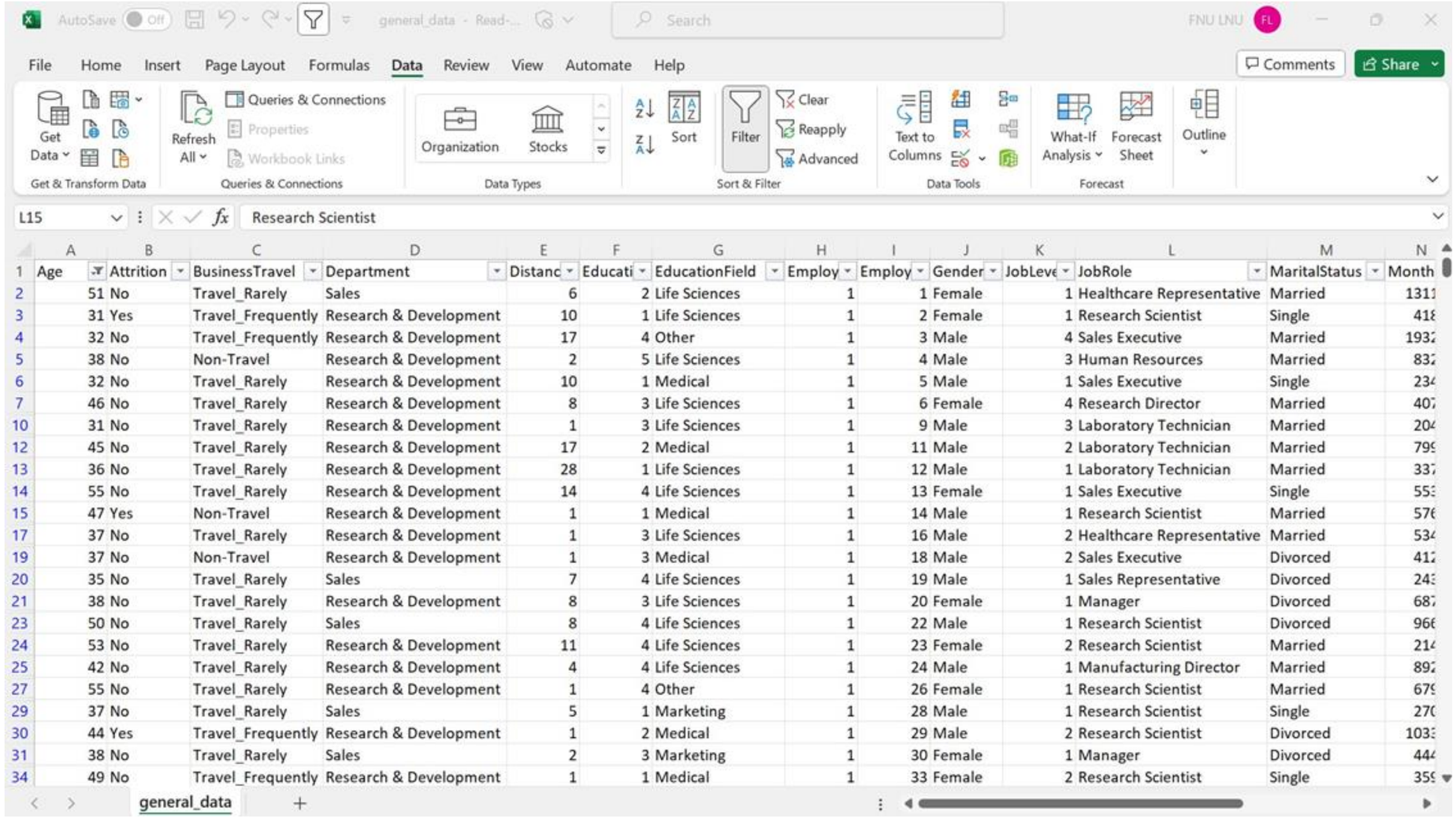
PSYLIQ

DATA ANALYSIS INTERNSHIP

Task 1: HR Data Analysis

Prepared by: Abhay Kumar

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



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Age	Attrition	BusinessTravel	Department	Distanc	Educati	EducationField	Employ	Employ	Gender	JobLeve	JobRole	MaritalStatus	Month
2	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	1	Female	1	Healthcare Representative	Married	1311
3	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1	2	Female	1	Research Scientist	Single	418
4	32	No	Travel_Frequently	Research & Development	17	4	Other	1	3	Male	4	Sales Executive	Married	1932
5	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	4	Male	3	Human Resources	Married	837
6	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	5	Male	1	Sales Executive	Single	234
7	46	No	Travel_Rarely	Research & Development	8	3	Life Sciences	1	6	Female	4	Research Director	Married	407
10	31	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	9	Male	3	Laboratory Technician	Married	204
12	45	No	Travel_Rarely	Research & Development	17	2	Medical	1	11	Male	2	Laboratory Technician	Married	799
13	36	No	Travel_Rarely	Research & Development	28	1	Life Sciences	1	12	Male	1	Laboratory Technician	Married	337
14	55	No	Travel_Rarely	Research & Development	14	4	Life Sciences	1	13	Female	1	Sales Executive	Single	553
15	47	Yes	Non-Travel	Research & Development	1	1	Medical	1	14	Male	1	Research Scientist	Married	576
17	37	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	16	Male	2	Healthcare Representative	Married	534
19	37	No	Non-Travel	Research & Development	1	3	Medical	1	18	Male	2	Sales Executive	Divorced	412
20	35	No	Travel_Rarely	Sales	7	4	Life Sciences	1	19	Male	1	Sales Representative	Divorced	243
21	38	No	Travel_Rarely	Research & Development	8	3	Life Sciences	1	20	Female	1	Manager	Divorced	687
23	50	No	Travel_Rarely	Sales	8	4	Life Sciences	1	22	Male	1	Research Scientist	Divorced	966
24	53	No	Travel_Rarely	Research & Development	11	4	Life Sciences	1	23	Female	2	Research Scientist	Married	214
25	42	No	Travel_Rarely	Research & Development	4	4	Life Sciences	1	24	Male	1	Manufacturing Director	Married	897
27	55	No	Travel_Rarely	Research & Development	1	4	Other	1	26	Female	1	Research Scientist	Married	679
29	37	No	Travel_Rarely	Sales	5	1	Marketing	1	28	Male	1	Research Scientist	Single	270
30	44	Yes	Travel_Frequently	Research & Development	1	2	Medical	1	29	Male	2	Research Scientist	Divorced	1033
31	38	No	Travel_Rarely	Sales	2	3	Marketing	1	30	Female	1	Manager	Divorced	444
34	49	No	Travel_Frequently	Research & Development	1	1	Medical	1	33	Female	2	Research Scientist	Single	359

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

Average of MonthlyIncome	
JobRole	Total
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**

