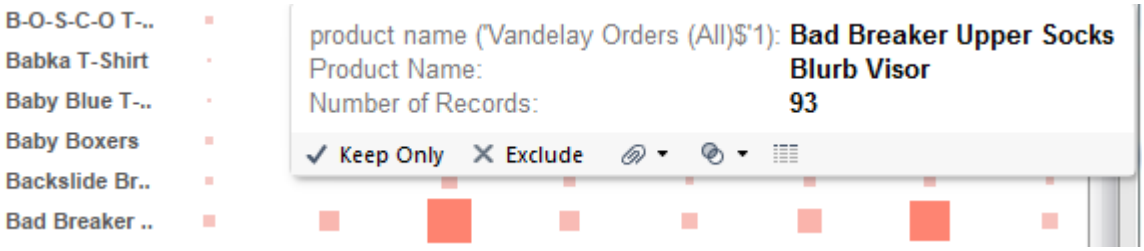
19) Click the Size icon in the Marks area and move the slider about two-thirds of the way to the right.



20) It's now very easy to see the product combinations that are most popular.



21) If you want to see detailed information about a product combination, hover your mouse over a square.



22) Save your Tableau workbook.