HR Data Analysis

Using Microsoft Excel and Power BI

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

Following the below steps to filter the dataset to only show Employees Age 30 and above :-

Go to Data Tab => Click on Filter => Select the column header arrow => Select Number Filters, and then select a comparison, like greater than or equal to => Enter the filter criteria(which is 30 in this case) and select OK.

Age	Attrition	Busines	Departr	Distanc	Educati	Educati	Employ	Employ	Gender	JobLeve	JobRole	Marital	Monthl	NumCo	Over18	Percent	Standar	StockOr	To
51	No	Travel_Ra	Sales	6	2	Life Scienc	1	1	Female	1	Healthcar	Married	131160	1	Y	11	8	0	
31	Yes	Travel_Fre	Research	10	1	Life Scienc	1	2	Female	1	Research	Single	41890	0	Y	23	8	1	
32	No	Travel_Fre	Research	17	4	Other	1	3	Male	4	Sales Exec	Married	193280	1	Y	15	8	3	
38	No	Non-Trave	Research	2	5	Life Scienc	1	4	Male	3	Human Re	Married	83210	3	Y	11	8	3	
32	No	Travel_Ra	Research	10	1	Medical	1	5	Male	1	Sales Exec	Single	23420	4	Y	12	8	2	
46	No	Travel_Ra	Research	8	3	Life Scienc	1	6	Female	4	Research I	Married	40710	3	Y	13	8	0	
31	No	Travel_Ra	Research	1	3	Life Scienc	1	9	Male	3	Laborator	Married	20440	0	Y	21	8	0	
45	No	Travel_Ra	Research	17	2	Medical	1	11	Male	2	Laborator	Married	79910	0	Y	13	8	2	
36	No	Travel_Ra	Research	28	1	Life Scienc	1	12	Male	1	Laborator	Married	33770	0	Y	12	8	2	
55	No	Travel_Ra	Research	14	4	Life Scienc	1	13	Female	1	Sales Exec	Single	55380	0	Y	17	8	0	
47	Yes	Non-Trave	Research	1	1	Medical	1	14	Male	1	Research	Married	57620	1	Y	11	8	2	
37	No	Travel_Ra	Research	1	3	Life Scienc	1	16	Male	2	Healthcar	Married	53460	4	Y	11	8	0	
37	No	Non-Trave	Research	1	3	Medical	1	18	Male	2	Sales Exec	Divorced	41270	2	Y	13	8	1	
35	No	Travel_Ra	Sales	7	4	Life Scienc	1	19	Male	1	Sales Repr	Divorced	24380	7	Y	16	8	0	
38	No	Travel_Ra	Research	8	3	Life Scienc	1	20	Female	1	Manager	Divorced	68700	1	Y	11	8	1	
50	No	Travel_Ra	Sales	8	4	Life Scienc	1	22	Male	1	Research	Divorced	96670	3	Y	23	8	0	
53	No	Travel_Ra	Research	11	4	Life Scienc	1	23	Female	2	Research	Married	21480	3	Y	11	8	0	
42	No	Travel_Ra	Research	4	4	Life Scienc	1	24	Male	1	Manufact	Married	89260	1	Y	14	8	0	N
55	No	Travel_Ra	Research	1	4	Other	1	26	Female	1	Research	Married	67000	2	Y	11	8	0	

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

JobRole ▼	Average of MonthlyIncome
Healthcare Representative	60983.74046
Human Resources	58528.07692
Laboratory Technician	66314.05405
Manager	63395.88235
Manufacturing Director	69183.72414
Research Director	65473.125
Research Scientist	64975.68493
Sales Executive	65186.68712
Sales Representative	65370.96386
Grand Total	**65029.31293**

