Task 1 : HR Data Analysis

Introduction

Hello everyone! I am thrilled to announce that I have secured a Data Analyst Internship at Psyliq. I've been tasked with an exciting challenge — an HR Data Analysis Assessment comprising 17 questions.

In this article, I will showcase my approach to solving the Task 1.

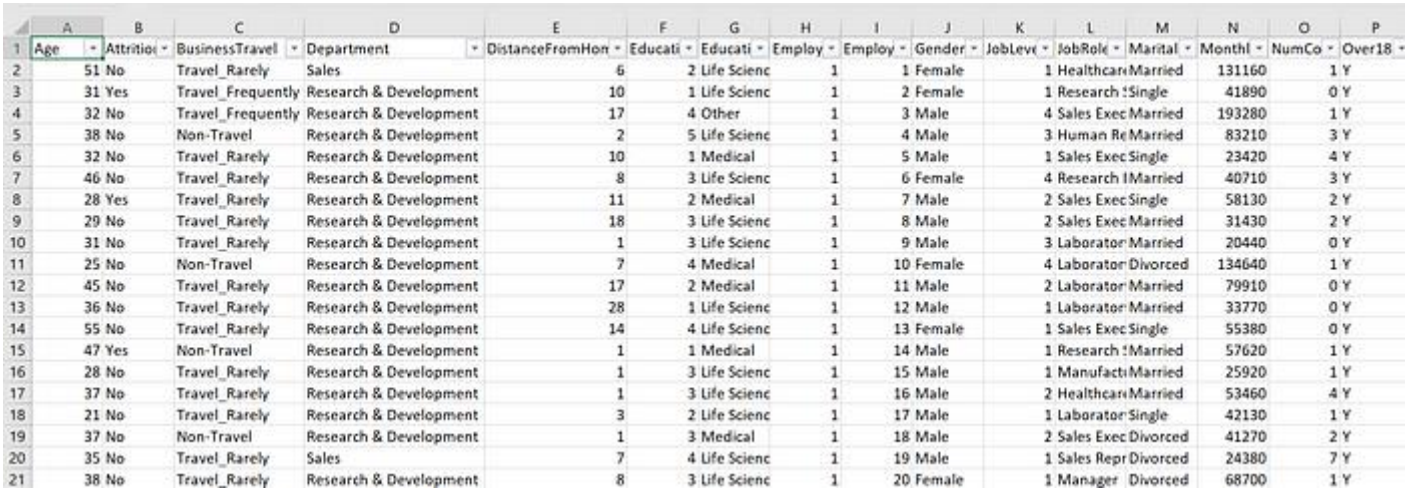
Let's dive in!

Task Questions

1. **Using Excel, how would you filter the dataset to only show employees aged 30 and above?**

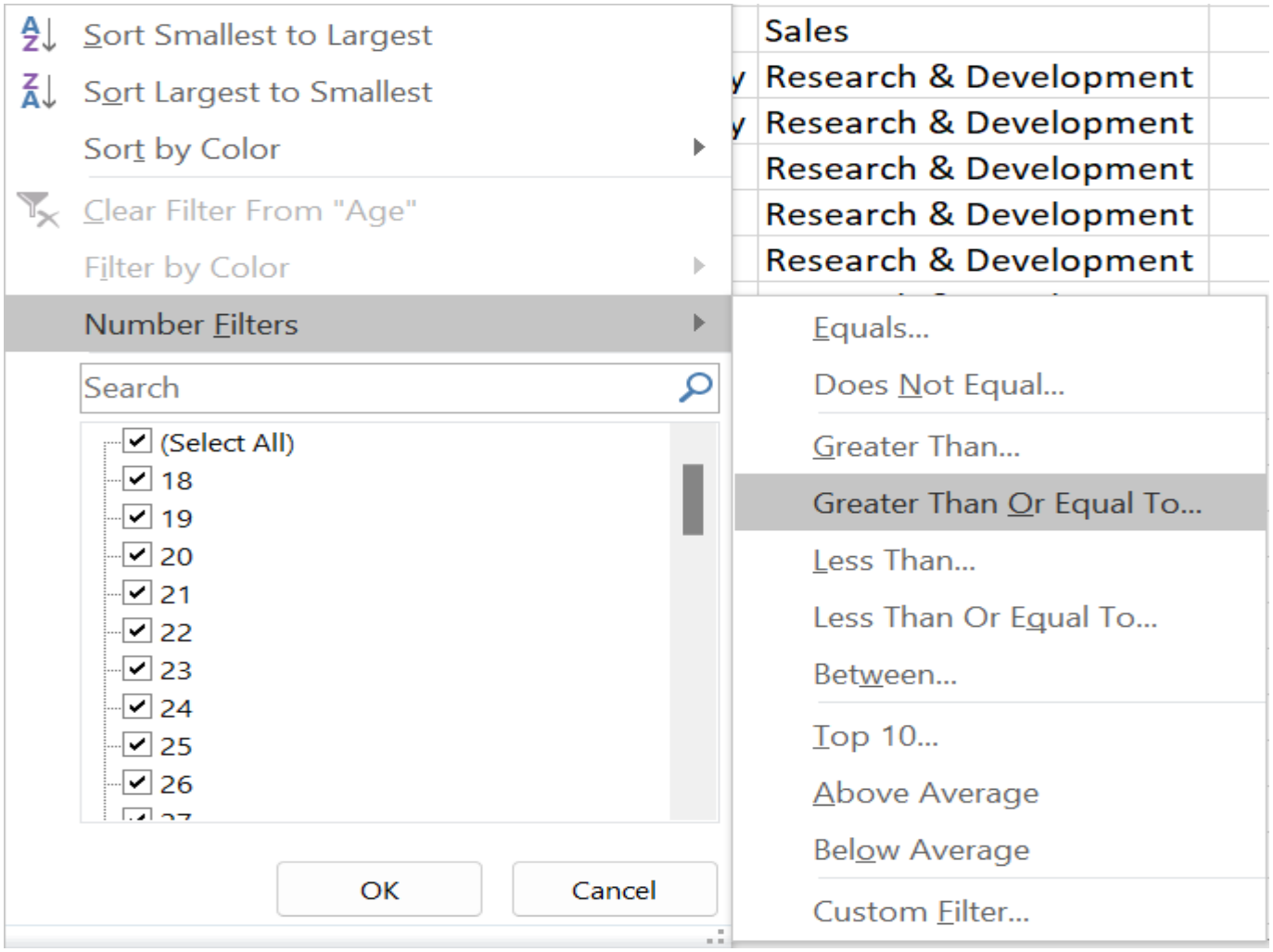
Before performing any analysis on Excel. I have created a copy of the dataset. Because It helps to preserve the integrity of the original data and serves as a precautionary measure in case any mistakes are made during the analysis process.

Now moving to answer this question. I have opened the **general data** Excel file. Then I applied a filter on the header by using the shortcut **Ctrl + Shift + L.**



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Age	Attritio	BusinessTravel	Department	DistanceFromHon	Educati	Educati	Employ	Employ	Gender	JobLeve	JobRole	Marital	Monthl	NumCo	Over18
2	51	No	Travel_Rarely	Sales	6	2	Life Scienc	1	1	Female	1	Healthcar	Married	131160	1	Y
3	31	Yes	Travel_Frequently	Research & Development	10	1	Life Scienc	1	2	Female	1	Research	Single	41890	0	Y
4	32	No	Travel_Frequently	Research & Development	17	4	Other	1	3	Male	4	Sales Exec	Married	193280	1	Y
5	38	No	Non-Travel	Research & Development	2	5	Life Scienc	1	4	Male	3	Human Re	Married	83210	3	Y
6	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	5	Male	1	Sales Exec	Single	23420	4	Y
7	46	No	Travel_Rarely	Research & Development	8	3	Life Scienc	1	6	Female	4	Research I	Married	40710	3	Y
8	28	Yes	Travel_Rarely	Research & Development	11	2	Medical	1	7	Male	2	Sales Exec	Single	58130	2	Y
9	29	No	Travel_Rarely	Research & Development	18	3	Life Scienc	1	8	Male	2	Sales Exec	Married	31430	2	Y
10	31	No	Travel_Rarely	Research & Development	1	3	Life Scienc	1	9	Male	3	Laborator	Married	20440	0	Y
11	25	No	Non-Travel	Research & Development	7	4	Medical	1	10	Female	4	Laborator	Divorced	134640	1	Y
12	45	No	Travel_Rarely	Research & Development	17	2	Medical	1	11	Male	2	Laborator	Married	79910	0	Y
13	36	No	Travel_Rarely	Research & Development	28	1	Life Scienc	1	12	Male	1	Laborator	Married	33770	0	Y
14	55	No	Travel_Rarely	Research & Development	14	4	Life Scienc	1	13	Female	1	Sales Exec	Single	55380	0	Y
15	47	Yes	Non-Travel	Research & Development	1	1	Medical	1	14	Male	1	Research	Married	57620	1	Y
16	28	No	Travel_Rarely	Research & Development	1	3	Life Scienc	1	15	Male	1	Manufacti	Married	25920	1	Y
17	37	No	Travel_Rarely	Research & Development	1	3	Life Scienc	1	16	Male	2	Healthcar	Married	53460	4	Y
18	21	No	Travel_Rarely	Research & Development	3	2	Life Scienc	1	17	Male	1	Laborator	Single	42130	1	Y
19	37	No	Non-Travel	Research & Development	1	3	Medical	1	18	Male	2	Sales Exec	Divorced	41270	2	Y
20	35	No	Travel_Rarely	Sales	7	4	Life Scienc	1	19	Male	1	Sales Repr	Divorced	24380	7	Y
21	38	No	Travel_Rarely	Research & Development	8	3	Life Scienc	1	20	Female	1	Manager	Divorced	68700	1	Y

Then, Click on the filter button and then choose option **Number Filters,** and then choose **Greater than or equal to** option.



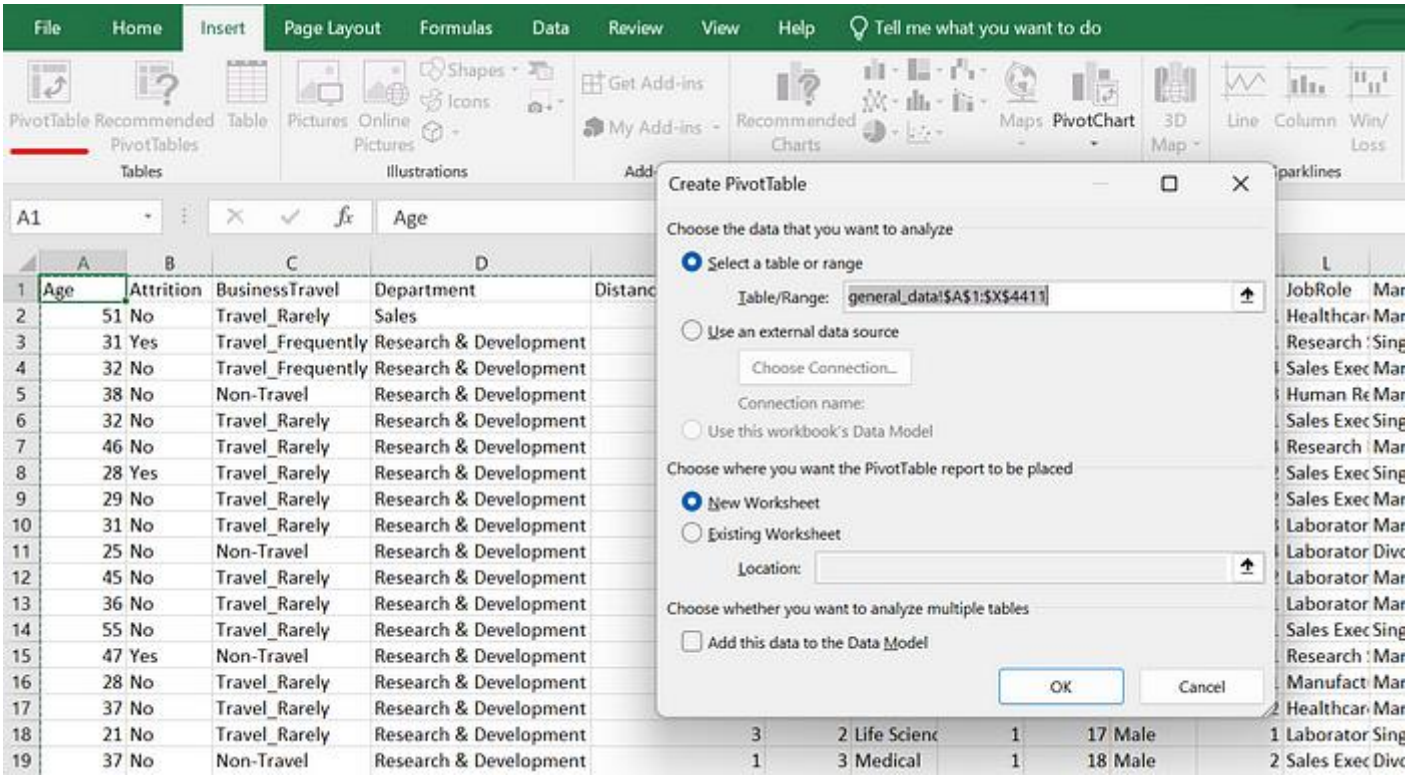
Then Custom AutoFilter Dialogue box is opened. Beside **is greater than or equal to** field I have filled **30**(as this is my condition) and then clicked ok.

