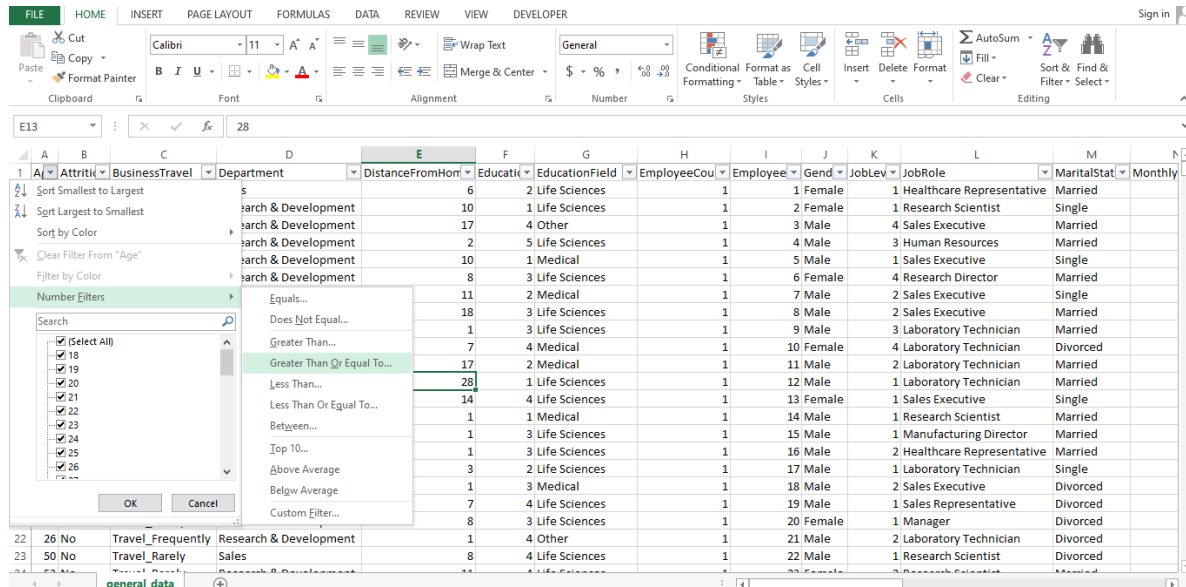


# **HR DATA ANALYSIS**

**First, I'll manage my table...the spaces, for that will  
click: Ctrl + A, Alt + HOA and Alt + HOI**

# 1. Apply Filter first by clicking Ctrl + Shift + L

From, Arrow in front of column Age,  
Select, Number Filters, Select greater than or equal to  
and Mention 30 there,




We'll get the desired output i.e. only the rows where  
age >= 30

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Attrition	BusinessTravel	Department	DistanceFromHome	EducationField	EmployeeCount	Employee	Gender	JobLevel	JobRole	MaritalStatus		
2	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	1	Female	1	Healthcare Representative	Married
3	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1	2	Female	1	Research Scientist	Single
4	32	No	Travel_Frequently	Research & Development	17	4	Other	1	3	Male	4	Sales Executive	Married
5	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	4	Male	3	Human Resources	Married
6	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	5	Male	1	Sales Executive	Single
7	46	No	Travel_Rarely	Research & Development	8	3	Life Sciences	1	6	Female	4	Research Director	Married
10	31	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	9	Male	3	Laboratory Technician	Married
12	45	No	Travel_Rarely	Research & Development	17	2	Medical	1	11	Male	2	Laboratory Technician	Married
13	36	No	Travel_Rarely	Research & Development	28	1	Life Sciences	1	12	Male	1	Laboratory Technician	Married
14	55	No	Travel_Rarely	Research & Development	14	4	Life Sciences	1	13	Female	1	Sales Executive	Single
15	47	Yes	Non-Travel	Research & Development	1	1	Medical	1	14	Male	1	Research Scientist	Married
17	37	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	16	Male	2	Healthcare Representative	Married
19	37	No	Non-Travel	Research & Development	1	3	Medical	1	18	Male	2	Sales Executive	Divorced
20	35	No	Travel_Rarely	Sales	7	4	Life Sciences	1	19	Male	1	Sales Representative	Divorced
21	38	No	Travel_Rarely	Research & Development	8	3	Life Sciences	1	20	Female	1	Manager	Divorced
23	50	No	Travel_Rarely	Sales	8	4	Life Sciences	1	22	Male	1	Research Scientist	Divorced
24	53	No	Travel_Rarely	Research & Development	11	4	Life Sciences	1	23	Female	2	Research Scientist	Married
25	42	No	Travel_Rarely	Research & Development	4	4	Life Sciences	1	24	Male	1	Manufacturing Director	Married
27	55	No	Travel_Rarely	Research & Development	1	4	Other	1	26	Female	1	Research Scientist	Married
29	37	No	Travel_Rarely	Sales	5	1	Marketing	1	28	Male	1	Research Scientist	Single
30	44	Yes	Travel_Frequently	Research & Development	1	2	Medical	1	29	Male	2	Research Scientist	Divorced
31	38	No	Travel_Rarely	Sales	2	3	Marketing	1	30	Female	1	Manager	Divorced

## 2. Average monthly income by Job Role

I've done with the help of pivot table in which I've kept JobRole in rows and Average of MonthlyIncome in Values.

