



# Applied Data Analysis

## Module 2 Lab: Business Data Walkthrough

### Learning Objectives

- Rearrange and sort data in Excel.
- Use Excel's COUNTIF and COUNTIFS functions to turn categorical data into numerical counts.
- Create and interpret bar charts in Excel.
- Find proportions and percentages in Excel.
- Create and interpret percentage stacked column charts in Excel.
- Ensure that visualizations are properly labeled and readable.

### Data Set

Mod2Lab.csv

**Note:** We're using a large data set for this lab, so make sure you copy and paste **all 2001 rows** of data into a new Excel Online worksheet.

### What You'll Need

To complete the lab, you will need the online version of Microsoft Excel.

### Overview

Imagine you're a data analyst, and your client is a company that asks you to put together a detailed report on the sales of four different leaf-blower products. In this lab, you'll take a deep dive into creating and using the formulas, algorithms, and visualizations that you briefly saw in the videos for this module.

### Exercise 1: Adoption Rates by Product

When you're working as a data analyst, a significant part of your job will be preparing visualizations and reports for your clients. So we'll start things off by creating a couple of useful visualizations/charts to more easily digest our data.

1. Open the data set in Excel Online. The data set should show purchase information on four different leaf-blower products, plus age and home ownership status, for 2000 different individuals. Here's a quick snapshot of the first few rows of the data set:

	A	B	C	D	E	F	G
1	id	Leaf Blower	Leaf Destroyer	Leaf Storm-nado	Leaf A-leaf-eator	age	home
2	1	No	Yes	Yes	No	56.82509	Own
3	2	No	Yes	No	No	55.83387	Own
4	3	No	No	Yes	No	32.01325	Rent
5	4	No	Yes	No	No	31.88865	Rent
6	5	No	Yes	Yes	No	45.94029	Own
7	6	No	No	No	No	27.02899	Rent
8	7	No	No	Yes	No	53.25716	Rent
9	8	No	No	Yes	No	54.33622	Rent
10	9	No	No	No	Yes	25.02194	Rent
11	10	No	No	Yes	No	31.92264	Rent
12	11	No	No	No	No	18.85002	Rent

Each row represents one person (there should be 2000 different people). Here's what each column/variable represents:

**id** = a number representing the individual for ID purposes (i.e. Person 1, Person 2, etc.)

**Leaf Blower** = "No" means that individual did not purchase the Leaf Blower product; "Yes" means they did purchase it

**Leaf Destroyer** = "No" means that individual did not purchase the Leaf Destroyer product; "Yes" means they did purchase it

**Leaf Storm-nado** = "No" means that individual did not purchase the Leaf Storm-nado product; "Yes" means they did purchase it

**Leaf A-leaf-eator** = "No" means that individual did not purchase the Leaf A-leaf-eator product; "Yes" means they did purchase it

**age** = the exact age of that individual, in years (as a decimal)

**home** = whether that individual owns their own home or is a renter

So, for example, you can see from that first row of data that Person 1 did not purchase the Leaf Blower, did purchase both the Leaf Destroyer and Leaf Storm-nado, did not purchase the Leaf A-leaf-eator, is about 56.82 years old, and owns their home.

- Set up a new mini-table off to the side of your data set. This new table will show the number of people who purchased or did not purchase each product. Start the new table over in columns I, J, and K, like this:

	A	B	C	D	E	F	G	H	I	J	K
1	id	Leaf Blower	Leaf Destroyer	Leaf Storm-nado	Leaf A-leaf-eator	age	home		Product	Did Not Purchase	Purchased
2	1	No	Yes	Yes	No	56.82509	Own		Leaf Blower		
3	2	No	Yes	No	No	55.83387	Own		Leaf Destroyer		
4	3	No	No	Yes	No	32.01325	Rent		Leaf Storm-nado		
5	4	No	Yes	No	No	31.88865	Rent		Leaf A-leaf-eator		
6	5	No	Yes	Yes	No	45.94029	Own				
7	6	No	No	No	No	27.02899	Rent				

Here's a closer look at that new mini-table itself, with one column for the product names, one column for "Did Not Purchase," and one column for "Purchased":

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower		
Leaf Destroyer		
Leaf Storm-nado		
Leaf A-leaf-eator		

To fill in the first cell of this table, you'll count the number of "No" entries under "Leaf Blower" to find out how many people did not purchase the Leaf Blower, etc.

- Now, obviously we don't want to count each "Yes" and "No" over all 2000 data points. Who has time for that? Instead, use Excel's handy COUNTIF function, which counts the number of times a certain value appears in a range of data. Here's the syntax: **=COUNTIF(range, criteria)**. The range takes the form **firstcell:lastcell** (with a colon in between) and the criteria is whatever value you're looking for. **Note:** The criteria part **MUST** be in quotation marks if you're looking for a text value (instead of a number).

So, in this case, the first thing you want to find is the number of people who did *not* purchase the Leaf Blower, i.e. the number of times "No" shows up in the Leaf Blower column (column B). The data in the Leaf Blower column runs from cell B2 all the way down to B2001, so your range is B2:B2001.

In cell J2 (below "No"), type **=COUNTIF(B2:B2001, "No")**. Don't forget to put "No" in quotation marks!

J2

When you press Enter, cell J2 should now show the number of times that "No" showed up in column B.

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	
Leaf Destroyer		
Leaf Storm-nado		
Leaf A-leaf-eator		

There we go: 1802 people did not purchase the Leaf Blower.

4. Use the same syntax to find out how many “Yes” values show up in the Leaf Blower column. Click into cell K2 for this. Your range of values is still B2:B2001, but this time, your criteria is “Yes”:

K2            =COUNTIF(B2:B2001, "Yes")

Hit Enter.

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer		
Leaf Storm-nado		
Leaf A-leaf-eator		

198 people did purchase the Leaf Blower.

5. Repeat Steps 3 and 4 for the Leaf Destroyer, which is found in column C.

For the number of people who did not purchase the Leaf Destroyer, type this into cell J3 and hit Enter:

J3            =COUNTIF(C2:C2001, "No")

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer	672	
Leaf Storm-nado		
Leaf A-leaf-eator		

Notice that your range is now C2:C2001. Use that same range for the number of people who *did* purchase the Leaf Destroyer. This goes in cell K3:

K3            =COUNTIF(C2:C2001, "Yes")

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer	672	1328
Leaf Storm-nado		
Leaf A-leaf-eator		

6. Repeat those steps again for the Leaf Storm-nado. The range here runs from D2:D2001. For the "No" answers in that range, type this into cell J4:

J4                =COUNTIF(D2:D2001, "No")

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer	672	1328
Leaf Storm-nado	1036	
Leaf A-leaf-eator		

And for the "Yes" answers, type this into K4:

K4                =COUNTIF(D2:D2001, "Yes")

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer	672	1328
Leaf Storm-nado	1036	964
Leaf A-leaf-eator		

7. Repeat those steps one more time for the Leaf A-leaf-eator. This time, the range is E2:E2001. Type this into cell J5 to find the number of people who didn't purchase the Leaf A-leaf-eator:

J5                =COUNTIF(E2:E2001, "No")

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer	672	1328
Leaf Storm-nado	1036	964
Leaf A-leaf-eator	1638	

And finally, type this into K5 to find the number of people who *did* purchase the Leaf A-leaf-eator:

K5       $\text{fx}$       =COUNTIF(E2:E2001, "Yes")

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer	672	1328
Leaf Storm-nado	1036	964
Leaf A-leaf-eator	1638	362

- Now that you've got all your counts filled in, you can use this table to create a graph. Highlight the whole mini-table you created, including the category titles. (Do *not* highlight the entire spreadsheet.)

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer	672	1328
Leaf Storm-nado	1036	964
Leaf A-leaf-eator	1638	362

- With those data highlighted, click Insert > Column > Clustered Column (usually the icon on the left under 2-D Column).

Excel Online | OneDrive > Documents | Book 2

File Home Insert Data Review View Tell me what you want to do Open in Excel

Function Survey Table PivotTable Picture Shapes Office Add-ins Column Line Pie Bar Area Scatter Other Charts Hyperlink Comment

Functions Tables Illustrations Add-ins

2-D Column

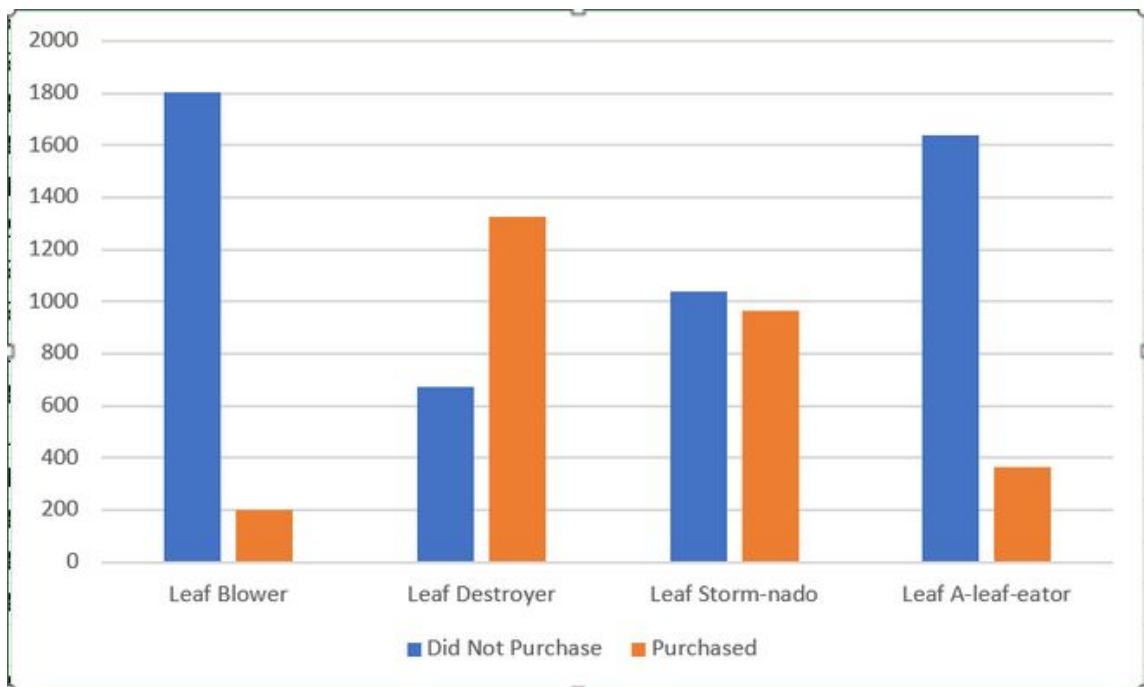
Clustered Column

Compare values across categories by using vertical rectangles.

Use it when the order of categories is not important or for displaying item counts such as a histogram.

