In-Class Exercise: Pivot Tables and Charts Tutorial

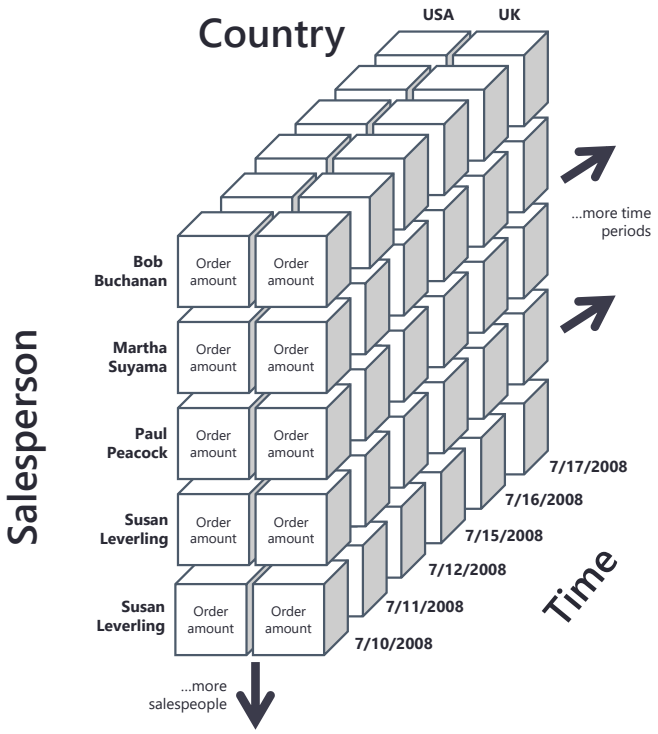
Consider the data cube is contained in the file "Salesperson Cube.xlsx." There is information about 800 sales transactions that occurred from 2008 to 2010 at a fictitious company. Here are the first few records:

	A	B	C	D	E	F	G	H	I	J	K
1	OrderID	SalesPerson ID	Salesperson LN	Salesperson FN	Salesperson Salary	Order Date	Order Month	Order Day	Order Year	Country	Order Amount
2	10248	101	Buchanan	Bob	$ 78,000.00	7/16/2008	7	16	2008	UK	$ 440.00
3	10249	102	Suyama	Martha	$ 82,000.00	7/10/2008	7	10	2008	UK	$ 1,863.40
4	10250	103	Peacock	Paul	$ 75,000.00	7/12/2008	7	12	2008	USA	$ 1,552.60
5	10251	104	Leverling	Susan	$ 65,000.00	7/15/2008	7	15	2008	USA	$ 654.06
6	10252	103	Peacock	Paul	$ 75,000.00	7/11/2008	7	11	2008	USA	$ 3,597.90
7	10253	104	Leverling	Susan	$ 65,000.00	7/16/2008	7	16	2008	USA	$ 1,444.80
8	10254	101	Buchanan	Bob	$ 78,000.00	7/23/2008	7	23	2008	UK	$ 556.62
9	10255	105	Dodsworth	Brian	$ 124,000.00	7/15/2008	7	15	2008	UK	$ 2,490.50
10	10256	104	Leverling	Susan	$ 65,000.00	7/17/2008	7	17	2008	USA	$ 517.80
11	10257	103	Peacock	Paul	$ 75,000.00	7/22/2008	7	22	2008	USA	$ 1,119.90
12	10258	106	Davolio	Angela	$ 142,000.00	7/23/2008	7	23	2008	USA	$ 1,614.88
13	10259	103	Peacock	Paul	$ 75,000.00	7/25/2008	7	25	2008	USA	$ 100.80
14	10260	103	Peacock	Paul	$ 75,000.00	7/29/2008	7	29	2008	USA	$ 1,504.65
15	10261	103	Peacock	Paul	$ 75,000.00	7/30/2008	7	30	2008	USA	$ 448.00
16	10262	107	Callahan	Art	$ 65,000.00	7/25/2008	7	25	2008	USA	$ 584.00

Keep in mind that this is the entire data set in a joined, cube format. It's not a typical, large-scale data cube – that would only contain summary data. In this tutorial, you'll be basically doing the summarization work of the dimensional engine by aggregating data as you need it. This works well for small data sets and gives you a lot of flexibility, but as your data set gets very large this would need to be implemented as a summarized cube.

