code.likeagirl.io

LinkedIn Data Analysis Case Study With Interactive Dashboard!

Uzaifa Memon

8–11 minutes

A step-by-step guide to analyzing a complex dataset and creating an interactive dashboard using Microsoft Excel.

LinkedIn is every job seeker’s go-to place and today I am presenting an extensive study of this platform. From the positions that have been listed to the managers who have listed

them, we will explore everything today. Last time, I published Netflix's data analysis case study and if you have read that, you'll know how deep this study is going to be. Only this time, I have curated an interactive dashboard for you, too!

Excel is way more than a mere spreadsheet and you're about to get to know that soon!

So, without wasting even a single second, let us begin analyzing the LinkedIn database!

Data source: https://www.kaggle.com/datasets/shashankshukla123123/linkedin-job-data/versions/2?resource=download

Objectives And Objective Questions

Photo by Alexander Shatov on Unsplash

In this guide, I'll be focusing on these questions to derive insights from the data I downloaded from the link I mentioned above.

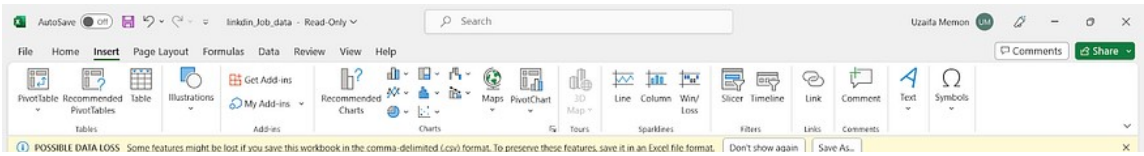
1. Which job is in high demand?
2. Which company has posted the highest number of jobs?
3. Which location has the highest number of job openings?
4. Which type of work is favoured by companies?
5. Which company has the highest employee count?
6. List hiring manager-wise company details.

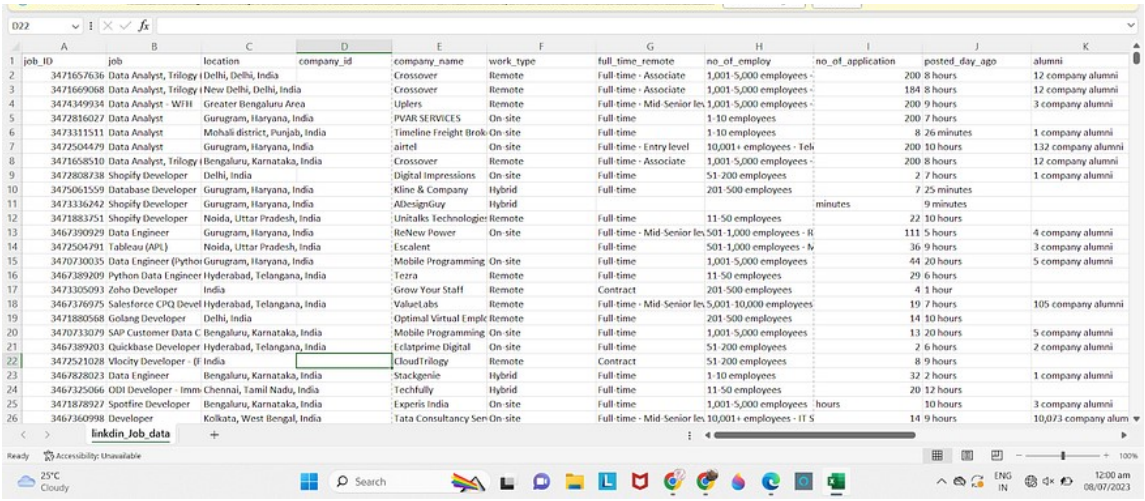
7. How many job postings have the highest number of applications?
8. Which company has the highest number of alumni?
9. Which company has the highest number of LinkedIn followers?

Step 1: Studying The Dataset

This is a crucial step in the extensive process of data analysis. Once you *know* what your data is about, you won't have a hard time processing the thousands of rows. I personally recommend it to everyone starting out with Excel for the purpose of data analysis. I realized a while later than starting with data analysis that studying data is the most crucial.

You can avoid this mistake right away!