Now, you can observe that the 'Age' field includes values equal to or above 30.

Age	Attritic	BusinessTravel	Department	DistanceFromHor	Educat	Educat	Emplo	Emplo	Gende	JobLev	JobRol	Marita	Month	NumCc	Over18	Percen	Standa
51	No	Travel_Rarely	Sales	6	2	Life Scienc	1	1	Female	1	Healthcar	Married	131160	1	Y	11	8
31	Yes	Travel_Frequently	Research & Development	10	1	Life Scienc	1	2	Female	1	Research	Single	41890	0	Y	23	8
32	No	Travel_Frequently	Research & Development	17	4	Other	1	3	Male	4	Sales Exec	Married	193280	1	Y	15	8
38	No	Non-Travel	Research & Development	2	5	Life Scienc	1	4	Male	3	Human Re	Married	83210	3	Y	11	8
32	No	Travel_Rarely	Research & Development	10	1	Medical	1	5	Male	1	Sales Exec	Single	23420	4	Y	12	8
46	No	Travel_Rarely	Research & Development	8	3	Life Scienc	1	6	Female	4	Research	Married	40710	3	Y	13	8
31	No	Travel_Rarely	Research & Development	1	3	Life Scienc	1	9	Male	3	Laborator	Married	20440	0	Y	21	8
45	No	Travel_Rarely	Research & Development	17	2	Medical	1	11	Male	2	Laborator	Married	79910	0	Y	13	8
36	No	Travel_Rarely	Research & Development	28	1	Life Scienc	1	12	Male	1	Laborator	Married	33770	0	Y	12	8
55	No	Travel_Rarely	Research & Development	14	4	Life Scienc	1	13	Female	1	Sales Exec	Single	55380	0	Y	17	8
47	Yes	Non-Travel	Research & Development	1	1	Medical	1	14	Male	1	Research	Married	57620	1	Y	11	8
37	No	Travel_Rarely	Research & Development	1	3	Life Scienc	1	16	Male	2	Healthcar	Married	53460	4	Y	11	8
37	No	Non-Travel	Research & Development	1	3	Medical	1	18	Male	2	Sales Exec	Divorced	41270	2	Y	13	8
35	No	Travel_Rarely	Sales	7	4	Life Scienc	1	19	Male	1	Sales Rep	Divorced	24380	7	Y	16	8
38	No	Travel_Rarely	Research & Development	8	3	Life Scienc	1	20	Female	1	Manager	Divorced	68700	1	Y	11	8
50	No	Travel_Rarely	Sales	8	4	Life Scienc	1	22	Male	1	Research	Divorced	96670	3	Y	23	8
53	No	Travel_Rarely	Research & Development	11	4	Life Scienc	1	23	Female	2	Research	Married	21480	3	Y	11	8
42	No	Travel_Rarely	Research & Development	4	4	Life Scienc	1	24	Male	1	Manufact	Married	89260	1	Y	14	8
55	No	Travel_Rarely	Research & Development	1	4	Other	1	26	Female	1	Research	Married	67990	3	Y	11	8
37	No	Travel_Rarely	Sales	5	1	Marketing	1	28	Male	1	Research	Single	27050	1	Y	11	8
44	Yes	Travel_Frequently	Research & Development	1	2	Medical	1	29	Male	2	Research	Divorced	103330	3	Y	14	8
38	No	Travel_Rarely	Sales	2	3	Marketing	1	30	Female	1	Manager	Divorced	44480	9	Y	12	8
49	No	Travel_Frequently	Research & Development	1	1	Medical	1	33	Female	2	Research	Single	35910	9	Y	13	8

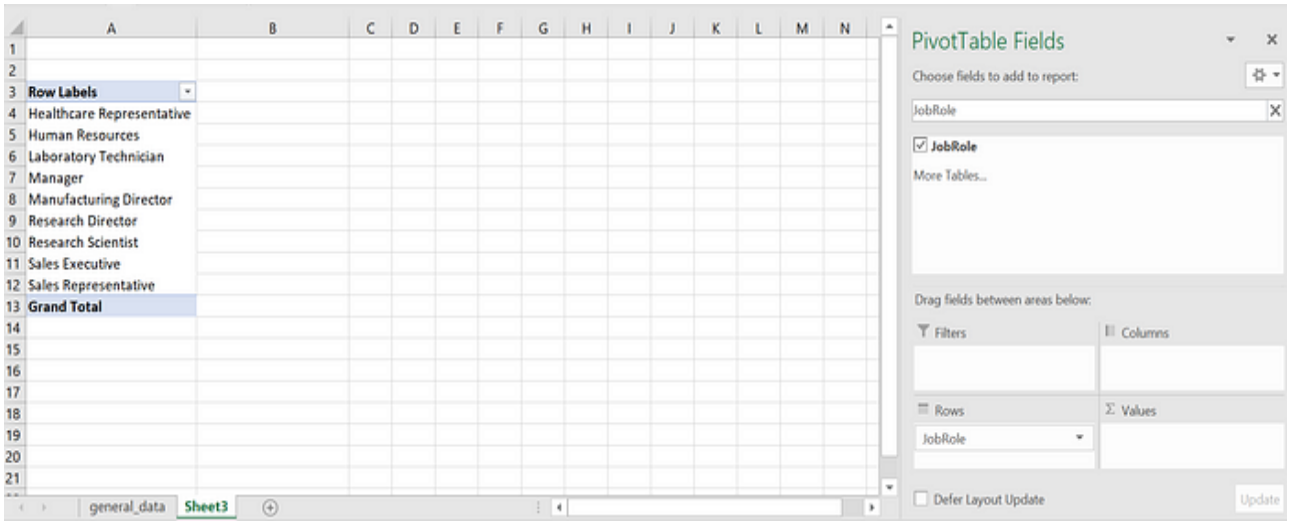
2. **Create a pivot table to summarize the average Monthly Income by Job Role.**

For Creating a pivot table for summarizing the Data. **Go to Insert tab in the Ribbon > Click on Pivot table as shown in the image > Create PivotTable dialogue box will open > Click Ok** (By default the Table/ Range and New Worksheet where you pivot table want has been selected).



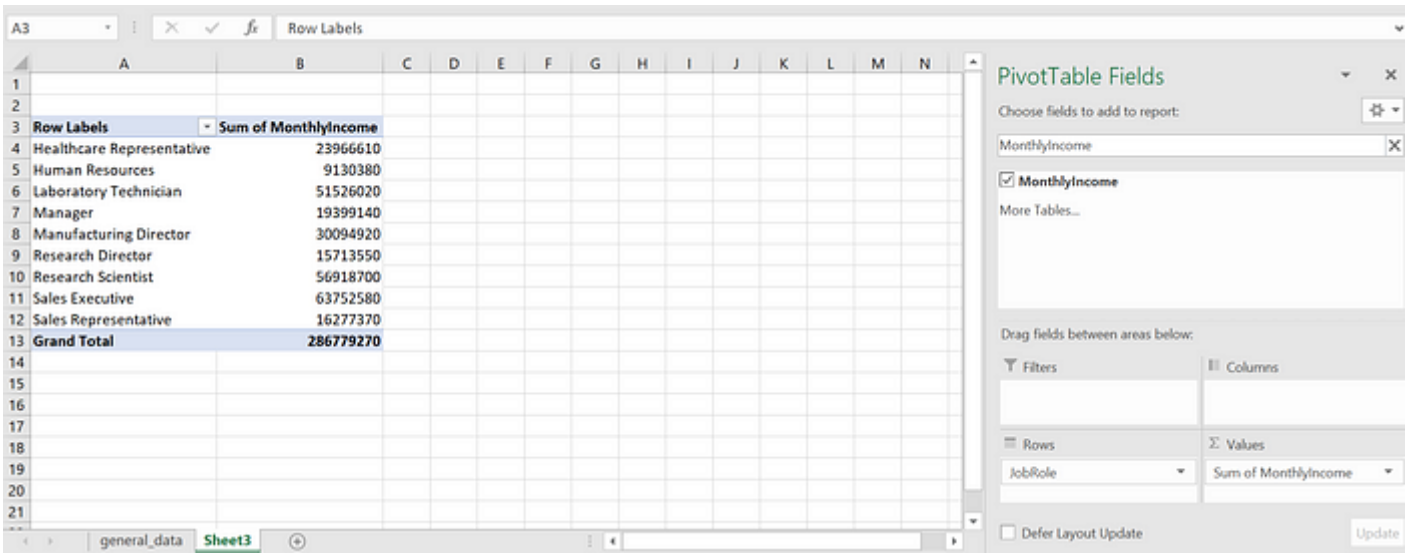
	A	B	C	D	L
1	Age	Attrition	BusinessTravel	Department	JobRole
2	51	No	Travel_Rarely	Sales	Healthcar
3	31	Yes	Travel_Frequently	Research & Development	Research
4	32	No	Travel_Frequently	Research & Development	Sales Exec
5	38	No	Non-Travel	Research & Development	Human Re
6	32	No	Travel_Rarely	Research & Development	Sales Exec
7	46	No	Travel_Rarely	Research & Development	Research
8	28	Yes	Travel_Rarely	Research & Development	Sales Exec
9	29	No	Travel_Rarely	Research & Development	Sales Exec
10	31	No	Travel_Rarely	Research & Development	Laborator
11	25	No	Non-Travel	Research & Development	Laborator
12	45	No	Travel_Rarely	Research & Development	Laborator
13	36	No	Travel_Rarely	Research & Development	Laborator
14	55	No	Travel_Rarely	Research & Development	Sales Exec
15	47	Yes	Non-Travel	Research & Development	Research
16	28	No	Travel_Rarely	Research & Development	Manufact
17	37	No	Travel_Rarely	Research & Development	Healthcar
18	21	No	Travel_Rarely	Research & Development	Laborator
19	37	No	Non-Travel	Research & Development	Sales Exec

In PivotTable Fields. First, I have selected the **Job role**. By Default JobRole will get inserted in the Rows field.



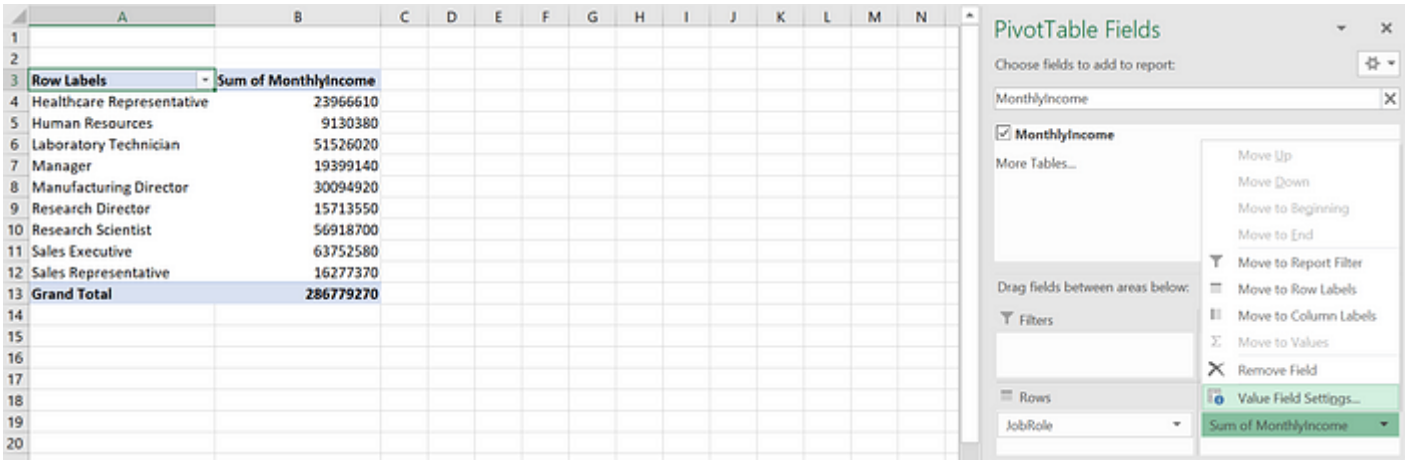
Pivot Table

Likewise, I have selected **Monthly Income**. By Default **Monthly Income** will get inserted in the values field.