The underlying star schema has three dimensions – Salesperson, Country, and Time. You can see data associated with each of those dimensions. For example, the Salesperson dimension consists of Salesperson LN, Salesperson FN, and Salesperson Salary.

So a diagram of our data cube looks would look like the image on the right.



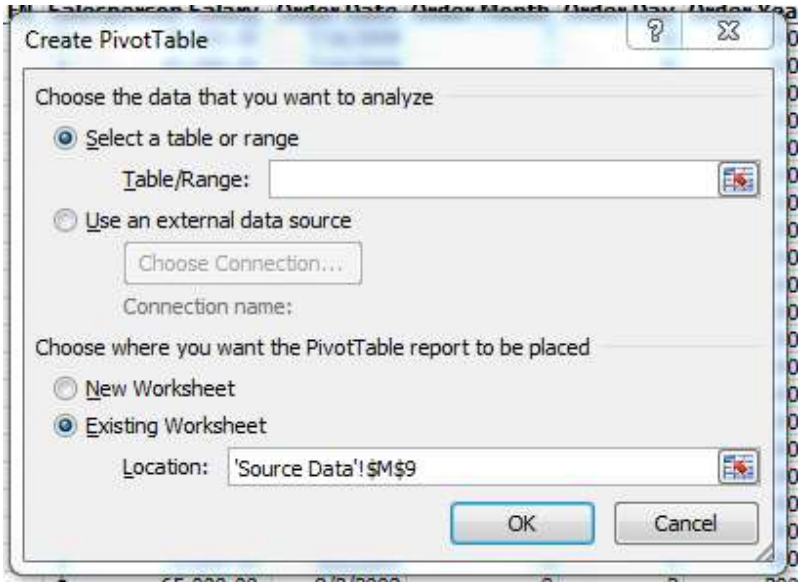
There are only two values for the Country dimension (USA and UK), but there are a lot for Salesperson and Time. In fact, there are too many to draw so we've used arrows to indicate that the cube keeps going in that direction for a while. But this gives you a feel for the structure of our data mart, and it also indicates to us that we will be able to aggregate and filter our results by salesperson, time, and country.

Notice time is expressed two different ways in our table. The first is with day, month, and year as separate attributes; the second as a standard date/time value (mm/dd/yyyy). It isn't necessary to do this – a single date attribute would work just fine – but this allows us to look at a particular month (or year) a

little more easily. Finally, order amount is our measured fact. So the indvidual order is the level of granularity for the fact table.

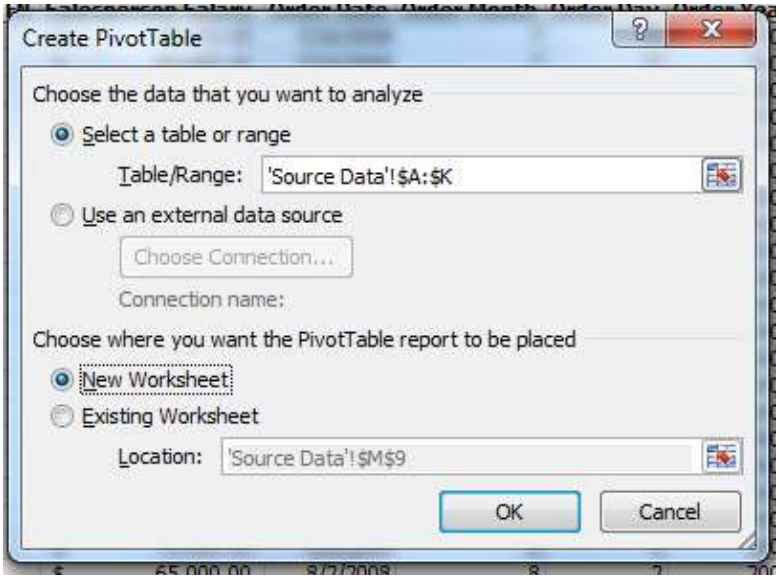
Create a basic pivot table

1) Open the spreadsheet and go to the Insert tab. Click “Pivot Table/Pivot Table.” You’ll see this dialog:



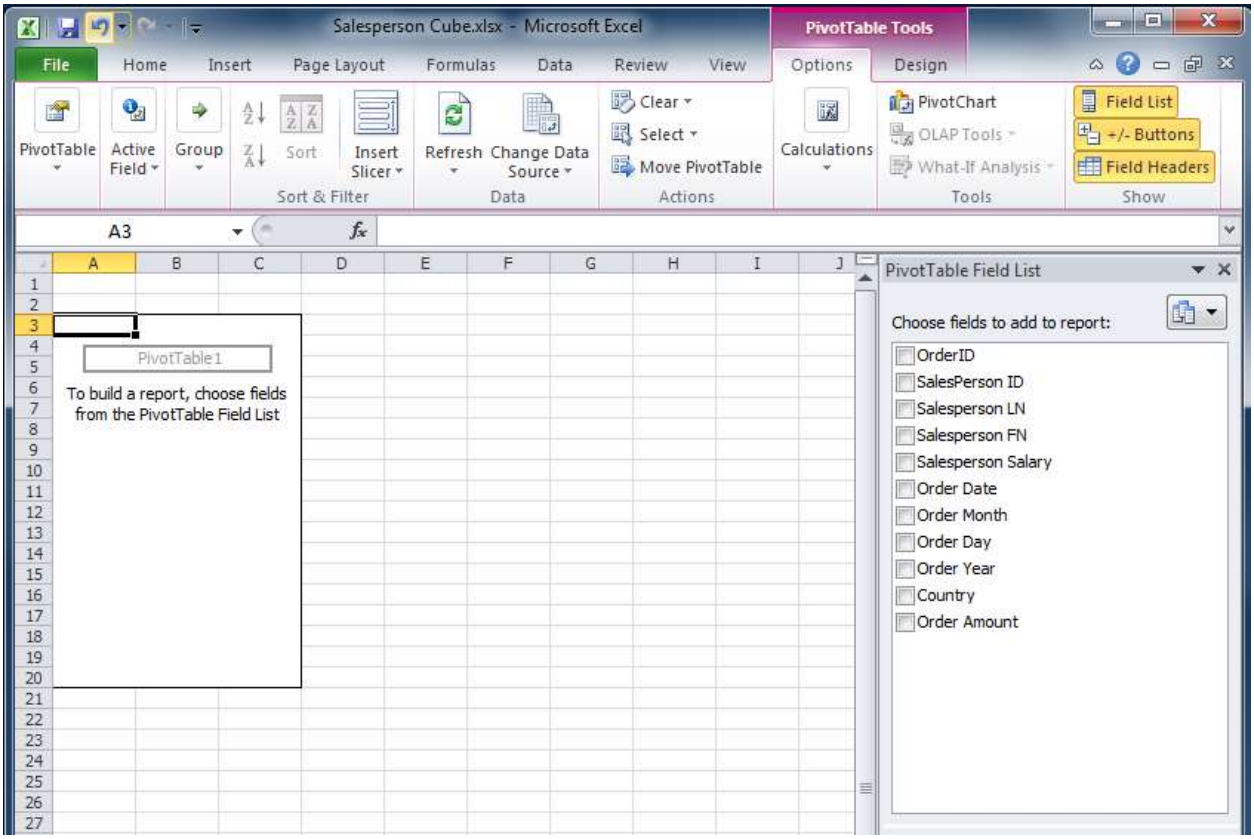
2) Select the entire table by clicking on column A (on the “A” at the top of the column) and dragging across to column K. It will place that range in the dialog box. S*elect the columns, even if they are already selected!*

Now place your PivotTable report on a new worksheet (so it’s easier to see). The dialog should look like this:

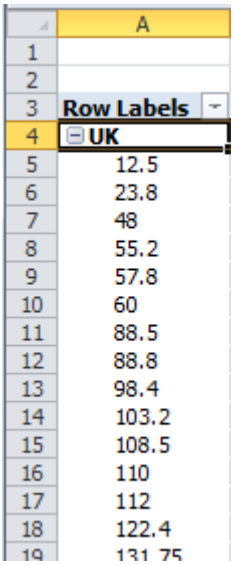


If it does, then click “OK.”

3) A new worksheet will be opened, and you’ll see the display below. There is a blank space on the left that will be populated with the data from your data cube, and a set of fields on the right.



4) Let’s figure out how much was sold in the in the UK versus the USA. Drag the Country and Order Amount fields into the Row Labels box at the bottom of the right panel. Now you’ll see the list pictured to the right.w



You’re looking at each sale, organized by country. However, this isn’t very helpful since you still have to manually add up each order to see which country had more sales.