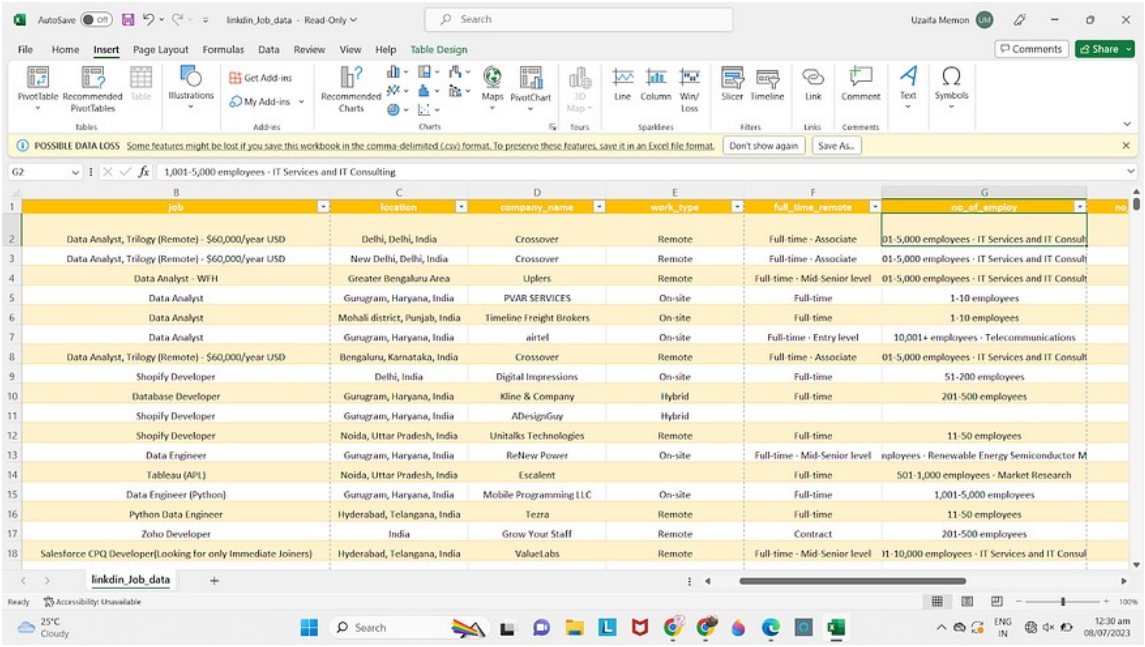




job_ID	job	location	company_id	company_name	work_type	full_time_remote	no_of_employ	no_of_application	posted_day_ago	alumni
3471657636	Data Analyst, Trilogy (	Delhi, Delhi, India		Crossover	Remote	Full-time - Associate	1,001-5,000 employees -		200 8 hours	12 company alumni
3471669068	Data Analyst, Trilogy (	New Delhi, Delhi, India		Crossover	Remote	Full-time - Associate	1,001-5,000 employees -		184 8 hours	12 company alumni
3474349934	Data Analyst - WFH	Greater Bengaluru Area		Uplers	Remote	Full-time - Mid-Senior le	1,001-5,000 employees		200 9 hours	3 company alumni
3472816027	Data Analyst	Gurugram, Haryana, India		PVAR SERVICES	On-site	Full-time	1-10 employees		200 7 hours	
3473311511	Data Analyst	Mohali district, Punjab, India		Timeline Freight Brok	On-site	Full-time	1-10 employees		8 26 minutes	1 company alumni
3472504479	Data Analyst	Gurugram, Haryana, India		airtel	On-site	Full-time - Entry level	10,001+ employees - Tel		200 10 hours	132 company alumni
3471658510	Data Analyst, Trilogy (	Bengaluru, Karnataka, India		Crossover	Remote	Full-time - Associate	1,001-5,000 employees -		200 8 hours	12 company alumni
3472808738	Shopify Developer	Delhi, India		Digital Impressions	On-site	Full-time	51-200 employees		2 7 hours	1 company alumni
3475061559	Database Developer	Gurugram, Haryana, India		Kline & Company	Hybrid	Full-time	201-500 employees		7 25 minutes	
3473336242	Shopify Developer	Gurugram, Haryana, India		ADesignGuy	Hybrid			minutes	9 minutes	
3471883751	Shopify Developer	Noida, Uttar Pradesh, India		Unitalks Technologies	Remote	Full-time	11-50 employees		22 10 hours	
3467390929	Data Engineer	Gurugram, Haryana, India		ReNew Power	On-site	Full-time - Mid-Senior le	501-1,000 employees - R		111 5 hours	4 company alumni
3472504791	Tableau (APL)	Noida, Uttar Pradesh, India		Escalent		Full-time	501-1,000 employees - M		36 9 hours	3 company alumni
3470730035	Data Engineer (Pytho	Gurugram, Haryana, India		Mobile Programming	On-site	Full-time	1,001-5,000 employees		44 20 hours	5 company alumni
3467389209	Python Data Engineer	Hyderabad, Telangana, India		Tezra	Remote	Full-time	11-50 employees		29 6 hours	
3473305093	Zoho Developer	India		Grow Your Staff	Remote	Contract	201-500 employees		4 1 hour	
3467376975	Salesforce CPQ Devel	Hyderabad, Telangana, India		ValueLabs	Remote	Full-time - Mid-Senior le	5,001-10,000 employees		19 7 hours	105 company alumni
3471880568	Golang Developer	Delhi, India		Optimal Virtual Emplo	Remote	Full-time	201-500 employees		14 10 hours	
3470733079	SAP Customer Data C	Bengaluru, Karnataka, India		Mobile Programming	On-site	Full-time	1,001-5,000 employees		13 20 hours	5 company alumni
3467389203	Quickbase Developer	Hyderabad, Telangana, India		Eclatprime Digital	On-site	Full-time	51-200 employees		2 6 hours	2 company alumni
3472521028	Vlocity Developer - (F	India		CloudTrilogy	Remote	Contract	51-200 employees		8 9 hours	
3467828023	Data Engineer	Bengaluru, Karnataka, India		Stackgenie	Hybrid	Full-time	1-10 employees		32 2 hours	1 company alumni
3467325066	ODI Developer - Imm	Chennai, Tamil Nadu, India		Techfully	Hybrid	Full-time	11-50 employees		20 12 hours	
3471878927	Spotfire Developer	Bengaluru, Karnataka, India		Experis India	On-site	Full-time	1,001-5,000 employees	hours	10 hours	3 company alumni
3467360998	Developer	Kolkata, West Bengal, India		Tata Consultancy Ser	On-site	Full-time - Mid-Senior le	10,001+ employees - IT S		14 9 hours	10,073 company alum

Raw dataset (Source: author)

After studying the data for a couple of minutes, turn it into a table to access the unlimited features Excel tables offer. Here, we are going to need the filter feature of the tables for undermining the outliers and missing data.



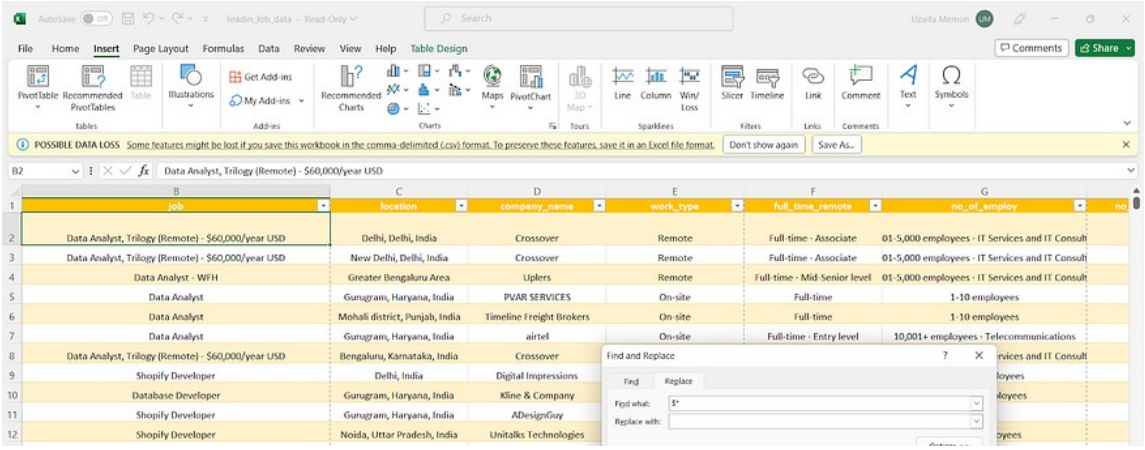
job	location	company_name	work_type	full_time_remote	no_of_employ
Data Analyst, Trilogy (Remote) - $60,000/year USD	Delhi, Delhi, India	Crossover	Remote	Full-time - Associate	01-5,000 employees - IT Services and IT Consult
Data Analyst, Trilogy (Remote) - $60,000/year USD	New Delhi, Delhi, India	Crossover	Remote	Full-time - Associate	01-5,000 employees - IT Services and IT Consult
Data Analyst - WFH	Greater Bengaluru Area	Uplers	Remote	Full-time - Mid-Senior level	01-5,000 employees - IT Services and IT Consult
Data Analyst	Gurugram, Haryana, India	PVAR SERVICES	On-site	Full-time	1-10 employees
Data Analyst	Mohali district, Punjab, India	Timeline Freight Brokers	On-site	Full-time	1-10 employees
Data Analyst	Gurugram, Haryana, India	airtel	On-site	Full-time - Entry level	10,001+ employees - Telecommunications
Data Analyst, Trilogy (Remote) - $60,000/year USD	Bengaluru, Karnataka, India	Crossover	Remote	Full-time - Associate	01-5,000 employees - IT Services and IT Consult
Shopify Developer	Delhi, India	Digital Impressions	On-site	Full-time	51-200 employees
Database Developer	Gurugram, Haryana, India	Kline & Company	Hybrid	Full-time	201-500 employees
Shopify Developer	Gurugram, Haryana, India	ADesignGuy	Hybrid		
Shopify Developer	Noida, Uttar Pradesh, India	Unitalks Technologies	Remote	Full-time	11-50 employees
Data Engineer	Gurugram, Haryana, India	ReNew Power	On-site	Full-time - Mid-Senior level	nployees - Renewable Energy Semiconductor M
Tableau (APL)	Noida, Uttar Pradesh, India	Escalent		Full-time	501-1,000 employees - Market Research
Data Engineer (Python)	Gurugram, Haryana, India	Mobile Programming LLC	On-site	Full-time	1,001-5,000 employees
Python Data Engineer	Hyderabad, Telangana, India	Tezra	Remote	Full-time	11-50 employees
Zoho Developer	India	Grow Your Staff	Remote	Contract	201-500 employees
Salesforce CPQ Developer(Looking for only Immediate Joiners)	Hyderabad, Telangana, India	ValueLabs	Remote	Full-time - Mid-Senior level	11-10,000 employees - IT Services and IT Consul