	A	B	C	D	E	F	G	H	I	J	K
1	id	Leaf Blower	Leaf Destroyer	Leaf Storm-nado					Product	Did Not Purchase	Purchased
2	1	No	Yes	Yes	No					1802	198
3	2	No	Yes	No	No					672	1328
4	3	No	No	Yes	No					1036	964
5	4	No	Yes	No	No					1638	362
6	5	No	Yes	Yes	No				45.94029	Own	
7	6	No	No	No	No				27.02899	Rent	
8	7	No	No	Yes	No				52.25716	Rent	

Once you click that icon, a fancy new bar chart should pop up, like this:



10. Looking good! But we can make this graph easier to read. Click on the graph itself, then click on Chart Tools in the ribbon. From there, select Chart > Data Labels > Outside End.



Excel Online | OneDrive > Documents | Chart Tools | Book 2

File Home Insert Data Review View Chart Tell me what you want to do Open in Excel

Switch Row/Column Select Data Data Column Line Pie Bar Area Scatter Other Charts Chart Title Axis Legend Data Labels Data Table Axes Gridlines

Change Chart Type

11

	A	B	C	D	E	F
1	id	Leaf Blower	Leaf Destroyer	Leaf Storm-nado	Leaf A-leaf-eator	age
2	1	No	Yes	Yes	No	2000
3	2	No	Yes	No	No	1800
4	3	No	No	Yes	No	1600
5	4	No	Yes	No	No	1400
6	5	No	Yes	Yes	No	1200
7	6	No	No	No	No	1000
8	7	No	No	Yes	No	800
9	8	No	No	Yes	No	
10	9	No	No	No	Yes	

None Turn off Data Labels for selection

Center Display Data Labels and position centered on the data point(s)

Inside End Display Data Labels and position inside the end of data point(s)

Inside Base Display Data Labels and position inside the base of data point(s)

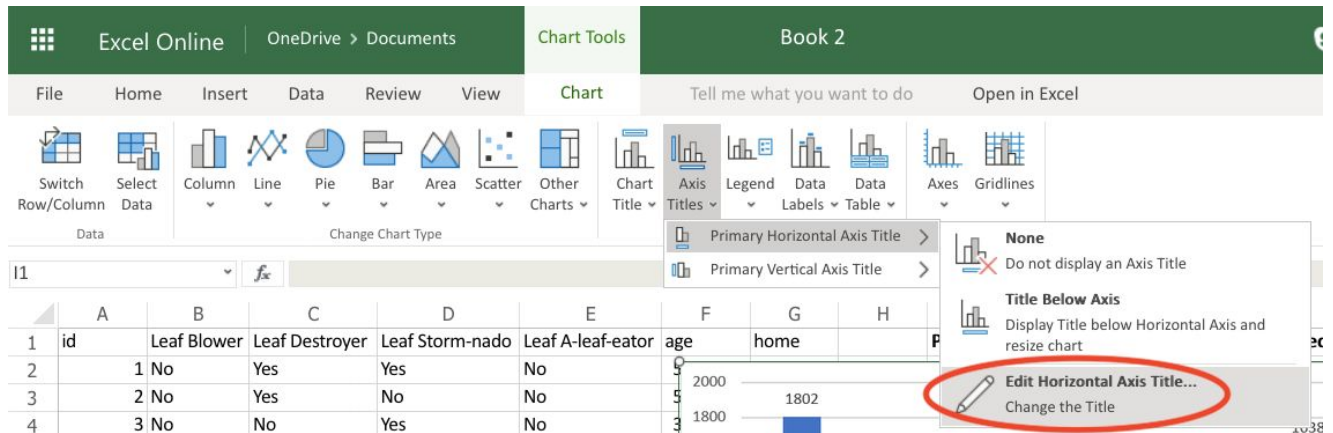
**Outside End** Display Data Labels and position outside the end of data point(s)

That should display the count for each category above the bars in your graph.



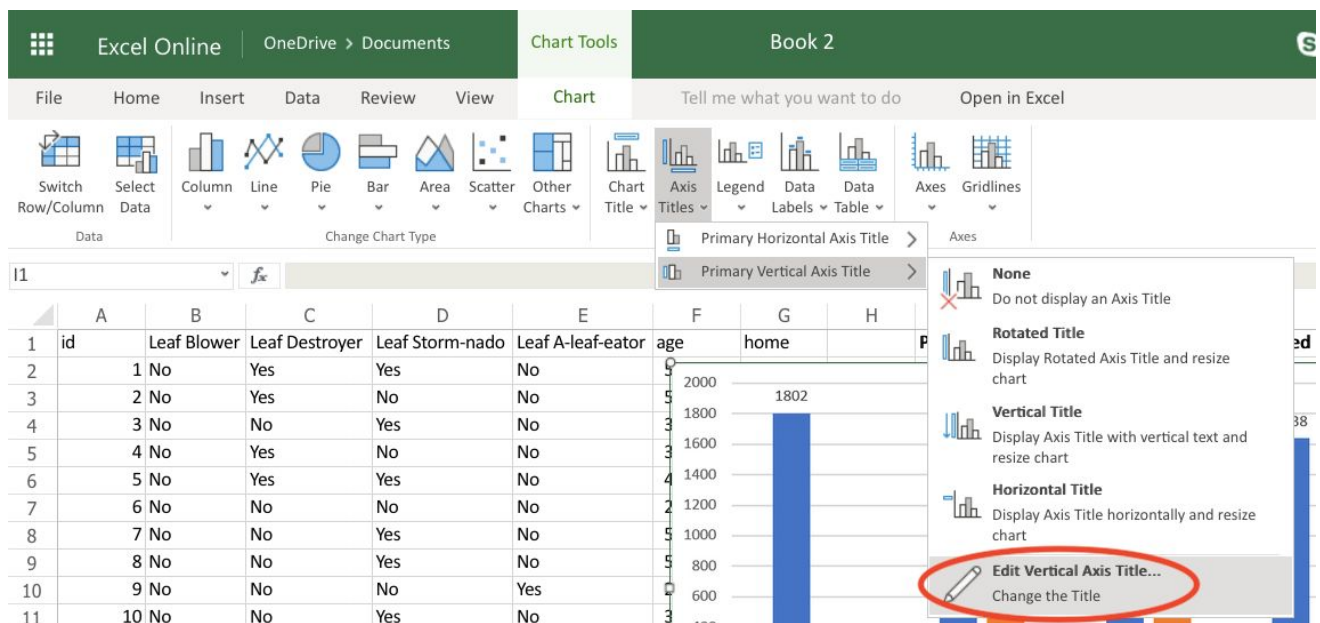
11. If you were presenting this chart to a client, you'd also want to add titles to each axis to make it even easier on the eyes. Click on Chart Tools > Axis Titles > Primary Horizontal Axis Title > Edit Horizontal Axis Title.





Type in "Product" in the text box and hit OK.

Then add a title for the vertical axis as well by clicking Chart Tools > Axis Titles > Primary Vertical Axis Title > Edit Vertical Axis Title.



For this axis, type "Count" and hit OK. Now your chart is all labeled!



Now it's a lot easier to see how each product stacked up against the others. Very few people bought the standard Leaf Blower product, for instance, but 1328 people decided to buy the fancier Leaf Destroyer.

12. Add one final title to really make your client happy. Click on Chart Tools > Chart Title > Edit Chart Title.

Excel Online | OneDrive > Documents | Chart Tools | Book 2

File Home Insert Data Review View | Chart

Tell me what you want to do

Switch Row/Column | Select Data | Column | Line | Pie | Bar | Area | Scatter | Other Charts

Data | Change Chart Type

Chart Title | Axis Titles | Legend | Data Labels | Data Table | Axes

None  
Do not display a chart Title

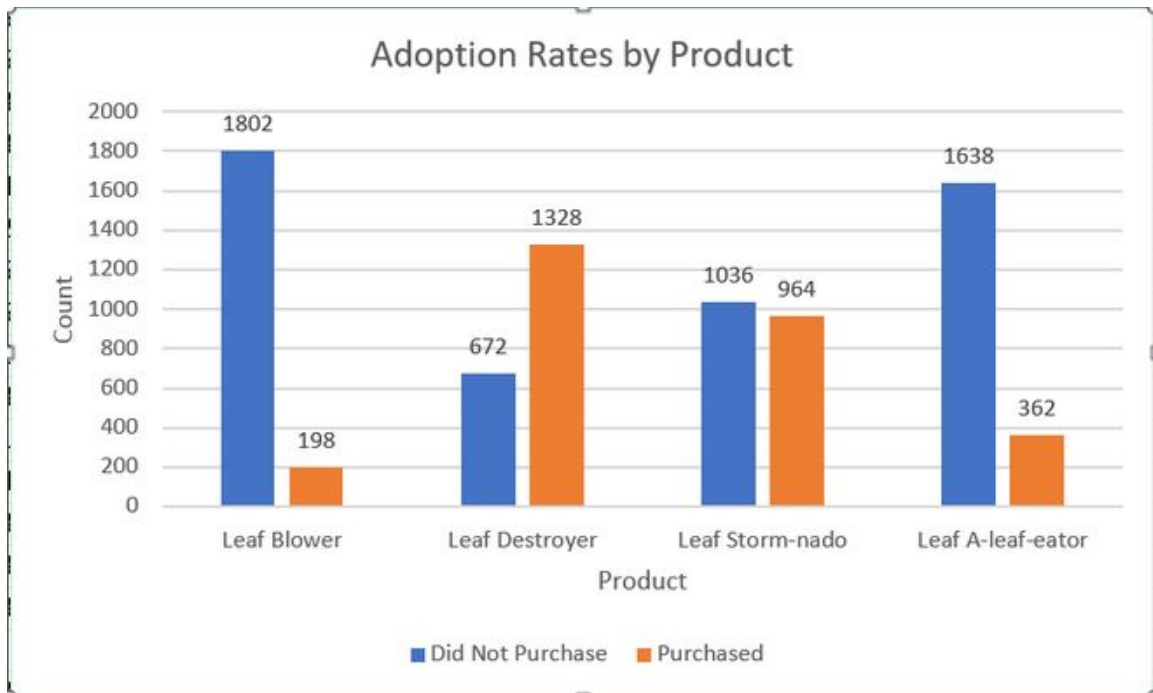
Centered Overlay Title  
Overlay centered Title on chart without resizing chart

Above Chart  
Display Title at top of chart area and resize chart

**Edit Chart Title...**  
Change the Title

	A	B	C	D	E
1	id	Leaf Blower	Leaf Destroyer	Leaf Storm-nado	Leaf A-leaf-eator
2	1	No	Yes	Yes	No
3	2	No	Yes	No	No
4	3	No	No	Yes	No
5	4	No	Yes	No	No
6	5	No	Yes	Yes	No
7	6	No	No	No	No

In the title box that pops up, type in “Adoption Rates by Product” and click OK. Boom, your first report is done!