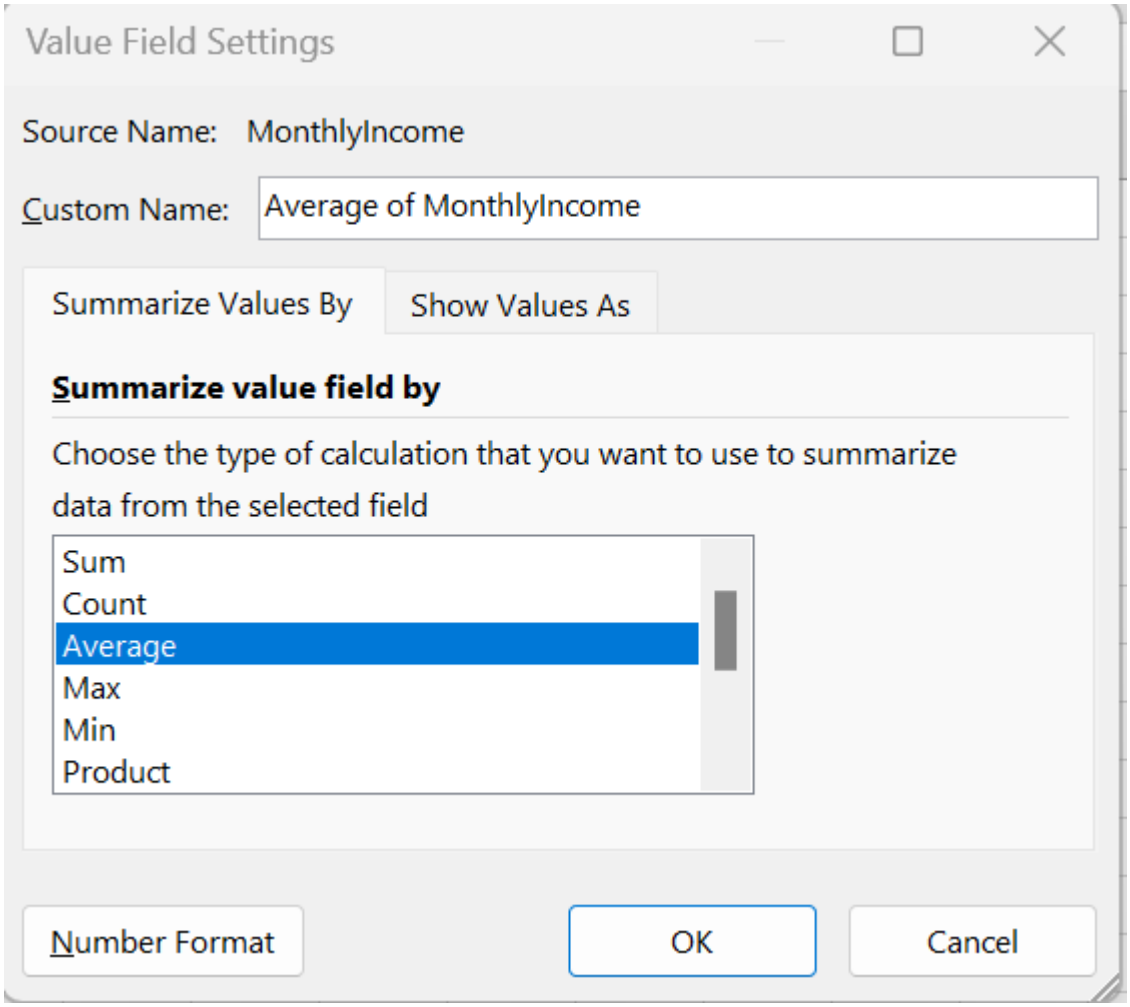
Pivot Table

After that, I have to change the summarization of the **Monthly Income** field from **Sum to Average** (as the task is to calculate the Average of Monthly Income). For that, I have to click the drop down option in the **Values field** and then Choose the **Value Field Settings**



Pivot Table

Then **Value Field Settings** dialogue box will open. After that choose the type of calculation you want. I want **Average** Calculation and then Click OK.



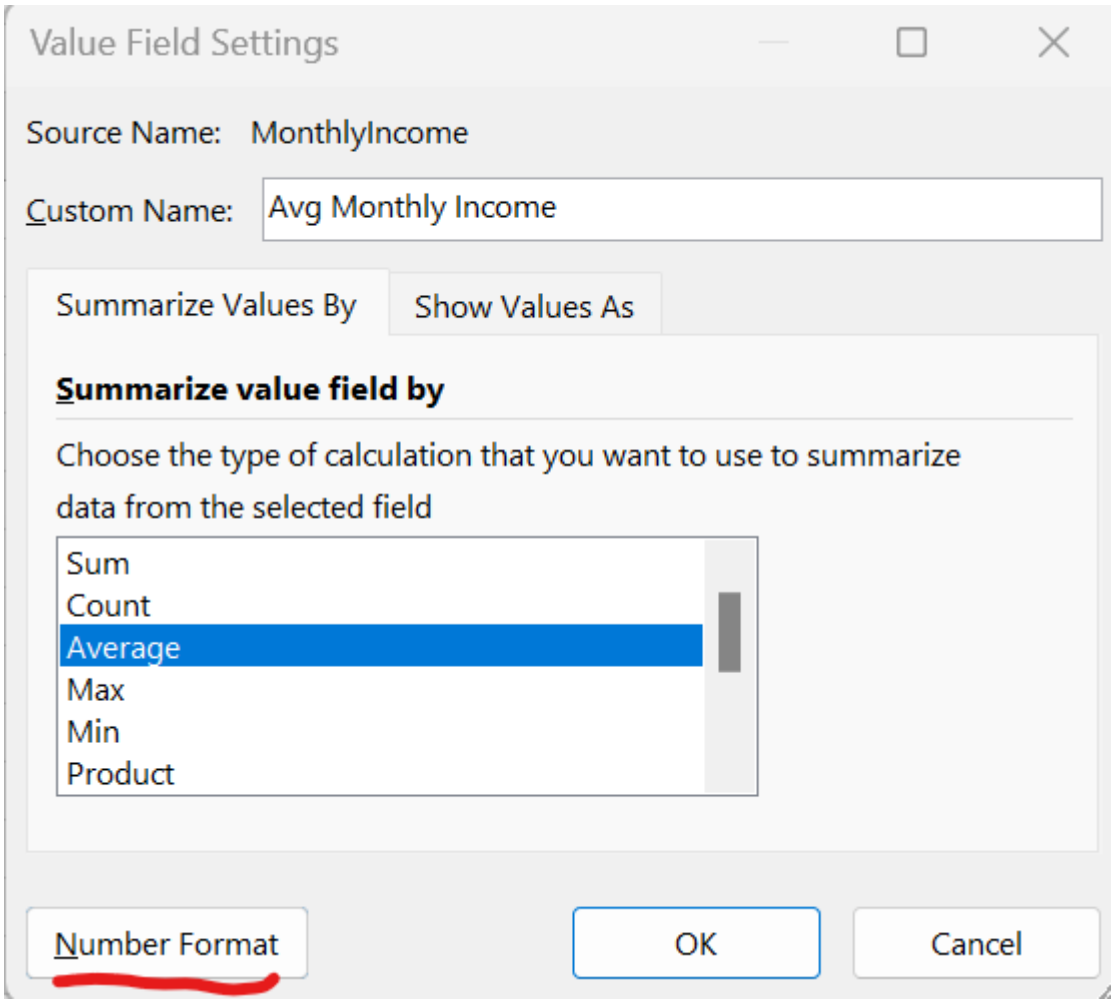
Value Field Settings

Now we can see, that **MonthlyIncome** values are summarized by **Average**.

Monthly Income by Job Role

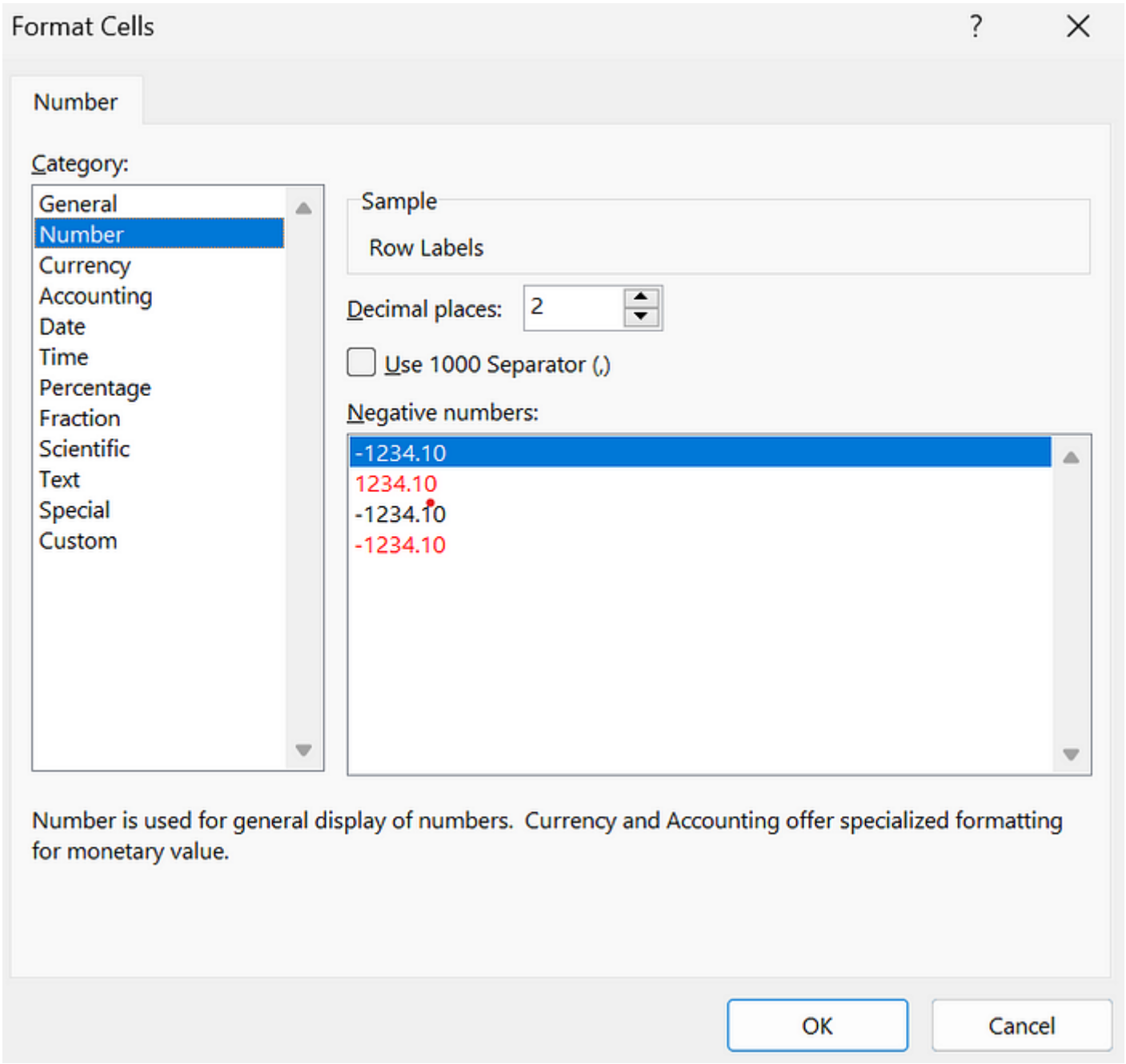
But there is a problem, It is not advisable to show more than 2 numbers after decimal so we will change the format. You just have to open the **Value Field Settings** Box again and then click on the **Number Format**.

Note: I have changed the name of the Column from Average of Monthly Income to Avg Monthly Income.



Value Field Settings

Format Cells Dialogue Box will open. Select **Number** from the Category. By default Decimal Places will be set to 2 and then Click OK.



Foramat Cells

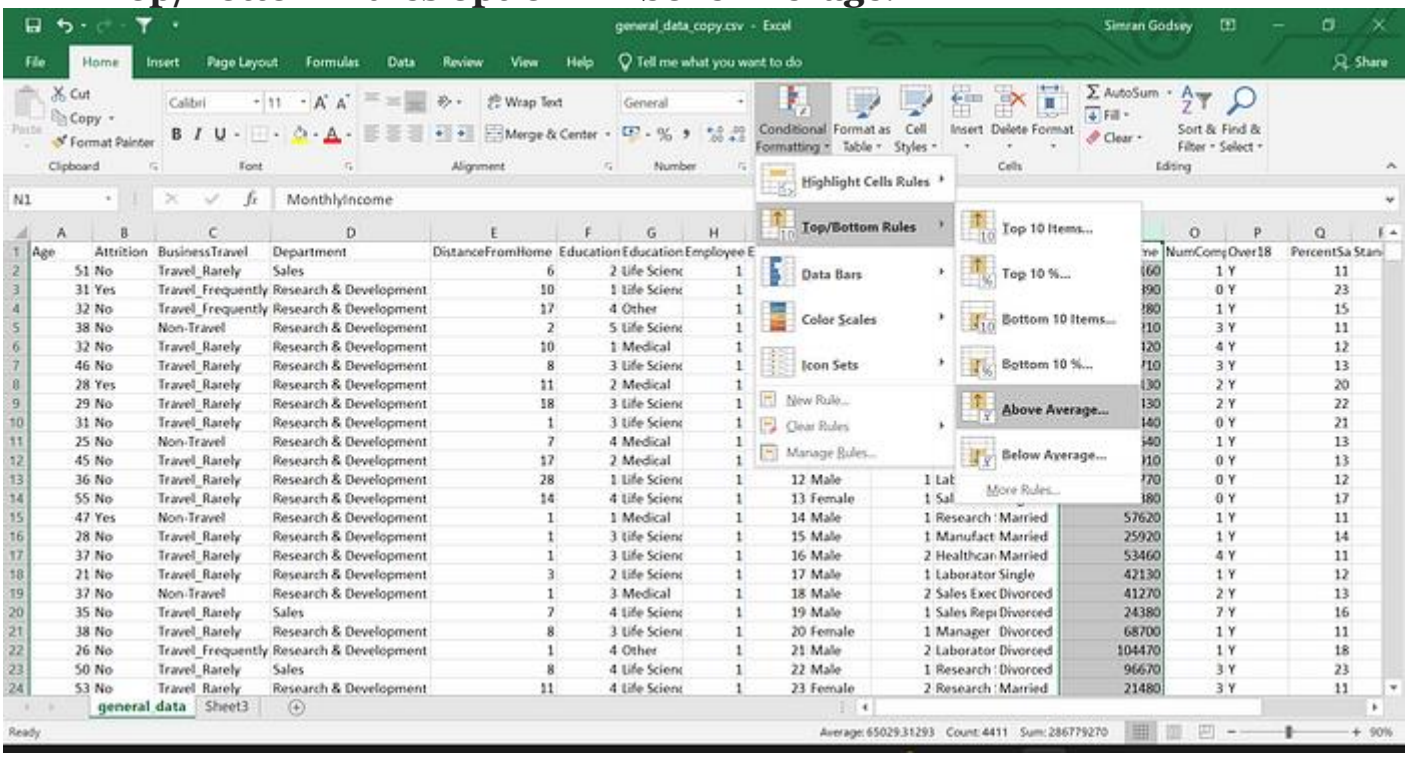
Voila! I have summarized the **Average Monthly Income** by **Job Role**