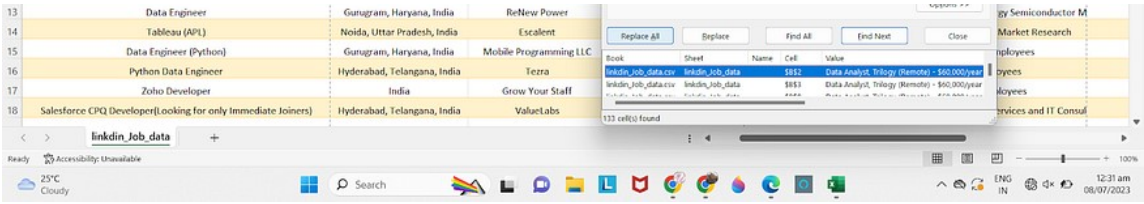
Raw data converted into a table

Step 2: Identifying The Outliers And Cleaning The Data

An outlier is any entry that doesn't align with the dataset. In other words, it is a trend breaker. In my case, the first outlier I found is in the job column. The jobs listed here do not follow a fixed pattern, rather they are sometimes accompanied by the salary or position which I do not want in this column.

So, I'm going to eliminate any character after the $ sign using Excel's infamous Find and Replace option. Please note that the * character is a wildcard. I'm using it along with the $ sign to eliminate any number of characters appearing after the $ sign.



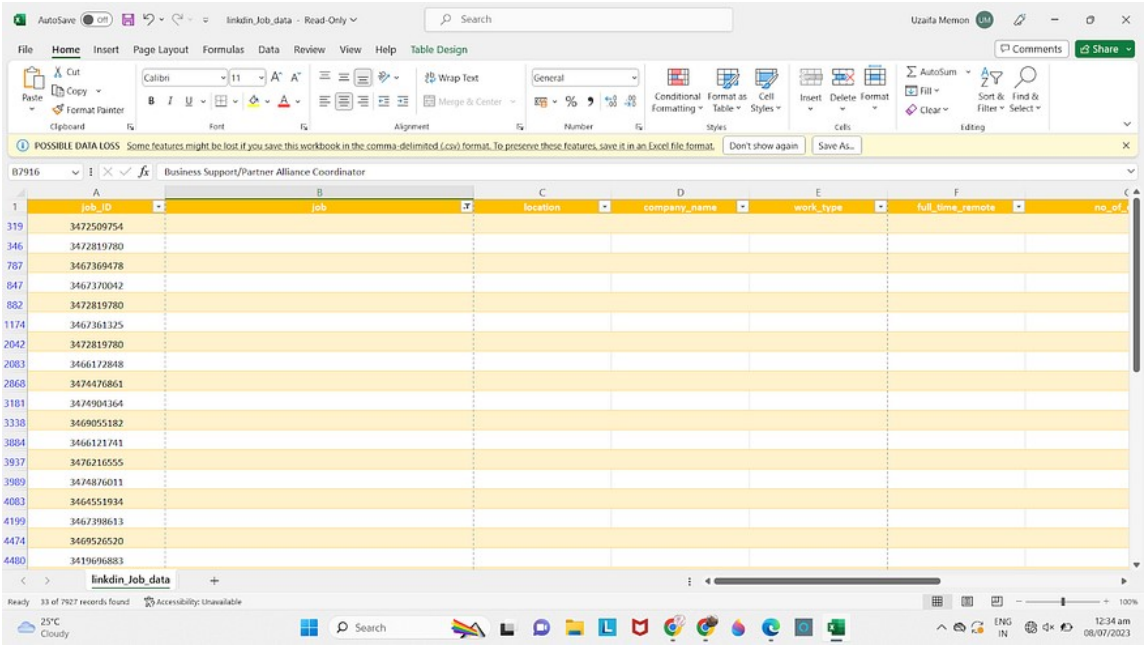


Removing the first outlier

Now, let us identify and deal with missing values in the dataset.

I'm going to use the filter feature of the table to find out blank values. (The filter option appears in each column in the form of the down-turned angle, simply click on it and you'll have the available filters.)

Since blank values are of no use to me in this particular dataset, I'll eliminate the rows that contain the blank cells.



Dealing with empty columns in the dataset

Continue the process for the remaining columns as well and you will have a complete dataset in a couple of minutes!

Step 3: Manipulate The Dataset

Now that all the outliers and missing values are dealt with, the next step is to manipulate the dataset to serve our purpose. Though this step is optional, it is vital for me to show this manipulation to you. There are two benefits of manipulating the dataset:

- You will have a sorted and clean dataset.
- Data visualization would become easier for you.