**PivotTable Fields** ▾ ×

Choose fields to add to report: 

- ☐ EmployeeID
- ☐ Gender
- ☐ JobLevel
- ☐ JobRole
- ☐ MaritalStatus
- ☒ MonthlyInco...
- ☐ NumCompani...
- ☐ Over18
- ☐ PercentSalary...
- ☐ StandardHours
- ☐ StockOptionL...
- ☐ TotalWorking...
- ☐ TrainingTimes...
- ☐ YearsAtComp...
- ☐ YearsSinceLast...
- ☐ YearsWithCurr...
- ☐ EnvironmentS...
- ☐ JobSatisfaction
- ☐ WorkLifeBalan...

MORE TABLES... ▾

Drag fields between areas below:

**FILTERS**

**COLUMNS**

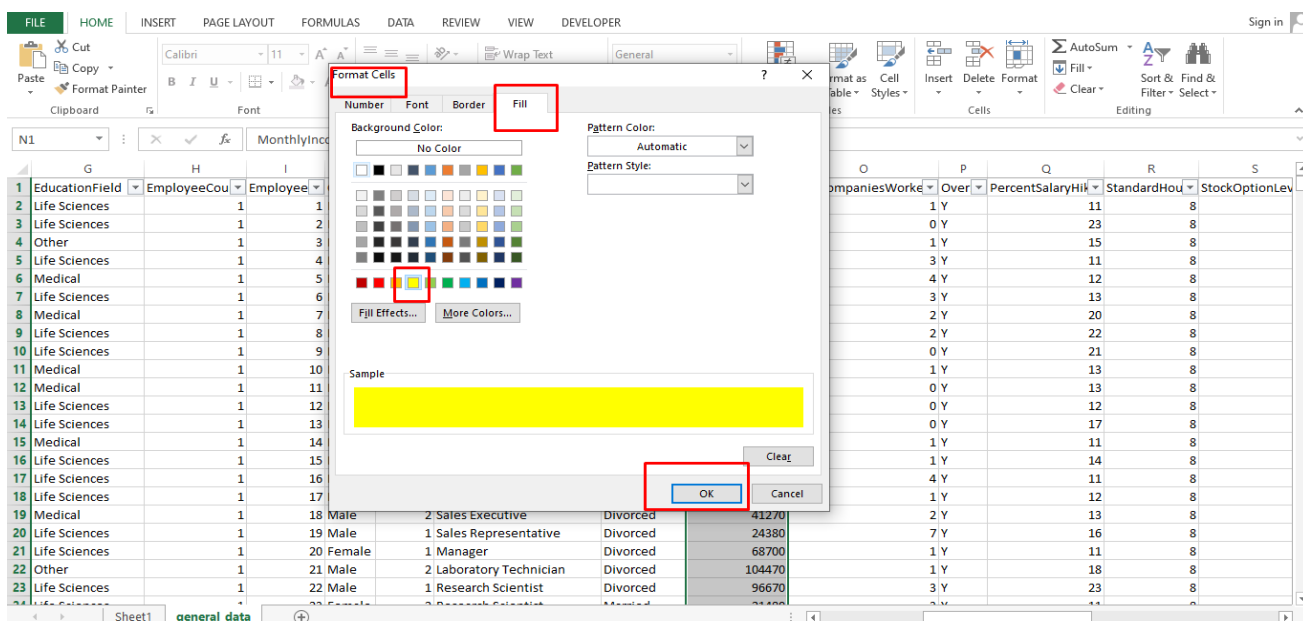
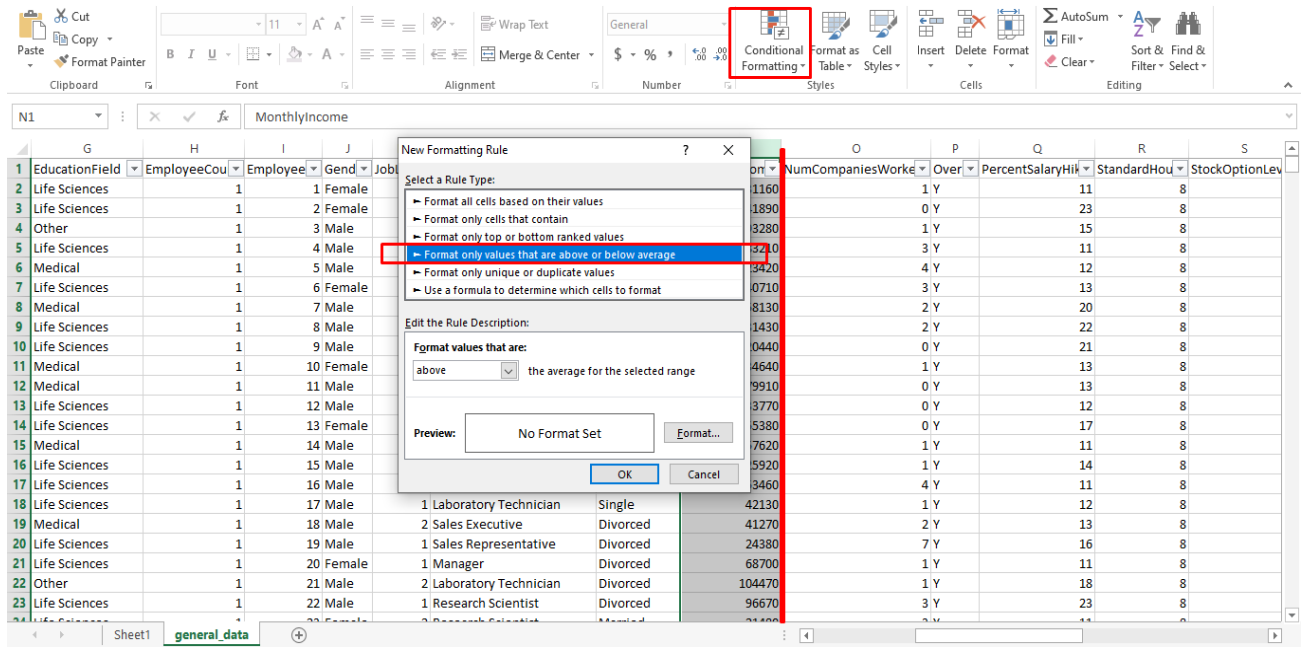
**ROWS**  
Age ▾