Job Role	Avg Monthly Income
Healthcare Representative	60983.74
Human Resources	58528.08
Laboratory Technician	66314.05
Manager	63395.88
Manufacturing Director	69183.72
Research Director	65473.13
Research Scientist	64975.68
Sales Executive	65186.69
Sales Representative	65370.96
Grand Total	**65029.31**

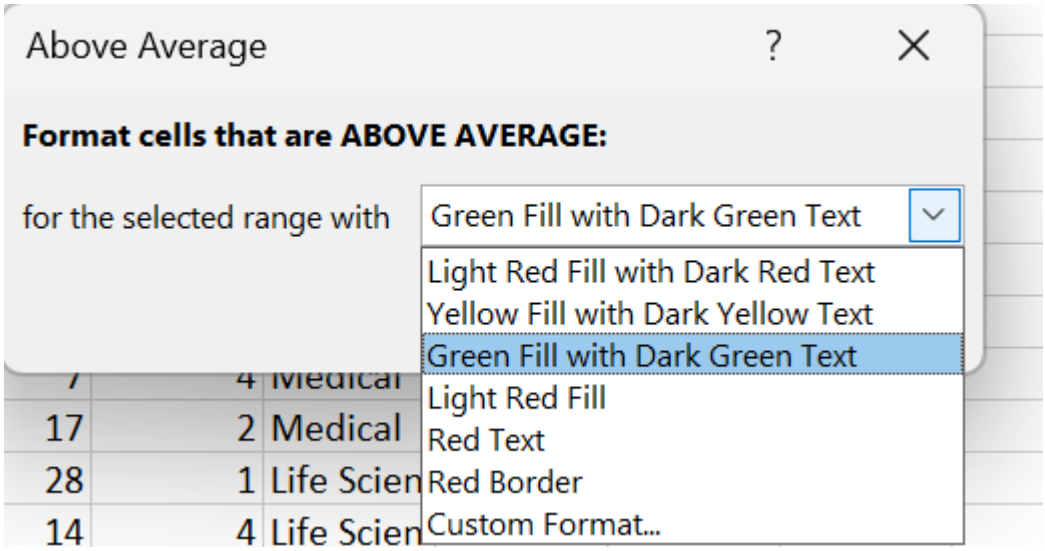
Avg Monthly Income By Job Role

3. Apply conditional formatting to highlight employees with Monthly Income above the company's average income.

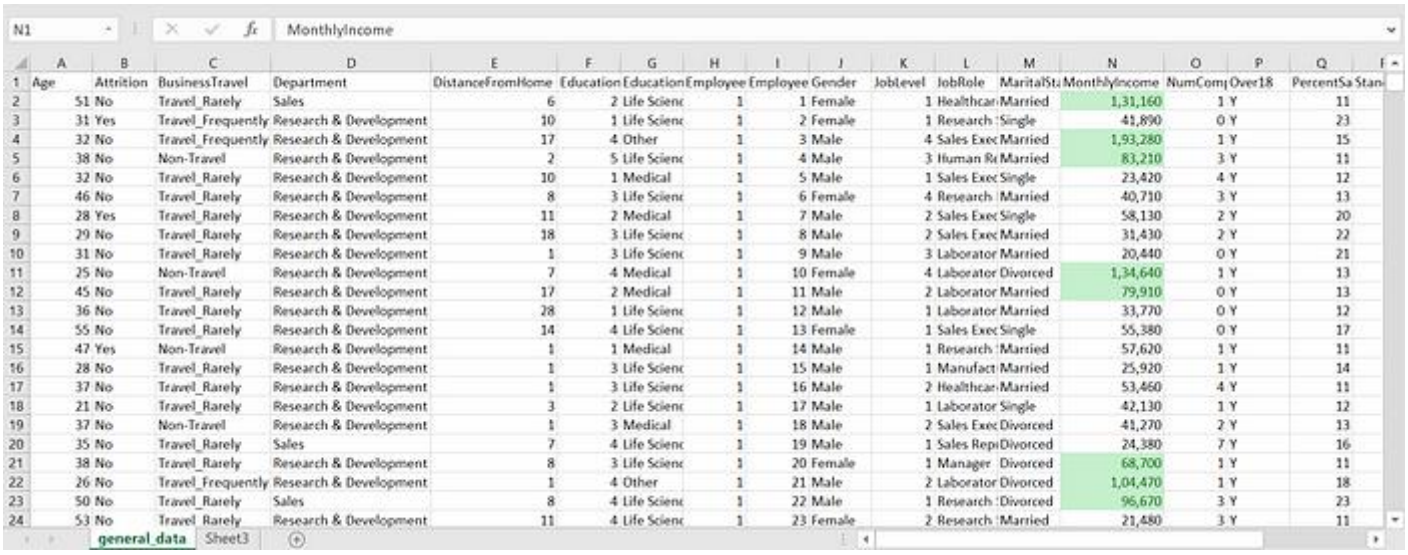
First Select the Monthly Income Column and **Go to the Home Tab in Ribbon > Click on the Conditional Formatting option > Top/Bottom Rules option > Above Average**.



Then Above Average box will open. Select the color range you want. Here I have selected Green Color and then clicked ok.



Here I have Changed the formatting of Values from General to Number. By clicking **Ctrl + Shift + 1**.



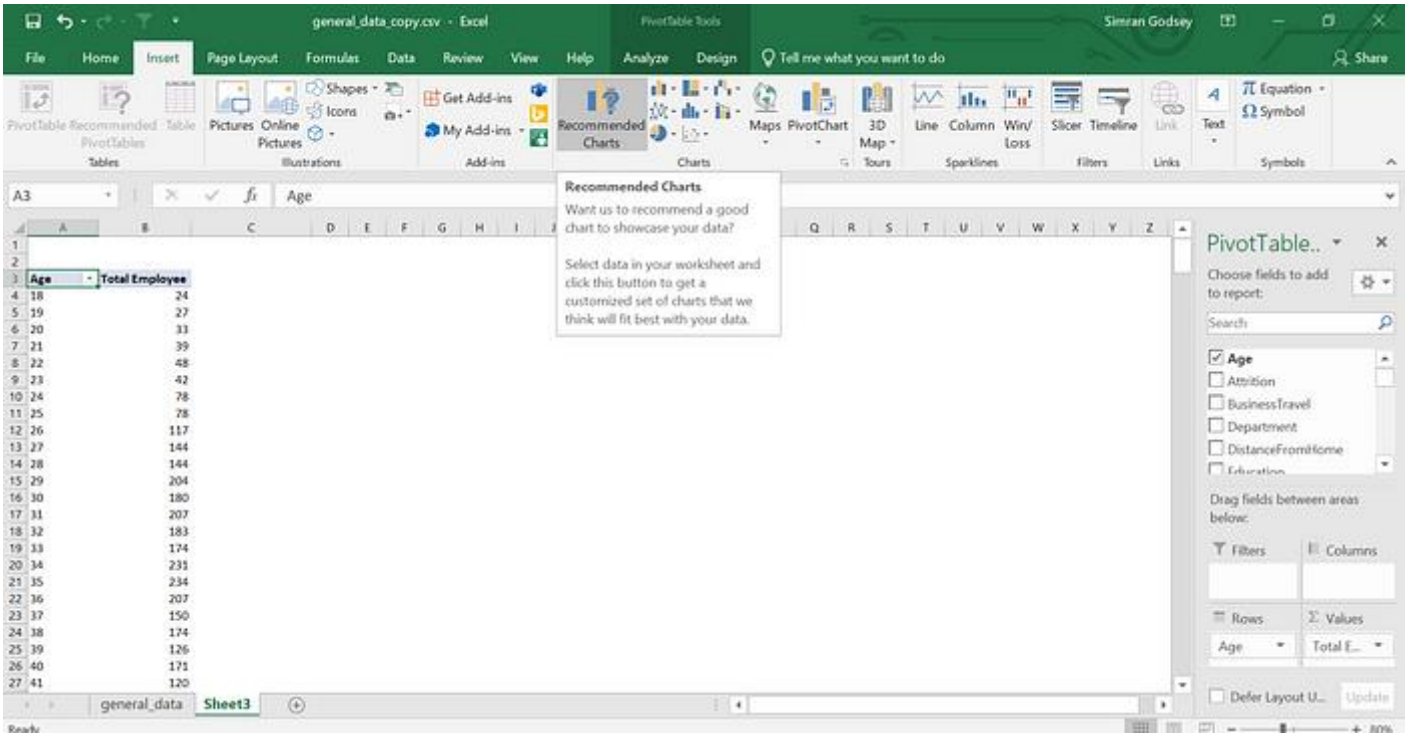
Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationEmployee	Employee	Gender	JobLevel	JobRole	MaritalSt	MonthlyIncome	NumCom	Over18	PercentSa	Stan
51	No	Travel_Rarely	Sales	6	2	Life Scienc	1	1	Female	1	Healthcar	Married	1,31,160	1	Y	11	
31	Yes	Travel_Frequently	Research & Development	10	1	Life Scienc	1	2	Female	1	Research	Single	41,890	0	Y	23	
32	No	Travel_Frequently	Research & Development	17	4	Other	1	3	Male	4	Sales Exec	Married	1,93,280	1	Y	15	
38	No	Non-Travel	Research & Development	2	5	Life Scienc	1	4	Male	3	Human Re	Married	83,210	3	Y	11	
32	No	Travel_Rarely	Research & Development	10	1	Medical	1	5	Male	1	Sales Exec	Single	23,420	4	Y	12	
46	No	Travel_Rarely	Research & Development	8	3	Life Scienc	1	6	Female	4	Research	Married	40,710	3	Y	13	
28	Yes	Travel_Rarely	Research & Development	11	2	Medical	1	7	Male	2	Sales Exec	Single	58,130	2	Y	20	
29	No	Travel_Rarely	Research & Development	18	3	Life Scienc	1	8	Male	2	Sales Exec	Married	31,430	2	Y	22	
31	No	Travel_Rarely	Research & Development	1	3	Life Scienc	1	9	Male	3	Laborator	Married	20,440	0	Y	21	
25	No	Non-Travel	Research & Development	7	4	Medical	1	10	Female	4	Laborator	Divorced	1,34,640	1	Y	13	
45	No	Travel_Rarely	Research & Development	17	2	Medical	1	11	Male	2	Laborator	Married	79,910	0	Y	13	
36	No	Travel_Rarely	Research & Development	28	1	Life Scienc	1	12	Male	1	Laborator	Married	33,770	0	Y	12	
55	No	Travel_Rarely	Research & Development	14	4	Life Scienc	1	13	Female	1	Sales Exec	Single	55,380	0	Y	17	
47	Yes	Non-Travel	Research & Development	1	1	Medical	1	14	Male	1	Research	Married	57,620	1	Y	11	
28	No	Travel_Rarely	Research & Development	1	3	Life Scienc	1	15	Male	1	Manufact	Married	25,920	1	Y	14	
37	No	Travel_Rarely	Research & Development	1	3	Life Scienc	1	16	Male	2	Healthcar	Married	53,460	4	Y	11	
21	No	Travel_Rarely	Research & Development	3	2	Life Scienc	1	17	Male	1	Laborator	Single	42,130	1	Y	12	
37	No	Non-Travel	Research & Development	1	3	Medical	1	18	Male	2	Sales Exec	Divorced	41,270	2	Y	13	
35	No	Travel_Rarely	Sales	7	4	Life Scienc	1	19	Male	1	Sales Repr	Divorced	24,380	7	Y	16	
38	No	Travel_Rarely	Research & Development	8	3	Life Scienc	1	20	Female	1	Manager	Divorced	68,700	1	Y	11	
26	No	Travel_Frequently	Research & Development	1	4	Other	1	21	Male	2	Laborator	Divorced	1,04,470	1	Y	18	
50	No	Travel_Rarely	Sales	8	4	Life Scienc	1	22	Male	1	Research	Divorced	96,670	3	Y	23	
53	No	Travel_Rarely	Research & Development	11	4	Life Scienc	1	23	Female	2	Research	Married	21,480	3	Y	11	