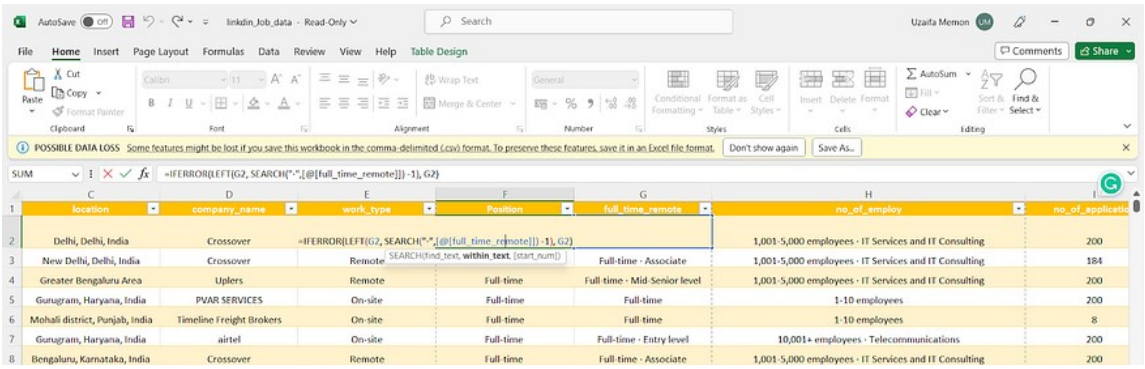
In the full_time_remote column, there is an irregular pattern that needs to be manipulated. I only want the position i.e. Full time, Contract, or Remote to appear in my

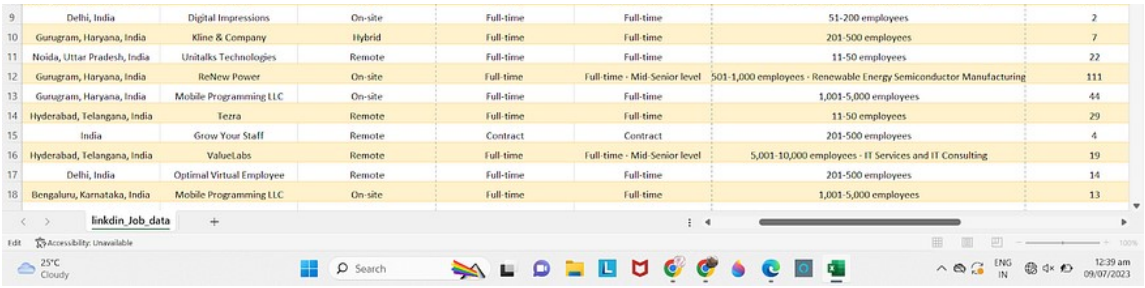
dashboard. So, I'll create a new column and extract the position from the full_time_remote column.

> Formula: =IFERROR(LEFT(G2, SEARCH("·", [@[full_time_remote]]) -1), G2)

Explanation: The SEARCH function will return the position of the specified character. The returned value would serve as a parameter for the LEFT function that will trim the characters before the specified character. Finally, if the SEARCH function doesn't find the specified character, I'll assign the original value to the cell.

If you find this messy or baffling, take a deep breath and read the explanation again. I know you have got this!

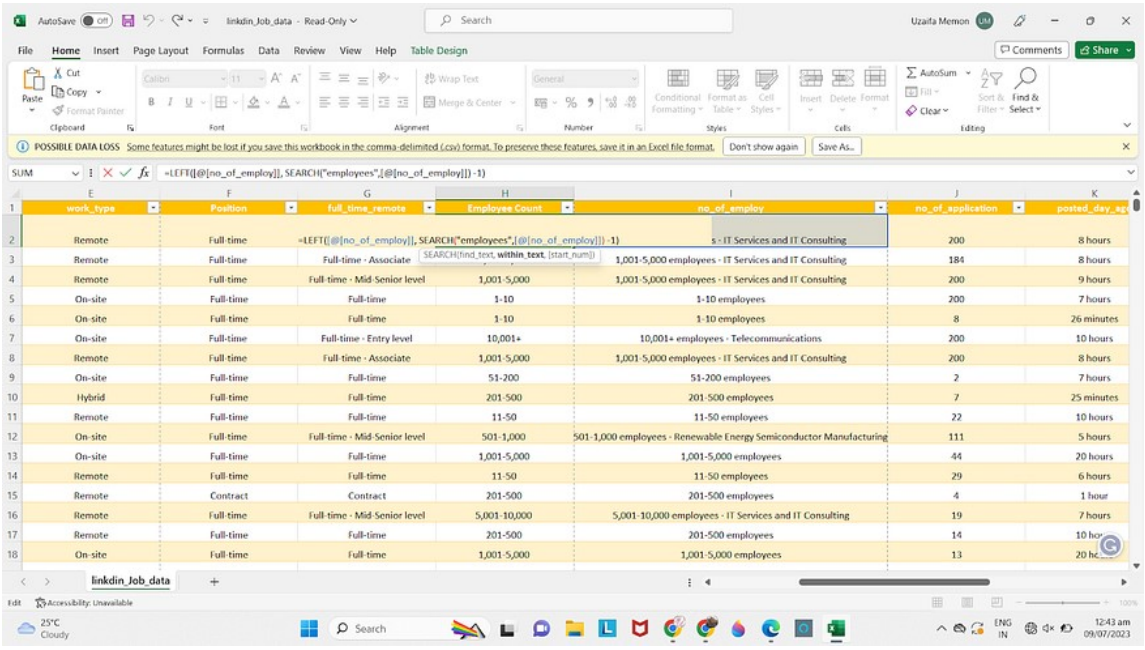




9	Delhi, India	Digital Impressions	On-site	Full-time	Full-time	51-200 employees	2
10	Gurugram, Haryana, India	Kline & Company	Hybrid	Full-time	Full-time	201-500 employees	7
11	Noida, Uttar Pradesh, India	Unitalks Technologies	Remote	Full-time	Full-time	11-50 employees	22
12	Gurugram, Haryana, India	ReNew Power	On-site	Full-time	Full-time · Mid-Senior level	501-1,000 employees · Renewable Energy Semiconductor Manufacturing	111
13	Gurugram, Haryana, India	Mobile Programming LLC	On-site	Full-time	Full-time	1,001-5,000 employees	44
14	Hyderabad, Telangana, India	Tezra	Remote	Full-time	Full-time	11-50 employees	29
15	India	Grow Your Staff	Remote	Contract	Contract	201-500 employees	4
16	Hyderabad, Telangana, India	ValueLabs	Remote	Full-time	Full-time · Mid-Senior level	5,001-10,000 employees · IT Services and IT Consulting	19
17	Delhi, India	Optimal Virtual Employee	Remote	Full-time	Full-time	201-500 employees	14
18	Bengaluru, Karnataka, India	Mobile Programming LLC	On-site	Full-time	Full-time	1,001-5,000 employees	13

Using search, left and if error formulae to extract the exact position

I'm going to use this same formula to retrieve the exact number of elements present in the no_of_employee column.