**VALUES**  
Σ Average of ... ▾

☐ De... **UPDATE**

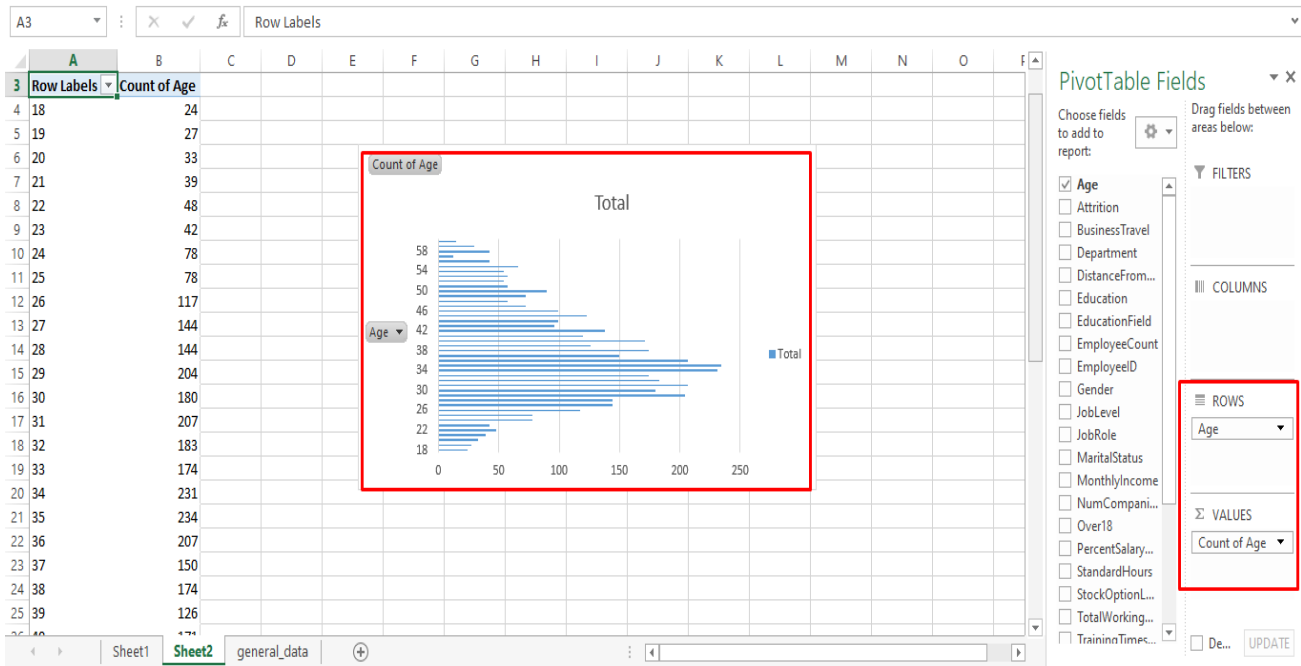
	A	B	C
1			
2			
3	Row Labels ▾	Average of MonthlyIncome	
4	Healthcare Representative	60983.74046	
5	Human Resources	58528.07692	
6	Laboratory Technician	66314.05405	
7	Manager	63395.88235	
8	Manufacturing Director	69183.72414	
9	Research Director	65473.125	
10	Research Scientist	64975.68493	
11	Sales Executive	65186.68712	
12	Sales Representative	65370.96386	
13	Grand Total	65029.31293	
14			

