HR DATA ANALYSIS

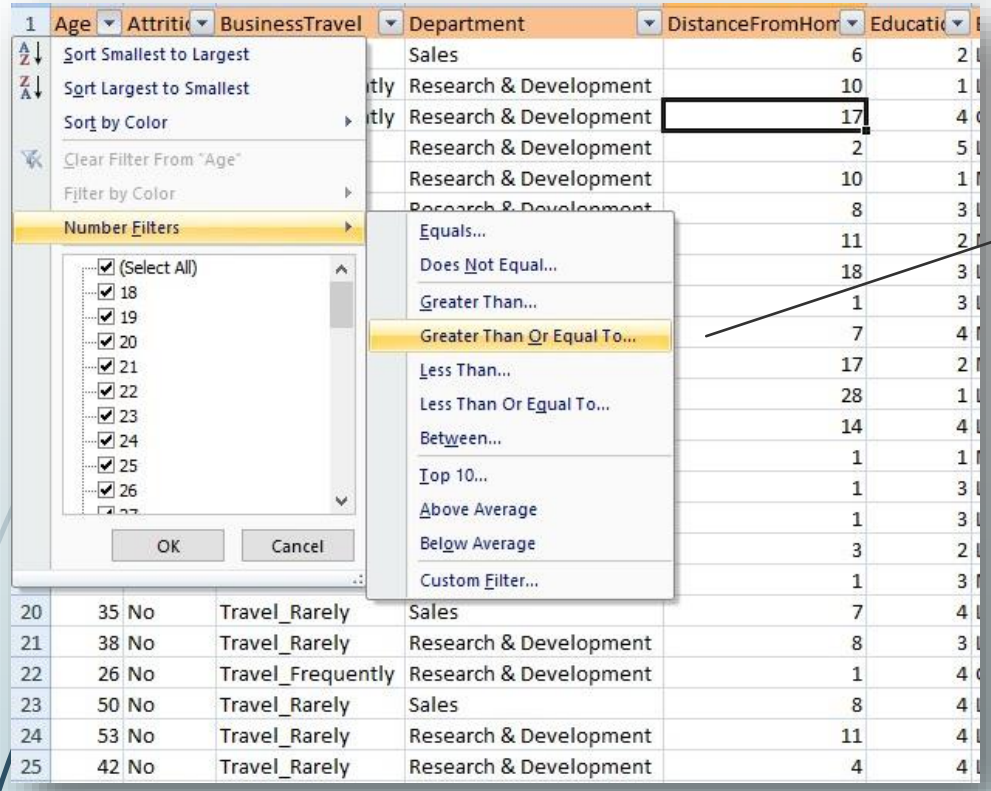
PSYLIQ Internship Project

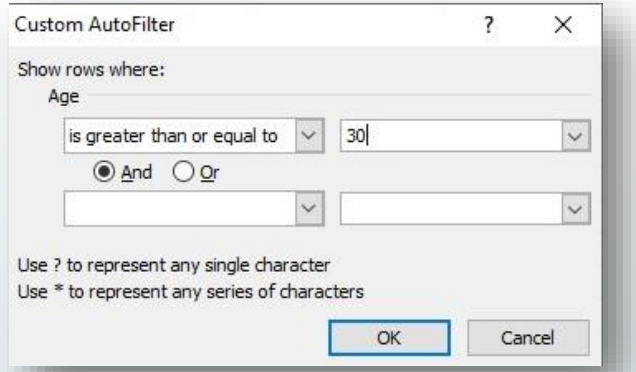
INTRODUCTION

This dataset is provided by PSYLIQ , In this project, I will use the HR data analysis data set to explore various aspects of human resource management and how they affect business outcomes. The data set contains information about 5000 employees from a large company, such as their age, gender, education, department, salary, performance rating, promotion status, and attrition rate.

I will use various tools and methods, such as EXCEL and POWERBI. I will perform data cleaning, data exploration, data visualization, data modelling, and data interpretation. I will also present my findings and insights in a clear and concise report.

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





	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education
2	51	No	Travel_Rarely	Sales	6	2
3	31	Yes	Travel_Frequently	Research & Development	10	1
4	32	No	Travel_Frequently	Research & Development	17	4
5	38	No	Non-Travel	Research & Development	2	5
6	32	No	Travel_Rarely	Research & Development	10	1
7	46	No	Travel_Rarely	Research & Development	8	3
10	31	No	Travel_Rarely	Research & Development	11	2
12	45	No	Travel_Rarely	Research & Development	18	3
13	36	No	Travel_Rarely	Research & Development	1	3
14	55	No	Travel_Rarely	Research & Development	7	4
15	47	Yes	Non-Travel	Research & Development	17	2
17	37	No	Travel_Rarely	Research & Development	28	1
19	37	No	Non-Travel	Research & Development	14	4
20	35	No	Travel_Rarely	Sales	1	1
21	38	No	Travel_Rarely	Research & Development	1	3
23	50	No	Travel_Rarely	Sales	1	3