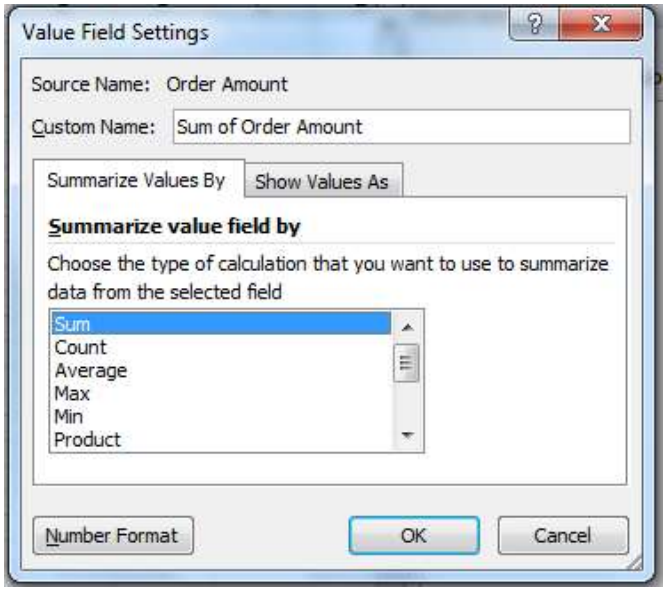
5) Now click and drag Order Amount to the Values box. You'll see this:

	A	B
1		
2		
3	**Row Labels**	**Sum of Order Amount**
4	UK	333330.91
5	USA	894996.49
6	(blank)	
7	**Grand Total**	**1228327.4**
8		

This shows us that $333,330.91 was sold in the UK, and $894,996.49 was sold in the USA.

6) Instead, let's say you wanted to see the number of orders, not the total amount. In the Values box, click Count of Order Amount and select "Value Field Settings."

7) Change the summarization from "Sum" to "Count."



8) Now you will see this:

	A	B
1		
2		
3	**Row Labels**	**Count of Order Amount**
4	UK	215
5	USA	584
6	(blank)	
7	**Grand Total**	**799**
8		

Which tells us is that there were 215 orders in the UK and 584 orders in the USA. It's just counting rows in for each value in the Country dimension.

Try it	Using the Value Field Settings dialog, answer the following questions: What is the amount of the largest single daily order placed in each country? Which country has larger daily orders on average?

Sorting and Multiple Levels of Categorization

1) Create a new pivot table in a new worksheet. Again, select the entire table (Columns A through K).

2) Drag Salesperson LN and Order Year to the Row Labels box and Order Amount to the Values box. This makes Order Amount the computed value, organized by Salesperson LN and Order Year.

3) Make sure the Value Field settings reflect computing the sum of Order Amount and not the count.

4) Your pivot table should look like this:

	A	B
1		
2		
3	**Row Labels**	**Sum of Order Amount**
4	**Buchanan**	**68792.25**
5	2008	17667.2
6	2009	31433.16
7	2010	19691.89
8	**Callahan**	**123032.67**
9	2008	19160.7
10	2009	56954.02
11	2010	46917.95
12	**Davolio**	**182500.09**
13	2008	30861.76
14	2009	95850.36
15	2010	55787.97
16	**Dodsworth**	**75048.04**
17	2008	9894.51
18	2009	24756.89
19	2010	40396.64
20	**Fuller**	**162503.78**
21	2008	17811.46
22	2009	71168.14
23	2010	73524.18

5) We can figure out which salesperson sold the most by sorting the records. Next to “Row Labels” at the top of the table there is a pull-down menu (). From that menu choose “More Sort Options…”

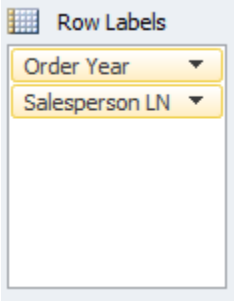
6) In the sort dialog, choose “Descending” and “Sum of Order Amount” as the value to sort. The dialog should look like this:

Then click “OK.”

7) We now see that Peacock sold the most ($225,763.68) over the years 2008-2010.

	A	B
1		
2		
3	**Row Labels**	**Sum of Order Amount**
4	**Peacock**	**225763.68**
5	2008	49945.11
6	2009	124655.56
7	2010	51163.01
8	**Leverling**	**201196.27**
9	2008	18223.96
10	2009	103719.07
11	2010	79253.24
12	**Davolio**	**182500.09**
13	2008	30861.76

8) However, let’s say we’re really interested in who sold the most in any particular year. We need to organize the table by year and then by salesperson, instead of the way it is now. To do that, drag the Salesperson LN field in the “Row Labels” box below Order Year, like this:



9) You'll notice the table has rearranged itself, and that it still sorts by sum of order amount within each year. We now see that Peacock was at the top in 2008 and 2009, but dropped to fourth in 2010. Leverling, who was in fourth place in 2008, was at the top in 2010. We also see that 2009 was the best year overall.

