

Data Science Salaries – Excel

In this project, we will be using pivot tables and slicers to create a dashboard of visualizations from a flat file to provide insights regarding various different data science job titles between 2020-2022.

Skills Demonstrated:

- Find and Replace
- Pivot Tables
- Slicers

Required Materials:

- The dataset `ds_salaries.csv` from either [Kaggle](#) or my [Project Portfolio Website](#).

Walkthrough:

1. First, open the file select the data. Go the insert tab and select Insert Table.

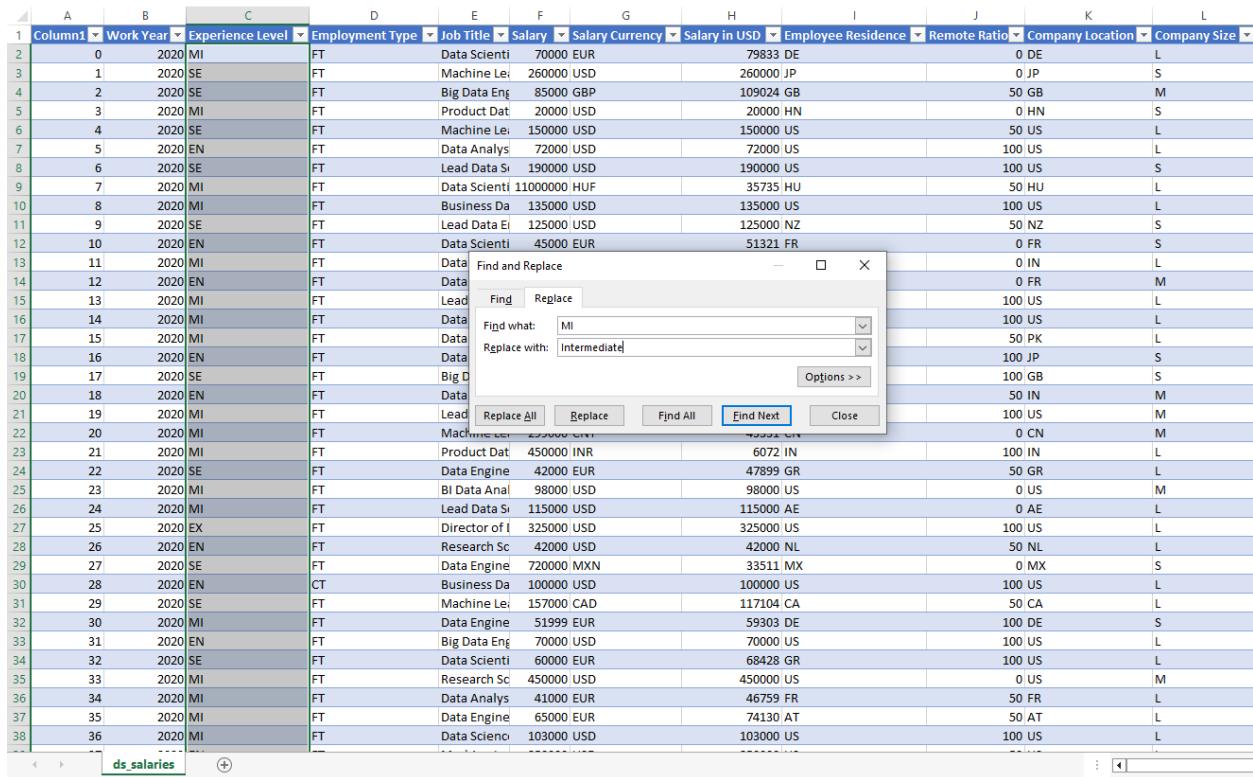
The screenshot shows the Microsoft Excel interface with the 'Insert' tab selected. The 'Table' button in the 'Tables' group is highlighted. A 'Create Table' dialog box is open, showing the range '\$A\$1:\$L\$608' and the checked option 'My table has headers'. The background spreadsheet displays data for various data science roles across different years and companies.