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

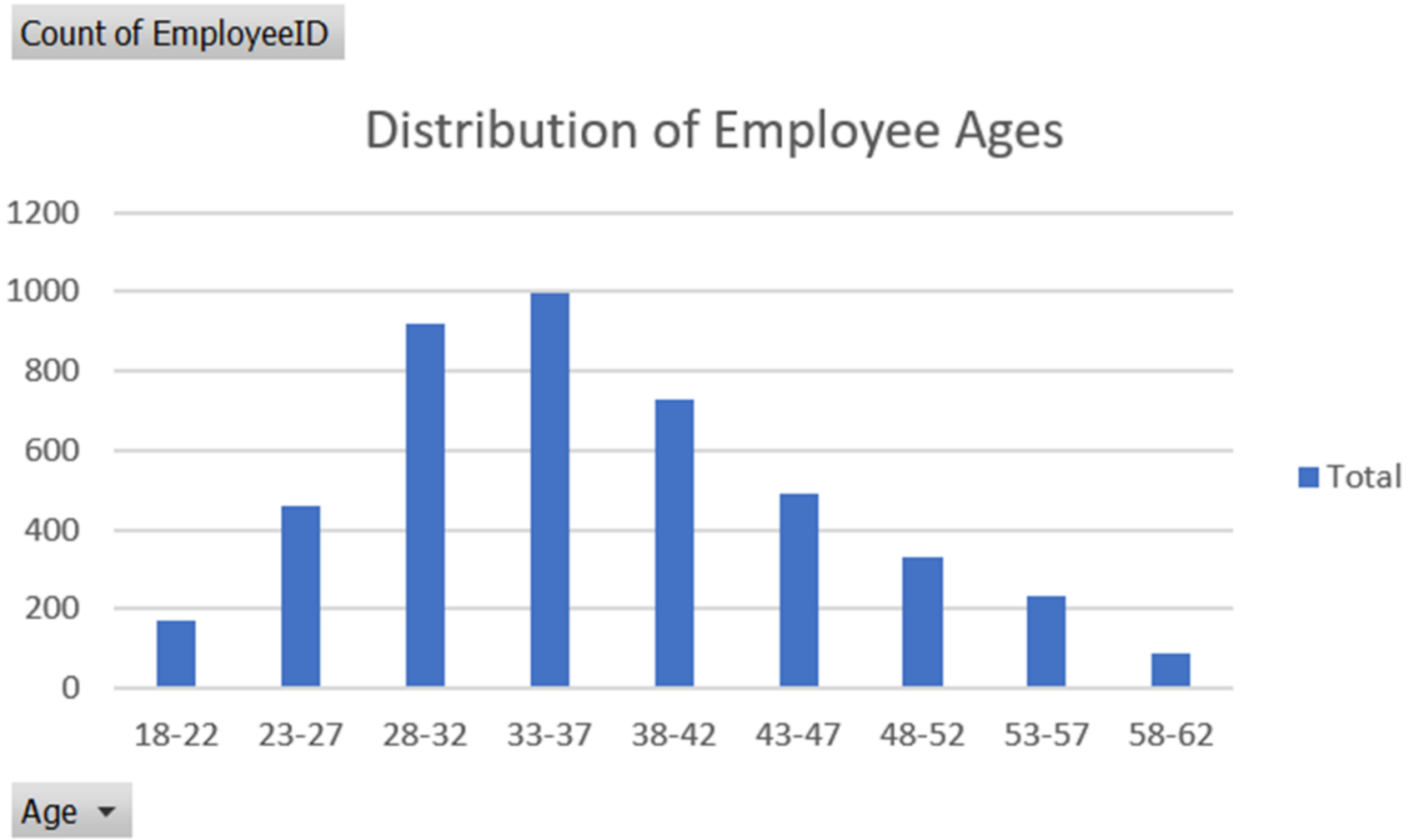
N1 | MonthlyIncome

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	Age	Attrition	BusinessTravel	Department	Dista	Edu	EducationField	Emp	Em	Gender	Jobl	JobRole	MaritalStatus	Monthl	NumCo	Ove
2	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	1	Female	1	Healthcare Representative	Married	131160	1	Y
3	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1	2	Female	1	Research Scientist	Single	41890	0	Y
4	32	No	Travel_Frequently	Research & Development	17	4	Other	1	3	Male	4	Sales Executive	Married	193280	1	Y
5	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	4	Male	3	Human Resources	Married	83210	3	Y
6	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	5	Male	1	Sales Executive	Single	23420	4	Y
7	46	No	Travel_Rarely	Research & Development	8	3	Life Sciences	1	6	Female	4	Research Director	Married	40710	3	Y
8	28	Yes	Travel_Rarely	Research & Development	11	2	Medical	1	7	Male	2	Sales Executive	Single	58130	2	Y
9	29	No	Travel_Rarely	Research & Development	18	3	Life Sciences	1	8	Male	2	Sales Executive	Married	31430	2	Y
10	31	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	9	Male	3	Laboratory Technician	Married	20440	0	Y
11	25	No	Non-Travel	Research & Development	7	4	Medical	1	10	Female	4	Laboratory Technician	Divorced	134640	1	Y