## Exercise 2: Product Rates by Purchased

Now we’ll create another similar graph that shows the exact same data, but from another perspective.

- With that same data set open, create another quick mini-table that shows the same categories you used in the previous exercise, but from the opposite direction. (Thankfully, you don’t need to repeat the COUNTIF functions.) In columns M, N, O, P, and Q, set up new columns labeled “Purchased,” “Leaf Blower,” “Leaf Destroyer,” “Leaf Storm-nado,” and “Leaf A-leaf-eator.” Below “Purchased,” set up rows for “No” and “Yes.” It should look like this:

M	N	O	P	Q
Purchased	Leaf Blower	Leaf Destroyer	Leaf Storm-nado	Leaf A-leaf-eator
No				
Yes				

- Copy all the counts you found in Exercise 1 into their proper cells in this new table. **Note:** Do NOT simply click on cell J2 from the old table and copy/paste it into cell N2 in the new table. If you try this, the formula may misunderstand your intentions and change the data range. Instead, do it like this: First, click into cell J2 of the old table. Then, highlight the formula up in the **function bar**, not in the cell itself.

J2					<b>f<sub>x</sub></b>	=COUNTIF(B2:B2001, "No")	
	E	F	G	H	I	J	K
1	leaf-eator	age	home		<b>Product</b>	<b>Did Not Purchase</b>	<b>Purchased</b>
2		56.82509	Own		=COUNTIF(B2:B2001, "No")		198
3		55.83387	Own		Leaf Destroyer	672	1328
4		32.01325	Rent		Leaf Storm-nado	1036	964
5		31.88865	Rent		Leaf A-leaf-eator	1638	362

Copy and paste that formula directly from the function bar, then paste *that* into cell N2 of the new table.

M	N	O	P	Q
<b>Purchased</b>	<b>Leaf Blower</b>	<b>Leaf Destroyer</b>	<b>Leaf Storm-nado</b>	<b>Leaf A-leaf-eator</b>
No	1802			
Yes				

That should give you the proper count for the “No” values in the Leaf Blower category.

- Carefully repeat Step 2 for all the other counts. Go nice and slowly, paying special attention to getting all the counts in their proper spots. Compare your final table to the old table you made in Exercise 1, just in case.

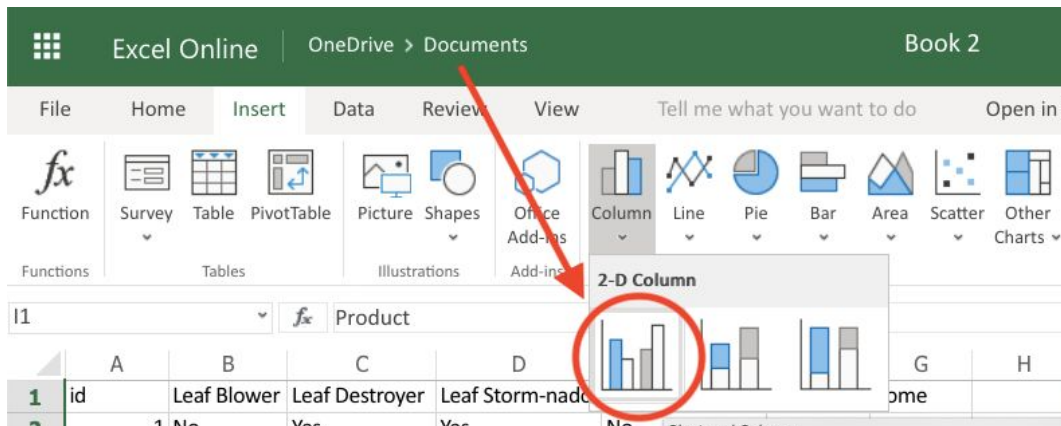
Note: If you’re struggling to get the right formulas copied and pasted, just go ahead and type the counts into the new table directly. Either way, the new table should look like this:

M	N	O	P	Q
<b>Purchased</b>	<b>Leaf Blower</b>	<b>Leaf Destroyer</b>	<b>Leaf Storm-nado</b>	<b>Leaf A-leaf-eator</b>
No	1802	672	1036	1638
Yes	198	1328	964	362

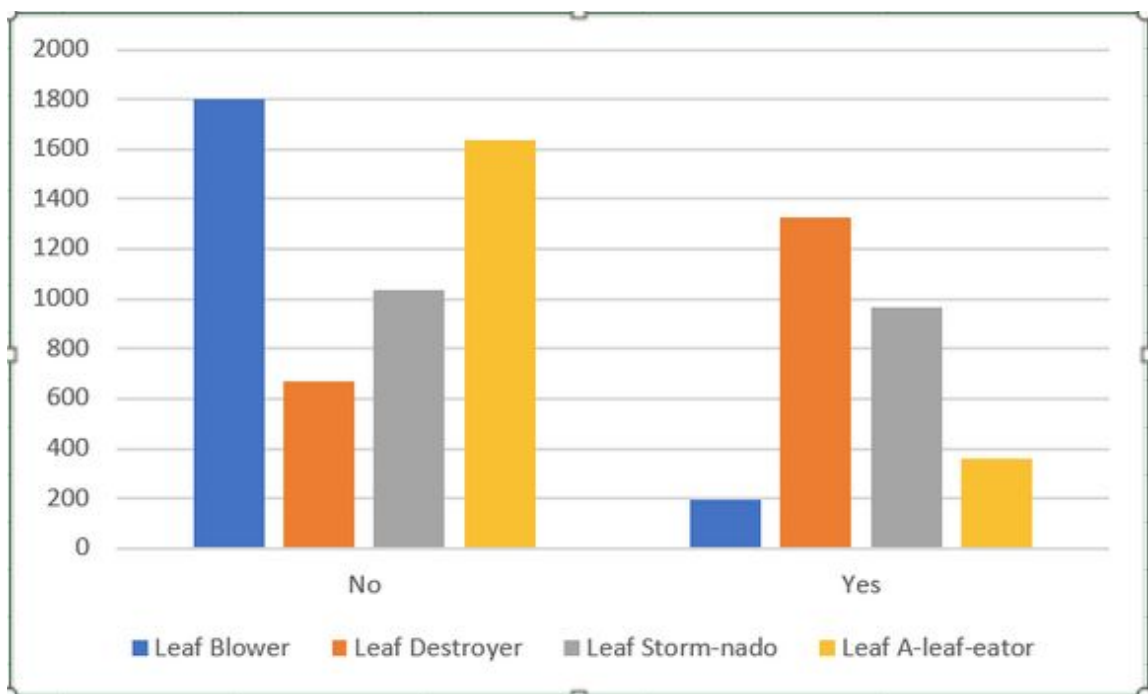
- Highlight all the data in this new table, including the category titles.

M	N	O	P	Q
<b>Purchased</b>	<b>Leaf Blower</b>	<b>Leaf Destroyer</b>	<b>Leaf Storm-nado</b>	<b>Leaf A-leaf-eator</b>
No	1802	672	1036	1638
Yes	198	1328	964	362

- With all those data highlighted, click Insert > Column > Clustered Column again (the exact same way you did in Exercise 1).



Behold! A new graph pops up.



- You'll still need to add a few helpful labels to make the graph easier to read, though. Follow the same steps you did in Exercise 1 to add the following (and be sure to click on the graph itself first):

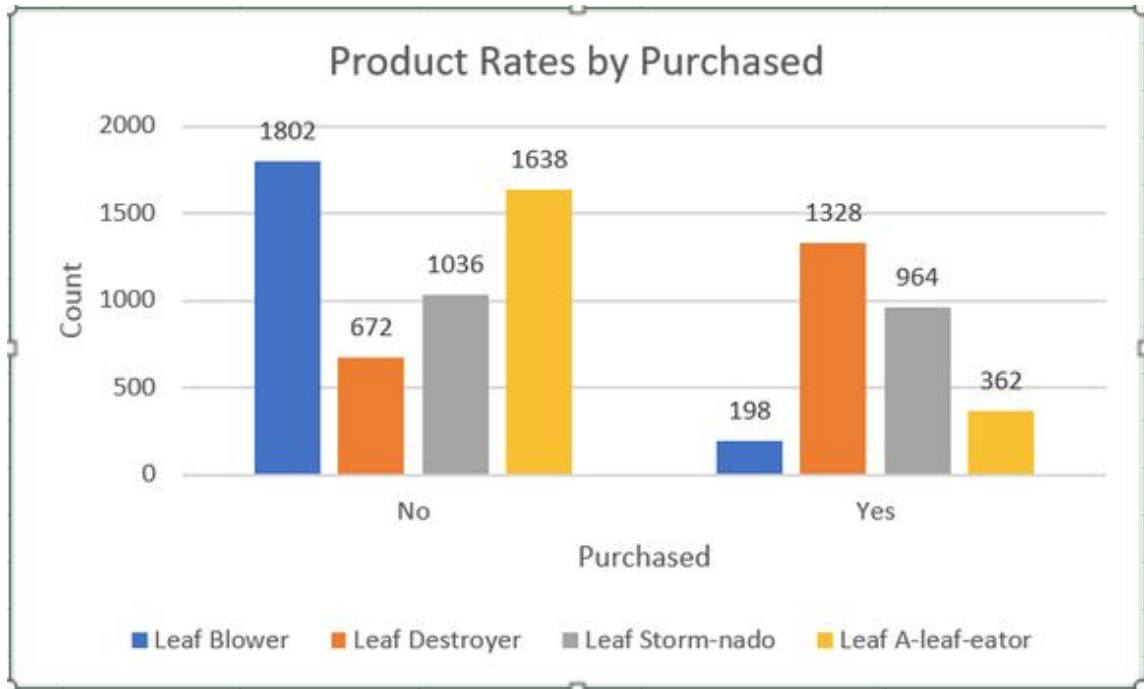
Counts above the bars: **Chart Tools > Data Labels > Outside End**

Horizontal axis title: **Chart Tools > Axis Titles > Primary Horizontal Axis Title > Edit Horizontal Axis Title**, then type "Purchased"

Vertical axis title: **Chart Tools > Axis Titles > Primary Vertical Axis Title > Edit Vertical Axis Title**, then type “Count”

Chart title: **Chart Tools > Chart Title > Edit Chart Title**, then type “Product Rates by Purchased”

7. Your second data visualization is done!



This chart shows the exact same data as the other chart we came up with in Exercise 1, but it shows the data in a new light, which can reveal new insights. For instance, now we can more easily compare the purchase rates of each product side by side.

### Exercise 3: Adoption Percentage

It's great to see how the raw counts stack up against each other, but how about percentages? Seeing what percentage/proportion of our customers purchased or didn't purchase each product can give us even more useful insights. We can even figure out whether these percentages are statistically significant, or just due to some random fluke.