work_year	experience	employee	job_title	salary	salary_currency	salary_in_currency	remote_work	company	company_size
2020	MI	FT	Data Scier	70000	EUR	79833	DE	0 DE	L
2020	SE	FT	Machine L	260000	USD	260000	JP	0 JP	S
2020	SE	FT	Big Data E	85000	GBP	109024	GB	50 GB	M
2020	MI	FT	Product D	20000	USD	20000	HN	0 HN	S
2020	SE	FT	Machine L	150000	USD	150000	US	50 US	L
2020	EN	FT	Data Anal	72000	USD	72000	US	100 US	L
2020	SE	FT	Lead Data	190000	USD	190000	US	100 US	S
2020	MI	FT	Data Scier	11000000	HUF	35735	HU	50 HU	L
2020	MI	FT	Business I	135000	USD	135000	US	100 US	L
2020	SE	FT	Lead Data	125000	USD	125000	NZ	50 NZ	S
2020	EN	FT	Data Scier	45000	EUR	51321	FR	0 FR	S
2020	MI	FT	Data Scier	3000000	INR	40481	IN	0 IN	L
2020	EN	FT	Data Scier	35000	EUR	39916	FR	0 FR	M
2020	MI	FT	Lead Data	87000	USD	87000	US	100 US	L
2020	MI	FT	Data Anal	85000	USD	85000	US	100 US	M
2020	MI	FT	Data Anal	8000	USD	8000	PK	0 PK	L
2020	EN	FT	Data Engir	4450000	JPY	41689	JP	0 JP	S
2020	SE	FT	Big Data E	100000	EUR	114047	PL	0 PL	M
2020	EN	FT	Data Scier	423000	INR	5707	IN	0 IN	L
2020	MI	FT	Lead Data	56000	USD	56000	PT	100 US	M
2020	MI	FT	Machine L	299000	CNY	43331	CN	0 CN	M
2020	MI	FT	Product D	450000	INR	6072	IN	100 IN	L
2020	SE	FT	Data Engir	42000	EUR	47899	GR	50 GR	L
2020	MI	FT	BI Data An	98000	USD	98000	US	0 US	M
2020	MI	FT	Lead Data	115000	USD	115000	AE	0 AE	L
2020	EX	FT	Director o	325000	USD	325000	US	100 US	L
2020	EN	FT	Research	42000	USD	42000	NL	50 NL	L
2020	SE	FT	Data Engir	720000	MXN	33511	MX	0 MX	S
2020	EN	CT	Business I	100000	USD	100000	US	100 US	L
2020	SE	FT	Machine L	157000	CAD	117104	CA	50 CA	L
2020	MI	FT	Data Engir	51999	EUR	59303	DE	100 DE	S
2020	EN	FT	Big Data E	70000	USD	70000	US	100 US	L
2020	SE	FT	Data Scier	60000	EUR	68428	GR	100 US	L
2020	MI	FT	Research	450000	USD	450000	US	0 US	M
2020	MI	FT	Data Anal	41000	EUR	46759	FR	50 FR	L
2020	MI	FT	Data Engir	65000	EUR	74130	AT	50 AT	L
2020	MI	FT	Data Scier	103000	USD	103000	US	100 US	L

This will apply formatting to our table, and make things look a little more professional. Additionally, I went ahead and changed the titles of some of the columns to continue with the professional look.

- Next, we will want to change the aliases of our data to make it more comprehensible. To do this, we will select cells C2 and D2 and use the shortcut CTRL+SHIFT+DOWN in order to select all data in the row except the titles. From here, we can use CTRL+F to use the Find and Replace feature and change the aliases of our data. Here were the following changes I made in my data, in order. (The order matters!)

Transformation	Column	Change
1	C	EN → Entry
2	C	MI → Intermediate
3	C	SE → Senior
4	C	EX → Expert
5	D	FT → Full Time
6	D	PT → Part Time
7	D	CT → Contract
8	D	FL → Freelance



Column1	Work Year	Experience Level	Employment Type	Job Title	Salary	Salary Currency	Salary in USD	Employee Residence	Remote Ratio	Company Location	Company Size
0	2020	MI	FT	Data Scienti	70000	EUR	79833	DE	0	DE	L
1	2020	SE	FT	Machine Lei	260000	USD	260000	JP	0	JP	S
2	2020	SE	FT	Big Data Eng	85000	GBP	109024	GB	50	GB	M
3	2020	MI	FT	Product Dat	20000	USD	20000	HN	0	HN	S
4	2020	SE	FT	Machine Lei	150000	USD	150000	US	50	US	L
5	2020	EN	FT	Data Analys	72000	USD	72000	US	100	US	L
6	2020	SE	FT	Lead Data Si	190000	USD	190000	US	100	US	S
7	2020	MI	FT	Data Scienti	1100000	HUF	35735	HU	50	HU	L
8	2020	MI	FT	Business Da	135000	USD	135000	US	100	US	L
9	2020	SE	FT	Lead Data Ei	125000	USD	125000	NZ	50	NZ	S
10	2020	EN	FT	Data Scienti	45000	EUR	51321	FR	0	FR	S
11	2020	MI	FT	Data				IN	0	IN	L
12	2020	EN	FT	Data				FR	0	FR	M
13	2020	MI	FT	Lead				US	100	US	L
14	2020	MI	FT	Data				US	100	US	L
15	2020	MI	FT	Data				PK	50	PK	L
16	2020	MI	FT	Data				JP	100	JP	S
17	2020	MI	FT	Data				GB	100	GB	S
18	2020	SE	FT	Big D				IN	50	IN	M
19	2020	EN	FT	Data				US	100	US	M
20	2020	MI	FT	Lead				CN	0	CN	M
21	2020	MI	FT	Machine Lei				US	0	US	M
22	2020	MI	FT	Product Dat	450000	INR	6072	IN	100	IN	L
23	2020	SE	FT	Data Engine	42000	EUR	47899	GR	50	GR	L
24	2020	MI	FT	BI Data Anal	98000	USD	98000	US	0	US	M
25	2020	MI	FT	Lead Data Si	115000	USD	115000	AE	0	AE	L
26	2020	MI	FT	Director of I	325000	USD	325000	US	100	US	L
27	2020	EX	FT	Research Sc	42000	USD	42000	NL	50	NL	L
28	2020	EN	FT	Research Sc	42000	USD	42000	NL	50	NL	L
29	2020	SE	FT	Data Engine	720000	MXN	33511	MX	0	MX	S
30	2020	EN	CT	Business Da	100000	USD	100000	US	100	US	L
31	2020	SE	FT	Machine Lei	157000	CAD	117104	CA	50	CA	L
32	2020	MI	FT	Data Engine	51999	EUR	59303	DE	100	DE	S
33	2020	EN	FT	Big Data Eng	70000	USD	70000	US	100	US	L
34	2020	SE	FT	Data Scienti	60000	EUR	68428	GR	100	US	L
35	2020	MI	FT	Research Sc	450000	USD	450000	US	0	US	M
36	2020	MI	FT	Data Analys	41000	EUR	46759	FR	50	FR	L
37	2020	MI	FT	Data Engine	65000	EUR	74130	AT	50	AT	L
38	2020	MI	FT	Data Scieno	103000	USD	103000	US	100	US	L