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

Age	Attrition	Business	Department	Distance	Education	Education	Employe	Employe	Gender	JobLevel	JobRole	MaritalStatus	MonthlyIncome	NumCo	Over18	Percent
51	No	Travel_Ra	Sales	6	2	Life Scienc	1	1	Female	1	Healthcare	Married	131160	1	Y	
31	Yes	Travel_Fre	Research &	10	1	Life Scienc	1	2	Female	1	Research S	Single	41890	0	Y	
32	No	Travel_Fre	Research &	17	4	Other	1	3	Male	4	Sales Exec	Married	193280	1	Y	
38	No	Non-Trave	Research &	2	5	Life Scienc	1	4	Male	3	Human Re	Married	83210	3	Y	
32	No	Travel_Ra	Research &	10	1	Medical	1	5	Male	1	Sales Exec	Single	23420	4	Y	
46	No	Travel_Ra	Research &	8	3	Life Scienc	1	6	Female	4	Research I	Married	40710	3	Y	
28	Yes	Travel_Ra	Research &	11	2	Medical	1	7	Male	2	Sales Exec	Single	58130	2	Y	
29	No	Travel_Ra	Research &	18	3	Life Scienc	1	8	Male	2	Sales Exec	Married	31430	2	Y	
31	No	Travel_Ra	Research &	1	3	Life Scienc	1	9	Male	3	Laborator	Married	20440	0	Y	
25	No	Non-Trave	Research &	7	4	Medical	1	10	Female	4	Laborator	Divorced	134640	1	Y	
45	No	Travel_Ra	Research &	17	2	Medical	1	11	Male	2	Laborator	Married	79910	0	Y	
36	No	Travel_Ra	Research &	28	1	Life Scienc	1	12	Male	1	Laborator	Married	33770	0	Y	
55	No	Travel_Ra	Research &	14	4	Life Scienc	1	13	Female	1	Sales Exec	Single	55380	0	Y	
47	Yes	Non-Trave	Research &	1	1	Medical	1	14	Male	1	Research S	Married	57620	1	Y	
28	No	Travel_Ra	Research &	1	3	Life Scienc	1	15	Male	1	Manufactu	Married	25920	1	Y	
37	No	Travel_Ra	Research &	1	3	Life Scienc	1	16	Male	2	Healthcare	Married	53460	4	Y	
21	No	Travel_Ra	Research &	3	2	Life Scienc	1	17	Male	1	Laborator	Single	42130	1	Y	
37	No	Non-Trave	Research &	1	3	Medical	1	18	Male	2	Sales Exec	Divorced	41270	2	Y	
35	No	Travel_Ra	Sales	7	4	Life Scienc	1	19	Male	1	Sales Repr	Divorced	24380	7	Y	

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

Age	Sum of EmployeeCount
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
42	138
43	96
44	99
45	123
46	99
47	72
48	57
49	72
50	90
51	57
52	54
53	57
54	54
55	66
56	42
57	12
58	42
59	30
60	15
Grand Total	**4410**

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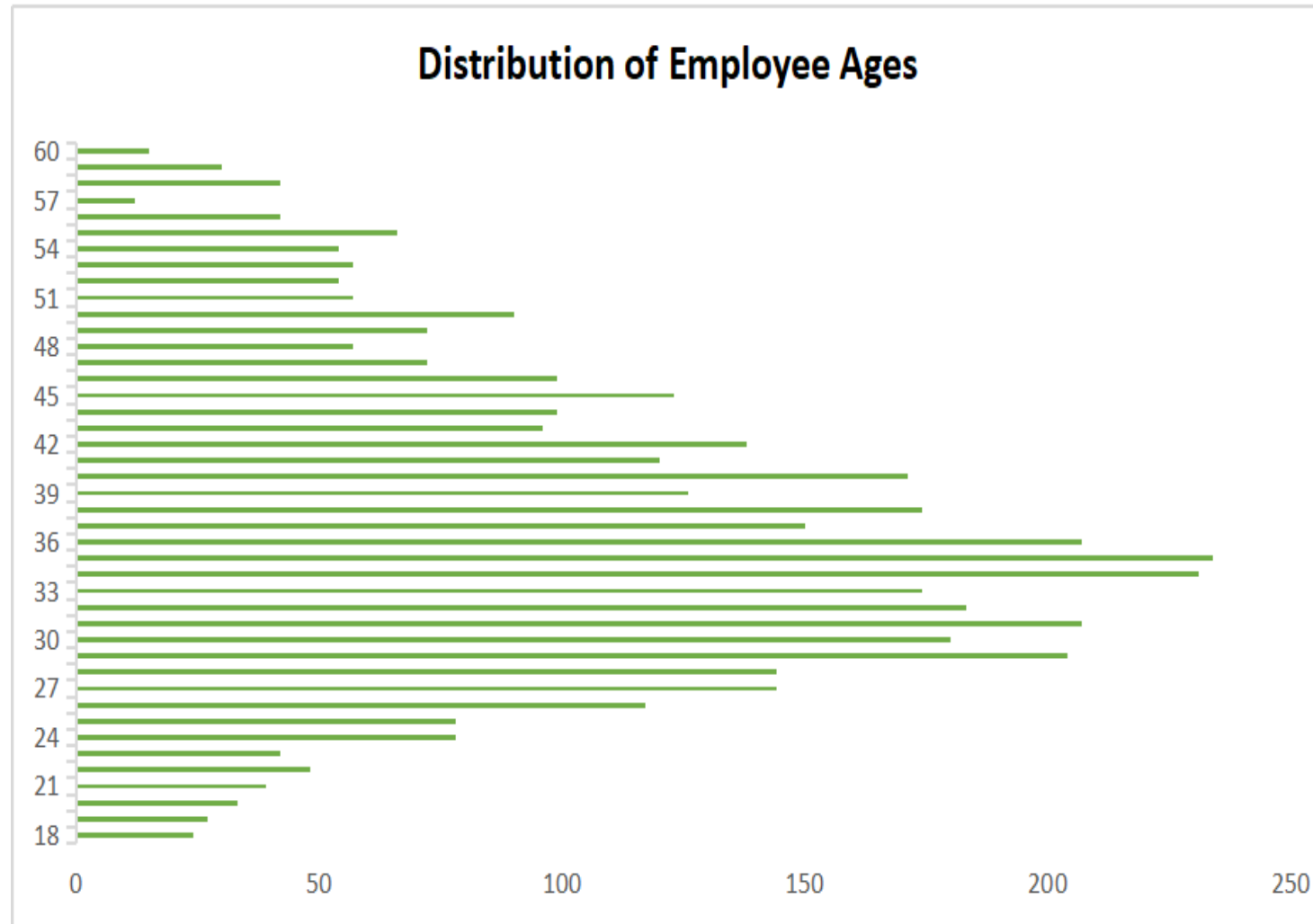
# 5. Identify and clean any missing or inconsistent data in the "Department" column.