1. Below the first mini-table you created in Exercise 1, set up another mini-table with the exact same labels and categories. But do not copy the entire table itself with the data counts—just the titles. Like this:



I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer	672	1328
Leaf Storm-nado	1036	964
Leaf A-leaf-eator	1638	362
Product	Did Not Purchase	Purchased
Leaf Blower		
Leaf Destroyer		
Leaf Storm-nado		
Leaf A-leaf-eator		

2. Now you want to populate this table with the *proportions* for each category. In other words, if 1802 people did not purchase the Leaf Blower, and there were 2000 total people surveyed in this sample data set, then the proportion is 1802 out of 2000, or 1802 divided by 2000. In the Leaf Blower/Did Not Purchase cell of your new table, type the following (this will divide the count of 1802 by the total of 2000):

J8       $\text{fx}$       =J2/2000

Hit Enter to find the proportion for that category. Excel will divide the value in cell J2 by 2000.

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer	672	1328
Leaf Storm-nado	1036	964
Leaf A-leaf-eator	1638	362
Product	Did Not Purchase	Purchased
Leaf Blower	0.901	
Leaf Destroyer		
Leaf Storm-nado		
Leaf A-leaf-eator		



Awesome. So 0.901 of our customers did not purchase the Leaf Blower (we'll convert these to percentages a little later).

Now find the proportion of people who *did* purchase the Leaf Blower. In cell K8, divide the value in K2 (that's 198) by the total of 2000.

K8       $\text{fx}$       =K2/2000

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer	672	1328
Leaf Storm-nado	1036	964
Leaf A-leaf-eator	1638	362
Product	Did Not Purchase	Purchased
Leaf Blower	0.901	0.099
Leaf Destroyer		
Leaf Storm-nado		
Leaf A-leaf-eator		

- Repeat Step 2 for the Leaf Destroyer product. Here are the “No” values:

J9       $\text{fx}$       =J3/2000

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer	672	1328
Leaf Storm-nado	1036	964
Leaf A-leaf-eator	1638	362
Product	Did Not Purchase	Purchased
Leaf Blower	0.901	0.099
Leaf Destroyer	0.336	
Leaf Storm-nado		
Leaf A-leaf-eator		

And here's what you'll enter for the "Yes" values:

K9  $\text{fx}$  =K3/2000

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer	672	1328
Leaf Storm-nado	1036	964
Leaf A-leaf-eator	1638	362
Product	Did Not Purchase	Purchased
Leaf Blower	0.901	0.099
Leaf Destroyer	0.336	0.664
Leaf Storm-nado		
Leaf A-leaf-eator		

- Do it all again for the Leaf Storm-nado. Here's the "No" category (remember, there were 2000 customers, so everything gets divided by 2000):

J10  $\text{fx}$  =J4/2000

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer	672	1328
Leaf Storm-nado	1036	964
Leaf A-leaf-eator	1638	362
Product	Did Not Purchase	Purchased
Leaf Blower	0.901	0.099
Leaf Destroyer	0.336	0.664
Leaf Storm-nado	0.518	
Leaf A-leaf-eator		

And here are the “Yes” values:

K10

fx

=K4/2000

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer	672	1328
Leaf Storm-nado	1036	964
Leaf A-leaf-eator	1638	362
Product	Did Not Purchase	Purchased
Leaf Blower	0.901	0.099
Leaf Destroyer	0.336	0.664
Leaf Storm-nado	0.518	0.482
Leaf A-leaf-eator		

- One more time: repeat the previous step for the Leaf A-leaf-eator product.

For “No”:

J11

fx

=J5/2000

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer	672	1328
Leaf Storm-nado	1036	964
Leaf A-leaf-eator	1638	362
Product	Did Not Purchase	Purchased
Leaf Blower	0.901	0.099
Leaf Destroyer	0.336	0.664
Leaf Storm-nado	0.518	0.482
Leaf A-leaf-eator	0.819	

For “Yes”:

K11	$\text{fx}$	=K5/2000
-----	-------------	----------

I	J	K
Product	Did Not Purchase	Purchased
Leaf Blower	1802	198
Leaf Destroyer	672	1328
Leaf Storm-nado	1036	964
Leaf A-leaf-eator	1638	362
Product	Did Not Purchase	Purchased
Leaf Blower	0.901	0.099
Leaf Destroyer	0.336	0.664
Leaf Storm-nado	0.518	0.482
Leaf A-leaf-eator	0.819	0.181

- Okay, now turn all those decimal values into percentages—this is something data analysts often do to make the data more digestible/readable for their clients. To do this, highlight all the decimals in the new table, click the Home tab in the ribbon, and select the percentage icon (%) in the Number section.

Excel Online | OneDrive > Documents | Book 2

File Home Insert Data Review View Tell me what you want to do Open in Excel

Undo Clipboard Font Alignment Percentage

K11 =K5/2000

	E	F	G	H	I	J	K
1	-nado	Leaf A-leaf-eator	age	home	Product	Did Not Purchase	Purchased
2	No	56.82509	Own		Leaf Blower	1802	198
3	No	55.83387	Own		Leaf Destroyer	672	1328
4	No	32.01325	Rent		Leaf Storm-nado	1036	964
5	No	31.88865	Rent		Leaf A-leaf-eator	1638	362
6	No	45.94029	Own				
7	No	27.02899	Rent		Product	Did Not Purchase	Purchased
8	No	53.25716	Rent		Leaf Blower	0.901	0.099
9	No	54.33622	Rent		Leaf Destroyer	0.336	0.664
10	Yes	25.02194	Rent		Leaf Storm-nado	0.518	0.482
11	No	31.92264	Rent		Leaf A-leaf-eator	0.819	0.181
12	No	18.85002	Rent				

Once you click it, Excel will convert all those values to percentages.

Product	Did Not Purchase	Purchased
Leaf Blower	90.10%	9.90%
Leaf Destroyer	33.60%	66.40%
Leaf Storm-nado	51.80%	48.20%
Leaf A-leaf-eator	81.90%	18.10%

Now you can see what percentage of customers purchased each product. For example, it looks like a much lower percentage of customers purchased the Leaf Blower (just under 10%) than any other product.

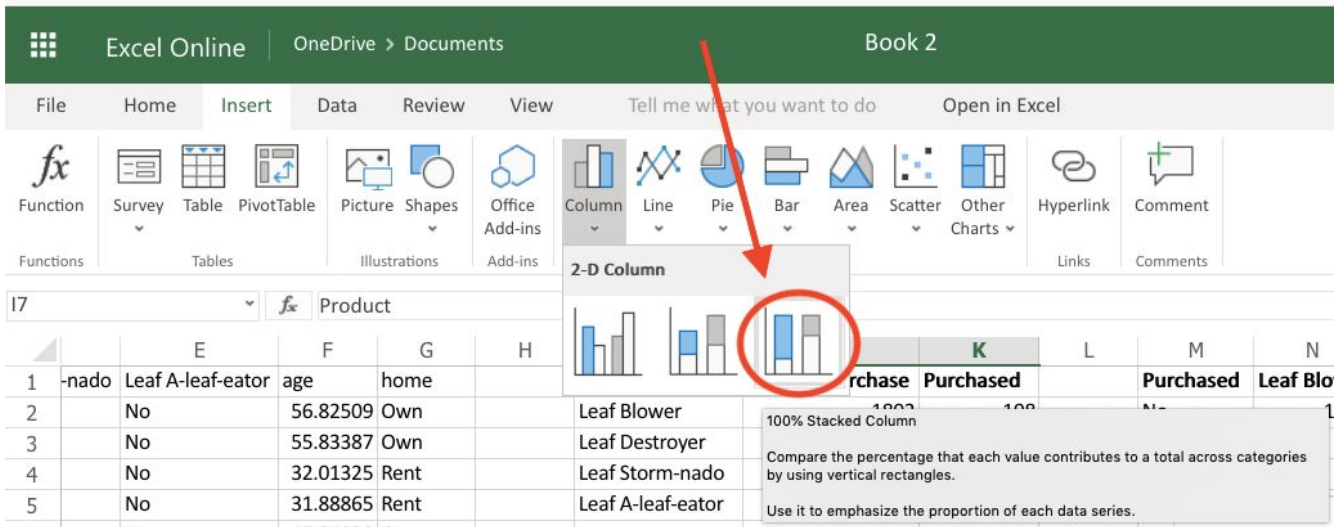
- One quick side note: Our data set is only a sample of 2000 individuals; it doesn't show the entire population of customers globally or nationally. That means our percentages are really just estimates based on our sample size. How do we know these differences aren't just the result of over- or under-estimating? Is the Leaf Blower *really* purchased at a far lower rate than the other products, or is it just a fluke with our particular slice of data? We won't get into the details of this idea for this course, but there's a statistical test called a "chi-squared" ( $\chi^2$ , after the Greek letter) independence test that can tell us whether the difference between any two rows in our percentage table (i.e. any two

products) is “statistically significant.” If so, it would mean that the difference is probably a real thing and not just a fluke. In this example, the results of that test would show us that every row is significantly different from every row **except** the Leaf Blower versus the Leaf A-leaf-eator, which do not significantly differ.

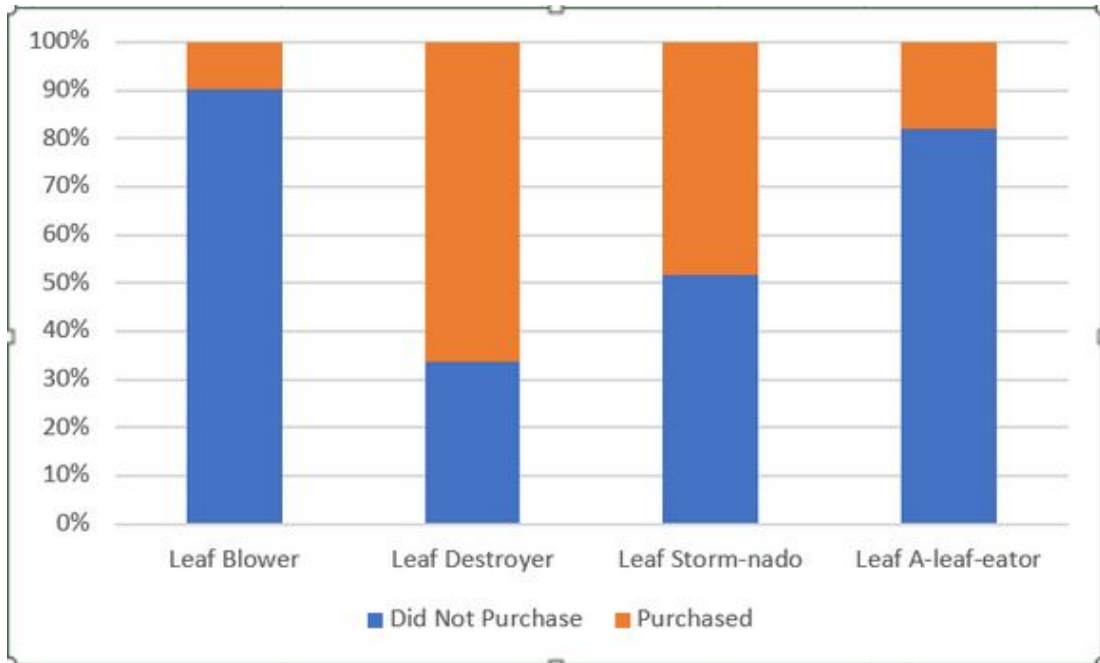
7. While you’re at it, you might as well make a quick chart showing the percentages, too. Highlight the new mini-table you created in this exercise, including the category titles.