With this, the data will be much easier to interpret.

Note: Make sure you change the formatting in each of the salary columns over to currency!

- The next set of pivot tables we will want to include are overall aggregates of our data. We will want the minimum, maximum, and average salary for the data we will be selecting. Create three pivot tables in this sheet, and then place the Salary in USD in values, and change the aggregate from Sum to the respective aggregate.

The screenshot displays a Google Sheet with a pivot table summarizing salary data. The pivot table is located in the range B34:U63. The data source is 'Salary in USD'. The pivot table structure is as follows:

Role	Sum of Salary in USD
Head of Data Science	\$386,875.00
Head of Machine Learning	\$79,039.00
Lead Data Analyst	\$276,609.00
Lead Data Engineer	\$838,347.00
Lead Data Scientist	\$345,570.00
Lead Machine Learning	\$87,932.00
Machine Learning Dev	\$257,582.00
Machine Learning Eng	\$4,300,086.00
Machine Learning Infra	\$303,435.00
Machine Learning Manager	\$117,104.00
Machine Learning Sci	\$1,267,300.00
Marketing Data Analyst	\$68,654.00
ML Engineer	\$705,024.00
NLP Engineer	\$37,236.00
Principal Data Analyst	\$245,000.00
Principal Data Engineer	\$985,000.00
Principal Data Scientist	\$1,506,697.00
Product Data Analyst	\$26,072.00
Research Scientist	\$1,744,312.00
Staff Data Scientist	\$105,000.00
Grand Total	\$68,164,807.00
Average of Salary in USD	\$112,297.87
Min of Salary in USD	\$2,859.00
Max of Salary in USD	\$600,000.00

The 'Value Field Settings' dialog is open, showing the 'Sum' aggregate. The 'PivotTable Fields' taskbar is also visible, showing the 'Salary in USD' field in the Values area.

- Next, we will be creating a new sheet for our next set of pivot tables that will revolve around pie chart or doughnut visuals (depending on if you're a pie or doughnut type of person 😊). We will want three pivot tables where the select columns are in the rows and values for the following columns:

- Experience
- Company size
- Employment type

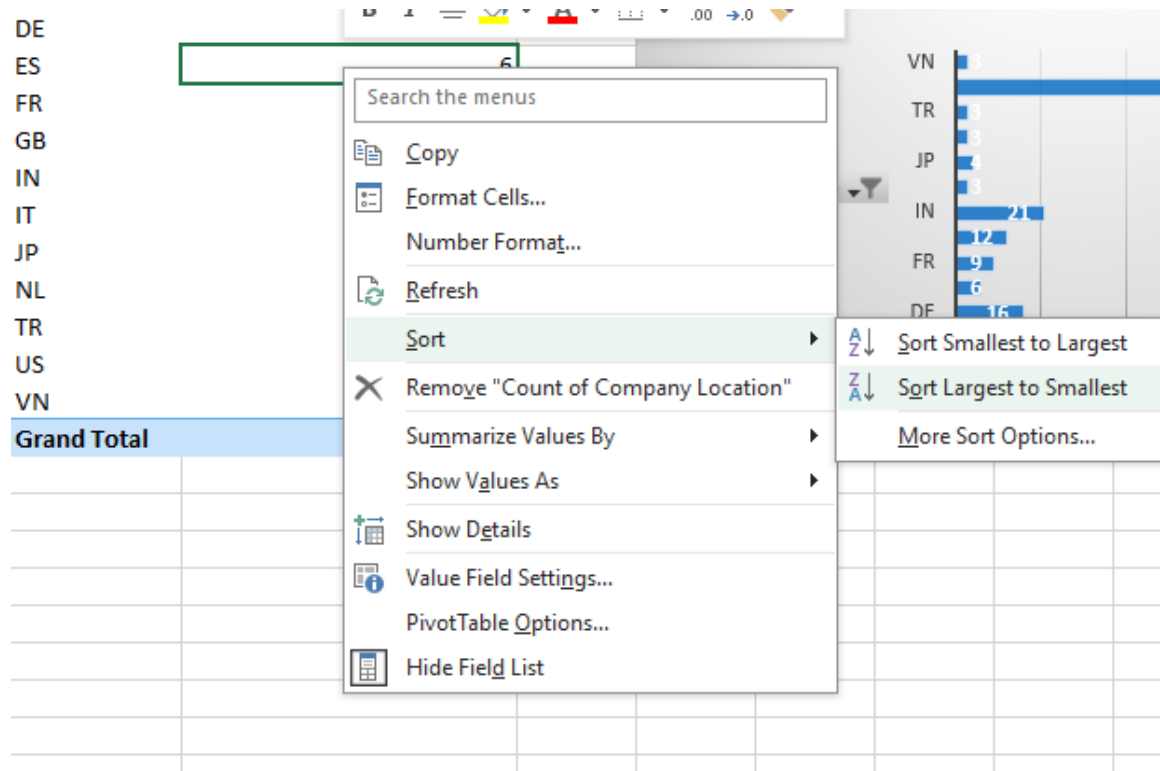
1	A	E	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
2																						
3	Row Labels	Count of Employment Type																				
4	Contract	3																				
5	Freelance	4																				
6	Full Time	588																				
7	Part Time	10																				
8	Grand Total	607																				
9																						
10	Row Labels	Count of Experience Level																				
11	Entry	88																				
12	Expert	26																				
13	Intermediate	213																				
14	Senior	280																				
15	Grand Total	607																				
16																						
17	Row Labels	Count of Company Size																				
18	L	198																				
19	M	326																				
20	S	83																				
21	Grand Total	607																				
22																						
23																						
24																						
25																						
26																						
27																						
28																						
29																						
30																						
31																						
32																						
33																						
34																						
35																						
36																						
37																						
38																						

- Finally we want insight as to where these opportunities are coming from, and which country provides the most of the opportunities we are looking for. We will be making another sheet for one pivot table that will provide the top ten countries based on the job location (You might notice in several of the images they refer to employee residence. This has been corrected in the final file). Place job location in the rows and values of this pivot table.

Next, we will want the top ten countries of this pivot table. Select the filter for this pivot table, value filters, then select top 10.

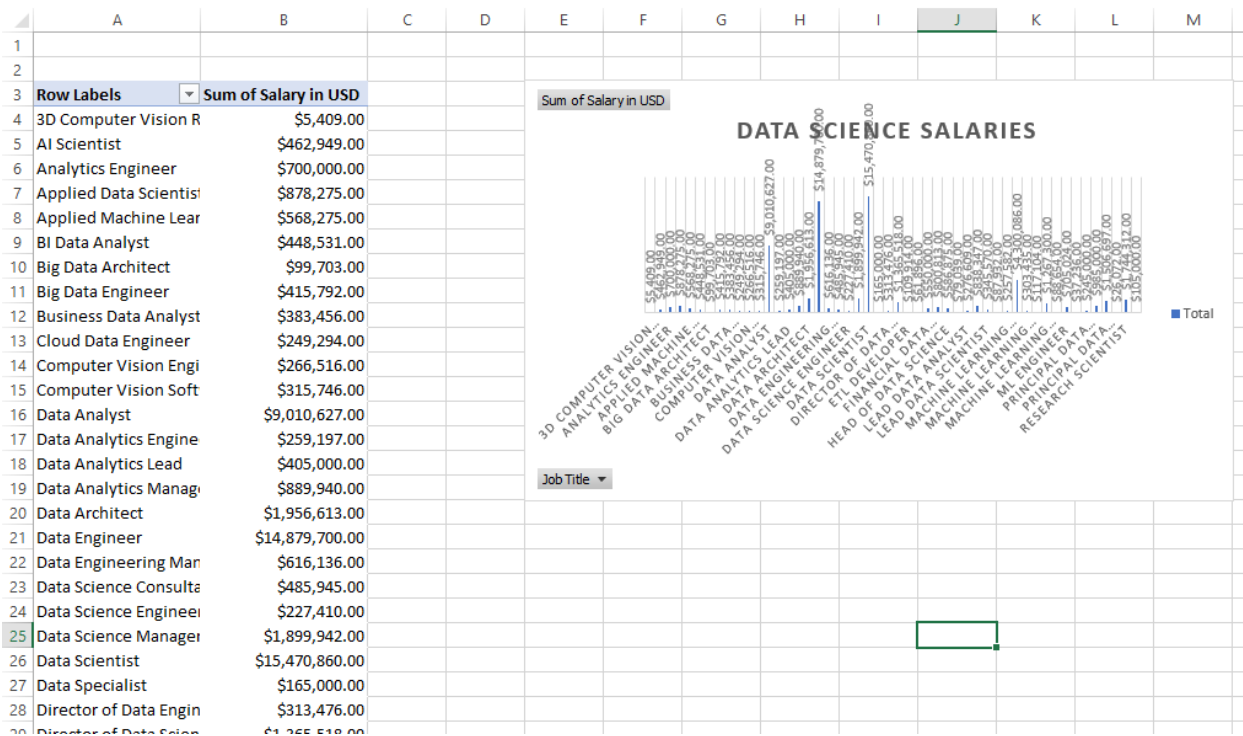
2					
3	Row Labels	Count of Company Location			Count of Company
4	Sort A to Z		5		
5	Sort Z to A		9		
6	More Sort Options...		16		
7	Clear Filter From "Employee Residence"		6		
8	Label Filters		9		
9	Value Filters		12		
10	Search				
11	(Select All)				
12	AE				
13	AR				
14	AT				
15	AU				
16	BE				
17	BG				
18	BO				
19	BR				
20	OK				
21	Cancel				