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



O13													
	G	H	I	J	K	L	M	N	O	P	Q	R	S
	EducationField	EmployeeCou	Employee	Gender	JobLev	JobRole	MaritalStat	MonthlyIncom	NumCompaniesWorked	Over18	PercentSalaryHike	StandardHours	StockOptionLevel
1	Life Sciences	1	1	Female	1	Healthcare Representative	Married	131160	1	Y	11	8	
2	Life Sciences	1	2	Female	1	Research Scientist	Single	41890	0	Y	23	8	
3	Other	1	3	Male	4	Sales Executive	Married	193280	1	Y	15	8	
4	Life Sciences	1	4	Male	3	Human Resources	Married	83210	3	Y	11	8	
5	Medical	1	5	Male	1	Sales Executive	Single	23420	4	Y	12	8	
6	Life Sciences	1	6	Female	4	Research Director	Married	40710	3	Y	13	8	
7	Medical	1	7	Male	2	Sales Executive	Single	58130	2	Y	20	8	
8	Life Sciences	1	8	Male	2	Sales Executive	Married	31430	2	Y	22	8	
9	Life Sciences	1	9	Male	3	Laboratory Technician	Married	20440	0	Y	21	8	
10	Medical	1	10	Female	4	Laboratory Technician	Divorced	134640	1	Y	13	8	
11	Medical	1	11	Male	2	Laboratory Technician	Married	79910	0	Y	13	8	
12	Life Sciences	1	12	Male	1	Laboratory Technician	Married	33770	0	Y	12	8	
13	Life Sciences	1	13	Female	1	Sales Executive	Single	55380	0	Y	17	8	
14	Medical	1	14	Male	1	Research Scientist	Married	57620	1	Y	11	8	
15	Life Sciences	1	15	Male	1	Manufacturing Director	Married	25920	1	Y	14	8	
16	Life Sciences	1	16	Male	2	Healthcare Representative	Married	53460	4	Y	11	8	
17	Life Sciences	1	17	Male	1	Laboratory Technician	Single	42130	1	Y	12	8	
18	Medical	1	18	Male	2	Sales Executive	Divorced	41270	2	Y	13	8	
19	Life Sciences	1	19	Male	1	Sales Representative	Divorced	24380	7	Y	16	8	
20	Life Sciences	1	20	Female	1	Manager	Divorced	68700	1	Y	11	8	
21	Other	1	21	Male	2	Laboratory Technician	Divorced	104470	1	Y	18	8	
22	Life Sciences	1	22	Male	1	Research Scientist	Divorced	96670	3	Y	23	8	
23	Life Sciences	1	23	Female	3	Research Scientist	Married	23420	4	Y	12	8	

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



## 5. Identify and clean any missing or inconsistent data in the "Department" column.

I could find only three types of data in it...so, no such inconsistency here.

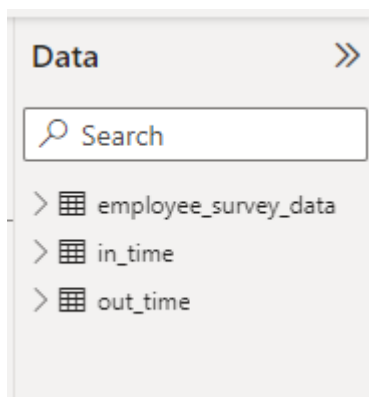
C	D	E	F
BusinessTravel	Department	DistanceFromHome	Education
Travel Frequently	Human Resources	6	
Travel Frequently	Human Resources	10	
Travel Frequently	Human Resources	17	
No Travel	Human Resources	2	
Travel Frequently	Human Resources	10	
Travel Frequently	Human Resources	8	
Travel Frequently	Human Resources	11	
Travel Frequently	Human Resources	18	
Travel Frequently	Human Resources	1	
No Travel	Human Resources	7	
Travel Frequently	Human Resources	17	
Travel Frequently	Human Resources	28	
Travel Frequently	Human Resources	14	
No Travel	Human Resources	1	
Travel Frequently	Human Resources	1	
Travel Frequently	Human Resources	1	
Travel Frequently	Human Resources	3	
No Travel	Human Resources	1	
Travel Frequently	Human Resources	7	
Travel Frequently	Human Resources	8	
Travel Frequently	Research & Development	1	