Product	Did Not Purchase	Purchased
Leaf Blower	90.10%	9.90%
Leaf Destroyer	33.60%	66.40%
Leaf Storm-nado	51.80%	48.20%
Leaf A-leaf-eator	81.90%	18.10%

8. With that table highlighted, click Insert > Column > 100% Stacked Column (usually the icon on the far right).



A new graph should pop up.



Beautiful—another helpful visualization to add to our report for this client. Now we can see that the Leaf Destroyer performed the best.

#### Exercise 4: Home Ownership Graphs

Another question we might ask as data analysts is: How do our customer's purchases differ by home ownership status? As usual, it's best to start with a graph—or really, a bunch of graphs.

- Below the percentage table you created in Exercise 3 (but still in columns I, J, and K), create a new mini-table that breaks down home ownership ("Own" or "Rent") by purchase ("Did Not Purchase" or "Purchased") for the Leaf Blower product only. It should look like this:

	I	J	K
12			
13		LEAF BLOWER	
14	Home Ownership	Did Not Purchase	Purchased
15	Own		
16	Rent		
17			

- Fill in the counts for this table using Excel's COUNTIFS function, along with the original data set. (Note the "S" at the end—this function is a bit fancier than the regular COUNTIF function that we used in an earlier exercise.)

The syntax is **=COUNTIFS(criteria\_range1, criteria1, criteria\_range2, criteria2)**. It's basically like stacking two COUNTIF functions on top of each other.



Both ranges take the form **firstcell:lastcell** (with a colon in between). The first criteria (criteria1) is, in this case, either a “No” or “Yes” (in quotation marks) from whichever product column we’re looking at. The second criteria (criteria2) is either “Own” or “Rent” (also in quotation marks) from the “home” column (column G).

Don’t worry, it sounds a lot worse than it really is in practice. To find the number of home owners who did not purchase a Leaf Blower, for instance, click into cell J15 and type this:

J15      **=COUNTIFS(B2:B2001, "No", G2:G2001, "Own")**

Essentially, you’re asking Excel to count the number of people who had both a “No” value in the Leaf Blower category (column B) *and* an “Own” value in the home category (column G), *at the same time*.

J15      **=COUNTIFS(B2:B2001, "No", G2:G2001, "Own")**

	A	B					G
1	id	Leaf Blower	Leaf Destroyer	Leaf Storm-nado	Leaf A-leaf-eator	age	home
2	1	No	Yes	Yes	No	56.82509	Own
3	2	No	Yes	No	No	55.83387	Own
4	3	No	No	Yes	No	32.01325	Rent
5	4	No	Yes	No	No	31.88865	Rent
6	5	No	Yes	Yes	No	45.94029	Own
7	6	No	No	No	No	27.02899	Rent
8	7	No	No	Yes	No	53.25716	Rent
9	8	No	No	Yes	No	54.33622	Rent
10	9	No	No	No	Yes	25.02194	Rent
11	10	No	No	Yes	No	31.92264	Rent
12	11	No	No	No	No	18.85002	Rent
13	12	Yes	Yes	No	No	19.11739	Rent

Hit Enter and the count should pop up in your new mini-table.

	LEAF BLOWER	
Home Ownership	Did Not Purchase	Purchased
Own	448	
Rent		

So 448 home owners did not purchase a Leaf Blower.

- Use COUNTIFS again to find the number of people who purchased a Leaf Blower (that's a "Yes" in the Leaf Blower range from B2:B2001) *and* own their homes (that's "Own" in the home range from G2:G2001).

All this should go in cell K15, like so. Notice how the syntax is identical to the previous step, except the first criteria is "Yes" instead of "No."

K15   =COUNTIFS(B2:B2001, "Yes", G2:G2001, "Own")

Hit Enter.

	LEAF BLOWER	
Home Ownership	Did Not Purchase	Purchased
Own	448	52
Rent		

So only 52 home owners purchased a shiny new Leaf Blower.

- Repeat the whole COUNTIFS thing again to find the number of people who did not purchase a Leaf Blower (that's a "No" in the Leaf Blower range from B2:B2001) *and* are renters (that's "Rent" in the home range from G2:G2001).

This should go in cell J16.

J16   =COUNTIFS(B2:B2001, "No", G2:G2001, "Rent")

Smash that Enter key.

	LEAF BLOWER	
Home Ownership	Did Not Purchase	Purchased
Own	448	52
Rent	1354	

Translation: 1354 renters didn't purchase a Leaf Blower.

- Use COUNTIFS one more time to find the number of people who purchased a Leaf Blower (that's a "Yes" in the Leaf Blower range from B2:B2001) *and* are renters (that's "Rent" in the home range from G2:G2001).

K16       $\text{fx}$       =COUNTIFS(B2:B2001, "Yes", G2:G2001, "Rent")

You know the drill. Press Enter.

	LEAF BLOWER	
Home Ownership	Did Not Purchase	Purchased
Own	448	52
Rent	1354	146

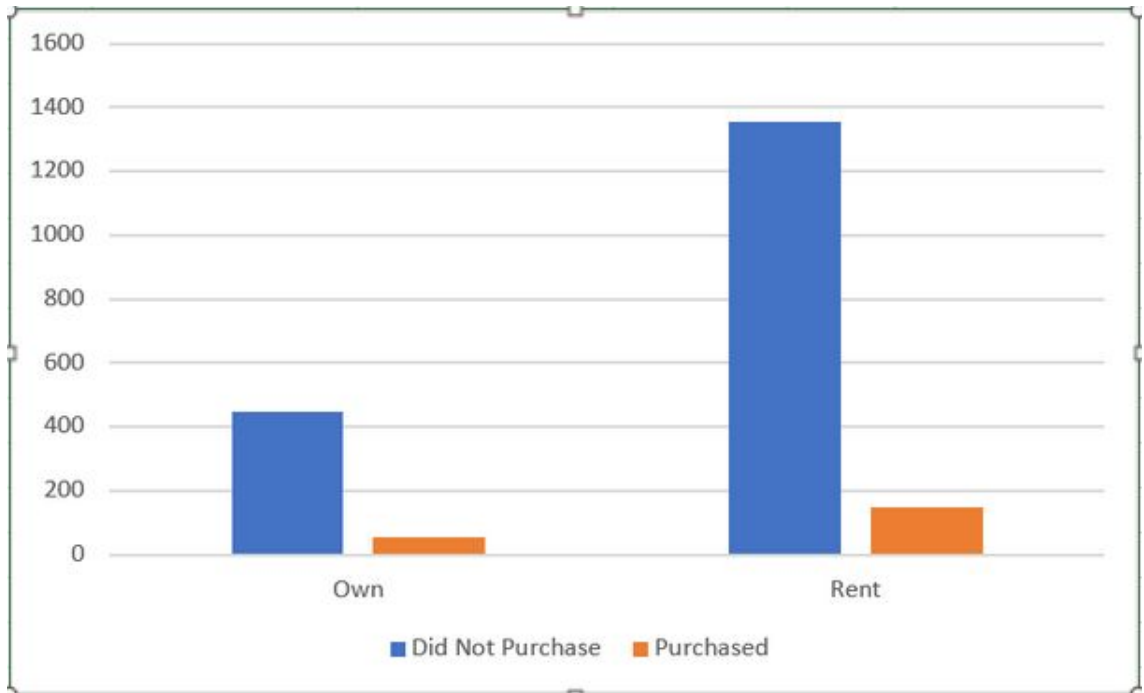
- Graph time. Highlight everything in this new mini-table EXCEPT for the “LEAF BLOWER” title (seriously, don’t highlight that—it’ll mess up the graph).

	LEAF BLOWER	
Home Ownership	Did Not Purchase	Purchased
Own	448	52
Rent	1354	146

Then click Insert > Column > Clustered Column (usually the icon on the left under 2-D Column).

The screenshot shows the Excel Online interface. The ribbon is set to 'Insert', and the 'Column' group is expanded. A red arrow points to the 'Clustered Column' icon. Below the ribbon, a small 3D bar chart is visible. To the right, a table titled 'LEAF BLOWER' is shown, which matches the data in the previous blocks.

A new chart should pop up.



7. Add some labels to your chart by following the same steps you did in Exercise 1:

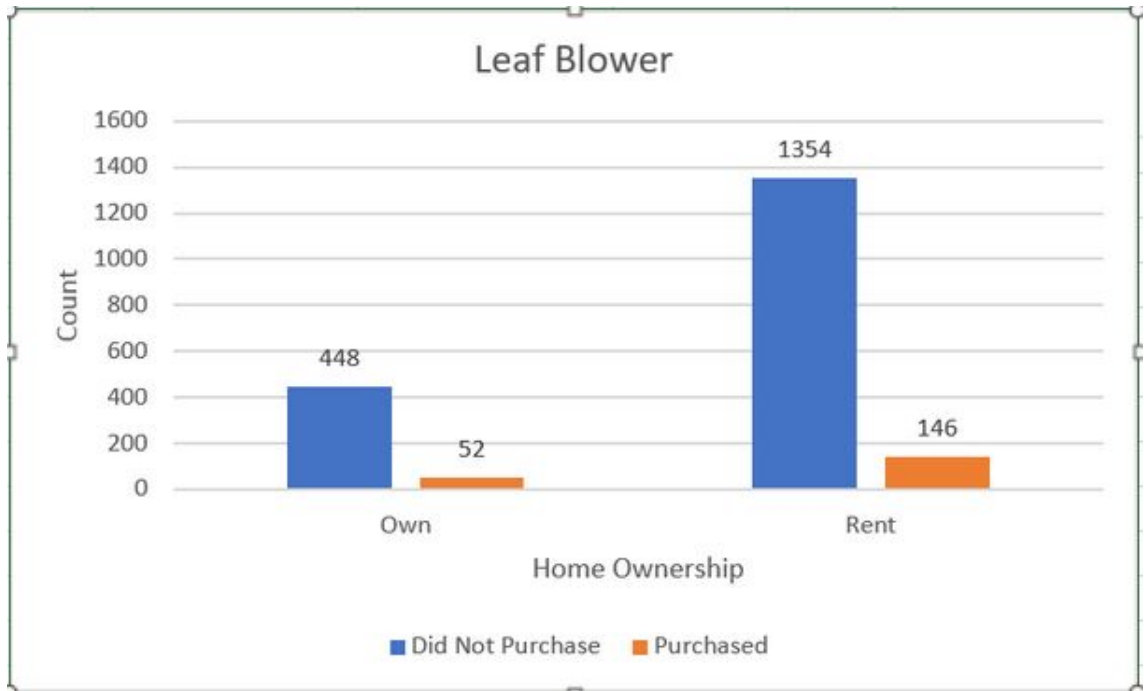
Counts above the bars: **Chart Tools > Data Labels > Outside End**

Horizontal axis title: **Chart Tools > Axis Titles > Primary Horizontal Axis Title > Edit Horizontal Axis Title**, then type "Home Ownership"

Vertical axis title: **Chart Tools > Axis Titles > Primary Vertical Axis Title > Edit Vertical Axis Title**, then type "Count"

Chart title: **Chart Tools > Chart Title > Edit Chart Title**, then type "Leaf Blower"

And boom, your chart looks great!