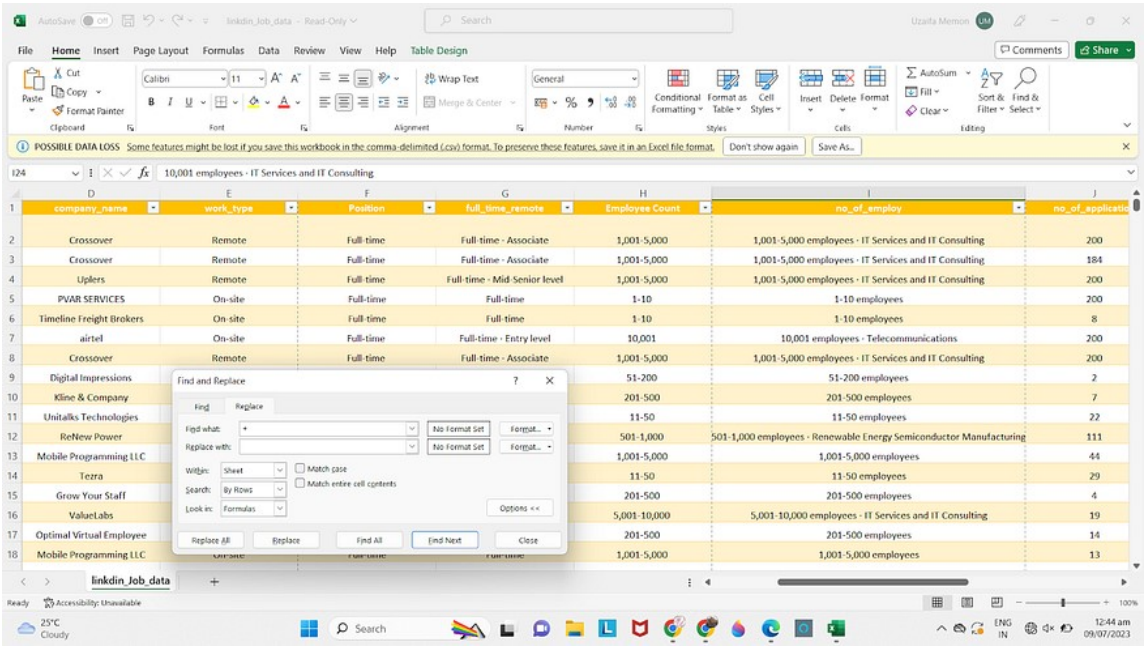


	E	F	G	H	I	J	K
1	work_type	Position	full_time_remote	Employee Count	no_of_employ	no_of_application	posted_day_a
2	Remote	Full-time	=LEFT([@[no_of_employ]], SEARCH("employees",[@[no_of_employ]])-1)		s - IT Services and IT Consulting	200	8 hours
3	Remote	Full-time	Full-time - Associate	SEARCH(find_text, within_text, [start_num])	1,001-5,000 employees - IT Services and IT Consulting	184	8 hours
4	Remote	Full-time	Full-time - Mid-Senior level	1,001-5,000	1,001-5,000 employees - IT Services and IT Consulting	200	9 hours
5	On-site	Full-time	Full-time	1-10	1-10 employees	200	7 hours
6	On-site	Full-time	Full-time	1-10	1-10 employees	8	26 minutes
7	On-site	Full-time	Full-time - Entry level	10,001+	10,001+ employees - Telecommunications	200	10 hours
8	Remote	Full-time	Full-time - Associate	1,001-5,000	1,001-5,000 employees - IT Services and IT Consulting	200	8 hours
9	On-site	Full-time	Full-time	51-200	51-200 employees	2	7 hours
10	Hybrid	Full-time	Full-time	201-500	201-500 employees	7	25 minutes
11	Remote	Full-time	Full-time	11-50	11-50 employees	22	10 hours
12	On-site	Full-time	Full-time - Mid-Senior level	501-1,000	501-1,000 employees - Renewable Energy Semiconductor Manufacturing	111	5 hours
13	On-site	Full-time	Full-time	1,001-5,000	1,001-5,000 employees	44	20 hours
14	Remote	Full-time	Full-time	11-50	11-50 employees	29	6 hours
15	Remote	Contract	Contract	201-500	201-500 employees	4	1 hour
16	Remote	Full-time	Full-time - Mid-Senior level	5,001-10,000	5,001-10,000 employees - IT Services and IT Consulting	19	7 hours
17	Remote	Full-time	Full-time	201-500	201-500 employees	14	10 ho
18	On-site	Full-time	Full-time	1,001-5,000	1,001-5,000 employees	13	20 ho

Using search, left and if error formulae to retrieve the exact number of employee

Here's a bit of a problem though. After retrieving the values before the hyphen (-), some cells still contain the + symbol. I don't

want this symbol to make my data look unprofessional. So, using the Find and Replace option, I'll eliminate the + symbol.



Eliminating the + symbol using Find and Replace option

Okay, so here we are done manipulating the data. Let us get the answers to our questions, derive some insights and curate a dashboard!

Step 4: Generating Insights From The Dataset

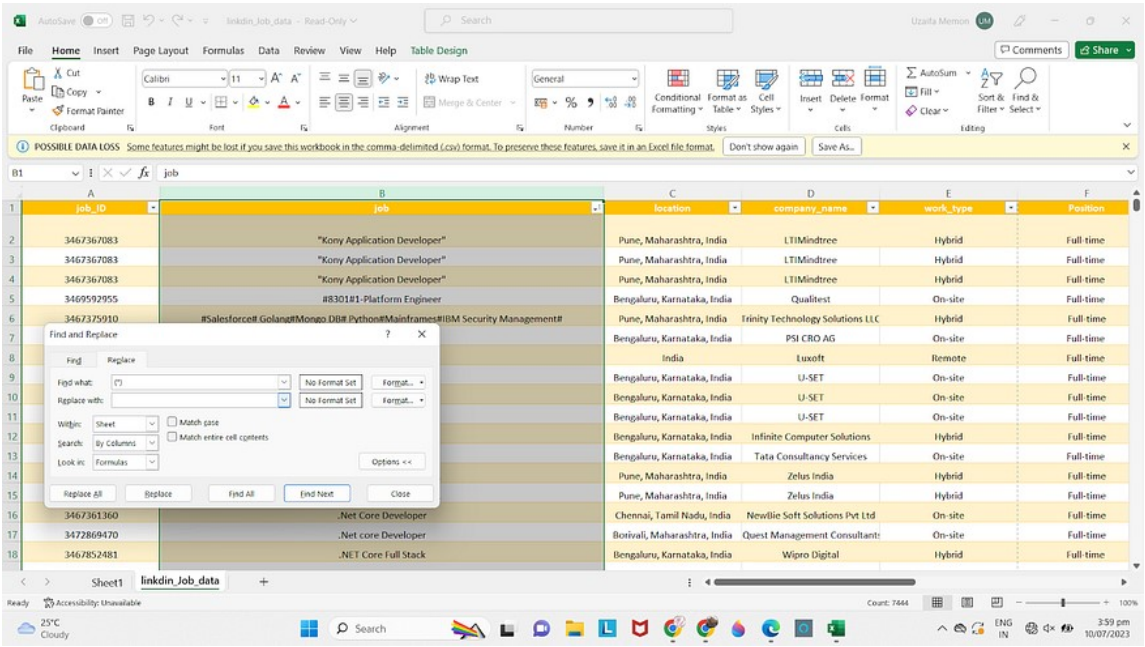
Question 1: Which job is in high demand?

To answer the question, I'll create a pivot

table.