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

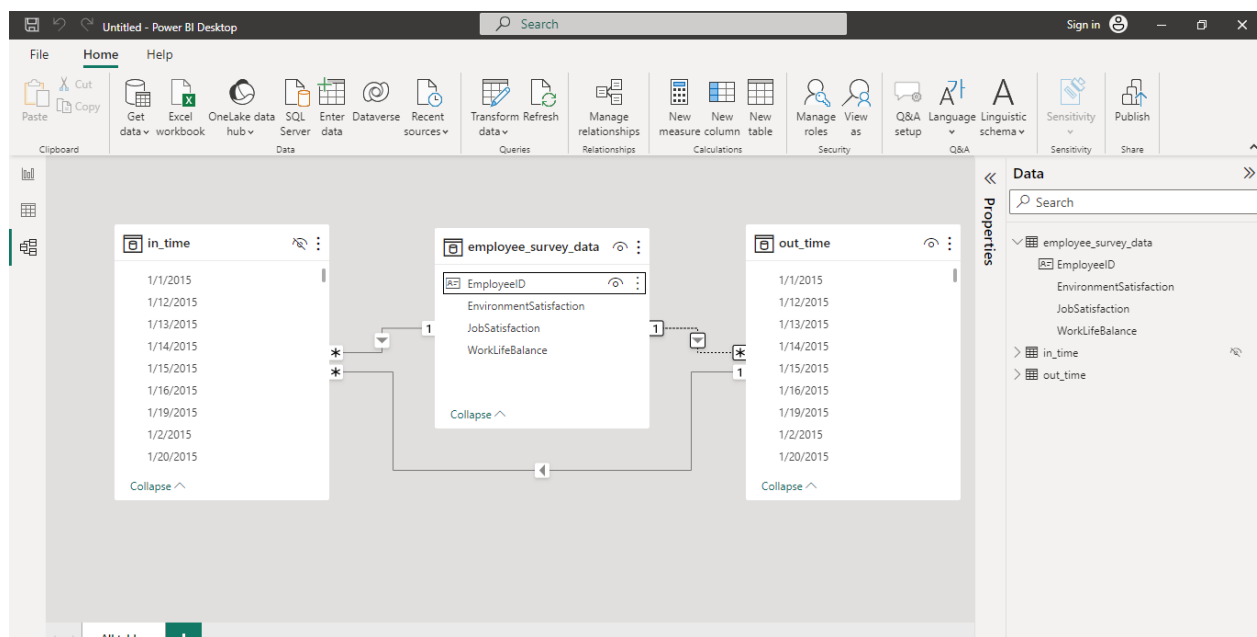
First, I made first row ad header in both in\_time and out\_time files:

Power Query Editor interface showing a table with 9 rows and 7 columns. The first column is 'Column1' with values 1-9. The next three columns are dates: '1/1/2015', '1/2/2015', and '1/5/2015'. The last three columns are times: '1/1/2015 9:43', '1/5/2015 10:08', '1/6/2015 9:54', '1/7/2015 9:43', '1/7/2015 9:44', '1/7/2015 9:45', '1/7/2015 9:46', '1/7/2015 9:47', '1/7/2015 9:48', '1/7/2015 9:49', '1/7/2015 9:50', '1/7/2015 9:51', '1/7/2015 9:52', '1/7/2015 9:53', '1/7/2015 9:54', '1/7/2015 9:55', '1/7/2015 9:56', '1/7/2015 9:57', '1/7/2015 9:58', '1/7/2015 9:59', '1/7/2015 10:00', '1/7/2015 10:01', '1/7/2015 10:02', '1/7/2015 10:03', '1/7/2015 10:04', '1/7/2015 10:05', '1/7/2015 10:06', '1/7/2015 10:07', '1/7/2015 10:08', '1/7/2015 10:09', '1/7/2015 10:10', '1/7/2015 10:11', '1/7/2015 10:12', '1/7/2015 10:13', '1/7/2015 10:14', '1/7/2015 10:15', '1/7/2015 10:16', '1/7/2015 10:17', '1/7/2015 10:18', '1/7/2015 10:19', '1/7/2015 10:20', '1/7/2015 10:21', '1/7/2015 10:22', '1/7/2015 10:23', '1/7/2015 10:24', '1/7/2015 10:25', '1/7/2015 10:26', '1/7/2015 10:27', '1/7/2015 10:28', '1/7/2015 10:29', '1/7/2015 10:30', '1/7/2015 10:31', '1/7/2015 10:32', '1/7/2015 10:33', '1/7/2015 10:34', '1/7/2015 10:35', '1/7/2015 10:36', '1/7/2015 10:37', '1/7/2015 10:38', '1/7/2015 10:39', '1/7/2015 10:40', '1/7/2015 10:41', '1/7/2015 10:42', '1/7/2015 10:43', '1/7/2015 10:44', '1/7/2015 10:45', '1/7/2015 10:46', '1/7/2015 10:47', '1/7/2015 10:48', '1/7/2015 10:49', '1/7/2015 10:50', '1/7/2015 10:51', '1/7/2015 10:52', '1/7/2015 10:53', '1/7/2015 10:54', '1/7/2015 10:55', '1/7/2015 10:56', '1/7/2015 10:57', '1/7/2015 10:58', '1/7/2015 10:59', '1/7/2015 11:00', '1/7/2015 11:01', '1/7/2015 11:02', '1/7/2015 11:03', '1/7/2015 11:04', '1/7/2015 11:05', '1/7/2015 11:06', '1/7/2015 11:07', '1/7/2015 11:08', '1/7/2015 11:09', '1/7/2015 11:10', '1/7/2015 11:11', '1/7/2015 11:12', '1/7/2015 11:13', '1/7/2015 11:14', '1/7/2015 11:15', '1/7/2015 11:16', '1/7/2015 11:17', '1/7/2015 11:18', '1/7/2015 11:19', '1/7/2015 11:20', '1/7/2015 11:21', '1/7/2015 11:22', '1/7/2015 11:23', '1/7/2015 11:24', '1/7/2015 11:25', '1/7/2015 11:26', '1/7/2015 11:27', '1/7/2015 11:28', '1/7/2015 11:29', '1/7/2015 11:30', '1/7/2015 11:31', '1/7/2015 11:32', '1/7/2015 11:33', '1/7/2015 11:34', '1/7/2015 11:35', '1/7/2015 11:36', '1/7/2015 11:37', '1/7/2015 11:38', '1/7/2015 11:39', '1/7/2015 11:40', '1/7/2015 11:41', '1/7/2015 11:42', '1/7/2015 11:43', '1/7/2015 11:44', '1/7/2015 11:45', '1/7/2015 11:46', '1/7/2015 11:47', '1/7/2015 11:48', '1/7/2015 11:49', '1/7/2015 11:50', '1/7/2015 11:51', '1/7/2015 11:52', '1/7/2015 11:53', '1/7/2015 11:54', '1/7/2015 11:55', '1/7/2015 11:56', '1/7/2015 11:57', '1/7/2015 11:58', '1/7/2015 11:59', '1/7/2015 12:00'.