Dealing with missing or inconsistent data values in categorical columns(like Department Column) is a lot easier. Simply I need to replace the missing or inconsistent value with a constant value or the most popular category. This is a good approach when my data size is small, though it does add bias. But the most preferable approach is to model the missing value in a categorical column(Department Column) as a new category called Unknown as our Dataset is large.

In the given Dataset, Department Column don't have any missing or inconsistent value so there is no need to change anything in the Department Column.

Age	Department
51	Sales
31	Research & Development
32	Research & Development
38	Research & Development
32	Research & Development
46	Research & Development
28	Research & Development
29	Research & Development
31	Research & Development
25	Research & Development
45	Research & Development
36	Research & Development
55	Research & Development
47	Research & Development
28	Research & Development
37	Research & Development
21	Research & Development
37	Research & Development
35	Sales
38	Research & Development
26	Research & Development
50	Sales
53	Research & Development
42	Research & Development
29	Research & Development
55	Research & Development

6. In Power BI, establish a relationship between the "EmployeeID" in the employee data and the "EmployeeID" in the time tracking data.
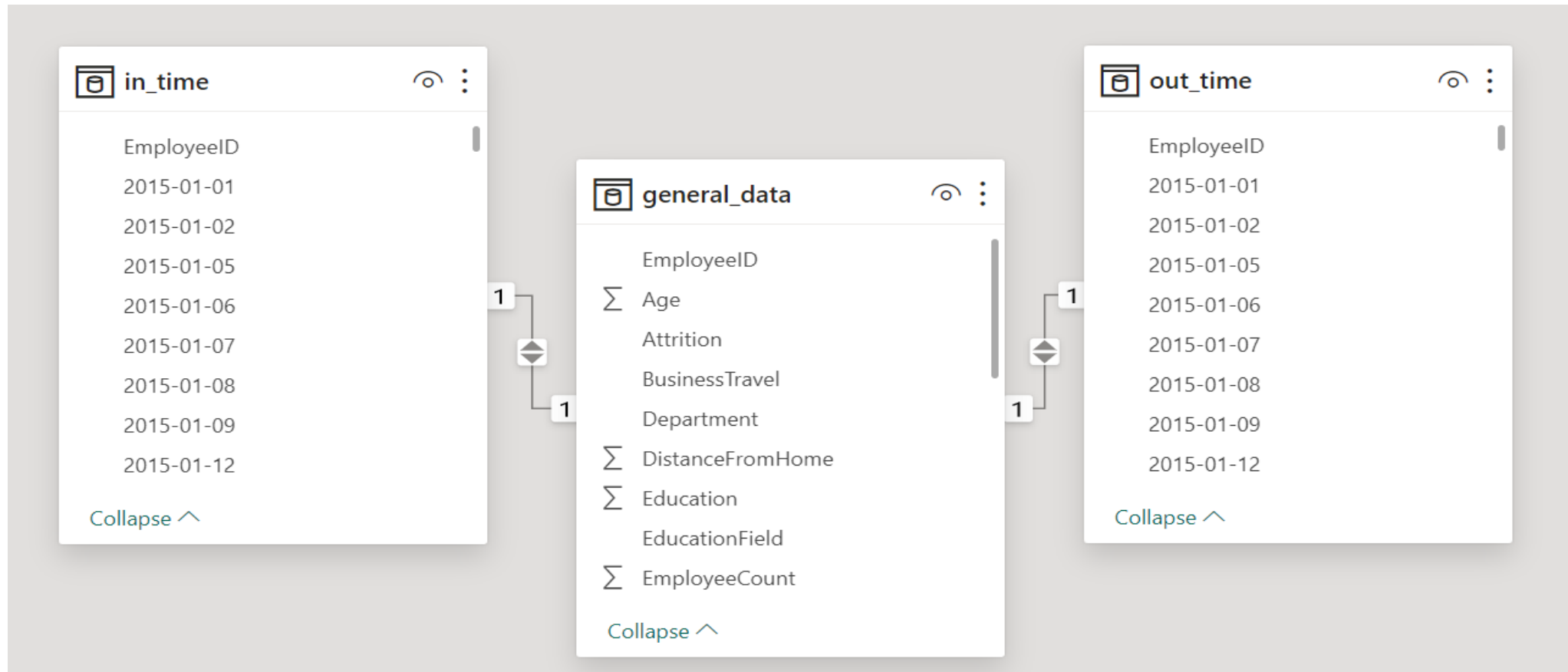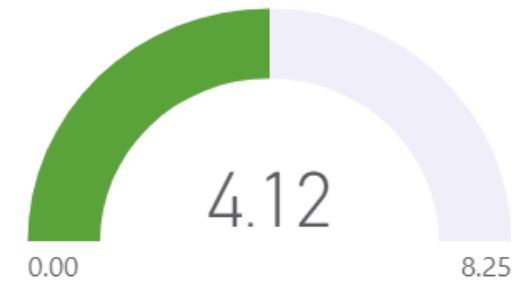
7. Using DAX, create a calculated column that calculates the average years an employee has spent with their current manager.