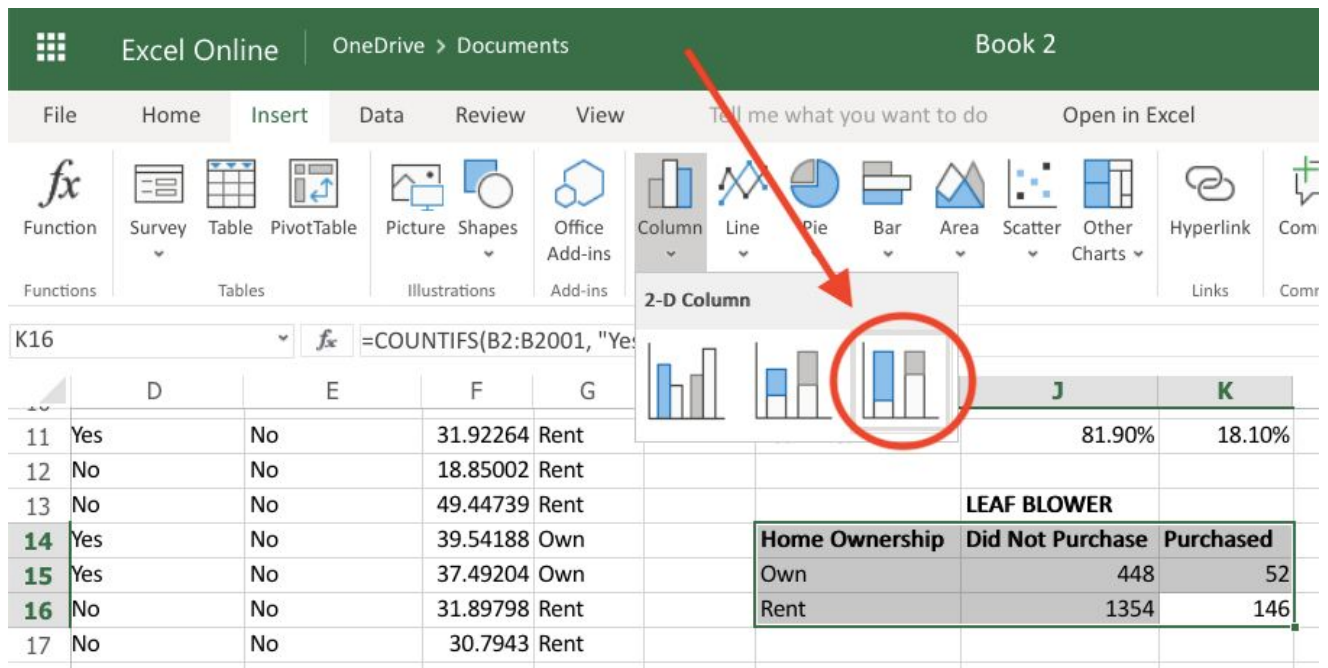


8. Going back to our original question: How do our customer's purchases differ by home ownership status? Well, according to the graph, there was clearly a higher number of renters purchasing (and not purchasing) the Leaf Blower, but how can we compare them?

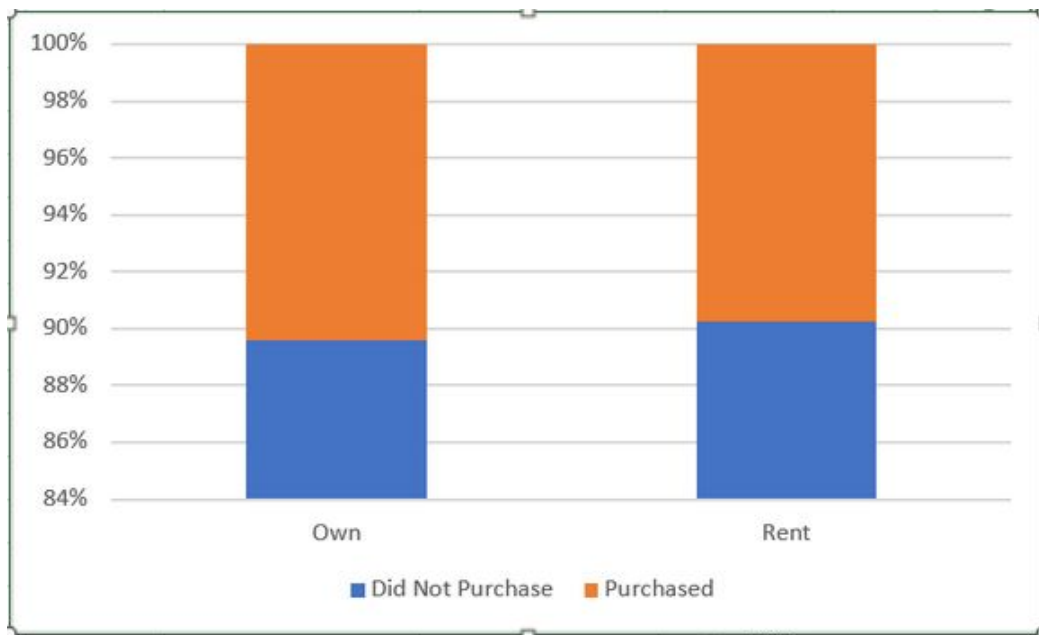
Yep, it's percentages to the rescue again. This time, we'll just run a quick chart in Excel Online to check the percentages. Once again, highlight everything in your new mini-table EXCEPT for the "LEAF BLOWER" title.

	LEAF BLOWER	
Home Ownership	Did Not Purchase	Purchased
Own	448	52
Rent	1354	146

Then click Insert > Column > 100% Stacked Column (on the far right).



Another glorious chart pops up.



Although the actual counts were wildly different, when you chart them as percentages like this, you can see that the rates are actually pretty similar by home ownership. In other words, there's not a huge difference between the proportion of home owners who purchased a Leaf Blower and the proportion of renters who did. The "Did Not Purchase" and "Purchased" bars are roughly the same size in both groups.

- Note: If we ran a chi-squared ( $\chi^2$ ) independence test on these results (remember when we mentioned this earlier?), we'd see that the difference here is not statistically significant. There's no real difference between the groups!

- Now let's look at how the home ownership stats stacked up for the Leaf Destroyer model. Below the table you created in Step 1 (still in columns I, J, and K), create a new mini-table. This one should break down home ownership ("Own" or "Rent") by purchase ("Did Not Purchase" or "Purchased") for the Leaf *Destroyer* product. It should look like this:

	I	J	K
13		<b>LEAF BLOWER</b>	
14	<b>Home Ownership</b>	<b>Did Not Purchase</b>	<b>Purchased</b>
15	Own	448	52
16	Rent	1354	146
17			
18		<b>LEAF DESTROYER</b>	
19	<b>Home Ownership</b>	<b>Did Not Purchase</b>	<b>Purchased</b>
20	Own		
21	Rent		
22			

- Use the COUNTIFS function once again to populate this table. The syntax is still **=COUNTIFS(criteria\_range1, criteria1, criteria\_range2, criteria2)**. Both ranges take the form **firstcell:lastcell** (with a colon in between). The first criteria (criteria1) is either a "No" or "Yes" (in quotation marks) from the Leaf Destroyer column (column C). The second criteria (criteria2) is either "Own" or "Rent" (also in quotation marks) from the "home" column (column G).

Basically, you'll repeat Steps 2, 3, 4, and 5 from this current exercise, but replace all the B ranges with C ranges (since the Leaf Destroyer data runs from C2:C2001).

For people who did not purchase a Leaf Destroyer and own a home:

J20      **=COUNTIFS(C2:C2001, "No", G2:G2001, "Own")**

For people who purchased a Leaf Destroyer and own a home:

K20      **=COUNTIFS(C2:C2001, "Yes", G2:G2001, "Own")**

For people who did not purchase a Leaf Destroyer and rent:



J21       $\text{fx}$       =COUNTIFS(C2:C2001, "No", G2:G2001, "Rent")

For people who purchased a Leaf Destroyer and rent:

K21       $\text{fx}$       =COUNTIFS(C2:C2001, "Yes", G2:G2001, "Rent")

And here's what the final table should look like:

	LEAF DESTROYER	
Home Ownership	Did Not Purchase	Purchased
Own	4	496
Rent	668	832

11. Let's do another graph. Highlight everything in this new mini-table EXCEPT for the "LEAF DESTROYER" title.

	LEAF DESTROYER	
Home Ownership	Did Not Purchase	Purchased
Own	4	496
Rent	668	832

Click Insert > Column > Clustered Column (on the left).

The screenshot shows the Excel Online interface with the 'Insert' tab selected. The 'Column' group is expanded, and the 'Clustered Column' chart icon is highlighted with a red circle and a red arrow. Below the ribbon, a small data table is visible with columns B, C, D, E, and H. Row 12 shows 'No' for C, D, and E. Row 13 shows 'Yes' for C, 'No' for D, and '49.44739 Rent' for E.