(A pivot table is generally, Excel's new feature that lets you work with a single row, column, or rows of them and provides the feature to manipulate this data.)

But before I create a chart and add the Top 10 value filter to it, I have to make some changes. Notice that the jobs listed here contain several brackets which are disrupting the dataset. I'll eliminate them using the (yes, you're right) Find and Replace option.

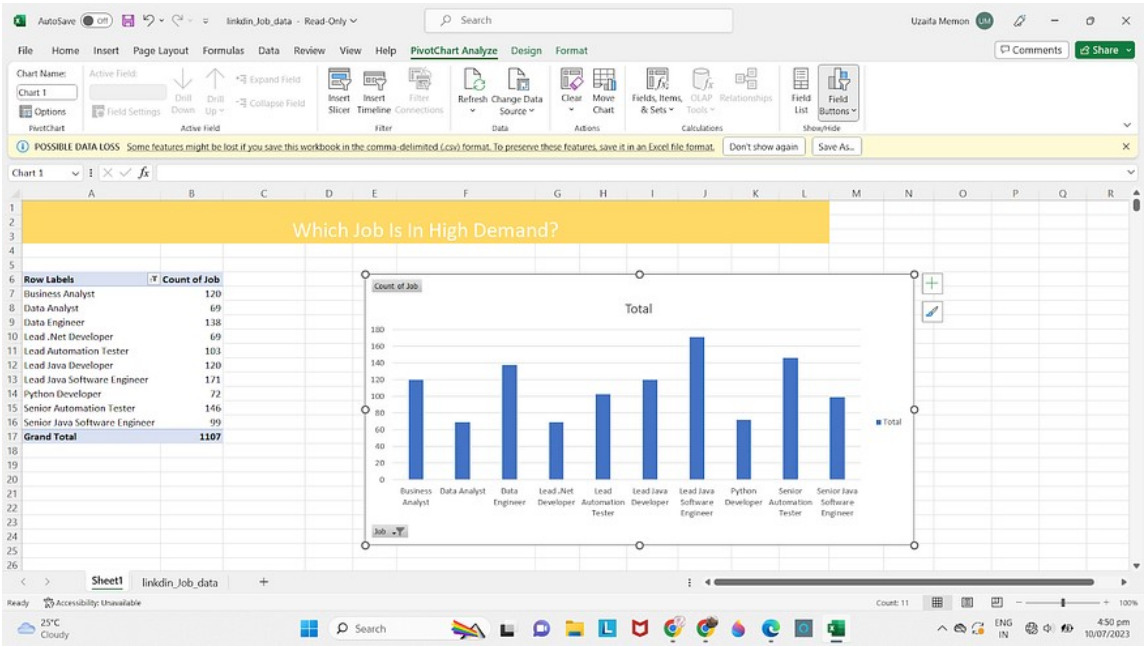


Eliminating the characters from (to) characters.

Now, I'm simply going to create a pivot table chart, and woah, we're done.

Lastly, apply the formula PROPER to the Job column to beautify your chart and you will have your answer. Note that the PROPER function is used to apply the sentence case.

Answer 1: Lead Java Software Engineer is trending on LinkedIn these days.



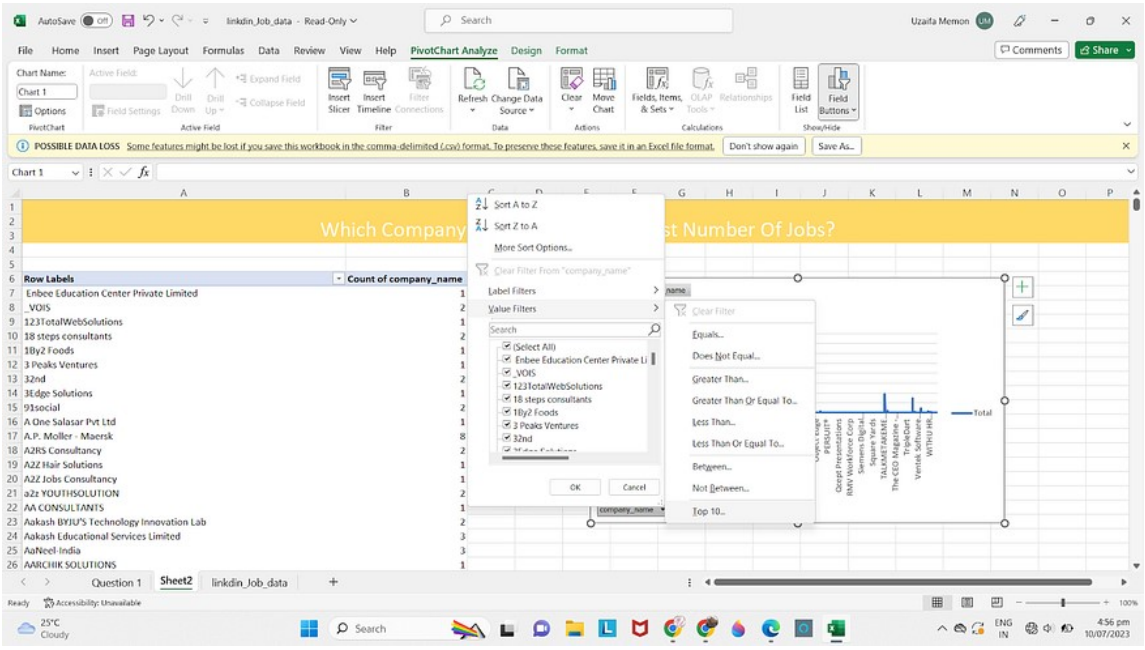
Answer to question 1

Question 2: Which company has posted the highest number of jobs?

Again, follow the same format (create a pivot table first and then a chart). Apply the Top 10

value filter and move on to the next question!

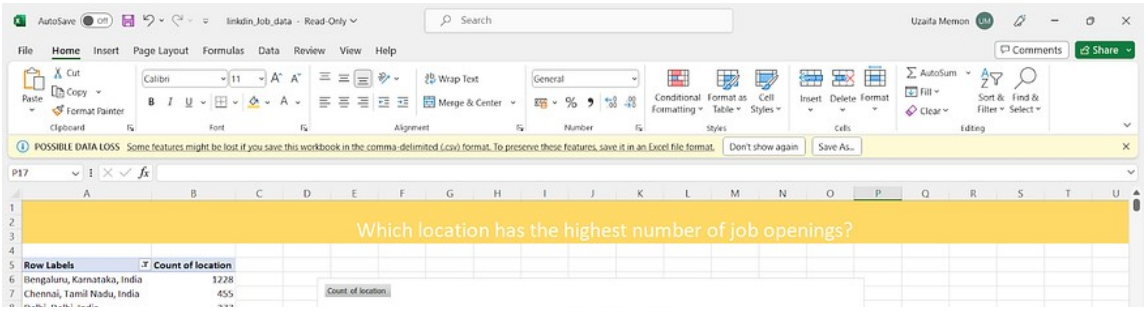
Answer 2: Tata Consultancy Services (TCS) is the most active company on LinkedIn with respect to posting a ginormous amount of jobs.