From here, we want to sort the values from largest to smallest. Right click the table, hover sort, then select Sort Largest to Smallest.

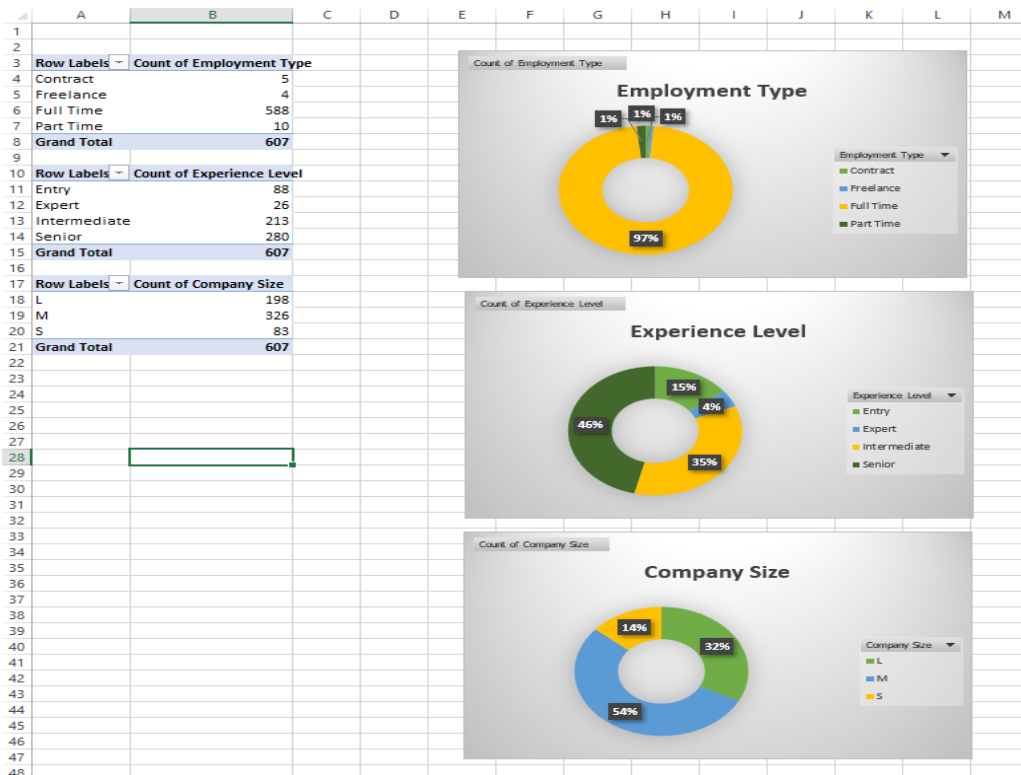


With this we now have the data we need to create our visualizations.

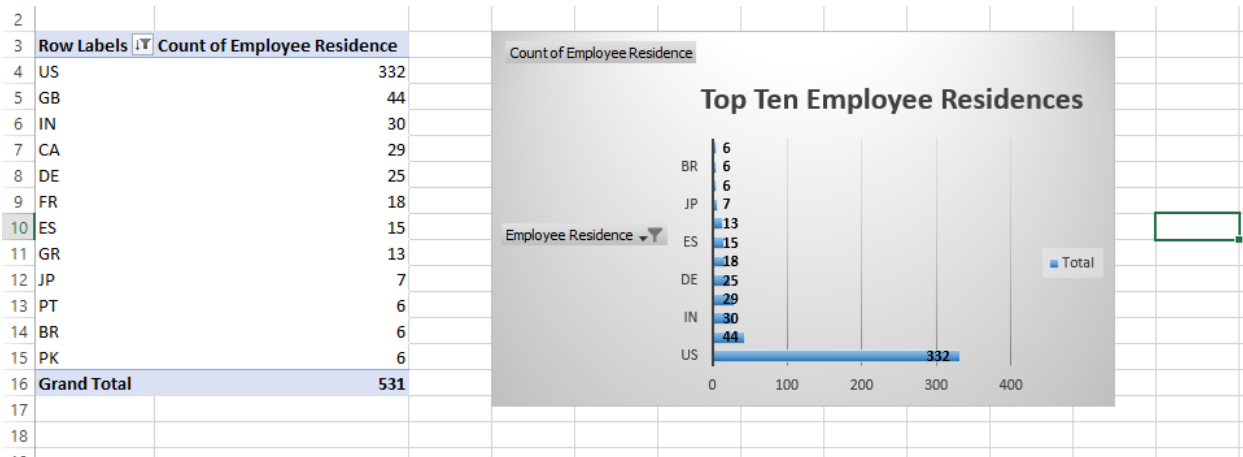
6. In our first pivot table sheet, we will select our job role salary pivot table, and create a vertical bar chart. Choosing a style for this chart might help with visibility initially. When we apply our slicers, we won't have to worry about as many different entries.



