

Lesson 1, Assignment 1 – Data Scientist vs Analyst Job Skills 2013 -2015

In this assignment, you will take skills data listed on job postings in the Seattle Metro area from 2013 and 2015 supplied by Labor Insights Burning Glass and analyze growth rates and changes in skills demanded for both Data Analysis and Data Science.

By completing this two-part assignment you will practice using VLOOKUP, CONDITIONAL FORMATTING and PIVOT TABLES in Microsoft Excel as well as learn more about the types of skills required in the data analytics field.

To complete this assignment, you need to have Microsoft XL 2016 loaded on your Windows based computer and the XL workbook titled – Lesson 1 Assignment 1 Workbook with Data Scientist vs Analyst 2013 -15 job skills data.xlsx

Part 1 – Creating a worksheet to compare growth rates and mix of co-occurring skills:

1. Open this XL workbook titled – Lesson 1 Assignment 1 Workbook with Data Scientist vs. Analyst 2013 – 15 job skills data.xlsx.
2. Note at the bottom there are four tabs – DS 2013 | DS 2015 | BI – DA 2013 | BI – DA 2015. Two with DS or Data Science skills data and two with DA-BI or Data Analytics/Business Intelligence skills data. Each of the two worksheets show data from 2013 and 2015, respectively.
3. To the right of the last tab (BI-DA 2015) you will see a (+) sign. Click on the (+) to add a new worksheet.
4. Double click on the tab of the new worksheet and rename the worksheet FROM [Sheet1] TO [Skill Comparison]
5. Click on the BI-DA 2015 tab, to open the worksheet with the most recent list of co-occurring skills ranked by highest frequency of job postings that had Data Analysis and/or Business Intelligence listed as a skill.
6. Highlight column A by clicking on the A – which includes a list of software “Skills” -- and select copy, then click on the [Skill Comparison] tab. Go to the top of the worksheet, click on a to select column A and select paste. You should see a copy of all the co-occurring skills in Column A of your new worksheet.
7. Now you are going to add headers for each column on row 1 as follows:
 - Column B, row 1 - BI-DA 2013
 - Column C, row 1 – BI-DA 2015
 - Column D, row 1 - % Change
 - Column E, row 1 – <leave blank>
 - Column F, row 1 – DS 2013
 - Column G, row 1 - DS 2015
 - Column H, row 1 - % Change

8. Add a Bottom Border in Row 1 from cell A1 to cell H1 and select Bold font from the menu above

So that your worksheet looks like this:

	A	B	C	D	E	F	G	H
1	Skills	BI-DA 2013	BI-DA 2015	% Change		DS 2013	DS 2015	% Change
2	All Postings							
3	SQL							
4	Microsoft Excel							
5	SQL Server							

9. Populate data covering co-occurring skills starting in column B2, by using the VLOOKUP function. Type in the following function expression

=VLOOKUP(A2,'BI-DA 2013'!\$A\$2:\$B\$200,2,FALSE)

This function does a vertical lookup to find the skill listed in column A of the current worksheet, by searching through columns A and B on the sheet containing BI and Data Analytic job postings skills and data. When it finds the matching skill, it selects the data in column 2 (or column B) and copies it from the source worksheet to the Skill Comparison worksheet. Thus even though the data is sorted in a different order on source data worksheet as compared to the Skill Comparison worksheet, the VLOOKUP function extracts the proper number of job postings for that skill and presents it in the order we have it listed in the Skill Comparison worksheet.

Notice:

- When the data is not found in the BI-DA worksheet, VLOOKUP enters an # NA in the cell.
- Within the formula when referring to the source data, we entered \$A\$2:\$B\$200, Adding the "\$" before the cell reference numbers and letters, will enable us to copy this formula down the column of our list of skills and have VLOOKUP refer each time to fixed source of data located on this other sheet in this specific range of column and rows.

10. Select cell B2 on the Skill Comparison worksheet and drag the formula down to 200. The full column of 2013 data regarding number of postings having the skill listed in column A will appear.

11. Repeat steps 9 and 10, but instead use the following formula to source data from the BI-DA 2015 worksheet and populate column C with data.

=VLOOKUP(A2,'BI-DA 2015'!\$A\$2:\$B\$200,2,FALSE)

12. Repeat steps 9 and 10, but instead use the following formula to source data from the DS 2013 worksheet and populate column F with data:

=VLOOKUP(A2,'DS 2013'!\$A\$2:\$B\$200,2,FALSE)

13. Repeat steps 9 and 10, but instead use the following formula to source data from the DS 2015 worksheet and populate column G with data:

=VLOOKUP(A2,'DS 2015'!\$A\$2:\$B\$200,2,FALSE)

14. Now we need to calculate the % change columns, starting with column D. Enter the following formula into cell D2 of the Skill Comparison Worksheet to calculate the % change in number of postings that had the skill listed in 2013 versus 2015.