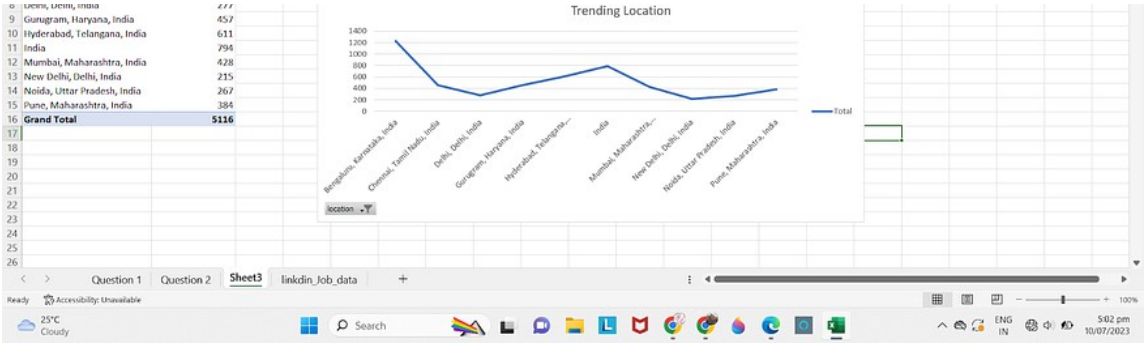


Answer to question 2

Question 3: Which location has the highest number of job openings?

Answer 3: Bengaluru, Karnataka, India has the highest number of job openings.

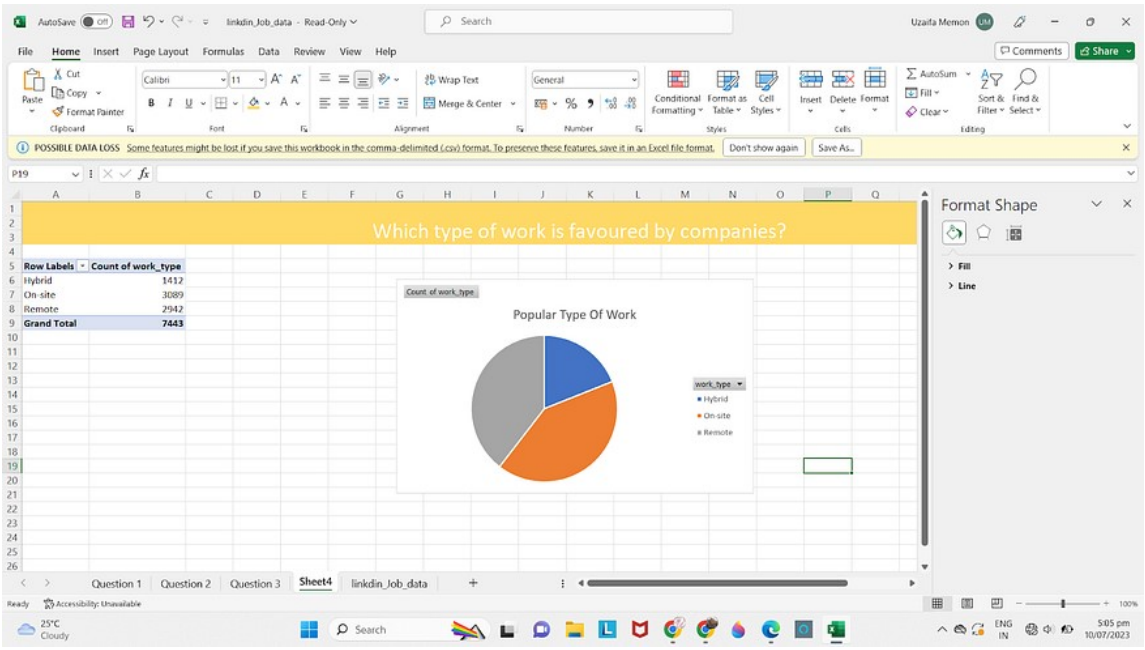




Answer to question 3

Question 4: Which type of work is favoured by companies?

Answer 4: Companies are currently preferring remote work (Yo, freelancers!)

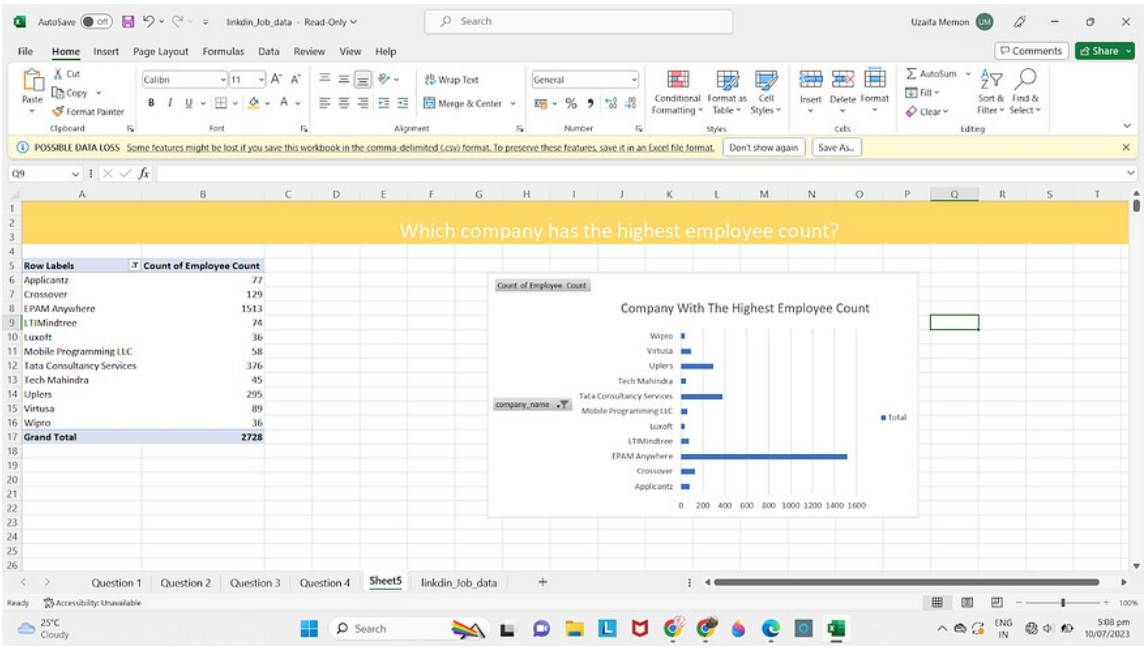


Answer to question 4

Question 5: Which company has the highest employee count?