12	45	No	Travel_Rarely	Research & Development	17	2	Medical	1	11	Male	2	Laboratory Technician	Married	79910	0	Y
13	36	No	Travel_Rarely	Research & Development	28	1	Life Sciences	1	12	Male	1	Laboratory Technician	Married	33770	0	Y
14	55	No	Travel_Rarely	Research & Development	14	4	Life Sciences	1	13	Female	1	Sales Executive	Single	55380	0	Y
15	47	Yes	Non-Travel	Research & Development	1	1	Medical	1	14	Male	1	Research Scientist	Married	57620	1	Y
16	28	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	15	Male	1	Manufacturing Director	Married	25920	1	Y
17	37	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	16	Male	2	Healthcare Representative	Married	53460	4	Y
18	21	No	Travel_Rarely	Research & Development	3	2	Life Sciences	1	17	Male	1	Laboratory Technician	Single	42130	1	Y
19	37	No	Non-Travel	Research & Development	1	3	Medical	1	18	Male	2	Sales Executive	Divorced	41270	2	Y
20	35	No	Travel_Rarely	Sales	7	4	Life Sciences	1	19	Male	1	Sales Representative	Divorced	24380	7	Y
21	38	No	Travel_Rarely	Research & Development	8	3	Life Sciences	1	20	Female	1	Manager	Divorced	68700	1	Y
22	26	No	Travel_Frequently	Research & Development	1	4	Other	1	21	Male	2	Laboratory Technician	Divorced	104470	1	Y
23	50	No	Travel_Rarely	Sales	8	4	Life Sciences	1	22	Male	1	Research Scientist	Divorced	96670	3	Y
24	53	No	Travel_Rarely	Research & Development	11	4	Life Sciences	1	23	Female	2	Research Scientist	Married	21480	3	Y