`1 Average Year with Current Manager = CALCULATE(AVERAGE(general_data[YearsWithCurrManager]))`

TotalWorkingYears	TrainingTimesLastYear	YearsAtCompany	YearsSinceLastPromotion	YearsWithCurrManager	Roll 3-month average of Monthly Income	Average Year with Current Manager
	3	4	0	0	92780	0
	2	2	0	0	47890	0
5	2	5	0	2	41270	2
1	4	9	0	8	85780	8
	0	7	0	1	47880	1
0	2	4	0	3	29330	3
	3	5	0	4	56170	4
	3	6	0	4	75100	4
9	0	5	0	2	28370	2
	2	6	0	4	132060	4
1	5	3	0	2	20010	2
0	3	1	0	1	30680	1
0	2	8	0	7	26700	7
0	3	8	0	5	41520	5
0	2	5	0	3	30580	3
	3	3	0	2	44500	2
	4	3	0	2	23890	2
2	3	7	0	7	48980	7
1	2	18	0	11	168720	11
	5	2	0	2	32100	2
9	3	18	0	13	36880	13
8	4	5	0	3	52280	3
	5	6	0	4	115100	4
	2	3	0	2	46490	2
8	3	8	0	1	96990	1
	2	5	0	3	28510	3
8	4	11	0	9	34520	9
	3	3	0	2	26570	

Average of YearsWithCurrManager

4.12

0.00 — 8.25

EmployeeID

1 — 4410

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

Sum of EmployeeCount	Department			
MaritalStatus	Human Resources	Research & Development	Sales	Grand Total
Divorced	21	621	339	981
Married	96	1350	573	2019
Single	72	912	426	1410
Grand Total	189	2883	1338	4410

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

EmployeeID	MonthlyIncome	JobSatisfaction
1	131160	4
2	41890	2
3	193280	2
4	83210	4
5	23420	1
6	40710	2
7	58130	3
8	31430	2
9	20440	4
10	134640	1
11	79910	4
12	33770	4
13	55380	1
14	57620	2
15	25920	4
16	53460	4
17	42130	3
18	41270	4
19	24380	2
20	68700	1
21	104470	2
22	96670	2
23	21480	3
24	89260	3
25	65130	4
26	67990	4