12. When your new chart pops into existence, do all the usual stuff to create chart titles and axis titles:

Counts above the bars: **Chart Tools > Data Labels > Outside End**

Horizontal axis title: **Chart Tools > Axis Titles > Primary Horizontal Axis Title > Edit Horizontal Axis Title**, then type “Home Ownership”

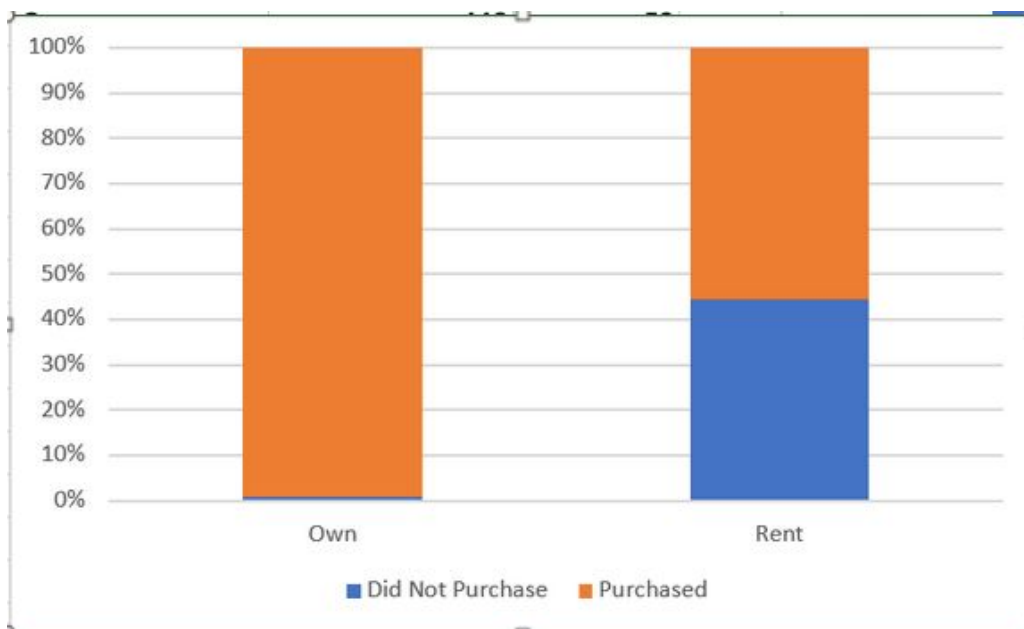
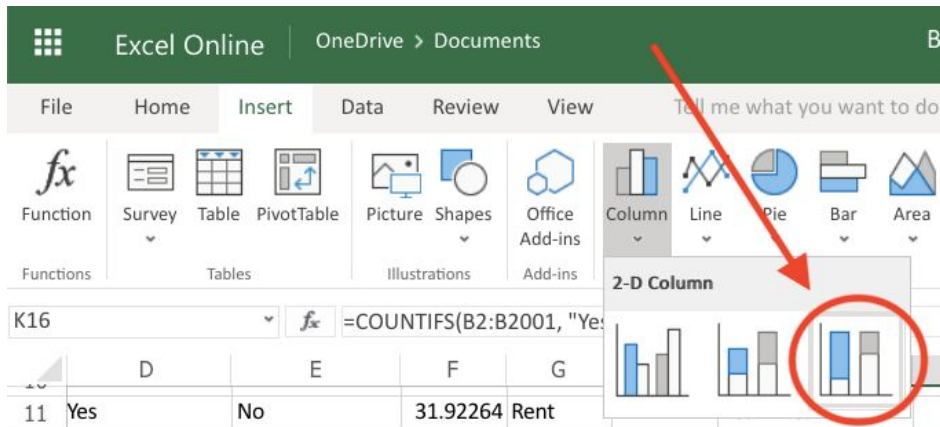
Vertical axis title: **Chart Tools > Axis Titles > Primary Vertical Axis Title > Edit Vertical Axis Title**, then type “Count”

Chart title: **Chart Tools > Chart Title > Edit Chart Title**, then type “Leaf Destroyer”

Gaze upon thy mighty chart:



- There does seem to be a pretty huge difference here (look at that tiny, tiny bar of 4 people!), but we can run another quick percentage chart to be sure. Once again, highlight everything in your new mini-table EXCEPT for the “LEAF DESTROYER” title, then click Insert > Column > 100% Stacked Column (on the far right).



Wow—the percentages are wildly different this time. It looks like there is a very large difference in the proportion of home owners versus renters who purchased the Leaf Destroyer.

- Note: A chi-squared ( $\chi^2$ ) independence test with these results would show that the difference here *is* statistically significant this time. There's a real difference between the groups.

14. Let's run through creating the charts again, this time for the Leaf Storm-nado model. Below the tables you made earlier (still in columns I, J, and K), create a new mini-table. This one should break down home ownership ("Own" or "Rent") by purchase ("Did Not Purchase" or "Purchased") for the Leaf Storm-nado product.

	LEAF STORM-NADO	
Home Ownership	Did Not Purchase	Purchased
Own		
Rent		

15. As usual, use the COUNTIFS function to populate the table. The syntax is still **=COUNTIFS(criteria\_range1, criteria1, criteria\_range2, criteria2)**. Both ranges take the form **firstcell:lastcell** (with a colon in between). The first criteria (criteria1) is either a "No" or "Yes" (in quotation marks) from the Leaf Storm-nado column (column D). The second criteria (criteria2) is either "Own" or "Rent" (also in quotation marks) from the "home" column (column G).

Again, think of it like this: Repeat Steps 2, 3, 4, and 5 from this current exercise, but replace all the B ranges with D ranges (since the Leaf Storm-nado data runs from D2:D2001).

For people who did not purchase a Leaf Storm-nado and own a home:

J25  =COUNTIFS(D2:D2001, "No", G2:G2001, "Own")

For people who purchased a Leaf Storm-nado and own a home:

K25  =COUNTIFS(D2:D2001, "Yes", G2:G2001, "Own")

For people who did not purchase a Leaf Storm-nado and rent:

J26  =COUNTIFS(D2:D2001, "No", G2:G2001, "Rent")

For people who purchased a Leaf Storm-nado and rent:

K26  =COUNTIFS(D2:D2001, "Yes", G2:G2001, "Rent")

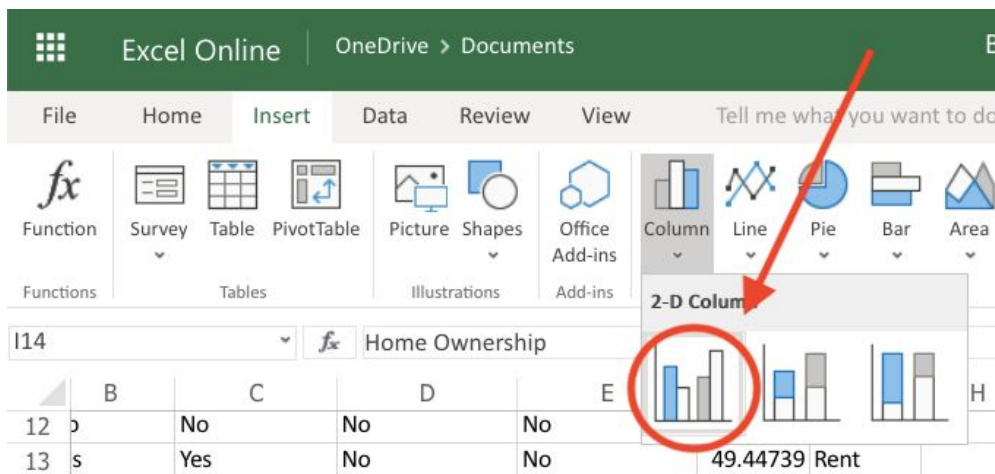
Final table:

	LEAF STORM-NADO	
Home Ownership	Did Not Purchase	Purchased
Own	206	294
Rent	830	670

16. Yep, you guessed it: It's time for another chart. Highlight everything in this new mini-table EXCEPT for the "LEAF STORM-NADO" title.

LEAF STORM-NADO		
Home Ownership	Did Not Purchase	Purchased
Own	206	294
Rent	830	670

Click Insert > Column > Clustered Column (on the left).



17. We know we're starting to sound like a broken record, but when the new chart pops up, repeat all the usual steps to create chart titles and axis titles:

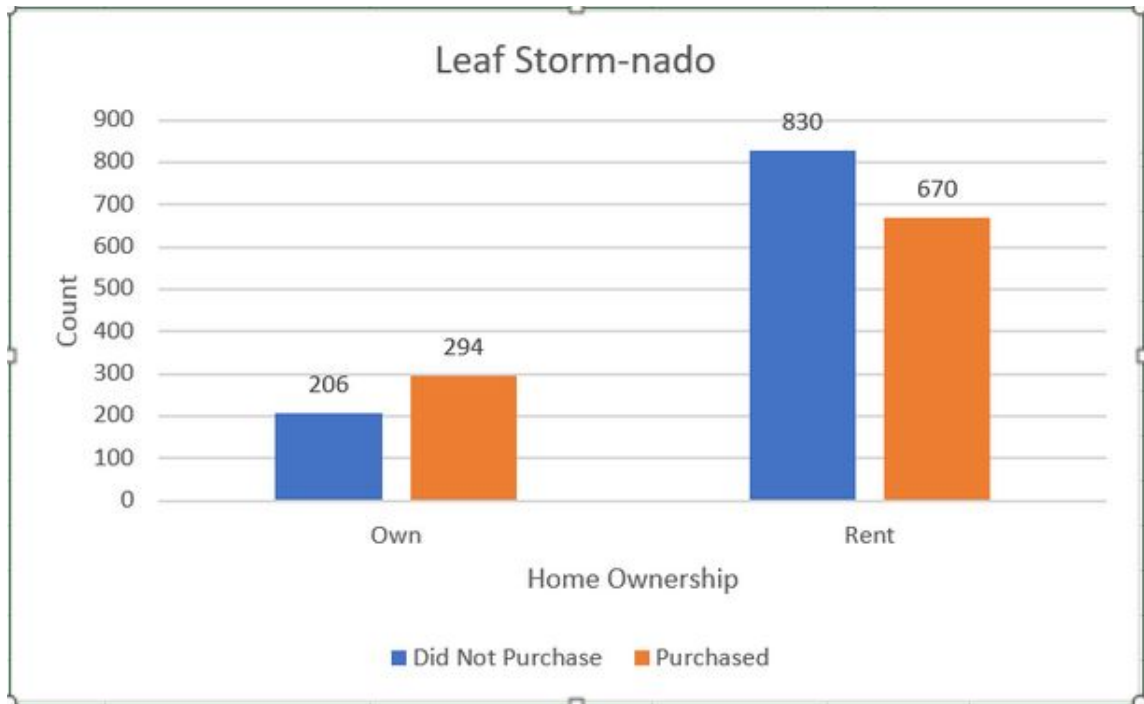
Counts above the bars: **Chart Tools > Data Labels > Outside End**

Horizontal axis title: **Chart Tools > Axis Titles > Primary Horizontal Axis Title > Edit Horizontal Axis Title**, then type "Home Ownership"

Vertical axis title: **Chart Tools > Axis Titles > Primary Vertical Axis Title > Edit Vertical Axis Title**, then type "Count"

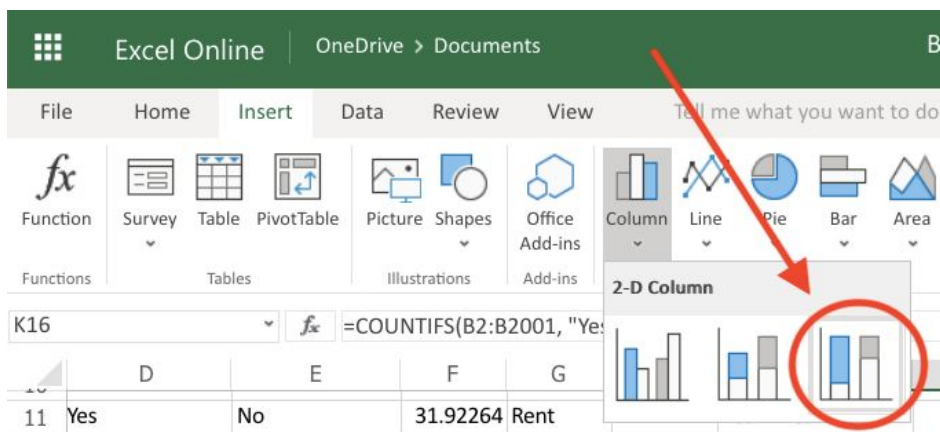
Chart title: **Chart Tools > Chart Title > Edit Chart Title**, then type "Leaf Storm-nado"

And here's your chart:



18. Hmm... it's tough to say whether there's a significant difference here. The counts are pretty different for home owners versus renters, but the counts within those categories are grouped fairly close together. A bird's-eye view of the percentages could help.