Later, uploaded all these three files.



And checked their relation which was found to be one-one connection.



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

**Avg\_year\_with\_current\_manager=CALCULATE(AVERAGE(general\_data[YearsWithCurrManager])**

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

Sum of EmployeeCount	Column Labels			
Row Labels	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.**  
By selecting the respected column and then, making such adjustments, we can do it so.



FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW DEVELOPER

Clipboard: O1

New Formatting Rule

Select a Rule Type:

- Format all cells based on their values
- Format only cells that contain
- Format only top or bottom ranked values
- Format only values that are above or below average
- Format only unique or duplicate values
- Use a formula to determine which cells to format

Edit the Rule Description:

Format values that are:

below the average for the selected range

Preview: AaBbCcYyZz

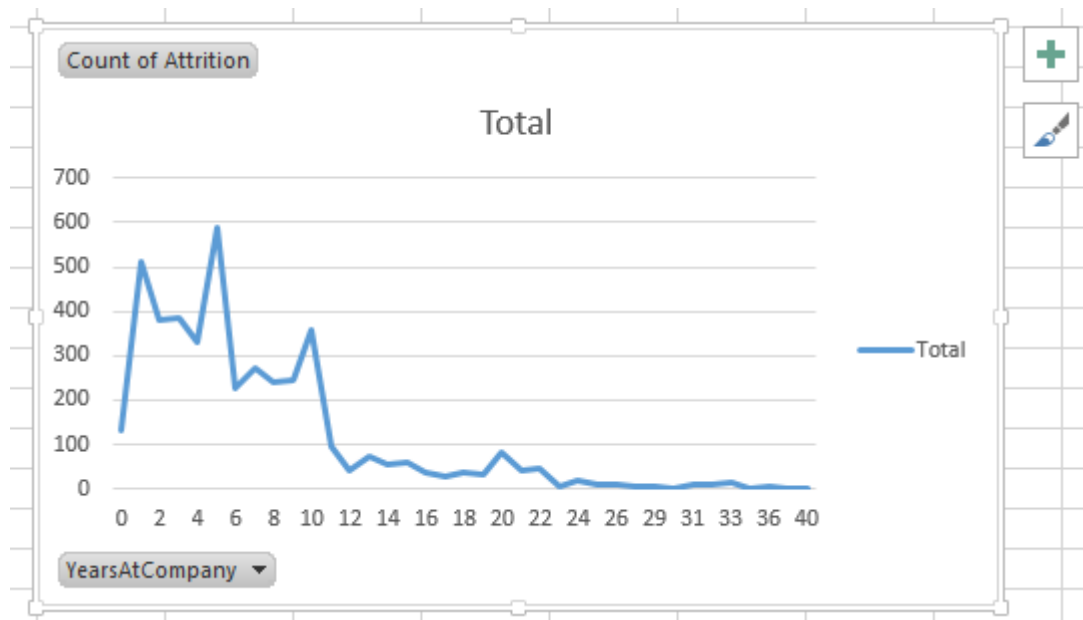
OK Cancel

JobLevel	JobRole	MaritalStatus	MonthlyIncome	JobSatisfaction	NumCompaniesWorked	Over18	PercentSalaryHike	StandardHours	StockOptionLevel	TotalWorkingYears	TrainingTimesLast
1	Healthcare Representative	Married	131160	4	1	Y	11	8	0	1	6
2	Research Scientist	Single	41890	2	0	Y	23	8	1	6	5