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

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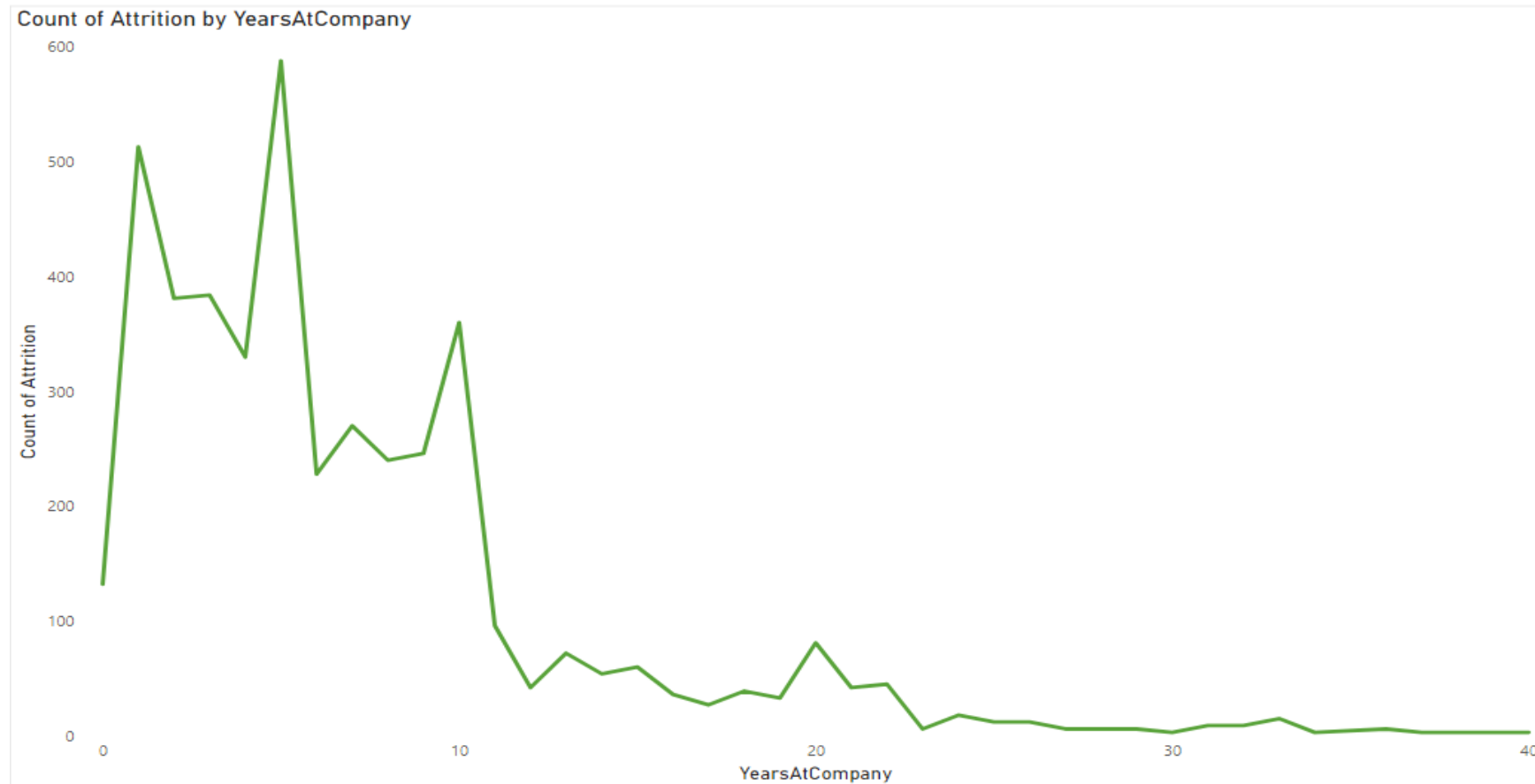
11. Describe how you would create a star schema for this dataset, explaining the benefits of doing so.

A star schema is a multi-dimensional data model used to organize data in a database so that it is easy to understand and analyze. Star schemas can be applied to data warehouses, databases, data marts, and other tools. The star schema design is optimized for querying large data sets.

A fact table sits at the center of a star schema database, and each star schema database only has a single fact table. The fact table contains the specific measurable (or quantifiable) primary data to be analyzed.

Dimension tables store supporting information to the fact table. Each star schema database has at least one dimension table, but will often have many. Each dimension table will relate to a column in the fact table with a dimension value, and will store additional information about that value.