=(C2-B2)/B2

15. Drag the formula down column D to populate the % change for all 200 skills

16. Repeat steps 14 and 15, but instead use the formula below to calculate the change Data Science skill job postings between 2013 and 2015:

=(G2-F2)/F2

17. Format columns D and H as percent. Click to select column D. Under Number on the ribbon, select Percentage. Repeat for column H

18. As there are many line items, we can use conditional formatting to color code the % change results and make it easier to see how the demand for skills that grew the fastest or in some cases declined. To do this:

- Highlight cells D2:D200
- On the home tab at the top, select conditional formatting button near the middle of the menu and then choose color scales. Select the color scale option in the upper left hand corner

This will make the cells which have changes that are close to average for the column be highlighted in yellow. As the change increases to well above the average of the column, the highlighted color will move to green. As the change is less than the average for the column or even negative, the cell will be highlighted in red.

19. Repeat step 18, for column H.

The first 5 lines of your worksheet should look as follows:

	A	B	C	D	E	F	G	H
1	Skills	BI-DA 2013	BI-DA 2015	% Change		DS 2013	DS 2015	% Change
2	All Postings	1,513	2,639	74.4%		230	864	276%
3	SQL	892	1,691	89.6%		127	485	282%
4	Microsoft Excel	532	1,041	95.7%		22	120	445%
5	SQL Server	244	524	114.8%		20	28	40%

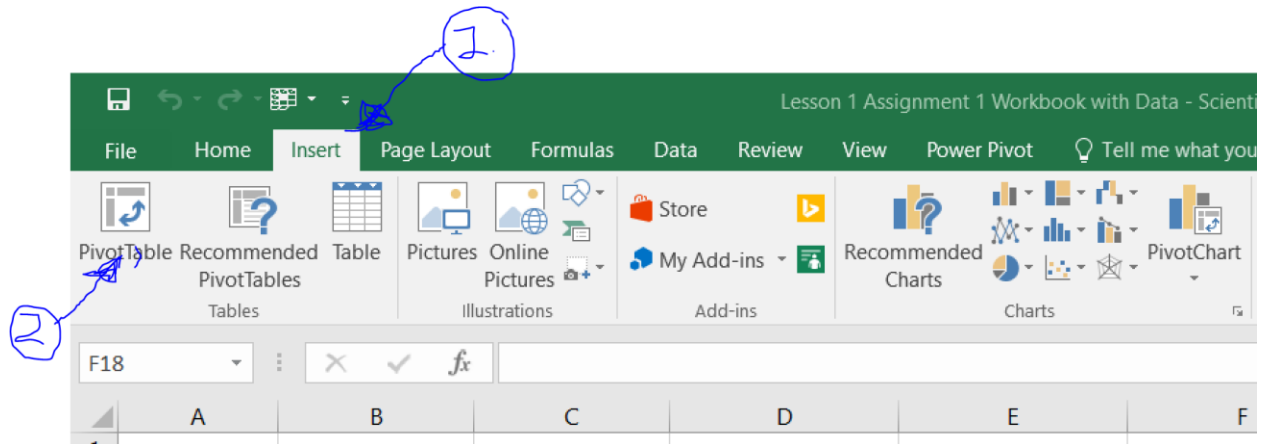
QUESTIONS:

- Looking at the number of job postings and growth
 - What was the % increase in job postings listing Business Intelligence or Data Analysis skills in 2015 versus 2013?
 - What was the % increase in job postings listing Data Science skills in 2015 versus 2013?
 - How many times larger is the demand for BI or Data Analysis skills as compared to Data Science skills in 2015?
 - How does the growth rate in demand for R programming skills among the postings that included business intelligence BI or data analysis DA skills compare to the overall growth rate for BI or DA skills?
Select the best answer
Demand for R programming skills:
 - grew 6.2 times faster than the overall demand for BI-DA skills
 - grew 2.5 times faster than the overall demand for BI-DA skills
 - grew 3.6 times faster than the overall demand for BI-DA skills
 - grew 1.6 times faster than the overall demand for BI-DA skills