Sheet1 | general_data

Ready | Accessibility: Unavailable | Average: 65029.31293 | Count: 4411 | Sum: 286779270 | 100%

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

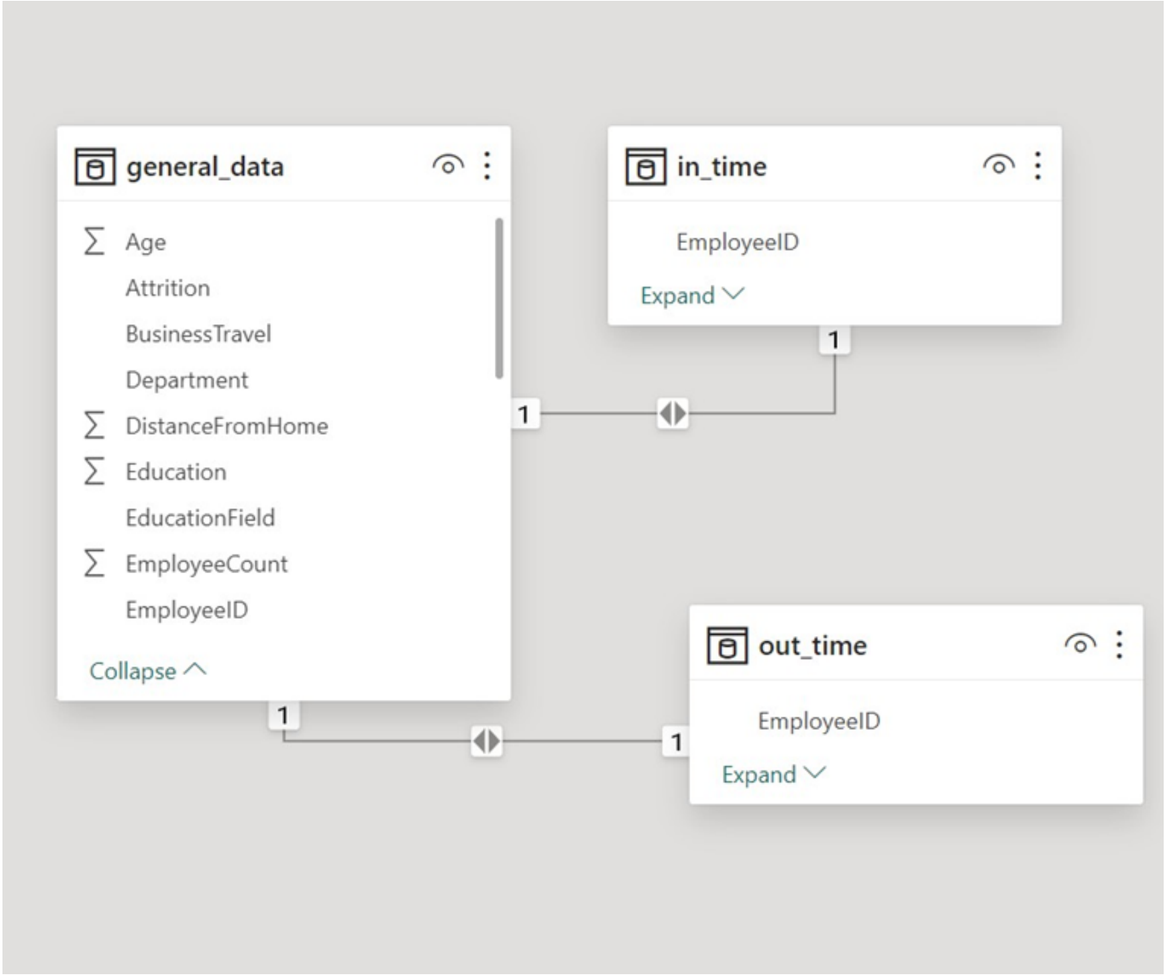


5 Identify and clean any missing or inconsistent data in the "Department" column.

- *I didn’t find any missing or inconsistent data in the "Department" column.*

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

- *Import both employee data and time tracking data into Power BI.In Power BI Desktop, go to the "Data" view by selecting the "Data" icon on the left side.Ensure that both tables have a column named "EmployeeID."*
- *Click on the "Model" icon on the left side to switch to the data model view.Drag the "EmployeeID" field from the employee data table and drop it onto the "EmployeeID" field in the time tracking data table.*