Here, fact table is General Data Table and dimesion tables are Manager Survey Data, Manager Survey Data, In_Time Data and Out_Time. Tables are connected through Employee ID.
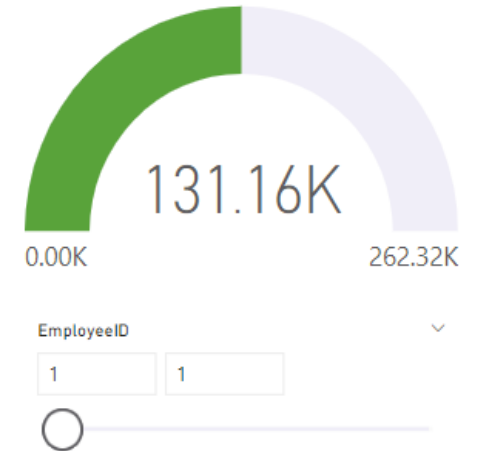
12. Using DAX, calculate the rolling 3-month average of Monthly Income for each employee.

```
1 Roll 3-month average of Monthly Income = CALCULATE(AVERAGE(general_data[MonthlyIncome]),FILTER(ALL(general_data[EmployeeID]),general_data[EmployeeID]=EARLIER
  (general_data[EmployeeID])))
```

dardHours	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	YearsAtCompany	YearsSinceLastPromotion	YearsWithCurrManager	Roll 3-month average of Monthly Income
8	0	1	6	1	0	0	131160
8	1	1	3	1	0	0	55610
8	1	1	2	1	0	0	53730
8	0	1	3	1	0	0	59680
8	2	1	2	1	0	0	76440
8	0	1	3	1	0	0	24260
8	0	1	4	1	0	0	12320
8	2	1	3	1	0	0	22930
8	3	1	2	1	0	0	106730
8	2	1	2	1	0	0	106500
8	2	1	2	1	0	0	56050
8	1	1	2	1	0	0	104000
8	1	1	5	1	0	0	20580
8	1	1	3	1	0	0	32120
8	1	1	3	1	0	0	160640
8	0	1	2	1	0	0	27070
8	1	1	3	1	0	0	63970
8	1	1	5	1	0	0	64100
8	0	1	1	1	0	0	52100
8	0	1	3	1	0	0	80950
8	0	1	3	1	0	0	41970
8	0	1	2	1	0	0	191970
8	1	1	5	1	0	0	28860
8	1	1	2	1	0	0	61800
8	1	1	4	1	0	0	25530
8	0	1	3	1	0	0	163070
8	0	1	2	1	0	0	25320