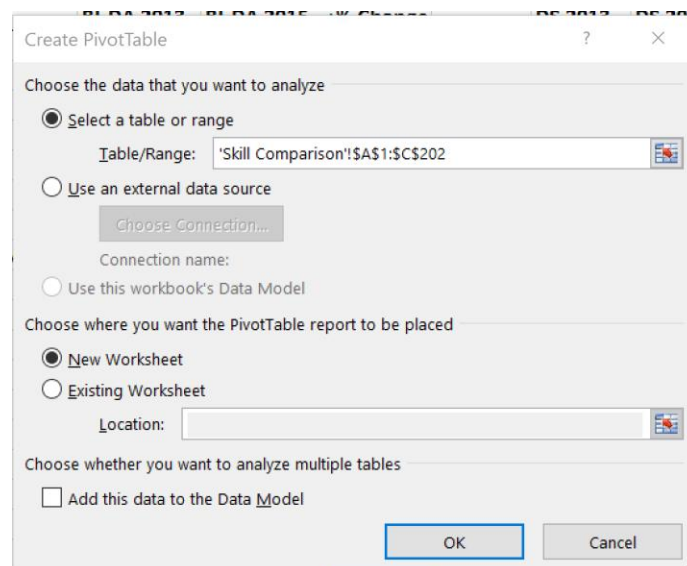
NOTE: Answer this question in the QUESTIONS Tab of the excel file provided

Part 2 – Creating a worksheet to compare growth rates and mix of co-occurring skills:

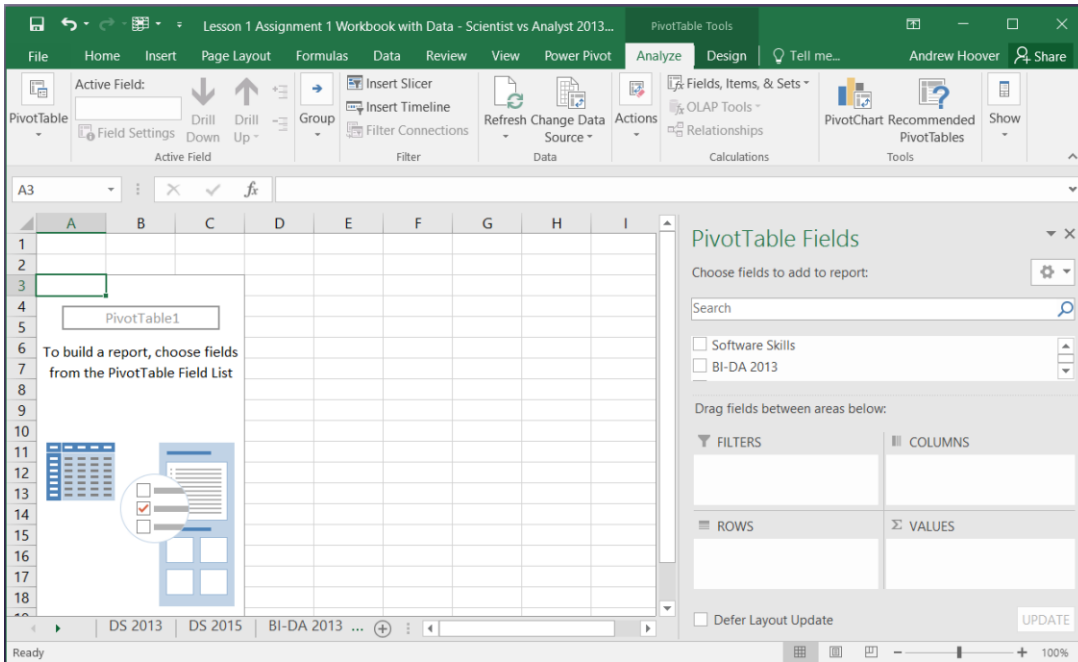
1. From the menu at the top of the screen, select INSERT
2. Double click on the Pivot Table icon