Next in the second sheet, create pie chart or doughnut chart visualizations of the three pivot tables. Again, styles can help improve the visualizations.

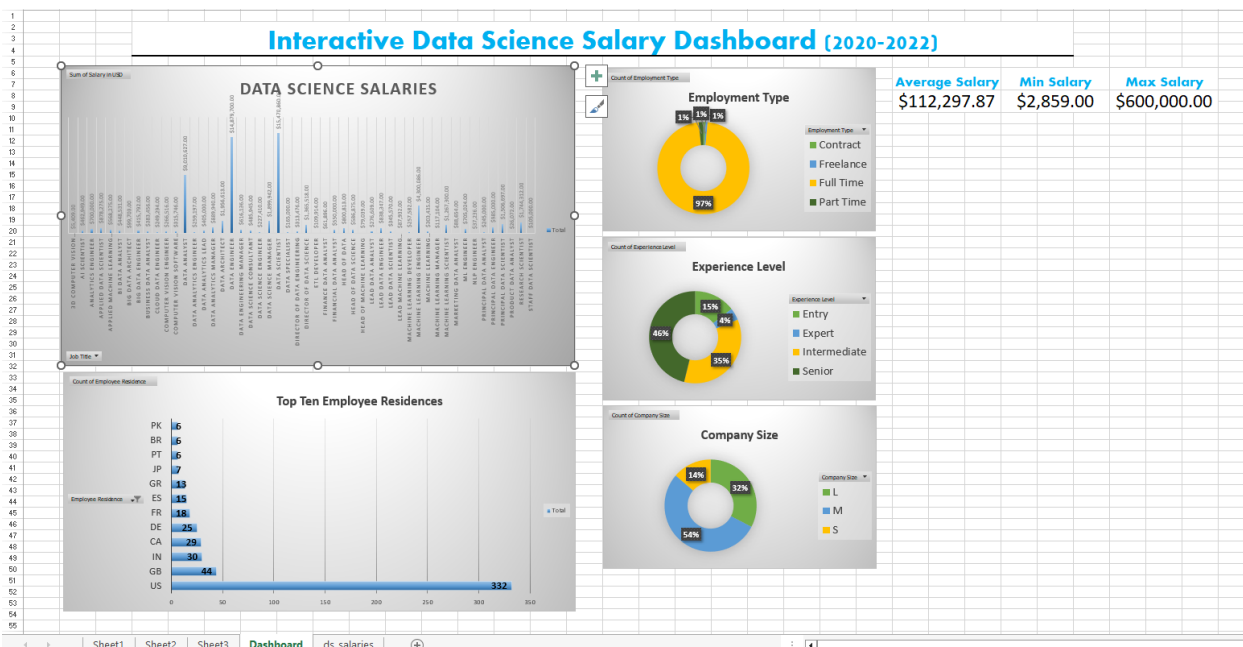


Finally we will make a horizontal bar chart for our ranked countries pivot table.