Sum of Roll 3-month average of Monthly Income

131.16K

0.00K   262.32K

EmployeeID

| 1 | 1 |

13. Create a hierarchy in Power BI that allows users to drill down from Department to Job Role to further narrow their analysis.
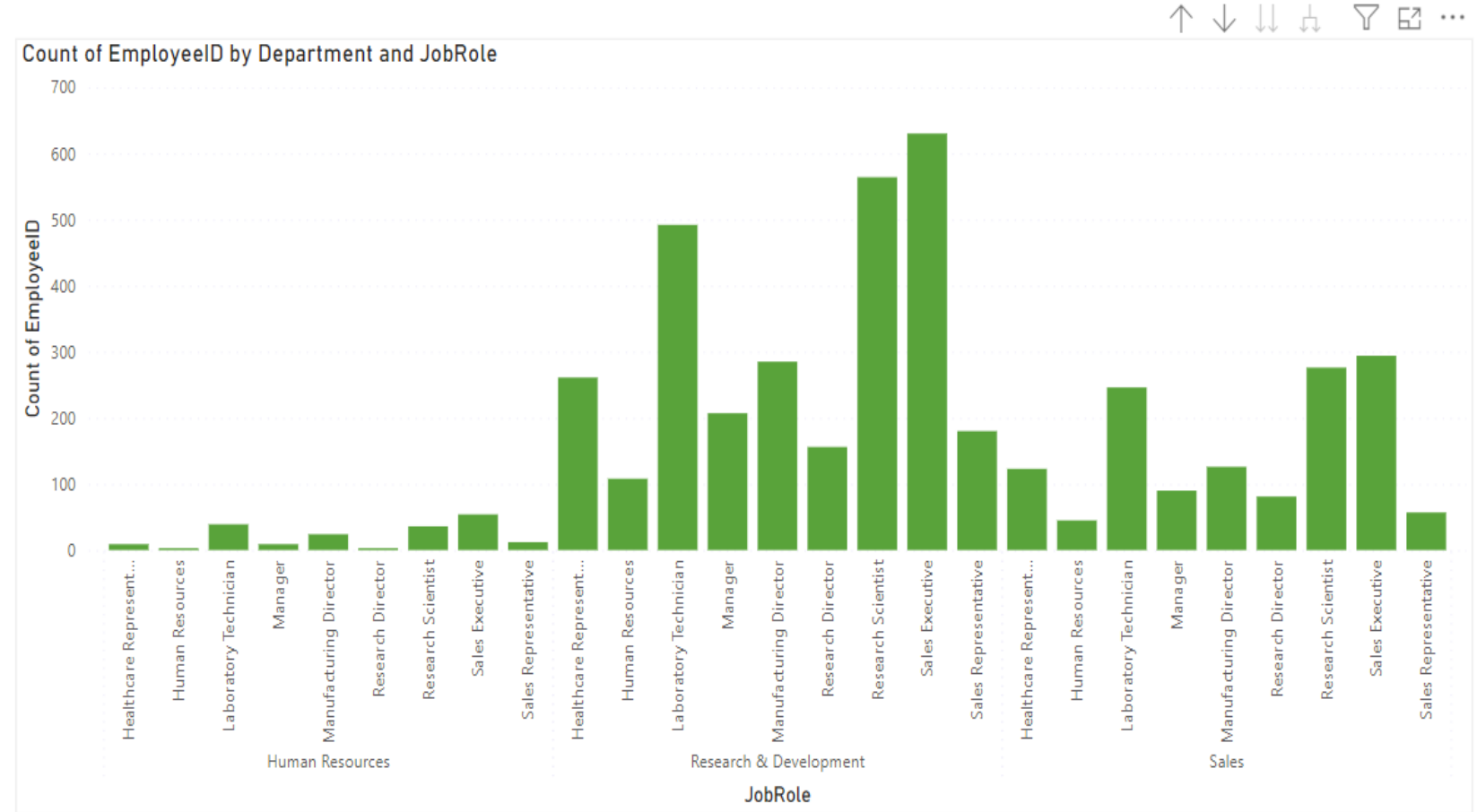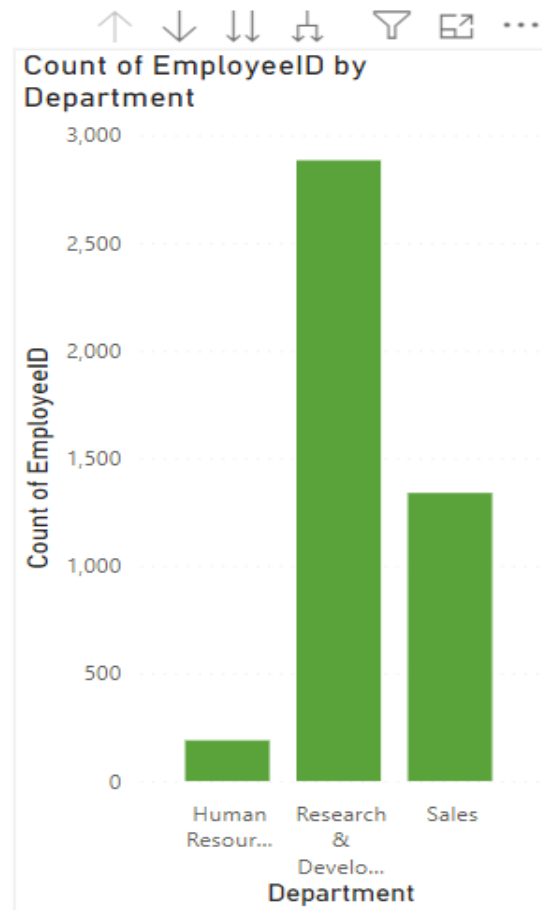
14. How can you set up parameterized queries in Power BI to allow users to filter data based 2 of 2 on the Distance from Home column?

In Power BI, we can set up parameterized queries using parameters in your queries. To filter data based on the distance from home column, follow these steps :-

❑ **Creating parameters -** Go to "Home" tab => Click on "Manage Properties" => Creating Parameters

❑ **Modify Queries -** Open Power Query Editor => Edit the Query => Use parameters

❑ **Load and use Parameters -** Close & Apply Changes => Creating Visualizations & Use parameters in slicer or other filters

❑ **Paramterize Distance Column -** If "DistanceFromHome" column is static and not a parameter, then we might to create a calculated column based on parameters.

❑ **Utilize Slicers -** Creating slicer for our report canvas linked to the parameters we created.