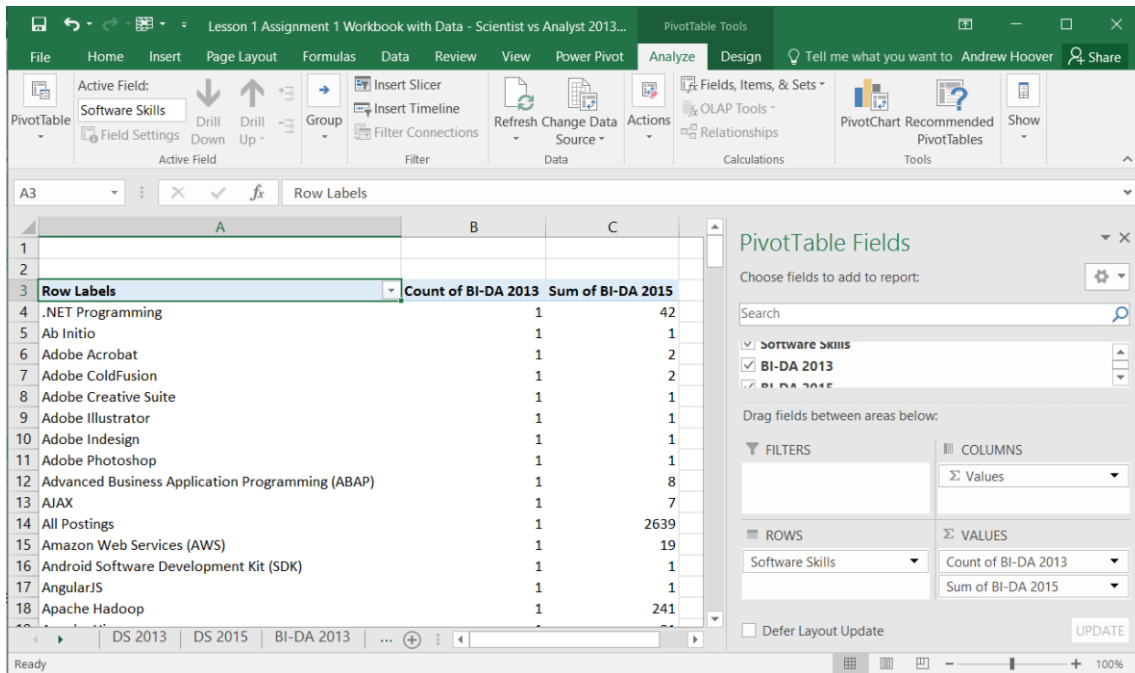
3. A Create Pivot Table Dialogue Box will appear. Click Select a table or range
4. In the Table/Range text box – place your cursor and click.
5. Go to the Skill Comparison worksheet and select the cell range A1: C200 and hit enter
6. This should take you back to the Create PivotTable Dialogue box
7. Under choose where you want the PivotTable report to be placed, select New Worksheet



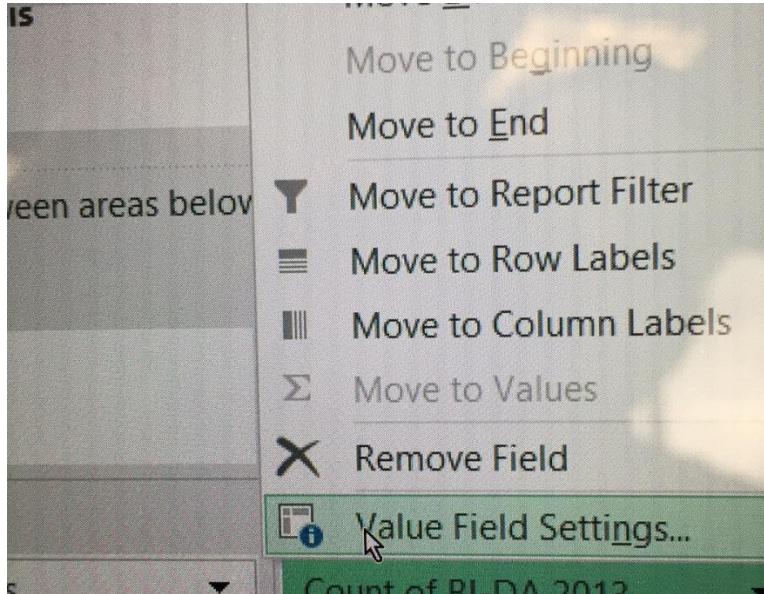
- A new worksheet will be created called Sheet 1 with Pivot Table Fields Dialogue box on the right



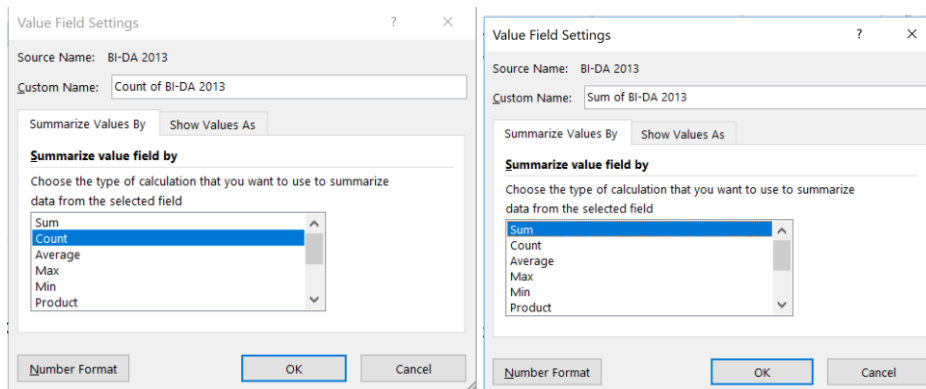
- Within the PivotTable Field dialogue box, click on the Skills field and drag it down to the Rows box
- Continue by clicking on and dragging down the BI-DA 2013 and BI-DA 2015 fields to the values box