	A	B
1		
2		
3	**Row Labels**	**Sum of Order Amount**
4	**2008**	**193316.54**
5	Peacock	49945.11
6	Davolio	30861.76
7	Callahan	19160.7
8	Leverling	18223.96
9	Fuller	17811.46
10	Buchanan	17667.2
11	King	15232.16
12	Suyama	14519.68
13	Dodsworth	9894.51
14	**2009**	**609190.76**
15	Peacock	124655.56
16	Leverling	103719.07
17	Davolio	95850.36
18	Fuller	71168.14
19	King	59827.19
20	Callahan	56954.02
21	Suyama	40826.37
22	Buchanan	31433.16
23	Dodsworth	24756.89
24	**2010**	**425820.1**
25	Leverling	79253.24
26	Fuller	73524.18
27	Davolio	55787.97
28	Peacock	51163.01
29	Callahan	46917.95
30	King	41903.64
31	Dodsworth	40396.64
32	Buchanan	19691.89
33	Suyama	17181.58
34	**(blank)**	
35	(blank)	
36	**Grand Total**	**1228327.4**

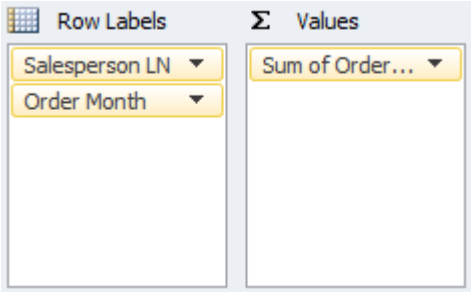
Try it

Find out which salespeople work in the UK and which ones work in the USA.

Find out if the best paid salespeople are also the best performers (all salary amounts are in US Dollars).

Filtering

1) Create a pivot table that looks at total sales by salesperson by month. Your row labels and computed value fields should look like this:



and your table will look like this:

	A	B
1		
2		
3	**Row Labels**	**Sum of Order Amount**
4	**Buchanan**	**68792.25**
5	1	9816.52
6	2	7980.06
7	3	4922.43
8	4	210
9	5	4500.27
10	6	2147.4
11	7	8362.02
12	8	2076.2
13	9	5596.4
14	10	8974.52
15	11	4385.33
16	12	9821.1
17	**Callahan**	**123032.67**
18	1	20456.89
19	2	6853.36
20	3	19712.83
21	4	18493.6
22	5	8157.9
23	6	1922.32

This is still giving us yearly totals for each salesperson. But let's say we just wanted to see sales during the holiday season (November and December). We'll need to apply a filter.

2) In the Pivot Table field list, click on the right side of the Order Month entry. A menu will appear.

3) From the menu choose "Label Filters/Between…" You're choosing "Label Filters" because Order Month is a Row Label.

   Another way of thinking about it is that you are constraining the Time Dimension, and dimensions are essentially labels in a pivot table.

4) In the Label Filter Dialog, enter 11 and 12 for the range, like this:

Click "OK."

5) You'll now see the total orders for only November and December:

	A	B
1		
2		
3	**Row Labels**	**Sum of Order Amount**
4	**Buchanan**	**14206.43**
5	11	4385.33
6	12	9821.1
7	**Callahan**	**13204.28**
8	11	6021.55
9	12	7182.73
10	**Davolio**	**36196.9**
11	11	15348.8
12	12	20848.1
13	**Dodsworth**	**7673.86**
14	11	7495.36
15	12	178.5
16	**Fuller**	**17946.3**
17	11	8221.3

Note that you could also have just checked the values off of the pull-down menu instead of entering a range (go back to the pull-down menu and you'll see what I mean). This would have been just as easy in this case because there were only 12 values to choose from. But if you had many possible values it is usually easier just to enter the range through the dialog box.

6) Notice you'll also see a filter icon next to Order Month in the Field List.

That indicates a filter has been applied to that label.