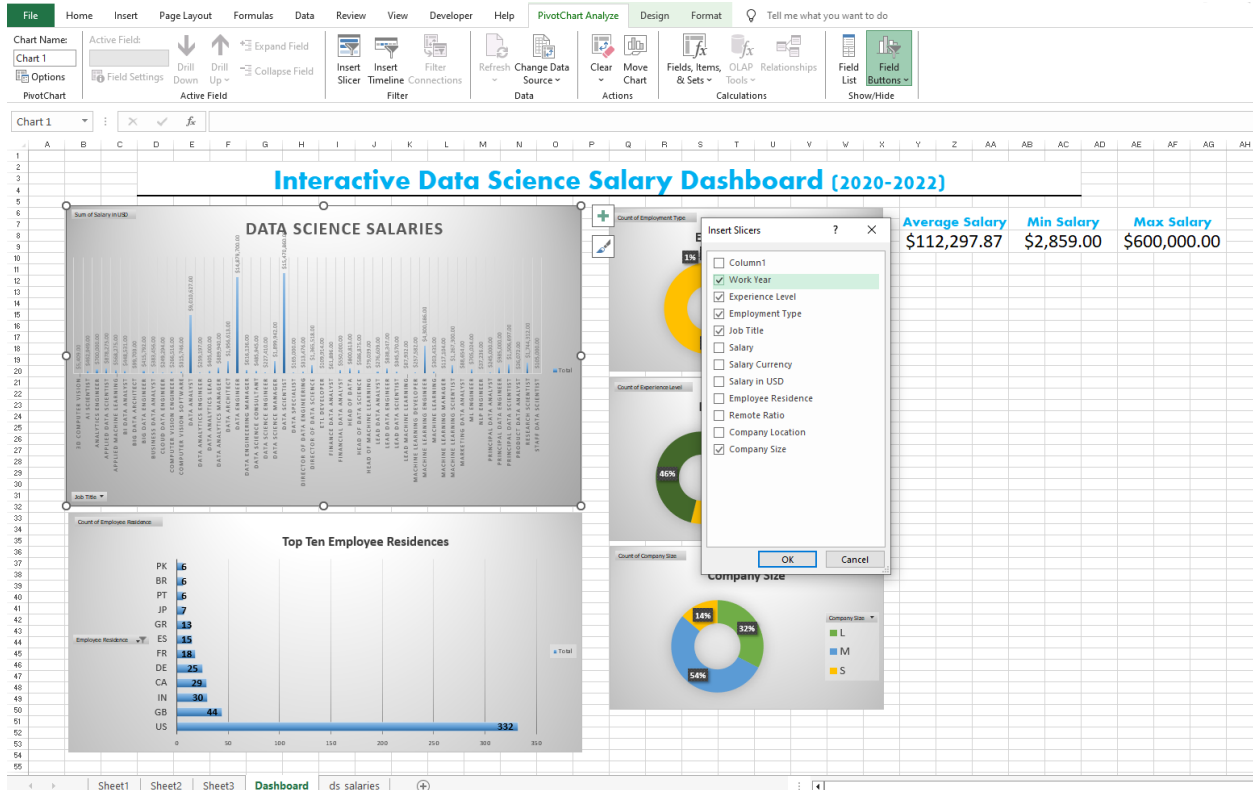


We are now ready to begin our dashboard.

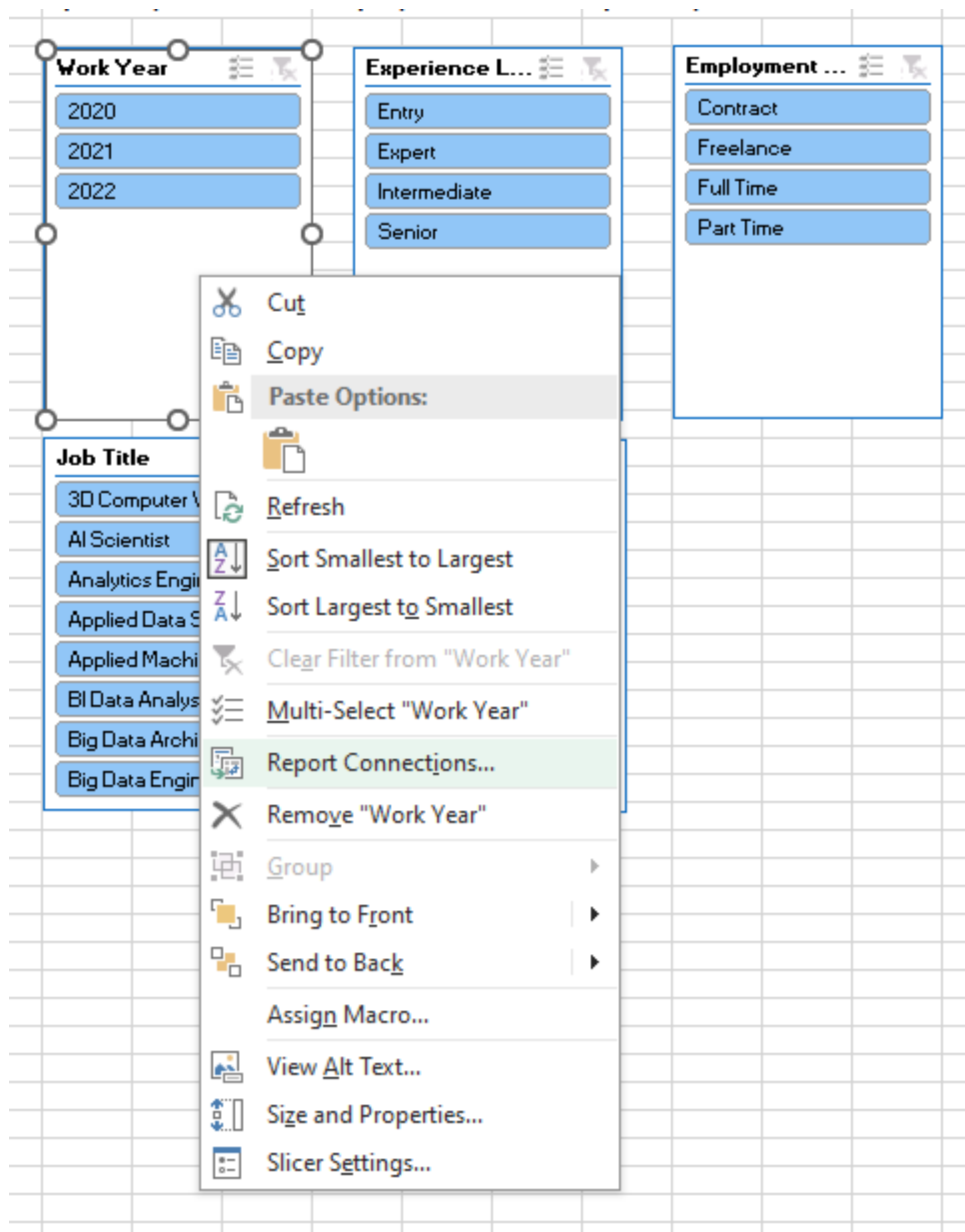
- Select a range of cells and use merge and center to create a title that spans across the screen. Use whatever style you would like for the theme you intend to set. Then, copy and paste all of your visualizations over. Include some room for your aggregates as well.