11. Notice in the lower right under Values, it says Count of BI-DA 2013, whereas below that there is Sum of BI-DA 2015. We want to sum, in other words, the sum of job postings listing the associated skills. Click the down arrowhead next to Sum of BI-DA 2013 and then select Value Field Settings.



12. In the value field settings dialogue box, change the selected field from Count to Sum



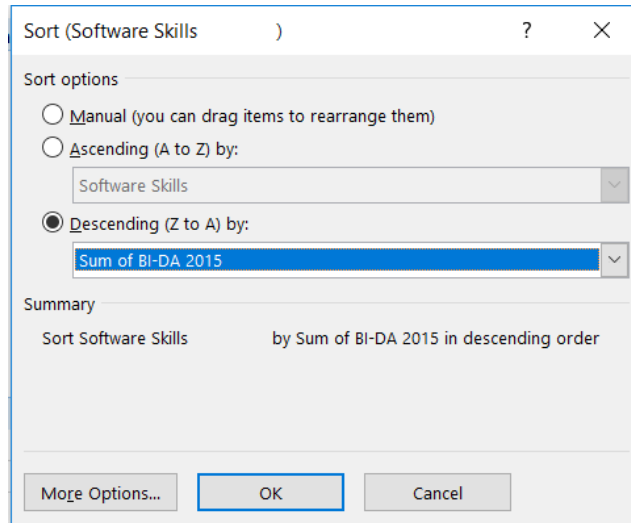
13. We now have essentially replicated the data we got from Skills Comparison Sheet into a Pivot Table. In this format, we can explore the data by pivoting or filtering the table and draw conclusions by visualizing the relevant data into a chart. In this example we are doing to use filtering and charting.
14. Click on the down arrow button next to Row Labels on the right side of Cell A3.

	A	B	C
1			
2			
3	Row Labels	Sum of BI-DA 2013	Sum of BI-DA 2015
4	.NET Programming	45	42

15. Uncheck the box next to (Select All)
16. We are going to focus on programming languages other than SQL -- which shows up on multiple rows -- to instead see the growth rates in demand and compare the demand for various programming language skills among the job postings that include either Business Intelligence or Data Analytics skills.
17. Under Value Filters in the dialogue box, put a check mark by clicking on the following programming languages C++, Java, Visual Basic, Microsoft C#, Objective C, Python, R, Ruby, SAS, S Plus and SPSS. You should have a table that looks like this after clicking OK in the Value Filters dialogue box.

	A	B	C
1			
2			
3	Row Labels	Sum of BI-DA 2013	Sum of BI-DA 2015
4	C++	39	44
5	JAVA	68	117
6	Microsoft C#	107	122
7	Objective C	1	3
8	Python	54	192
9	R	57	161
10	Ruby	30	63
11	SAS	175	304
12	S-Plus	6	3
13	SPSS	83	66
14	Visual Basic	19	36
15	Grand Total	639	1111

18. Now let's sort the results, so that the most frequently requested language is shown first. To do this, click on the cell A3 Row Labels filter icon. Select "More Sort Options". In the Sort dialogue box that comes up, click on "Descending (Z to A) by:" and then in the neighboring drop down box, select Sum of BI-DA 2015, so the languages will be sorted in order of highest to lowest demand. Click OK button at the bottom of the Sort Dialogue box.



19. Your table should look like this

	A	B	C
1			
2			
3	Row Labels	Sum of BI-DA 2013	Sum of BI-DA 2015
4	SAS	175	304
5	Python	54	192
6	R	57	161
7	Microsoft C#	107	122
8	JAVA	68	117
9	SPSS	83	66
10	Ruby	30	63
11	C++	39	44
12	Visual Basic	19	36
13	Objective C	1	3
14	S-Plus	6	3
15	Grand Total	639	1111

QUESTIONS:

1. Which of the top five languages more than doubled in terms of job postings that listed both the programming language along with Business Intelligence and/or Data Analysis skills?

Select all that apply:

- A. SAS
- B. Python
- C. R
- D. Microsoft C#
- E. Java

NOTE: Answer this question in the QUESTIONS Tab of the excel file provided