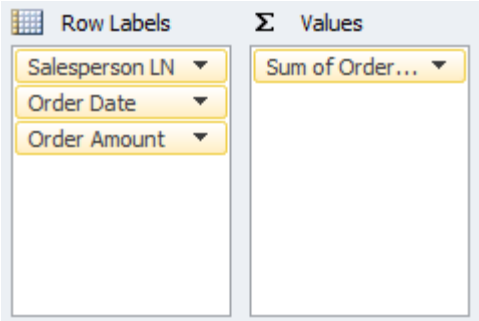
7) Now remove the filter by clicking on the filter icon and selecting “Clear Filter from ‘Order Amount’” from the drop-down menu.

8) Now let’s say we want to see which sales people have had a lot of low volume days. Let’s define “low volume” as less than $200 in a day. So first let’s create a pivot table that looks at total sales (Sum of Order Amount) per salesperson (Salesperson LN) per day (Order Date), like this:

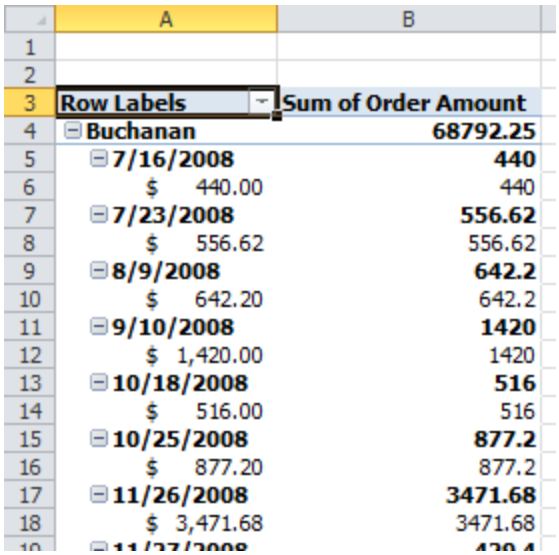
	A	B
1		
2		
3	**Row Labels**	**Sum of Order Amount**
4	**Buchanan**	**68792.25**
5	7/16/2008	440
6	7/23/2008	556.62
7	8/9/2008	642.2
8	9/10/2008	1420
9	10/18/2008	516
10	10/25/2008	877.2
11	11/26/2008	3471.68
12	11/27/2008	429.4
13	12/9/2008	9210.9
14	12/19/2008	103.2
15	1/2/2009	716.72
16	3/6/2009	713.3
17	3/21/2009	1249.1
18	3/25/2009	558
19	5/9/2009	946
20	5/30/2009	3554.27
21	6/30/2009	2147.4
22	7/11/2009	890

9) We need to apply a filter to order amount. However, we can’t apply a filter to a computed value (i.e., “Sum of Order Amount”). We need to also make “Order Amount” a row label; then we can apply a filter.

So drag the Order Amount field (it’s already checked but don’t worry about that) to Row Labels. You will now see this:

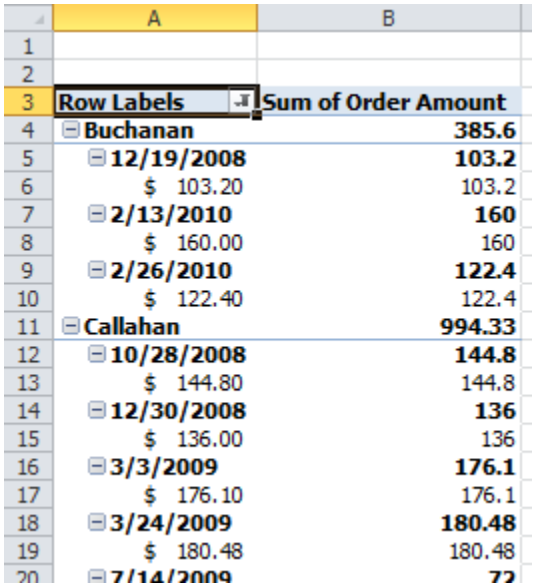


and your Pivot Table will look like this:



	A	B
1		
2		
3	**Row Labels**	**Sum of Order Amount**
4	**Buchanan**	**68792.25**
5	**7/16/2008**	**440**
6	$ 440.00	440
7	**7/23/2008**	**556.62**
8	$ 556.62	556.62
9	**8/9/2008**	**642.2**
10	$ 642.20	642.2
11	**9/10/2008**	**1420**
12	$ 1,420.00	1420
13	**10/18/2008**	**516**
14	$ 516.00	516
15	**10/25/2008**	**877.2**
16	$ 877.20	877.2
17	**11/26/2008**	**3471.68**
18	$ 3,471.68	3471.68
19	**11/27/2008**	**429.4**