4. Create a bar chart in Excel to visualize the distribution of employee ages.

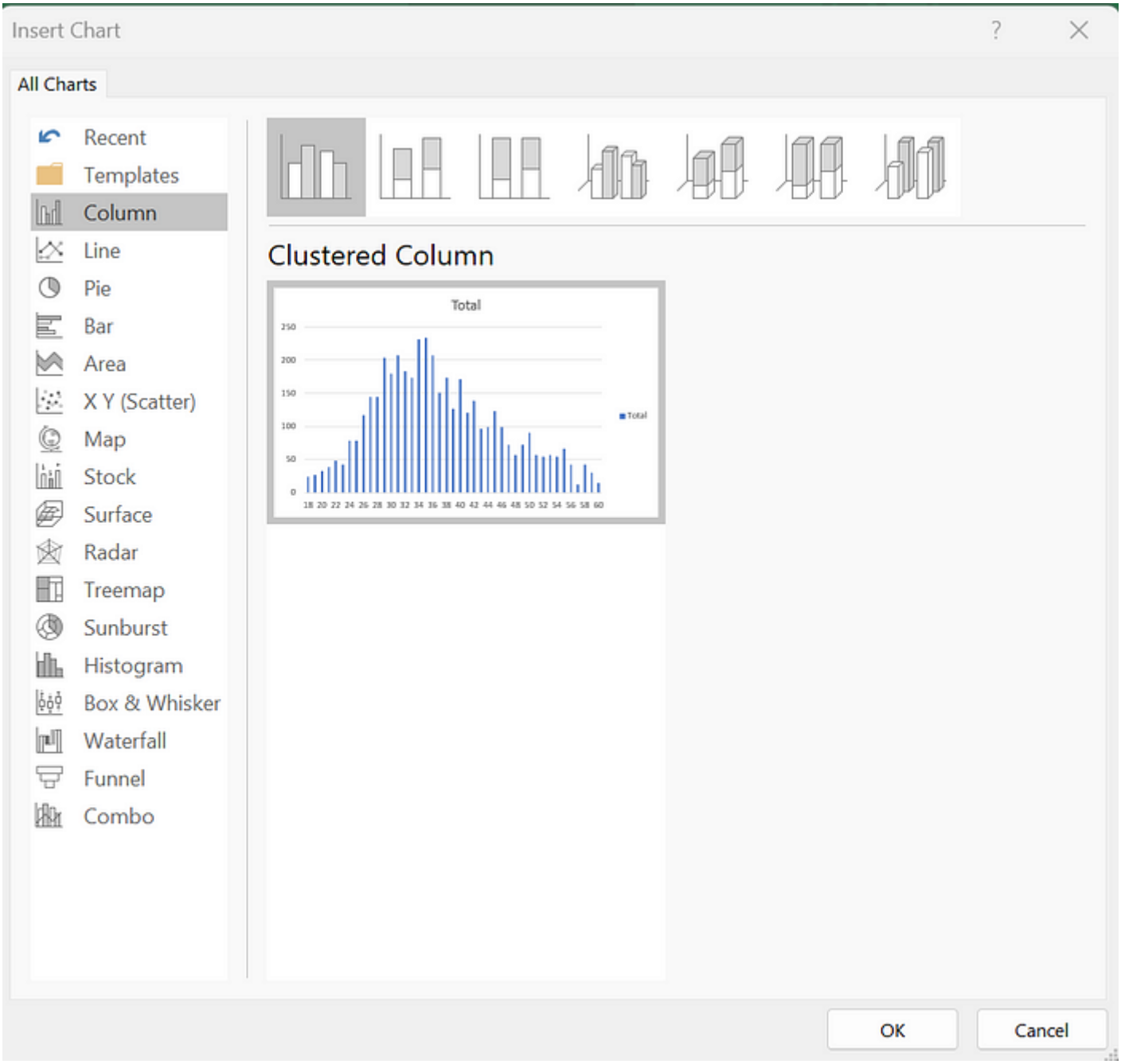
Select the **Age** and then **Employee Count** Column from the Pivot table Field.

Age	Total Employee
18	24
19	27
20	33
21	39
22	48
23	42
24	78
25	78
26	117
27	144
28	144
29	204
30	180
31	207
32	183
33	174
34	231
35	234
36	207
37	150
38	174
39	126
40	171
41	120

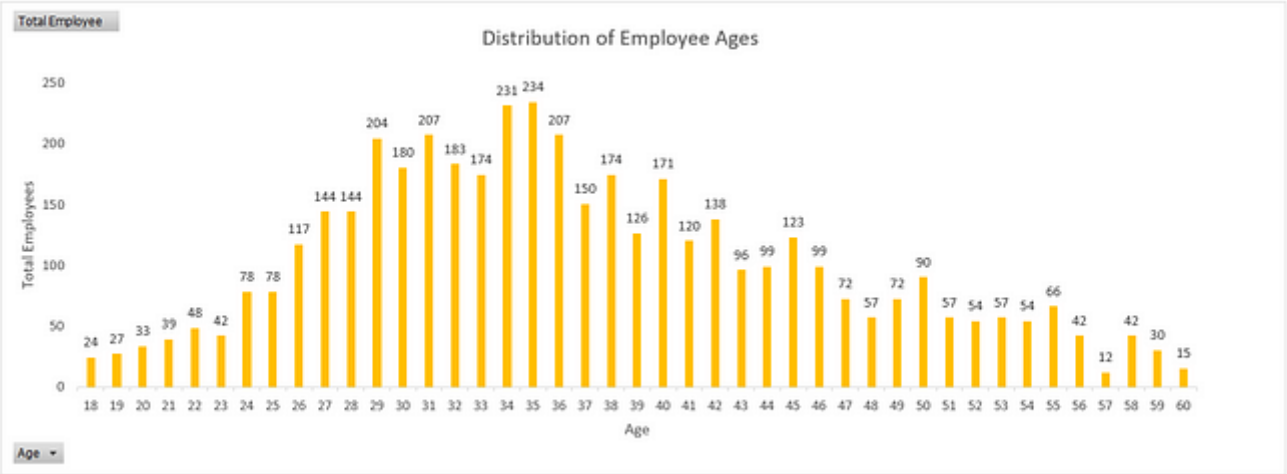
Select the Entire Pivot table and **Go to Insert tab > Recommended Charts** or you can select any chart you want.



Then Insert Chart dialogue box will open. Here you can select any chart according to your business needs. I am selecting Column Chart for now and then Clicked OK.

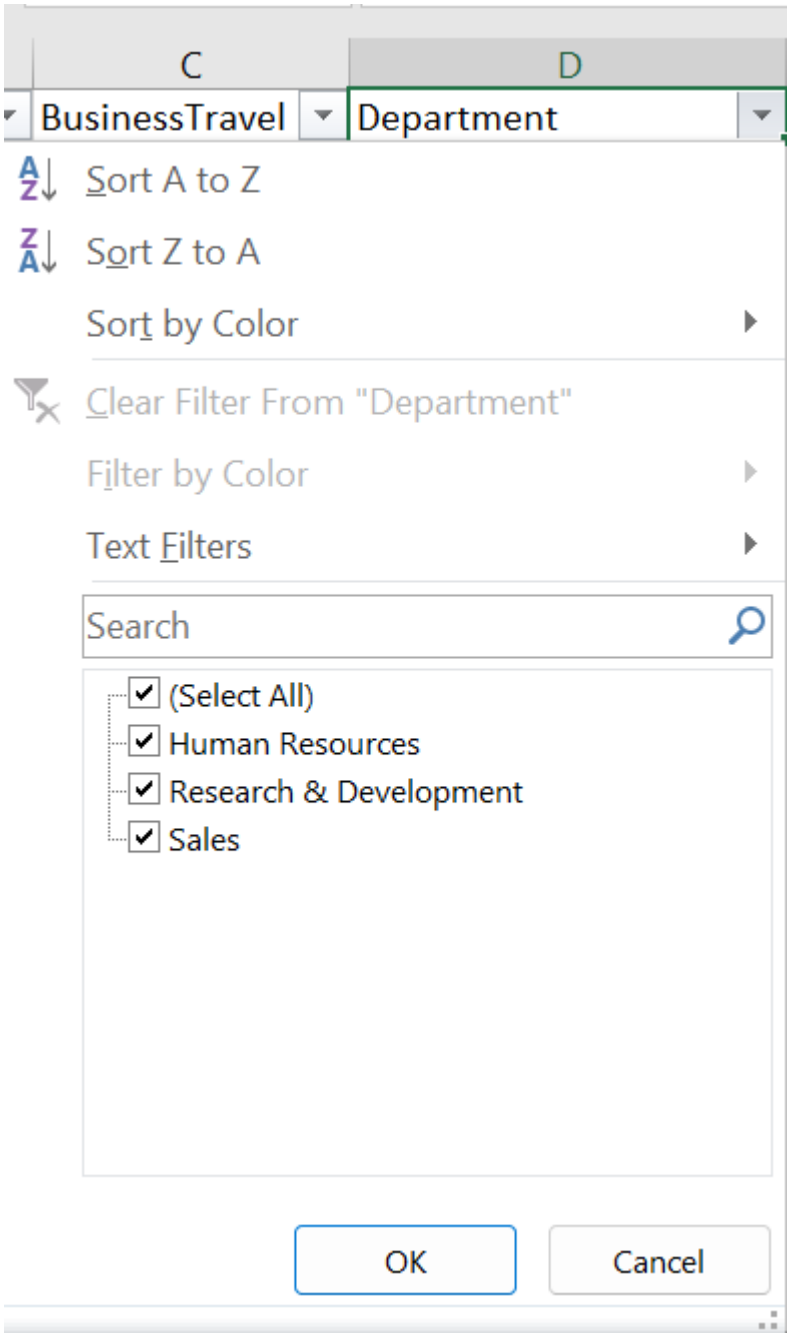


I have changed the color and added axis names, Chart Titles, and Data labels to make it visually appealing.



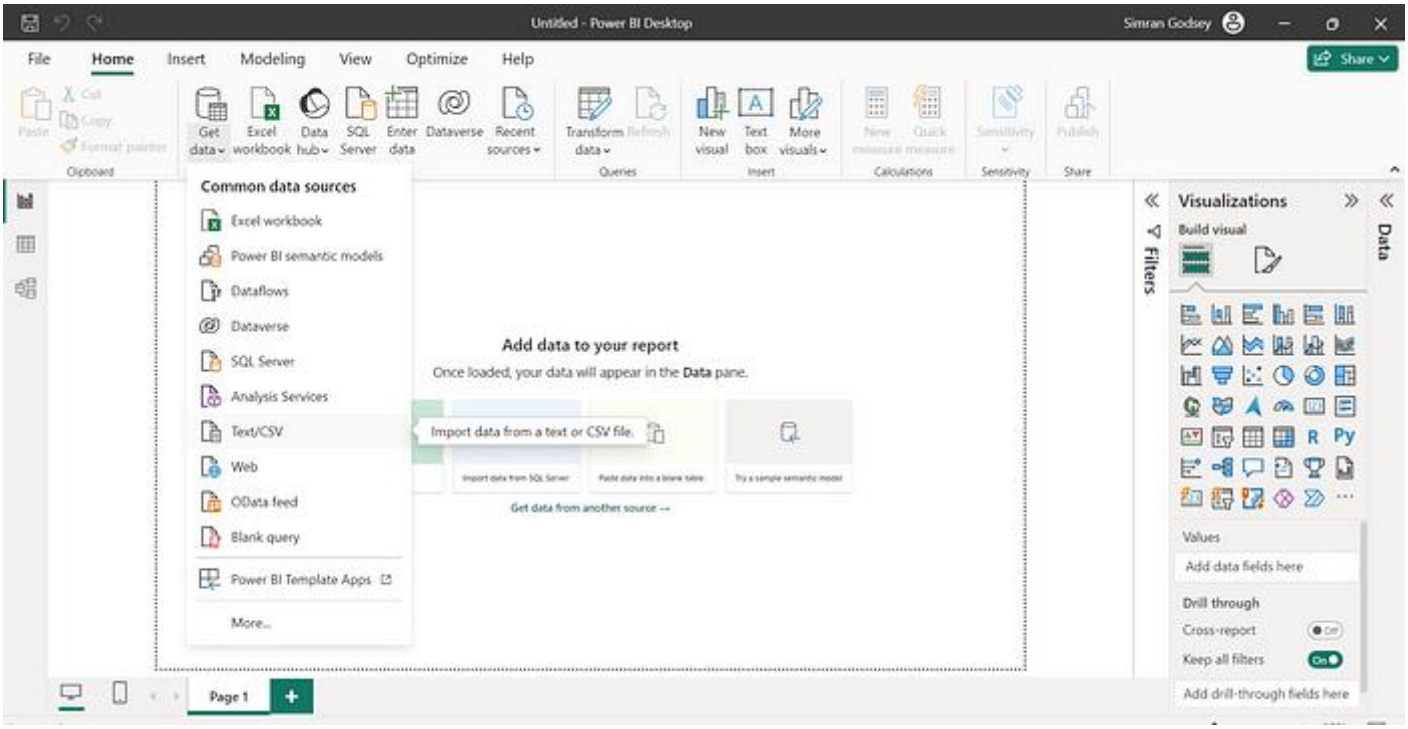
5. Identify and clean any missing or inconsistent data in the “Department” column.

After applying Filters on the Department Column. I have not seen any inconsistent data.