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

G4		=SUMIFS(MonthlyIncome,Department,"Sales",JobLevel,">=3")

|	A	B	C	D	E	F	G	H
1	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							
2								
3	EmployeeID	MonthlyIncome	Department	JobLevel		Department	Total Monthly Income for each Department	
4	1	131160	Sales	1		Sales	22974330	
5	2	41890	Research & Development	1		Research & Development	53502900	
6	3	193280	Research & Development	4		Human Resources	3259140	
7	4	83210	Research & Development	3				
8	5	23420	Research & Development	1				
9	6	40710	Research & Development	4				
10	7	58130	Research & Development	2				
11	8	31430	Research & Development	2				
12	9	20440	Research & Development	3				
13	10	134640	Research & Development	4				
14	11	79910	Research & Development	2				
15	12	33770	Research & Development	1				
16	13	55380	Research & Development	1				
17	14	57620	Research & Development	1				
18	15	25920	Research & Development	1				
19	16	53460	Research & Development	2				
20	17	42130	Research & Development	1				
21	18	41270	Research & Development	2				
22	19	24380	Sales	1				
23	20	68700	Research & Development	1				
24	21	104470	Research & Development	2				
25	22	96670	Sales	1				
26	23	21480	Research & Development	2				
27	24	89260	Research & Development	1				
28	25	65130	Research & Development	1				
29	26	67990	Research & Development	1				
30	27	162910	Research & Development	1				
31	28	27050	Sales	1				
32	29	103330	Research & Development	2				
33	30	44480	Sales	1				

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.

The What-If Analysis in Excel is a powerful tool to perform complex mathematical calculations, experiment with data, and try out different scenarios. There are three What-If Analysis tools available :-

❑ **Goal Seek** - It calculate input value based on given result. Goal Seek performs reverse calculations.

❑ **Scenario Manager -** Scenario Manager is used for a comparison of different scenarios.

❑ **Data Table -** Data Table is used for sensitivity analysis.

In the given dataset, for understanding the impact of 10% increase in Percent Salary Hike on Monthly Income I use Data Table tool of What-If Analaysis.

Firstly, I was calculated 10% Salary with the given salary as a Desired Salary then put the Desired Salary in blank cell in the same sheet. After that enter Monthly Income of all Employees in vertical cells and Percentage Increment in horizontal cell. Select the data and apply data table (Go to Data Tab => Click on What-If Analysis => Select Data Table). Put cell name of Percentage Hike in Row Input cell and cell name of Monthly Income in Column Input Data. Then click on OK. It will calculate 10% Percent Salary Hike for each employee based on their Monthly Income.
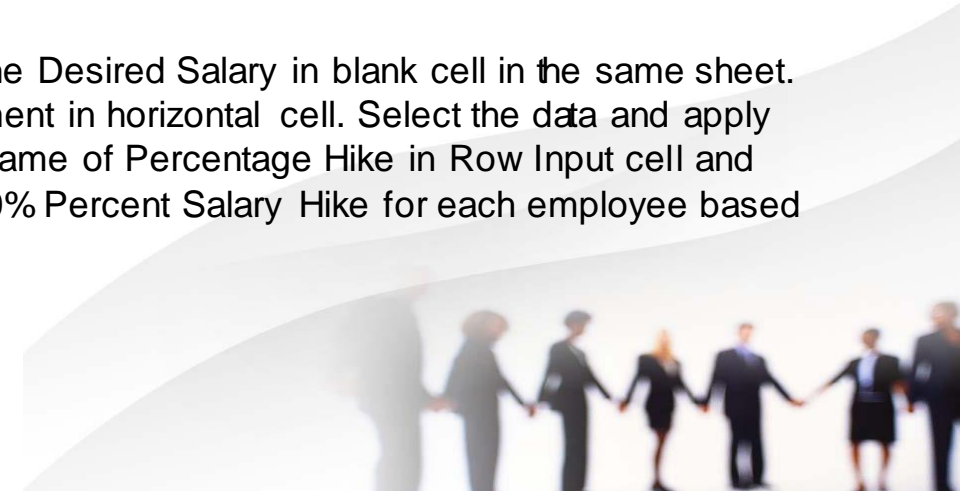
144276	**10%**
41890	46079
193280	212608
83210	91531
23420	25762
40710	44781
58130	63943
31430	34573
20440	22484
134640	148104
79910	87901
33770	37147
55380	60918
57620	63382
25920	28512
53460	58806
42130	46343
41270	45397
24380	26818
68700	75570
104470	114917
96670	106337
21480	23628
89260	98186
65130	71643
67990	74789
162910	179201

Monthly Income	131160
Percentage Hike	10%
Desired Salary	144276

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

- ❑ Understand the Database Schema

- ❑ Review Data Sources

- ❑ Compare Dataset

- ❑ Datatypes and Constraints

- ❑ Date and Time Format

- ❑ Applying filters and Conditional Formatting

- ❑ Use Data Validation

- ❑ Remove Duplicates

- ❑ Identify Patterns and Outliers

Thank You!