- *A window will appear. Confirm that the relationship is based on the correct fields. Adjust any additional settings as needed, such as cardinality and cross-filter direction.*
- *Click "OK" to close the relationship window.*
- *Save your Power BI file and apply the changes.*



7

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

- *AverageYearsWithManager = Average(‘general_data'[YearsWithCurr Manager])*

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

Count of EmployeeID	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**

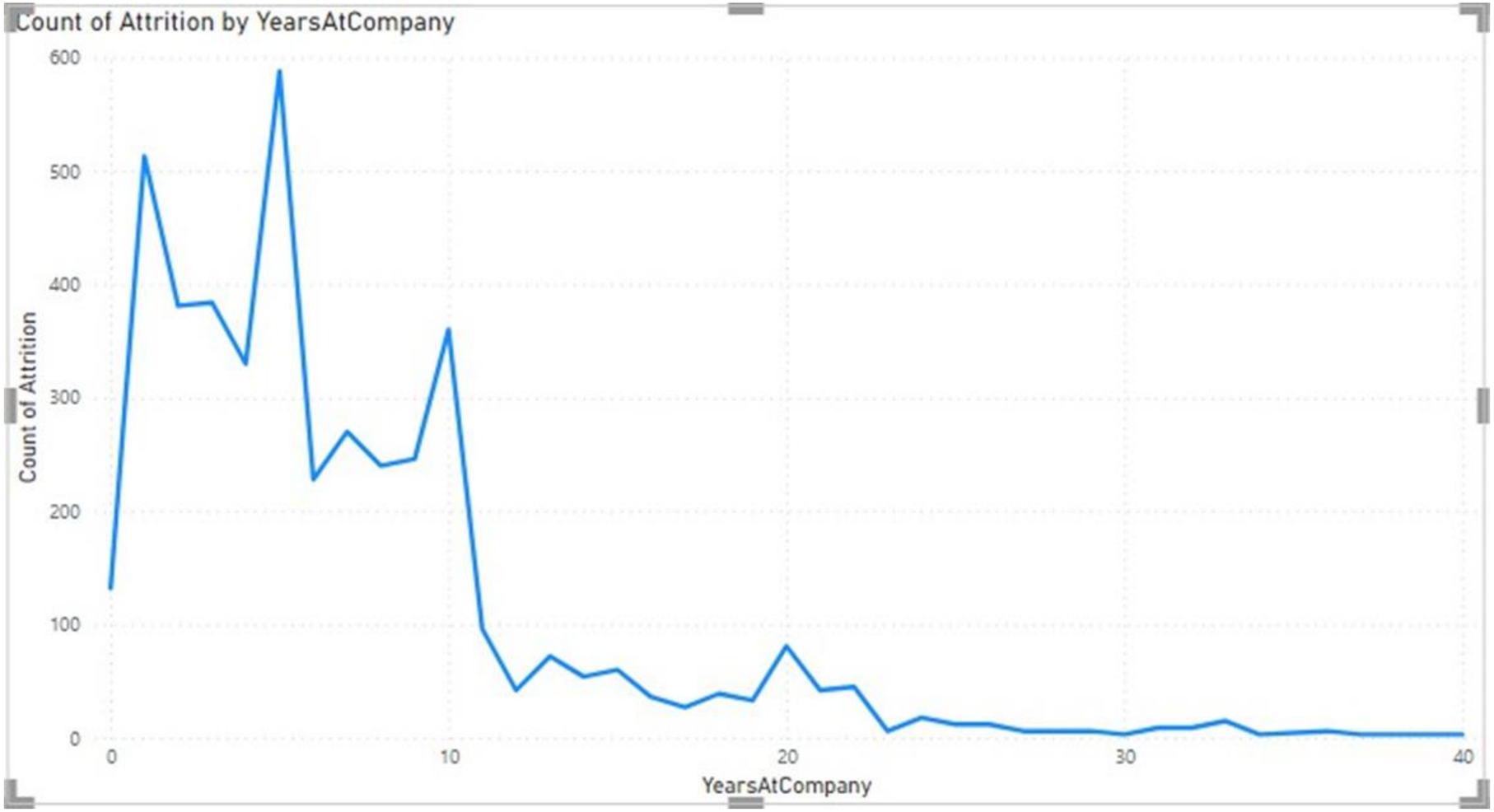
9 Apply conditional formatting to highlight employees with both above-average Monthly Income and above-average Job Satisfaction