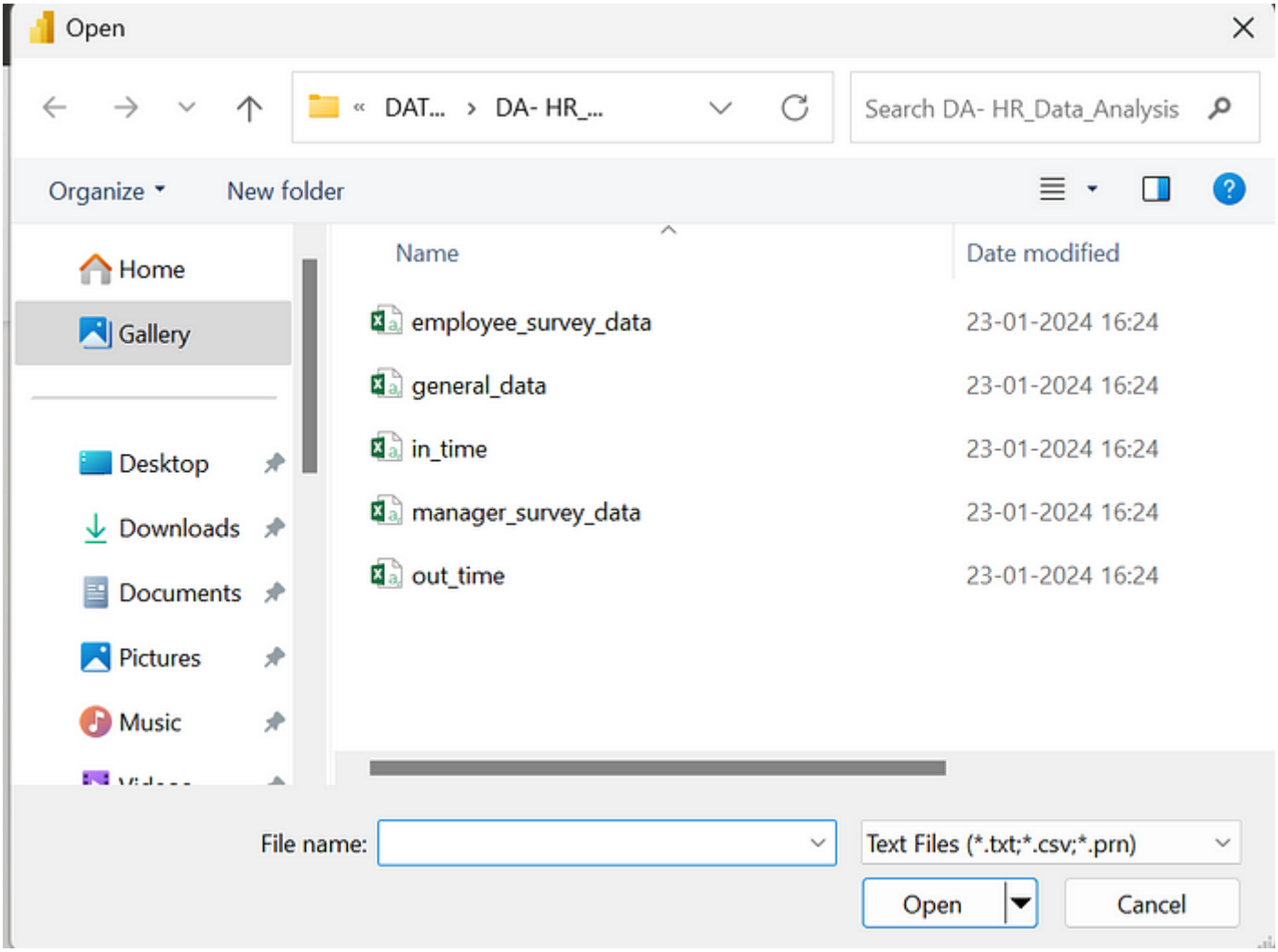


6. In Power BI, establish a relationship between the "EmployeeID" in the employee data and the "EmployeeID" in the time tracking data.

Open **Power BI desktop > Click Get Data Option in the Home tab > Select Text/CSV** because our file format is Comma Separated Value.

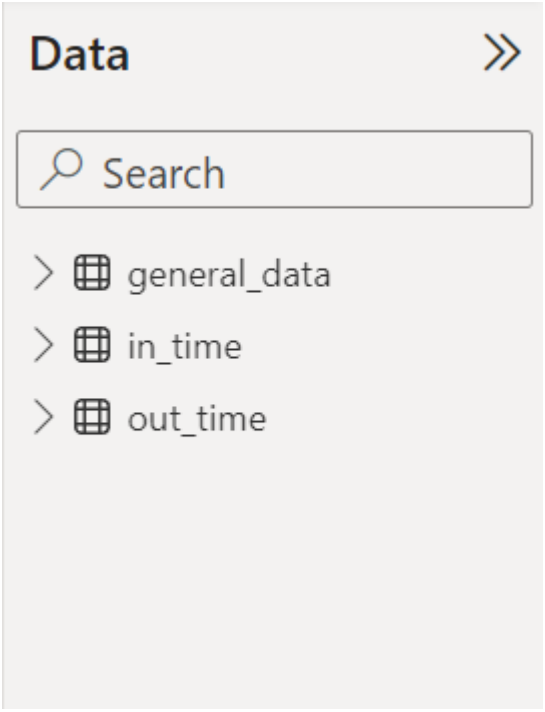


One dialogue Box will Open where you have to choose files that you want to import into Power BI one by one.

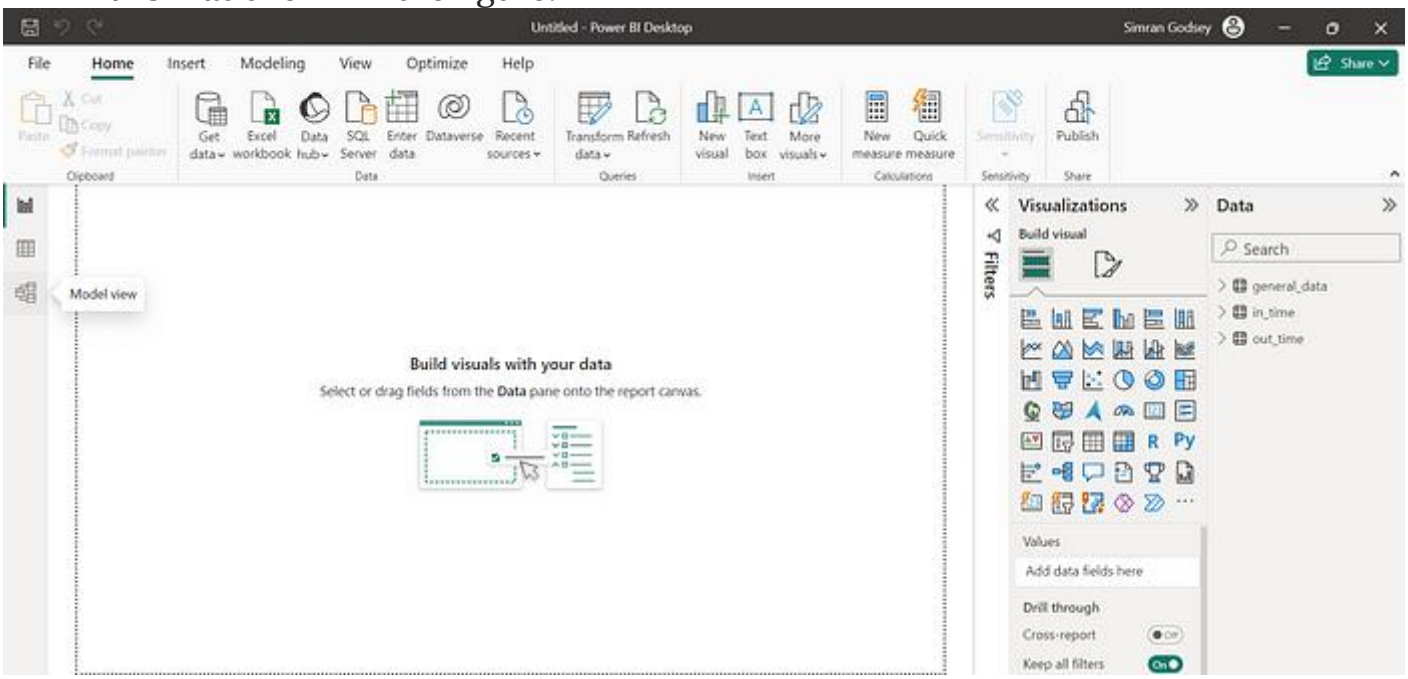


Load the data into the Power BI. By choosing the option **Load**. Here I am loading Three tables.

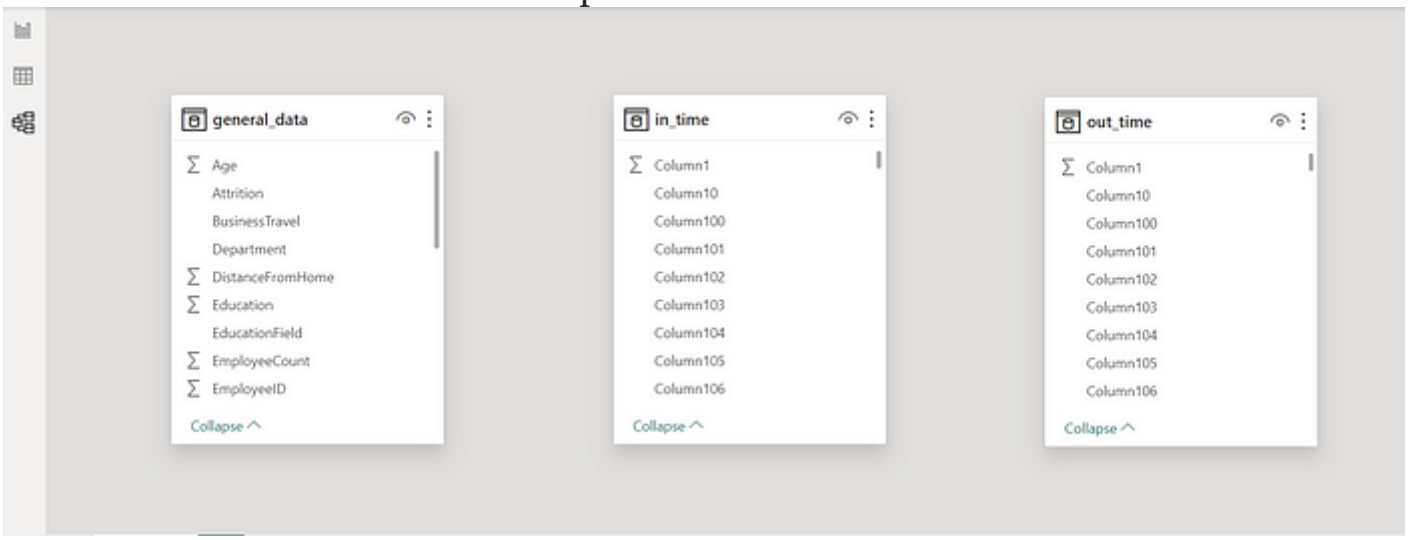
We can see in our **Data** Field we have successfully imported **three** tables.



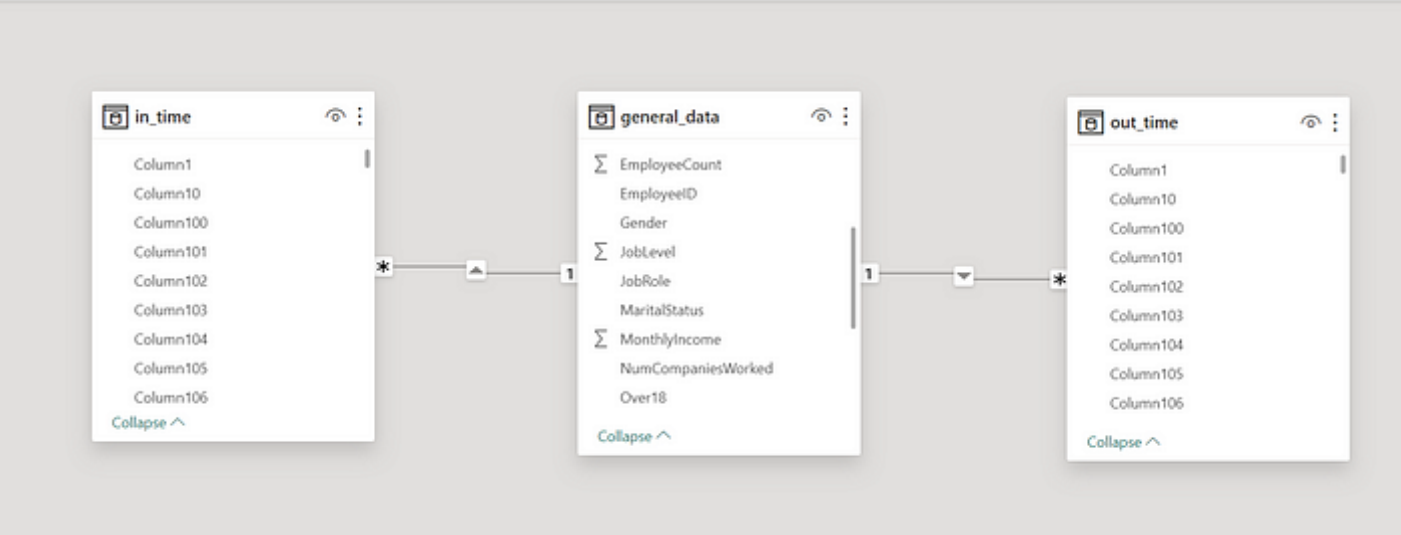
For building relationships between the tables. We have to go to the **Model View** as shown in the figure.