3	Sales Executive	Married	193280	2	1	Y	15	8	3	5	13
4	Human Resources	Married	83210	4	3	Y	11	8	3	9	28
5	Sales Executive	Single	23420	1	4	Y	12	8	2	5	10
6	Research Director	Married	40710	2	3	Y	13	8	0	10	6
7	Sales Executive	Single	58130	3	2	Y	20	8	1	10	37
8	Sales Executive	Married	31430	2	2	Y	22	8	3	10	21
9	Laboratory Technician	Married	20440	4	0	Y	21	8	0	16	37
10	Laboratory Technician	Divorced	134640	1	1	Y	13	8	1	6	21
11	Laboratory Technician	Married	79910	4	0	Y	13	8	2	16	37
12	Laboratory Technician	Married	33770	4	0	Y	12	8	2	10	6
13	Sales Executive	Single	55380	1	0	Y	17	8	0	10	21
14	Research Scientist	Married	57620	2	1	Y	11	8	2	16	37
15	Manufacturing Director	Married	25920	4	1	Y	14	8	0	7	3
16	Healthcare Representative	Married	53460	4	4	Y	11	8	0	15	10
17	Laboratory Technician	Single	42130	3	1	Y	12	8	3	8	6
18	Sales Executive	Divorced	41270	4	2	Y	13	8	1	37	21
19	Sales Representative	Divorced	24380	2	7	Y	16	8	0	10	6
20	Manager	Divorced	68700	1	1	Y	11	8	1	10	37
21	Laboratory Technician	Divorced	104470	2	1	Y	18	8	0	6	21
22	Research Scientist	Divorced	96670	2	3	Y	23	8	0	16	37

READY | Sheet1 | Sheet2 | general\_data

AVERAGE: 2.728246914 COUNT: 4411 SUM: 11977

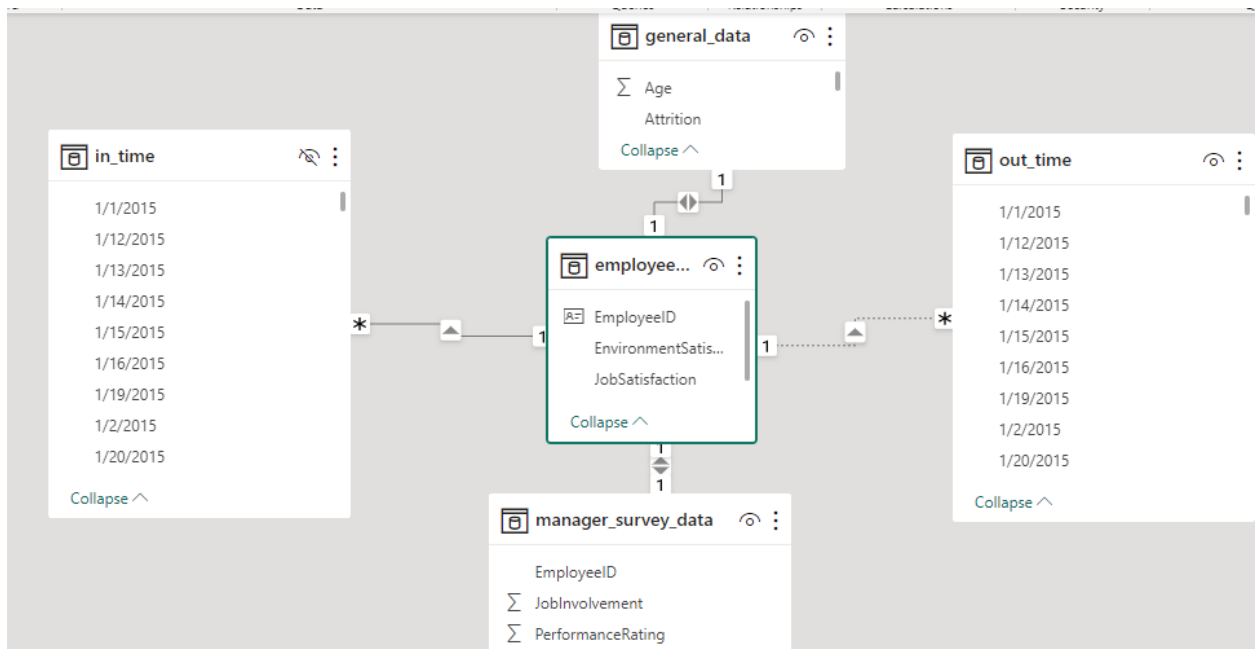
**10. In Power BI, create a line chart that visualizes the trend of Employee Attrition over the years.**