	A	B	C	D	E
	Employee	Environme	JobSatisfaction	WorkLifeBalance	
	1	3	4	2	
	2	3	2	4	
	3	2	2	1	
	4	4	4	3	
	5	4	1	3	
	6	3	2	2	
	7	1	3	1	
	8	1	2	3	
	9	2	4	3	
	10	2	1	3	
	11	3	4	3	
	12	NA	4	3	
	13	4	1	3	
	14	1	2	2	
	15	4	4	2	
	16	3	4	4	
	17	4	3	4	
	18	1	4	3	
	19	2	2	2	
	20	1	1	3	
	21	3	2	1	
	22	1	2	2	
	23	3	3	2	

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
	Age	Attritio	Busines	Departı	Distanc	Educati	Educati	Employ	Employ	Gender	JobLeve	JobRole	Marital	Monthl	NumCo	Over18	Percent	Standa	StockO	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	Married	40710	3	Y	13	8	0	
	28	Yes	Travel_Ra	Research	11	2	Medical	1	7	Male	2	Sales Exec	Single	58130	2	Y	20	8	1	
	29	No	Travel_Ra	Research	18	3	Life Scienc	1	8	Male	2	Sales Exec	Married	31430	2	Y	22	8	3	
	31	No	Travel_Ra	Research	1	3	Life Scienc	1	9	Male	3	Laborator	Married	20440	0	Y	21	8	0	
	25	No	Non-Trave	Research	7	4	Medical	1	10	Female	4	Laborator	Divorced	134640	1	Y	13	8	1	
	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	
	28	No	Travel_Ra	Research	1	3	Life Scienc	1	15	Male	1	Manufacti	Married	25920	1	Y	14	8	0	
	37	No	Travel_Ra	Research	1	3	Life Scienc	1	16	Male	2	Healthcar	Married	53460	4	Y	11	8	0	
	21	No	Travel_Ra	Research	3	2	Life Scienc	1	17	Male	1	Laborator	Single	42130	1	Y	12	8	3	
	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	
	26	No	Travel_Fre	Research	1	4	Other	1	21	Male	2	Laborator	Divorced	104470	1	Y	18	8	0	
	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	

Sheet1 general_data +

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.