- Select a pivot table, go to PivotChart analyze, and select Insert Slicer. We will now have the option to insert several slicers based on our dataset's columns. Here, we will select Work Year, Experience Level, Employment Type, Job Title, and Company size. Select OK.

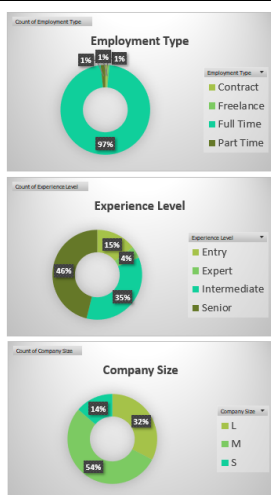
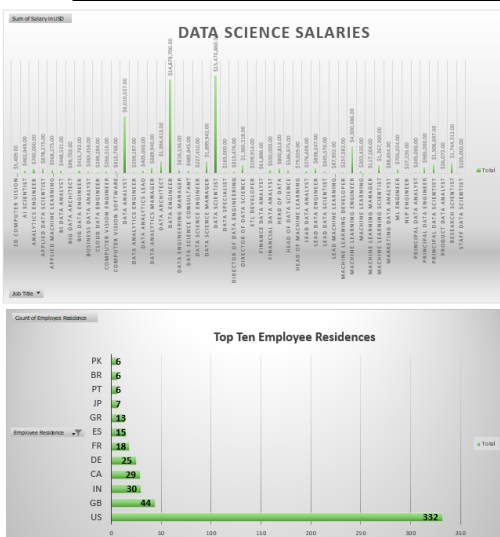


Though our slicers are in, they will only affect one visualization for the moment. Right click each slicer, and select Report Connections. This will allow us to select each of our pivot tables to affect with the slicer, and we will want to select all.



Once all slicers affect all pivot tables, we are done. Selecting slicers will now filter our results based on our selections. Feel free to adjust visualizations according to your desired style.

Interactive Data Science Salary Dashboard (2020-2022)



Average Salary \$112,297.87 Min Salary \$2,859.00 Max Salary \$600,000.00

Work Year

2020
2021
2022

Experience Level

Entry
Expert
Intermediate
Senior

Employment Type

Contract
Freelance
Full Time
Part Time

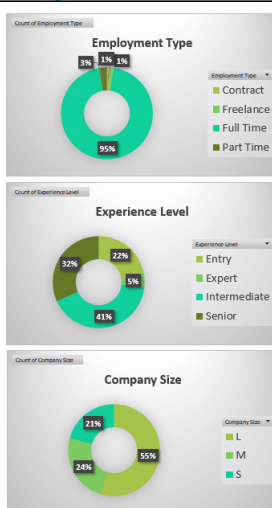
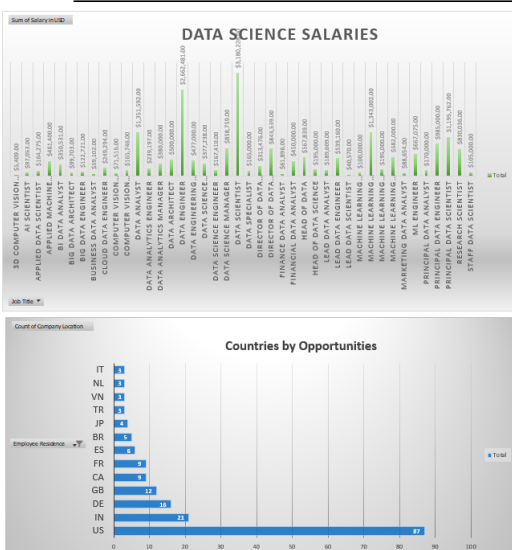
Job Title

3D Computer Vision
AI Scientist
Analytics Engineer
Applied Data Science
Applied Machine Learning
BI Data Analyst
Big Data Architect
Big Data Engineer

Company Size

L
M
S

Interactive Data Science Salary Dashboard (2020-2022)



Average Salary \$99,853.79 Min Salary \$2,859.00 Max Salary \$600,000.00

Work Year

2020
2021
2022

Experience Level

Entry
Expert
Intermediate
Senior

Employment Type

Contract
Freelance
Full Time
Part Time

Job Title

3D Computer Vision
AI Scientist
Analytics Engineer
Applied Data Science
Applied Machine Learning
BI Data Analyst
Big Data Architect
Big Data Engineer

Company Size

L
M
S