In the **model view**, we can see our three tables which are not connected because there is no relationship between them.

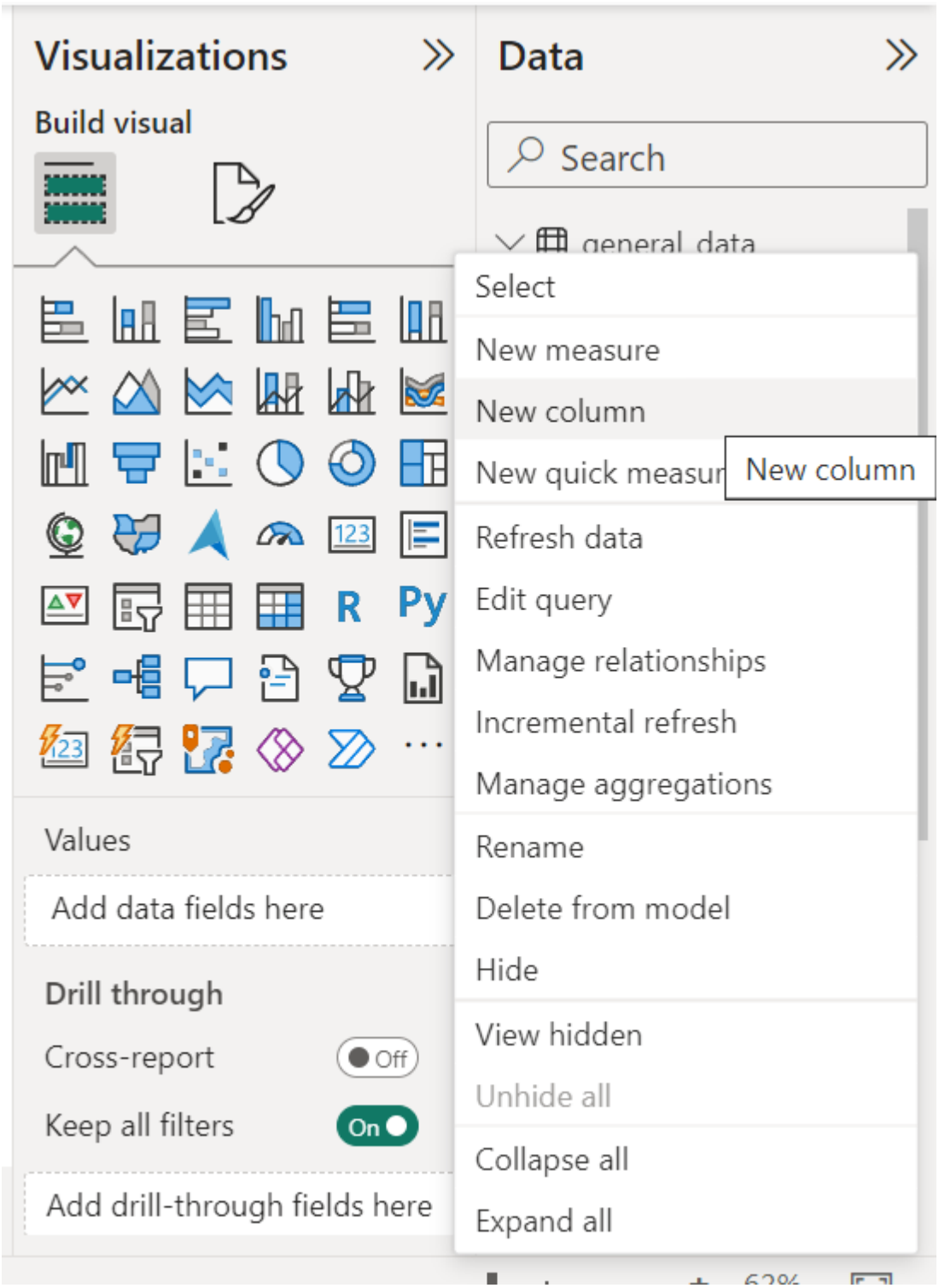


Now, I have established the relationship between the three tables. By selecting the EmployeeID present in the general_data table and dragging it into column 1 of the in_time table and the out_time table.

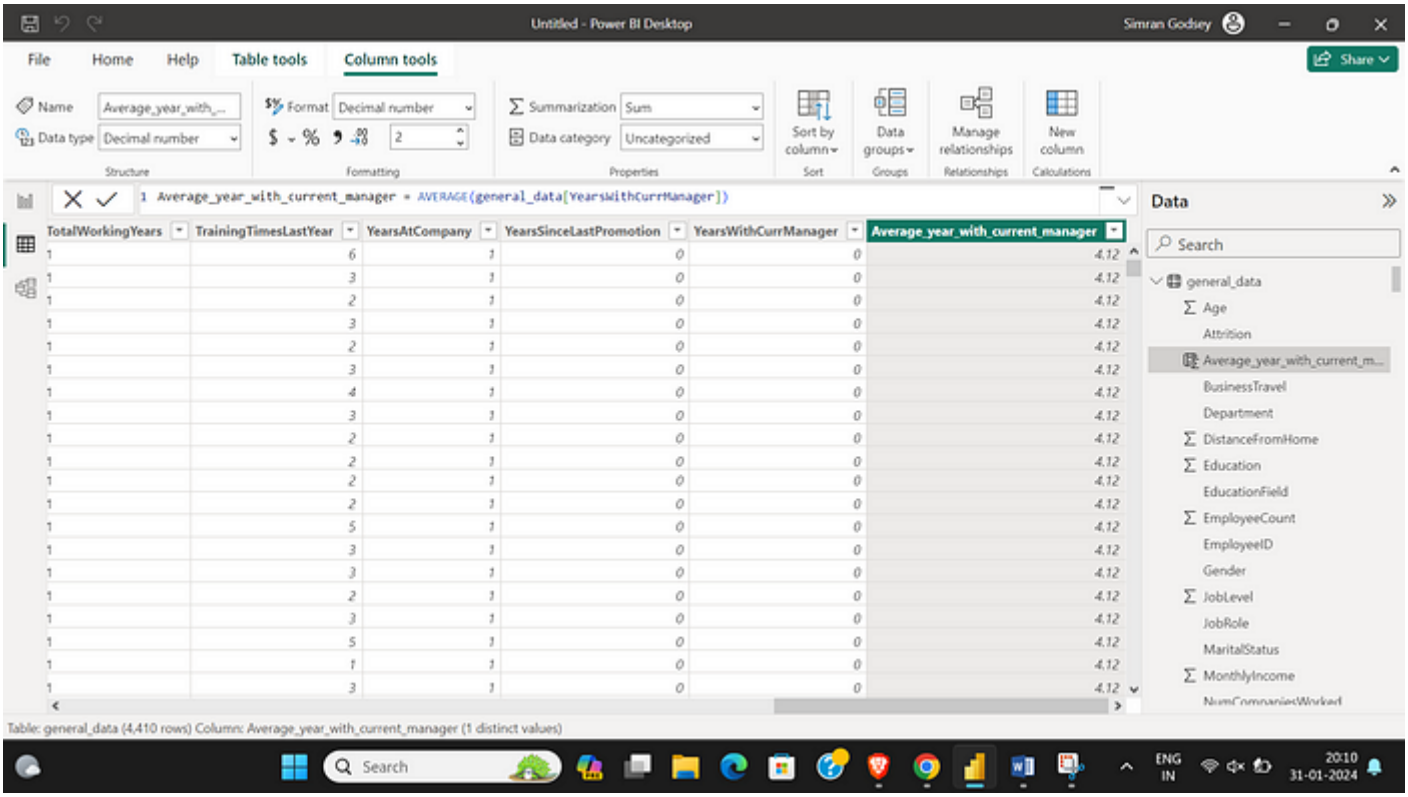


7. Using DAX, create a calculated column that calculates the average years an employee has spent with their current manager.

For creating a new calculated column. Select the **general_data** table and **right-click** on it. It will show a list of options where you have to select the **New Column**.



Then the formula bar will appear. Here I have used the **average** function to calculate the average of the column(YearWithCurrManager) and then give it a name of **Average_Year_wih_current_manager**.



8. Using Excel, create a pivot table that displays the count of employees in each Marital Status category, segmented by Department.

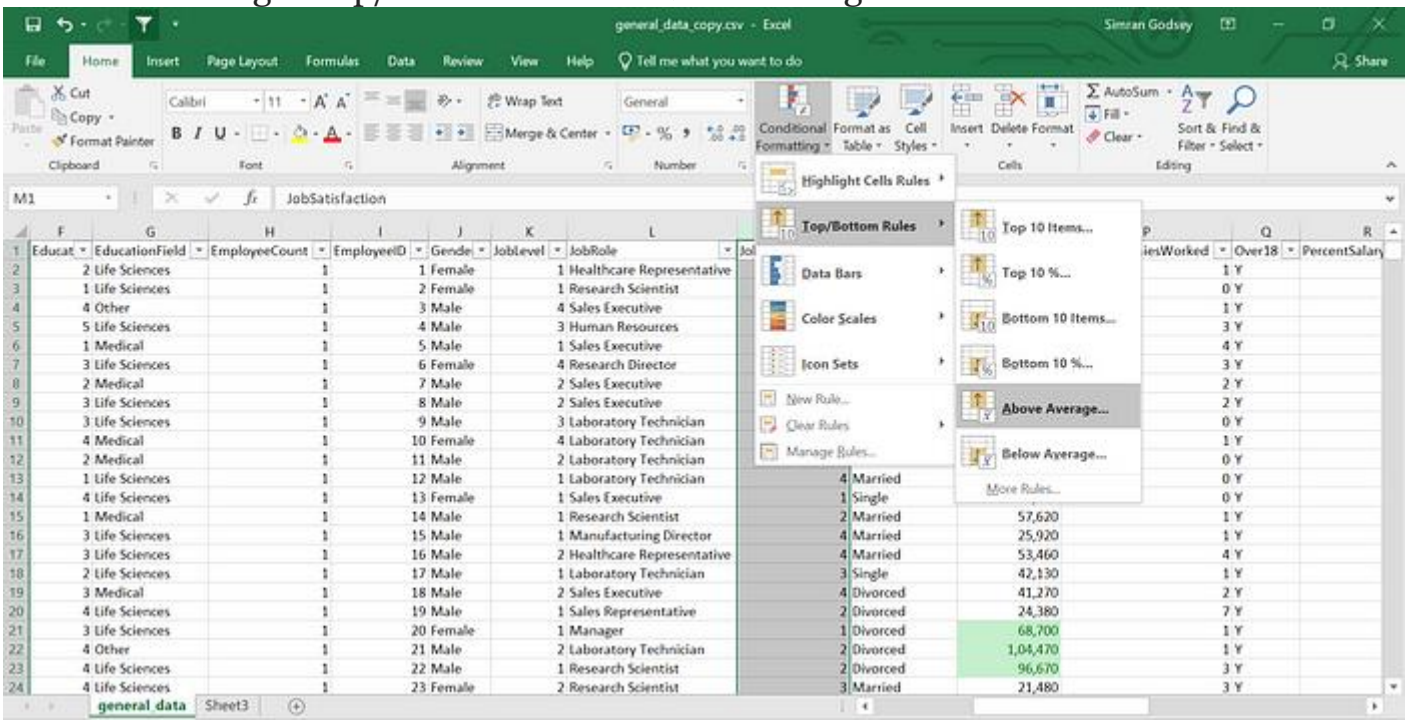
We have already created a pivot table earlier, we just have to change the **rows** and **values** field.

Select the **Department** column and put it into the **Rows** field >
Select **Marital Status** and put it into the **Column** field >
Select **Employee Count** Column and put it into the **Values** field.

Departments	Marital Status			
	Divorced	Married	Single	Grand Total
Human Resources	21	96	72	189
Research & Development	621	1350	912	2883
Sales	339	573	426	1338
Grand Total	**981**	**2019**	**1410**	**4410**

9. Apply conditional formatting to highlight employees with both above-average Monthly Income and above-average Job Satisfaction.

The job Satisfaction Column is in another workbook. So I copy-paste the Job Satisfaction column into the general_Data workbook and then Select the Job Satisfaction column > Go to the Home Tab > Select Conditional formatting > Top/Bottom Rules > Above Average.