Creating a star schema involves organizing your dataset into a central fact table and surrounding it with dimension tables, forming a star-like structure.

Benefits of Star Schema:

- *Star schemas are simple and easy to understand. Users can quickly grasp the structure, enhancing data comprehension and ease of use.*
- *Queries often perform better in a star schema because it reduces the number of joins needed. Dimension tables are small and typically have simple relationships with the fact table, leading to faster query response times.*
- *Star schemas are scalable and can handle large datasets efficiently. They provide a foundation for growth as more data is added to the system.*
- *Star schemas are flexible and can be easily modified or expanded. Adding new dimensions or measures does not significantly impact the existing structure, making it easier to maintain.*
- *By separating dimensions and facts, data integrity is improved. Changes in dimension attributes are less likely to impact the fact table, maintaining data consistency.*
- *Star schemas align well with business concepts, making it easier for business users to relate to and navigate the data. This alignment facilitates better decision-making.*
- *Many Business Intelligence (BI) tools are designed to work effectively with star schemas, simplifying the development of reports and dashboards.*

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

```
Rolling 3-Month Avg =
CALCULATE( AVERAGE(Employee[MonthlyIncome]), FILTER(
ALL(Employee),
Employee[EmployeeID] = EARLIER(Employee[EmployeeID]) &&
Employee[Date] <= EARLIER(Employee[Date]) &&
Employee[Date] > DATEADD(EARLIER(Employee[Date]), -3, MONTH) ) )
```

This formula calculates the average monthly income for each employee over the past 3 months based on the date column.

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?

- *Parameterized queries can be set up using Power Query, to allow users to filter data based on the Distance from Home column using parameters.*
- *Load your data, create a new parameter in the Power Query Editor for DistanceParameter, and set its datatype to decimal/whole number.*
- *In the Power Query Editor, locate the query that loads your data and add a filter step to filter the data based on the parameter. = Table.SelectRows(YourPreviousStep, each [DistanceFromHome] <= DistanceParameter)*
- *Click "Close & Apply" in the Home tab to apply the changes. 5)In your Power BI report, create a slicer visual or any other method for users to input the parameter value and manually update the filter condition in your visual to reference the parameter*

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.

Sum of MonthlyIncome	JobLevel			
Department	3	4	5	Grand Total
Human Resources	1648500	754800	855840	3259140
Research & Development	28117740	15277290	10107870	53502900
Sales	11792400	8753070	2428860	22974330
Grand Total	**41558640**	**24785160**	**13392570**	**79736370**

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.

To perform a What-If analysis:

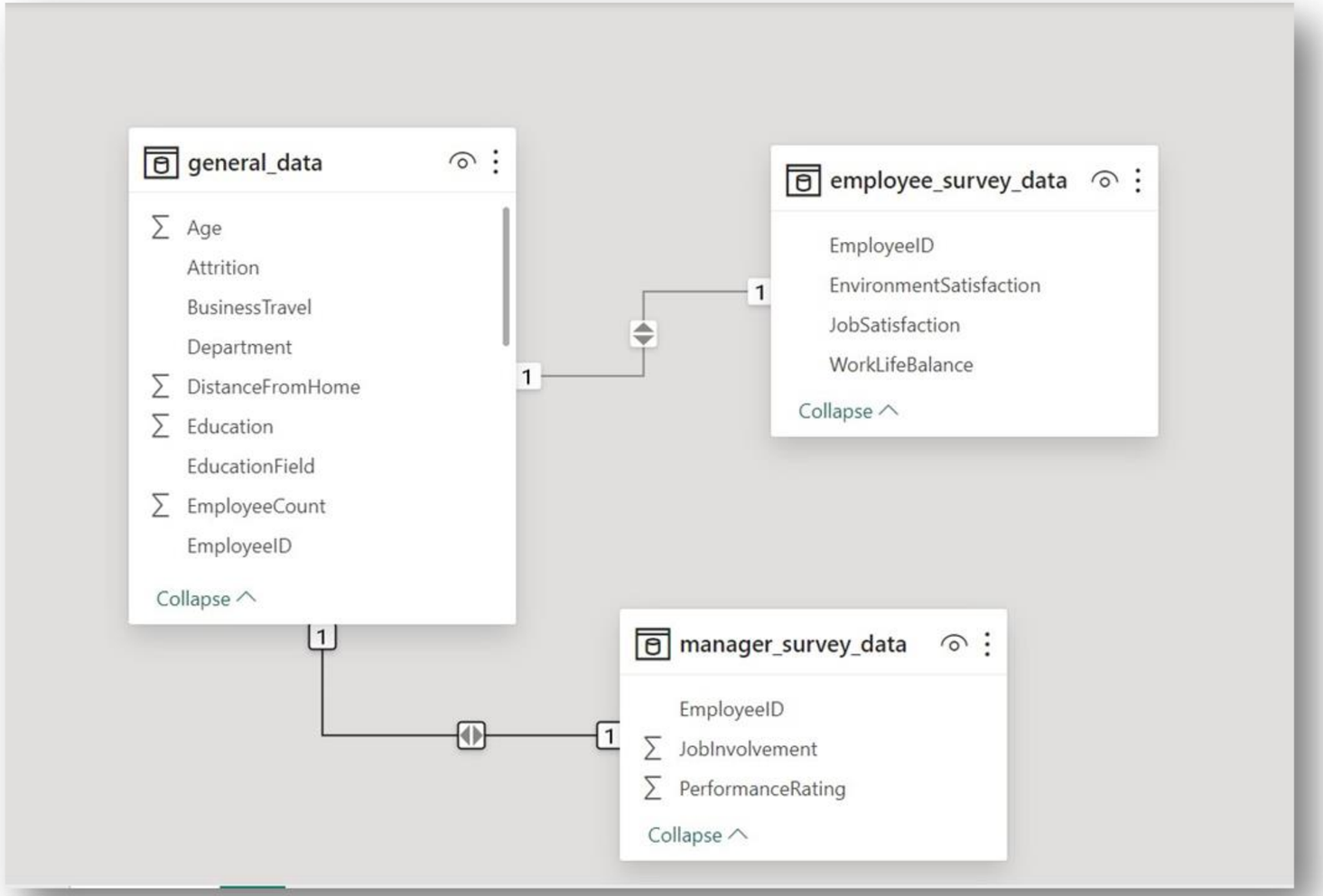
- *In a cell (let's say A1),enter the initial Percent Salary Hike value (e.g., 5%), or else, highlight the cells where the percent hike is stored.*
- *In another cell, calculate the Monthly Income based on the formula that includes the Percent Salary Hike.*
- Now that you have both the columns 'Monthly Income' and 'Percent Salary Hike', drag down the cell with the initial Percent Salary Hike to, say, A10. Excel will fill in the remaining values.
- We can see how Monthly Income changes with different Percent Salary Hike values.
- Likewise, you can also go to Data Tab in Excel and click on 'What-If analysis' and select 'Data Table'. Choose the cell with the formula for Monthly Income as the "Column input cell" and play with different Percent Salary Hike values, and Excel will show you the corresponding Monthly Income

17

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

- *Verifying if data adheres to a predefined schema involves checking whether the actual data in a dataset aligns with the expected structure and rules outlined in the predefined schema.*
- *As per current status of the data, there is a need to reorder EmployeeID column, changing data type of 'TotalWorkingHours' column, flling NA values and blank values from general_data, employee_survey_data, manager_survey_data.*
- *There is 'EmployeeID' label missing in the 'in-time' and 'out-time' data.*
- *It is very important to address these inconsistencies in the data files and check if data profiling is done carefully and the data is validated correctly.*
- *The dataset would then conform to the predefined schema following the resolution of all identified inconsistencies and the implementation of necessary data quality measures.*

Data Modeling for HR Data Analysis



HR Analysis Dashboard

JobRole	0	1	2	3	4
Sales Representative	0	12	3	9	12
Sales Executive	0	27	24	72	42
Research Scientist	0	48	48	33	30
Research Director	0	18	3	27	9
Manufacturing Director	0	21	9	3	15
Manager	1	11	3	15	12
Laboratory Technician	0	36	36	36	18
Human Resources	0	6	3	6	6
Healthcare Representative	0	18	9	18	12
Total	1	197	138	219	156