10) Now create a label filter for Order Amount, showing only values less than 200 (go back and look at steps 1 through 5 if don't remember how to do this). You'll now see fewer entries for each salesperson now:



	A	B
1		
2		
3	**Row Labels**	**Sum of Order Amount**
4	**Buchanan**	**385.6**
5	**12/19/2008**	**103.2**
6	$ 103.20	103.2
7	**2/13/2010**	**160**
8	$ 160.00	160
9	**2/26/2010**	**122.4**
10	$ 122.40	122.4
11	**Callahan**	**994.33**
12	**10/28/2008**	**144.8**
13	$ 144.80	144.8
14	**12/30/2008**	**136**
15	$ 136.00	136
16	**3/3/2009**	**176.1**
17	$ 176.10	176.1
18	**3/24/2009**	**180.48**
19	$ 180.48	180.48
20	**7/14/2009**	**72**

11) Now change the Value Field Settings for “Sum of Order Amount” to compute the count instead of the total. You’ll see counts instead of dollar amounts:

	A	B
1		
2		
3	**Row Labels**	**Count of Order Amount**
4	**Buchanan**	**3**
5	**12/19/2008**	**1**
6	$ 103.20	1
7	**2/13/2010**	**1**
8	$ 160.00	1
9	**2/26/2010**	**1**
10	$ 122.40	1
11	**Callahan**	**10**
12	**10/28/2008**	**1**
13	$ 144.80	1
14	**12/30/2008**	**1**
15	$ 136.00	1
16	**3/3/2009**	**1**
17	$ 176.10	1
18	**3/24/2009**	**1**
19	$ 180.48	1

12) It’s still a little difficult to read, so click on “Buchanan” and click on the “Collapse Entire Field” icon in the ribbon (). That gives you just what you want, and tells you that Leverling had the most days below $200:

	A	B
1		
2		
3	**Row Labels**	**Count of Order Amount**
4	**Buchanan**	**3**
5	**Callahan**	**10**
6	**Davolio**	**9**
7	**Dodsworth**	**7**
8	**Fuller**	**11**
9	**King**	**9**
10	**Leverling**	**14**
11	**Peacock**	**12**
12	**Suyama**	**6**
13	**Grand Total**	**81**
14		

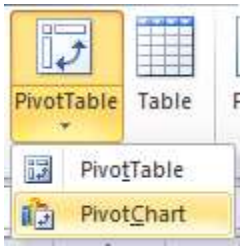
Try it

Find out the best two salespeople in the first two weeks of the month (Days 1 through 14). Is it different from the best two salespeople overall?

Simple Pivot Charts

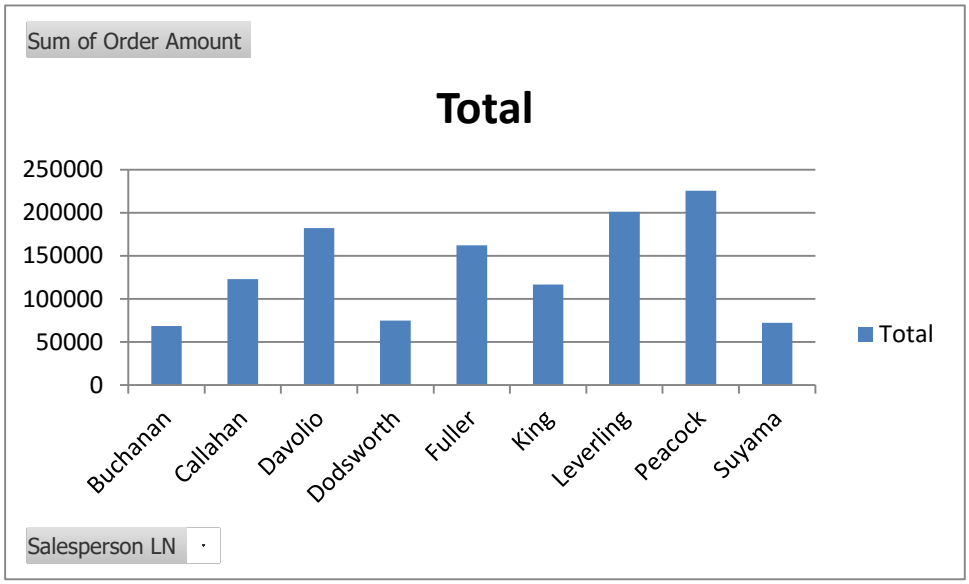
Just like we can present summarized data in tabular form using Pivot Tables, we can present that same data in graphical form using Pivot Charts. We'll get more into this later, but for now here's how to create a basic chart:

1) Select "PivotChart" from the Pivot Table drop-down menu in the Insert tab of the ribbon.