Once again, highlight everything in your new mini-table EXCEPT for the "LEAF STORM-NADO" title, then click Insert > Column > 100% Stacked Column (on the far right).





Hey, interesting. The percentages aren't terribly far apart, but home owners definitely tended to purchase the Leaf Storm-nado (we never get tired of saying that) a little more often than renters did.

- Note: A chi-squared ( $\chi^2$ ) independence test with these results would show that the difference here *is* statistically significant. There's a real difference between the groups.

19. Now it's time to repeat all those steps one more time for the Leaf A-leaf-eator product. Below the tables you made earlier (still in columns I, J, and K), create one final new mini-table. This one should break down home ownership ("Own" or "Rent") by purchase ("Did Not Purchase" or "Purchased") for the Leaf A-leaf-eator.

	LEAF A-LEAF-EATOR	
Home Ownership	Did Not Purchase	Purchased
Own		
Rent		

20. Use the COUNTIFS function to populate the table (this is the last time, we promise). The syntax is still **=COUNTIFS(criteria\_range1, criteria1, criteria\_range2, criteria2)**. Both ranges take the form **firstcell:lastcell** (with a colon in between). The first criteria (criteria1) is either a "No" or "Yes" (in quotation marks) from the Leaf A-leaf-eator column (column E this time). The second criteria (criteria2) is either "Own" or "Rent" (also in quotation marks) from the "home" column (column G).

Repeat Steps 2, 3, 4, and 5 from this current exercise, but replace all the B ranges with E ranges (since the Leaf A-leaf-eator data runs from E2:E2001).



For people who did not purchase a Leaf A-leaf-eator and own a home:

J30                =COUNTIFS(E2:E2001, "No", G2:G2001, "Own")

For people who purchased a Leaf A-leaf-eator and own a home:

K30                =COUNTIFS(E2:E2001, "Yes", G2:G2001, "Own")

For people who did not purchase a Leaf A-leaf-eator and rent:

J31                =COUNTIFS(E2:E2001, "No", G2:G2001, "Rent")

For people who purchased a Leaf A-leaf-eator and rent:

K31                =COUNTIFS(E2:E2001, "Yes", G2:G2001, "Rent")

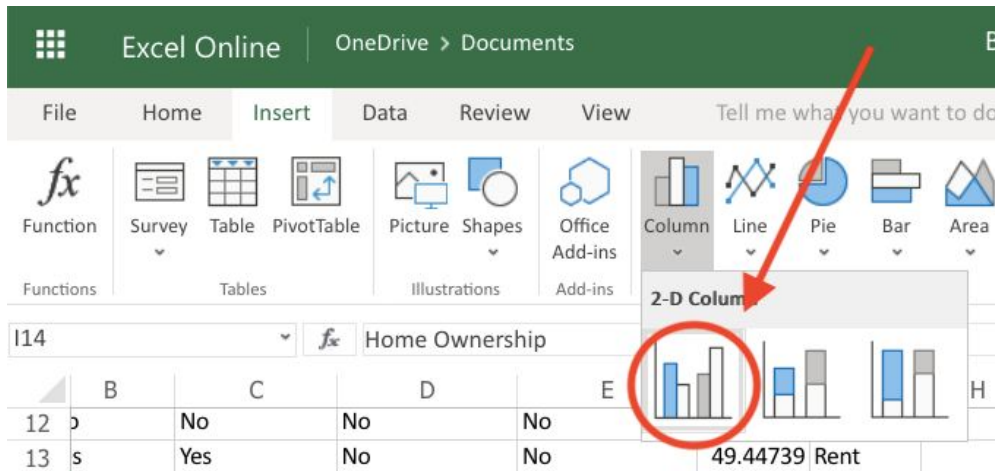
Final table:

	LEAF A-LEAF-EATOR	
Home Ownership	Did Not Purchase	Purchased
Own	492	8
Rent	1146	354

21. Let's throw this info into a chart again. Highlight everything in this new mini-table EXCEPT for the "LEAF A-LEAF-EATOR" title.

	LEAF A-LEAF-EATOR	
Home Ownership	Did Not Purchase	Purchased
Own	492	8
Rent	1146	354

Click Insert > Column > Clustered Column (on the left).



22. When your chart pops up, it's time to create chart titles and axis titles, as usual:

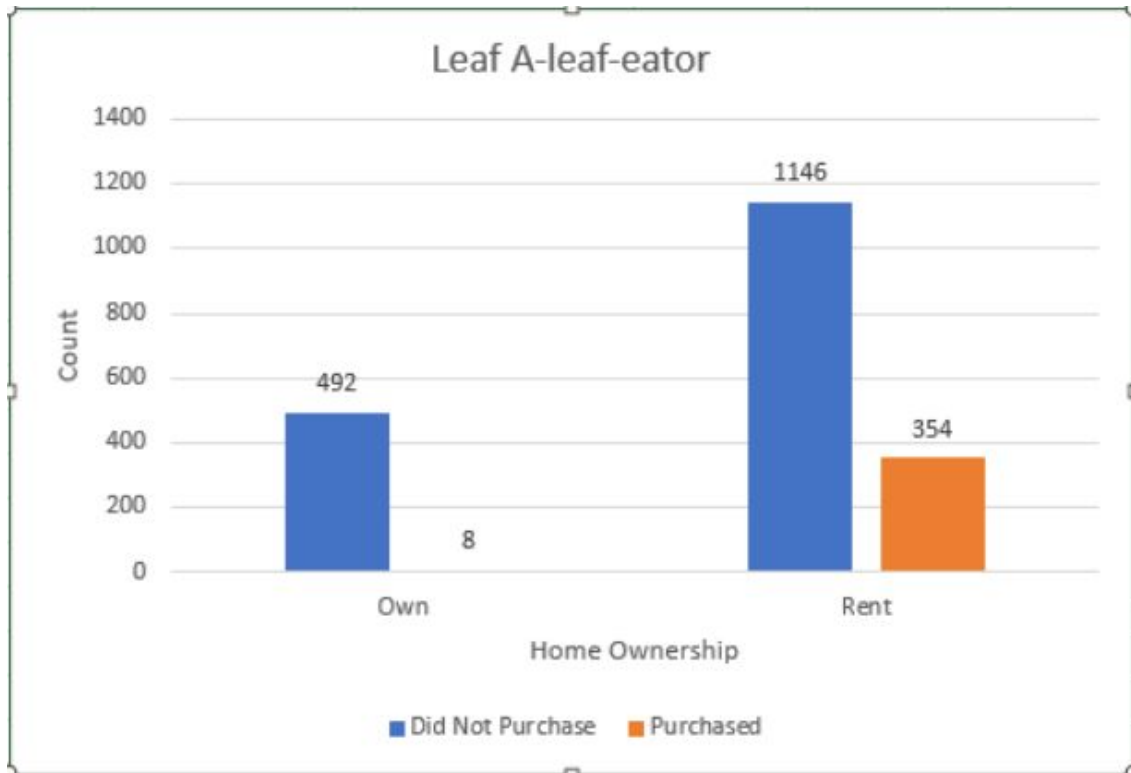
Counts above the bars: **Chart Tools > Data Labels > Outside End**

Horizontal axis title: **Chart Tools > Axis Titles > Primary Horizontal Axis Title > Edit Horizontal Axis Title**, then type "Home Ownership"

Vertical axis title: **Chart Tools > Axis Titles > Primary Vertical Axis Title > Edit Vertical Axis Title**, then type "Count"

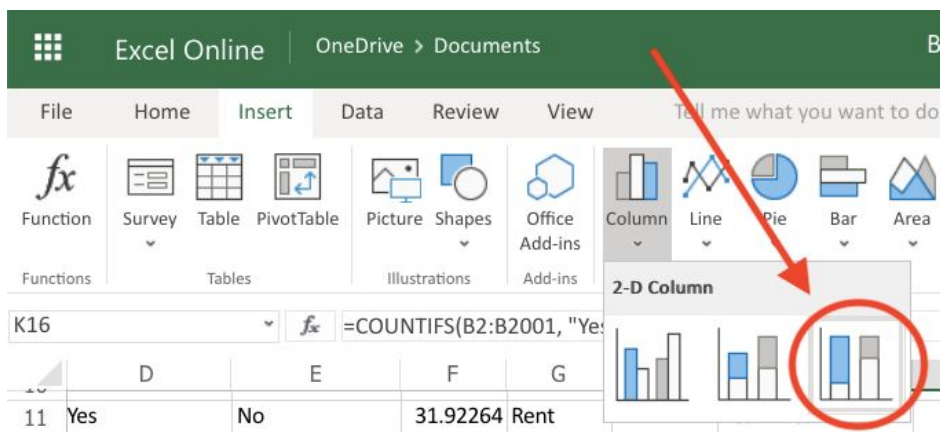
Chart title: **Chart Tools > Chart Title > Edit Chart Title**, then type "Leaf A-leaf-eator"

And here's your chart:



23. Looks like very few home owners purchased this product—you can't even see the orange bar representing those 8 people. To see how the rates of purchase for home owners and renters stack up against each other, create one final chart for the percentages.

Highlight everything in your new mini-table EXCEPT for the "LEAF A-LEAF-EATOR" title, then click Insert > Column > 100% Stacked Column (on the far right).





Now we can see that was a pretty significant difference between the two groups: Only a tiny sliver of home owners bought the Leaf A-leaf-eator, compared to over 20% of renters who bought the product.

- Note: A chi-squared ( $\chi^2$ ) independence test with these results would show that the difference here *is* statistically significant. There is indeed a real difference between the groups.