Answer 5: EPAM Anywhere has the highest

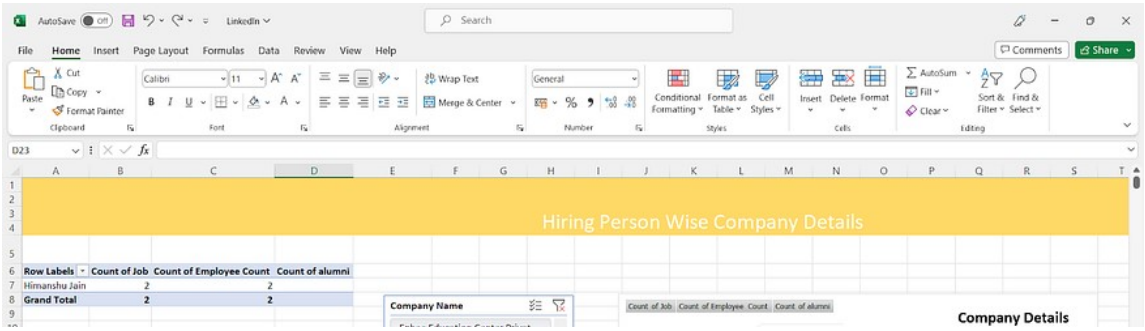
employee count with 1513 employees.

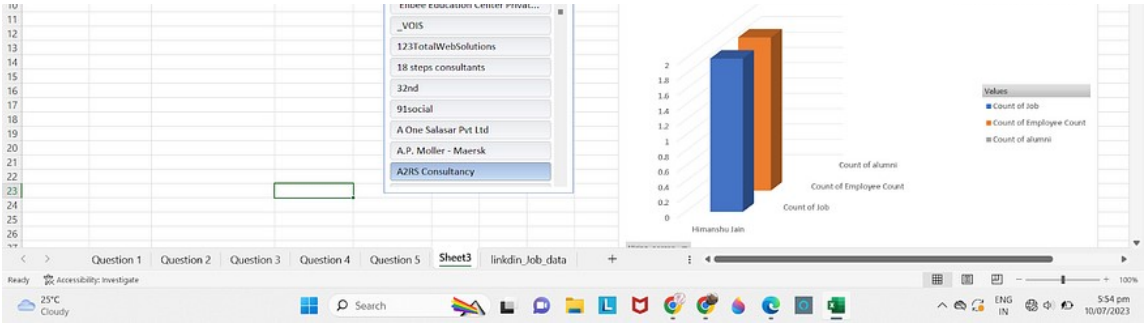


Answer to question 5

Question 6: List hiring manager-wise company details.

For this objective question, I'll be using a slicer. (A slicer is simply a filter you could apply on your pivot table to filter the values). I'll then create a two-dimensional chart to portray the employee and alumni count of the company along with the hiring person's name.





Answer to question 6

Please note that I'm leaving the remaining questions for you to practice and solve. Once you're done and ready to share, you could either let me know in the comment section of this post or on my email (memonuzaifa21@gmail.com).

Step 5: Creating the dashboard

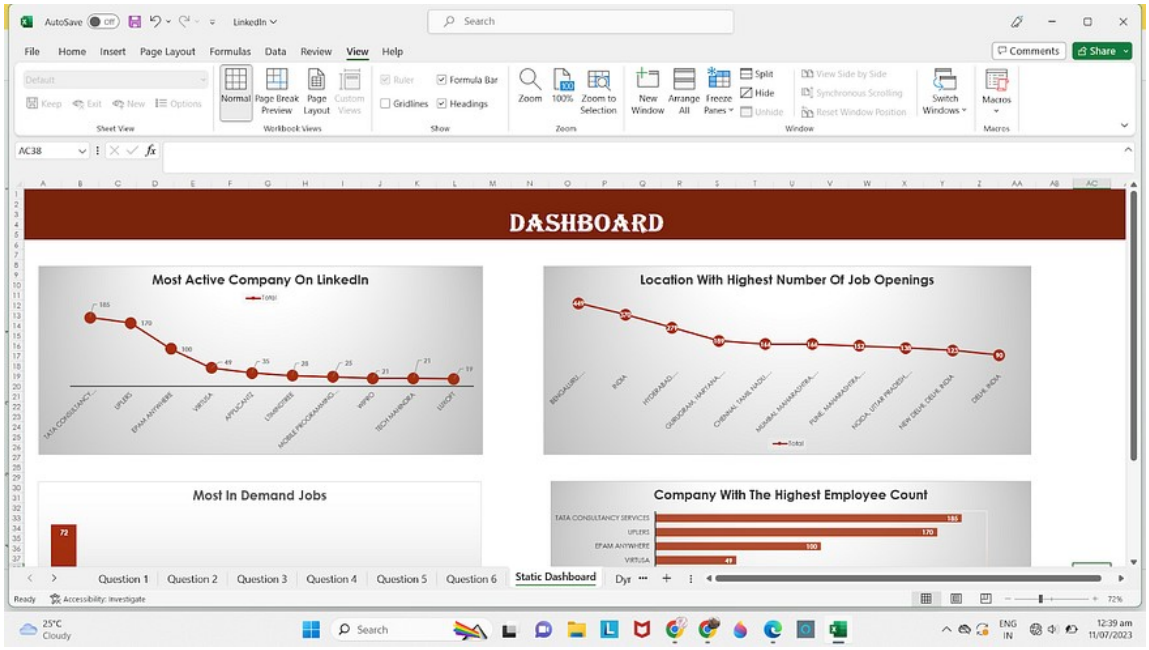
Ah, finally!

Here, I'll show you two types of dashboards that would help you portray the generated insights in a user-friendly manner. The first is a static dashboard in which I have simply cut the dashboards from their respective sheets and pasted them into this new dashboard

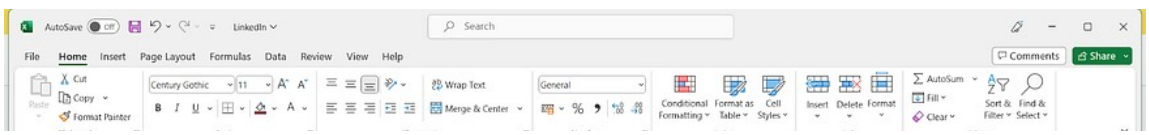
sheet.

Oh, and I have applied some formatting to the charts, disabled the field buttons from the PivotChart Analyze ribbon, and arranged the charts accordingly.

As for the dynamic dashboard, I have created another sheet, pasted the slicer, and attached this slicer to three different chars. You can link one slicer to multiple pivot tables using the Report Connections option from the Slicer ribbon.



Static Dashboard



Dynamic dashboard

Conclusion

Did you like the dashboards? I certainly hope you did and if you did, please let me know. Don’t worry I won’t take a lot of your time because I know you need a coffee break after this thorough case study.

Just two things.

Thank you for reading!

And stay tuned for more such case studies, tutorials, tips, and discussions!

Oh, and yes, in case you want to take a deeper look at these dashboards, you can find this

Excel file here.