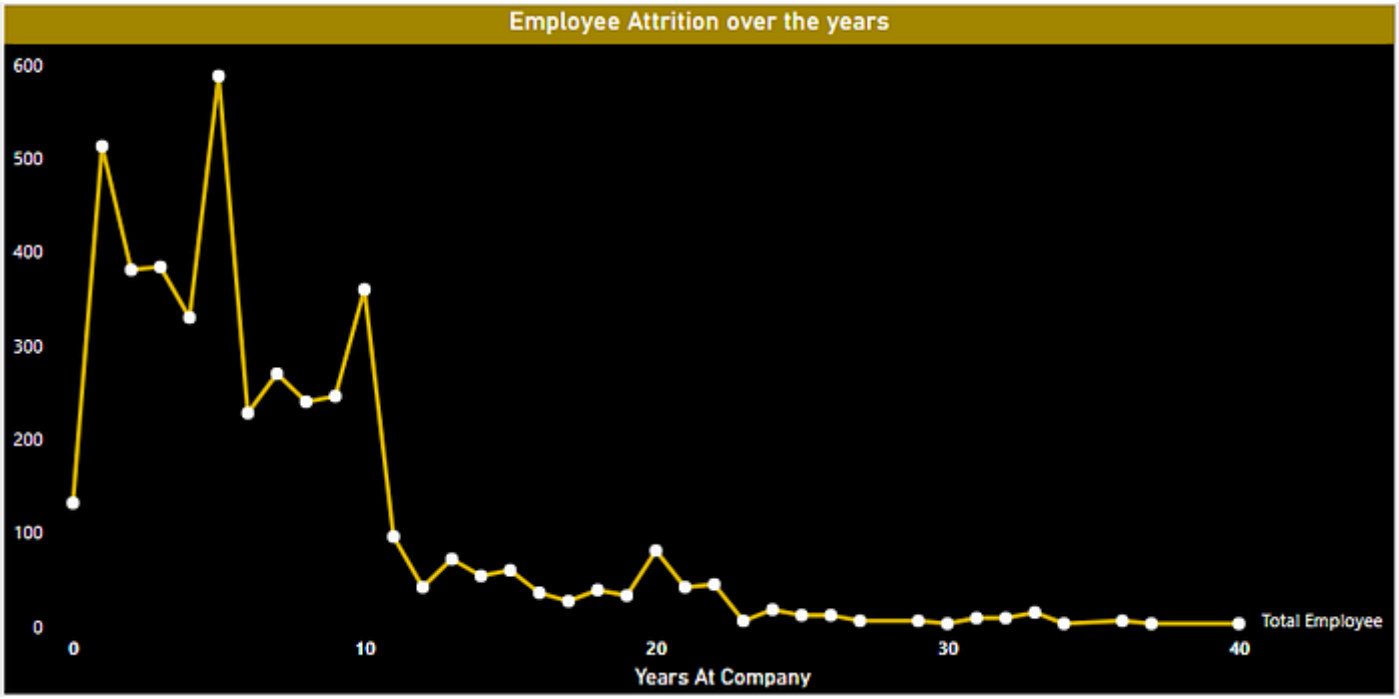


Above Average Box will appear. Select the color you want and then click ok.

M	N	O
JobSatisfaction	MaritalStatus	MonthlyIncor
4	Married	1,31,160
2	Single	41,890
2	Married	1,93,280
4	Married	83,210
1	Single	23,420
2	Married	40,710
3	Single	58,130
2	Married	31,430
4	Married	20,440
1	Divorced	1,34,640
4	Married	79,910
4	Married	33,770
1	Single	55,380
2	Married	57,620
4	Married	25,920
4	Married	53,460
3	Single	42,130
4	Divorced	41,270
2	Divorced	24,380
1	Divorced	68,700
2	Divorced	1,04,470
2	Divorced	96,670
3	Married	21,480

10. In Power BI, create a line chart that visualizes the trend of Employee Attrition over the years.

In Power BI, Select the **line chart** from the **Visualization Pane** and then Pull the **Years at company** column into the **X-axis** and **Employee ID** into the **Y-axis**. Your line chart will be ready. For **Formatting**, In the Visualization Pane Select the **Format Visual** option and Format your visual according to the business needs.



11. Describe how you would create a star schema for this dataset, explaining the benefits of doing so.

First of all, Let's understand about Star Schema.

Star schema contains one or more dimensions and fact tables. It is called a star schema because the entity-relationship diagram between dimensions and fact tables resembles a star where one fact table which is in the center is connected to multiple dimensions tables.

After understanding about star schema let's understand the fact and dimensions tables.

Fact tables: Fact tables are those tables that store quantitative data that you want to analyze, such as sales, transactions, and invoice tables.

Dimension tables: Dimension tables are those tables that store descriptive attributes that provide context for the data in the fact tables, such as customer demographics, and product information.

Steps to Create a Star Schema

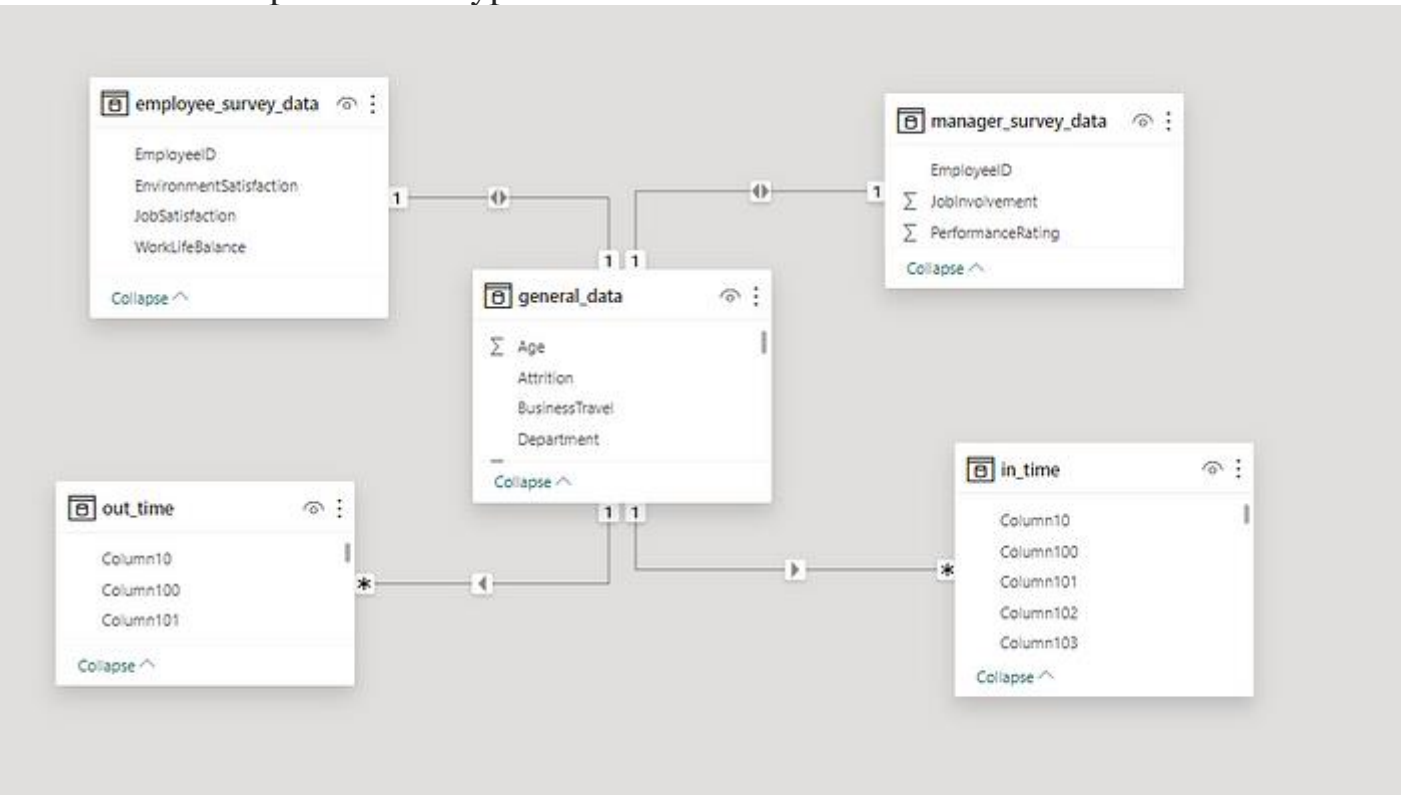
1. **Identify Fact tables and Dimensions table:** The dataset provided by the psyliq are general_data, manager_survey_data, employee_survey_data, in_time, out_time.

Here general_data is the fact table and other data are the dimensions table.

2. **Establish Relationship:** Go to the model view, where you can see all the tables imported into the Power BI. Drag the key column from the dimensions table into the Foreign key of the fact table.

3. **Configure Relationships:** Double-click on the relationship lines to open the Manage Relationships dialog. Make sure that the correct fields are matched. Choose the relationship type (single, both, or none). Set the cross-filter direction based on your analysis needs.

4. **Validate Data Model:** Switch to the Data view and ensure that relationships and data types are correct.



Benefits of creating Star Schema

1. **Simplicity:** The simple structure of star schemas makes easy to understand for both technical and non-technical users.
2. **Performance:** Star schemas are optimized for querying large datasets, making them ideal for data analysis.
3. **Scalability:** Star schemas can be easily extended to add new dimension tables or measures to the fact table.
4. **Flexible:** Star schemas can be used to model a wide variety of business data.

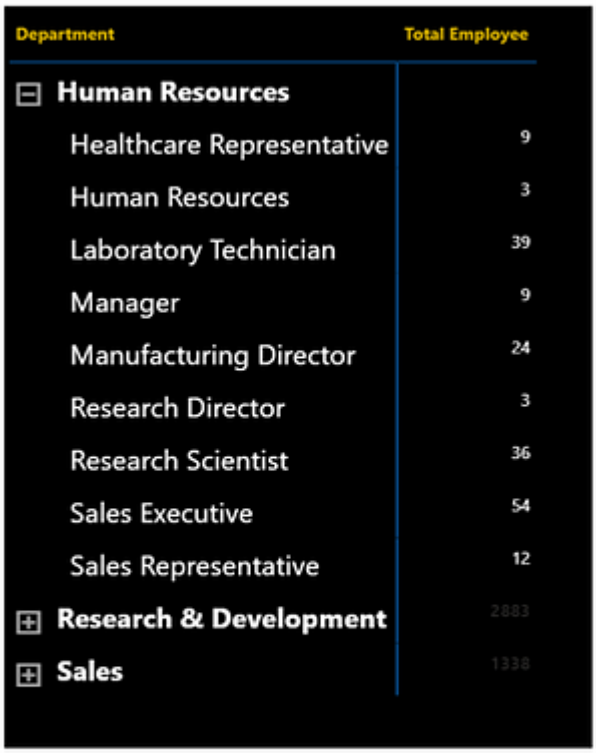
12. Using DAX, calculate the rolling 3-month average of Monthly Income for each employee

Rolling Avg = CALCULATE(
AVERAGE('YourTableName'[MonthlyIncome]), DATESINPERIOD(
'YourTableName'[Date], LASTDATE('YourTableName'[Date]), -3, MONTH
))

13. Create a hierarchy in Power BI that allows users to drill down from Department to Job Role to further narrow their analysis.

To create a hierarchy, I have used a Matrix visual. Matrix visual is best to drill down into the details of your data by clicking on specific cells. This allows you to explore deeper levels of analysis and identify trends within specific categories.

Here I have put the Department column first in the rows field and then added the JobRole Column for creating a hierarchy and then I put the Employee ID in the values field.



15. In Excel, calculate the total Monthly Income for each Department, considering only the employees with a Job Level greater than or equal to 3.

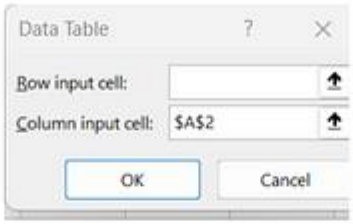
Go to the pivot table > Place Job Level Column in the rows field > Department in the column field > Monthly Income in the values field.

Job Level	Department			
	Human Resources	Research & Development	Sales	Grand Total
3	16,48,500	2,81,17,740	1,17,92,400	4,15,58,640
4	7,54,800	1,52,77,290	87,53,070	2,47,85,160
5	8,55,840	1,01,07,870	24,28,860	1,33,92,570
Grand Total	**32,59,140**	**5,35,02,900**	**2,29,74,330**	**7,97,36,370**

16. Explain how to perform a What-If analysis in Excel to understand the impact of a 10% increase in Percent Salary Hike on Monthly Income.

Create a new sheet and Copy the Monthly Income column in it. Create a new column named Percent_Hike and calculate the salary hike. The formula for a salary hike is (=A2*(1+10%). Then select the Monthly Income and Percent_hike Column and Go to Data Tab > What-if Analysis > select Data table > Do not give any Row input > In column input, select cell A2 and then Click OK.

MonthlyIncome	Percent_hike
131160	=A2*(1+10%)
41890	
193280	
83210	
23420	
40710	
58130	
31430	
20440	
134640	
79910	
33770	
55380	
57620	
25920	
53460	
42130	
41270	
24380	
68700	
104470	
96670	
21480	
89260	
65130	



MonthlyIncome	Percent_hike
131160	144276
41890	46079
193280	212608
83210	91531
23420	25762
40710	44781
58130	63943
31430	34573
20440	22484
134640	148104
79910	87901
33770	37147
55380	60918
57620	63382
25920	28512
53460	58806
42130	46343
41270	45397
24380	26818
68700	75570
104470	114917
96670	106337
21480	23628
89260	98186
65130	71643
67990	74789
162910	179201
27050	29755

17. Verify if the data adheres to a predefined schema. What actions would you take if you found inconsistencies?

Yes, the Data adheres to a predefined schema.

Action to take when Inconsistent are found:

1. Check the Data normalization, if not normalize the dataset.
2. Check the dataset's column and its data type.
3. Check the relationship between the tables.
4. Transform the dataset if any inconsistent data is found.
5. Replace the null values with 0 or mean or according to the client.

Thank you for reading. Feel free to share your thoughts, ask questions, and let's continue this journey together.

Additionally, if you'd like to connect with me, you can find my LinkedIn ID and GitHub profile below.

LinkedIn
GitHub

Happy learning!