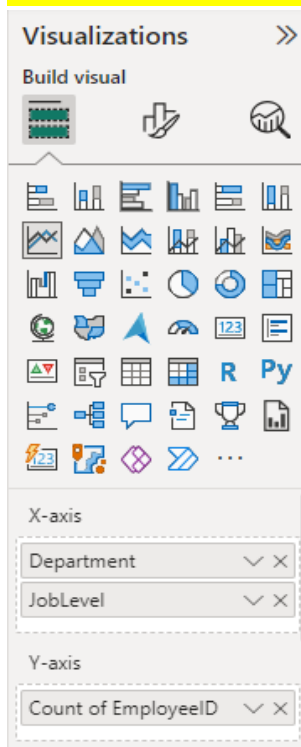
**11. Describe how you would create a star schema for this dataset, explaining the benefits of doing so.**

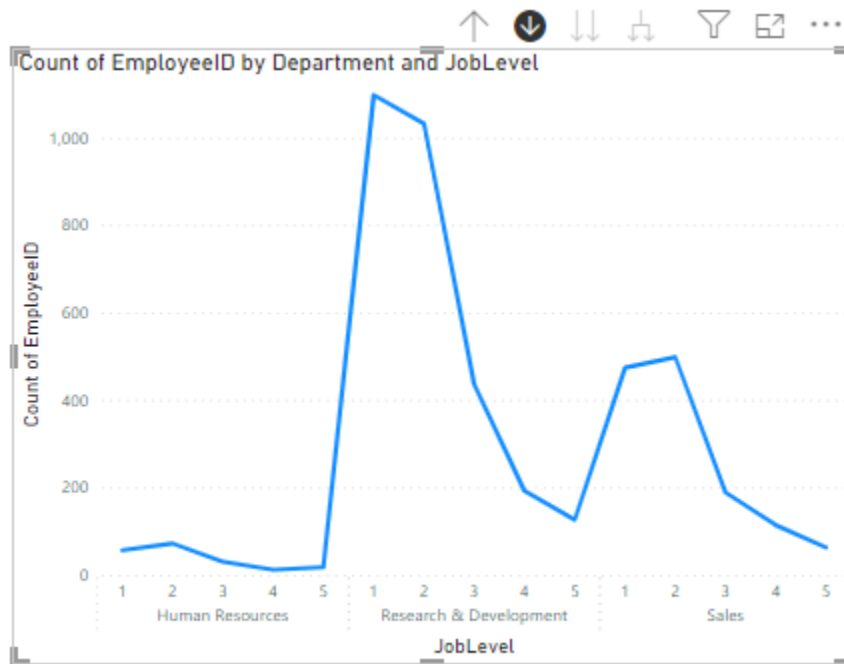
**Benefits of star schema:**

- 1. Centralized data:** the data is stored with main data and can be accessed easily.
- 2. Easy to interpret:** It is always easy to understand the data in a star schema as all the data is correlated.
- 3. Having different groups make data more concise and proper.**



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





## 14. How can you set up parameterized queries in Power BI to allow users to filter data based on the Distance from Home column?

The screenshot shows the Power BI Transform ribbon with the 'Query Properties' dialog box open for the 'DistanceFromHome' query. The dialog box has the following fields and options:

- Name:** DistanceFromHome
- Description:** (Empty text box)
- Enable load to report:** ☒
- Include in report refresh:** ☒

The 'Query Settings' pane on the right shows the 'DistanceFromHome1' step selected under the 'APPLIED STEPS' section.

**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.**

A	B
Row Labels	Sum of MonthlyIncome
Human Resources	3259140
Research & Development	53502900
Sales	22974330
(blank)	
<b>Grand Total</b>	<b>79736370</b>

**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.**

	A	B	C
1	MonthlyIncome	10% inc	
2	131160	144276	
3	41890	46079	
4	193280	212608	
5	83210	91531	
6	23420	25762	
7	40710	44781	
8	58130	63943	
9	31430	34573	
10	20440	22484	
11	134640	148104	
12	79910	87901	
13	33770	37147	
14	55380	60918	
15	57620	63382	
16	25920	28512	
17	53460	58806	
18	42130	46343	
19	41270	45397	
20	24380	26818	
21	68700	75570	
22	104470	114917	

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

**Overall Approach:**

**Handling inconsistencies involves a mix of data cleaning, transformation, documentation updates, stakeholder engagement, and continuous monitoring. Regular checks and proactive measures ensure data quality and compliance with the predefined schema, aligning data practices with the established standards**