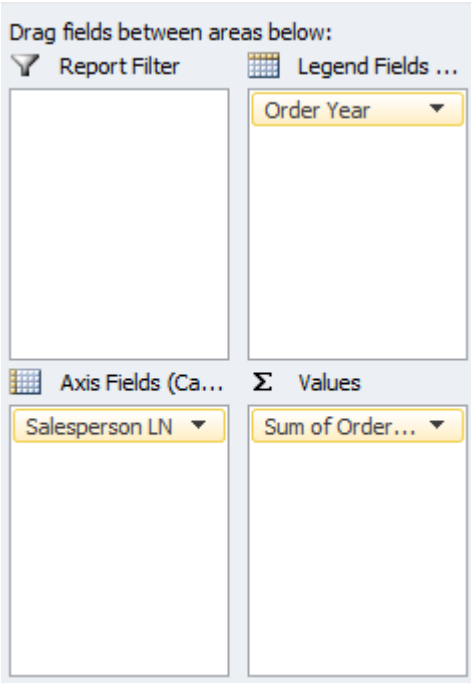
2) Make sure all the columns are selected (A through K) and you are placing the chart on a new worksheet.

3) We can quickly compare our salespersons' sales by placing Salesperson LN into the Axis Fields box and the Order Amount field into the Values box. Make sure it's computing the sum of the order amount. If that's done correctly, the chart will look like this:

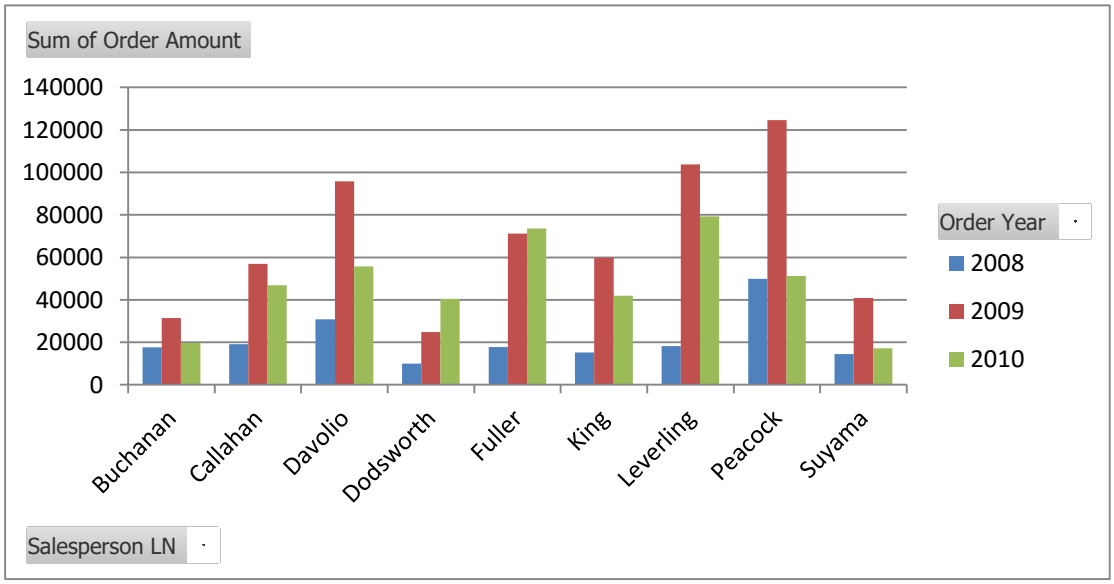


And we can visually verify what we already knew – that Peacock was our best salesperson.

4) Now let's split that up by year to find out what was the best year for each of our salespeople. Drag the "Order Year" field from the Pivot Table Field List to the Legend Fields area, like this:



5) Your chart now looks like this:



And you can see that for most of our salespeople, 2009 was the best year.

Try it	Find out in which month the company has it's highest level of sales. Then find the best month for the UK versus the USA.