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

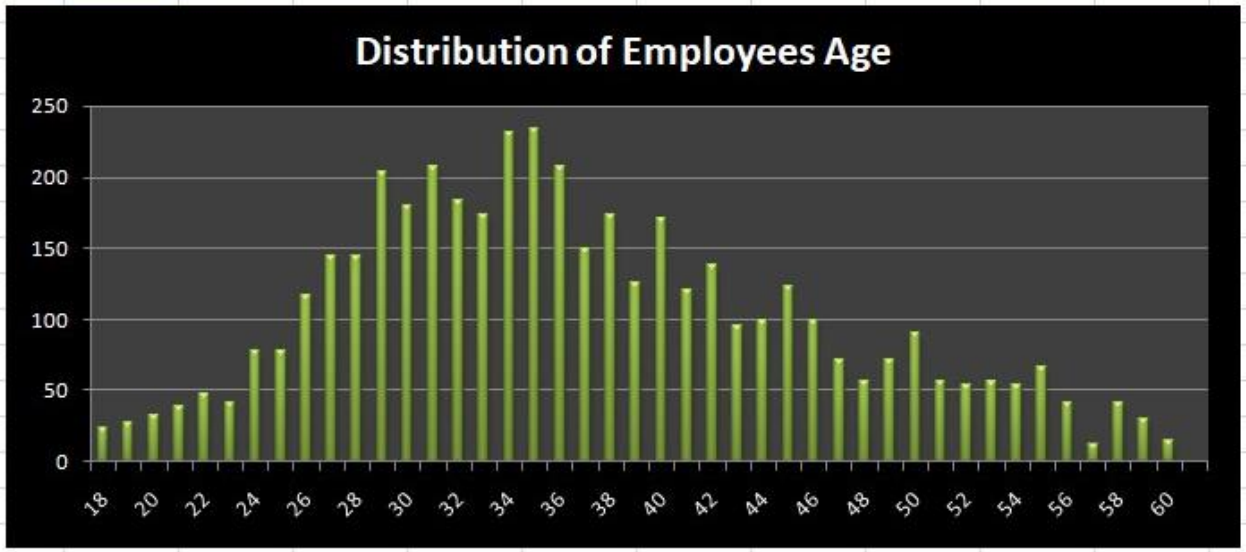
	A	B
2		
3	**Row Labels**	**Average of MonthlyIncome**
4	Healthcare Representative	60983.74
5	Human Resources	58528.08
6	Laboratory Technician	66314.05
7	Manager	63395.88
8	Manufacturing Director	69183.72
9	Research Director	65473.13
10	Research Scientist	64975.68
11	Sales Executive	65186.69
12	Sales Representative	65370.96
13	**Grand Total**	**65029.31**

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

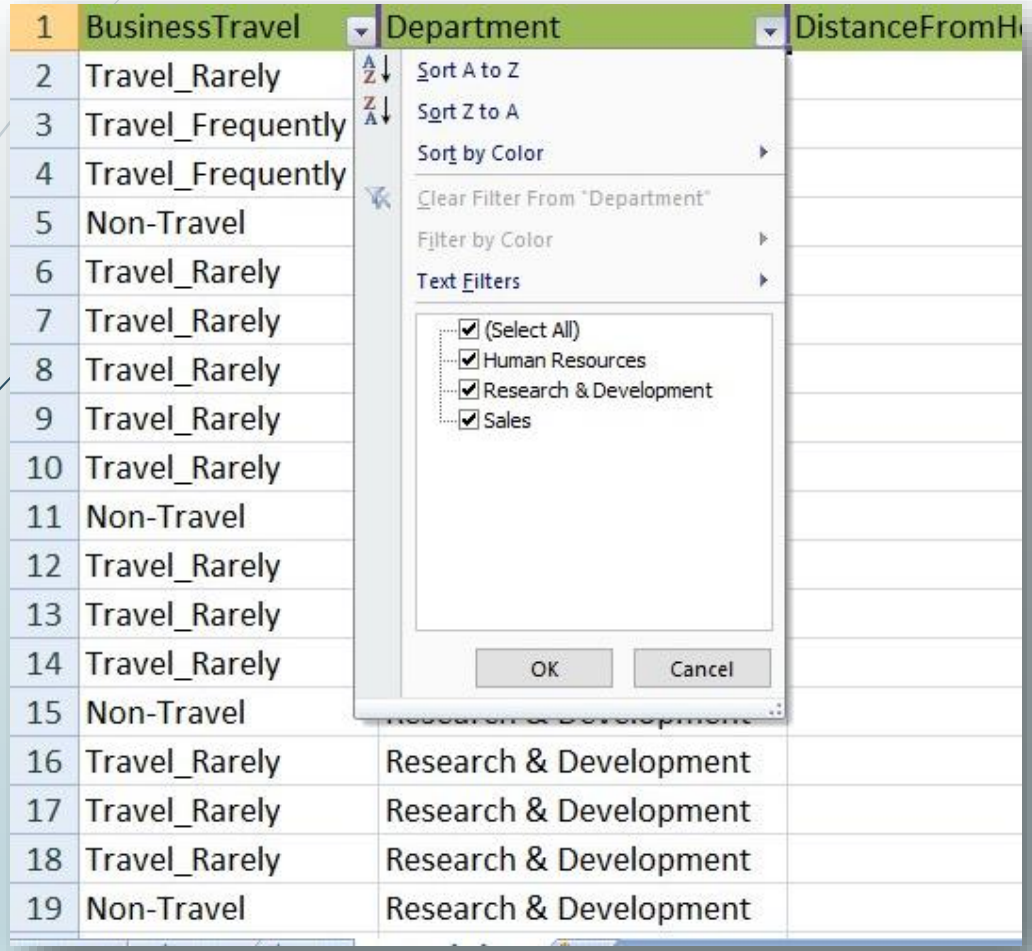
1	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeID	Gender	JobLevel	JobRole	MaritalStatus	MonthlyIncome	Nur
2	6	2	Life Sciences	1	1	Female	1	Healthcare Representative	Married	131160	
3	10	1	Life Sciences	1	2	Female	1	Research Scientist	Single	41890	
4	17	4	Other	1	3	Male	4	Sales Executive	Married	193280	
5	2	5	Life Sciences	1	4	Male	3	Human Resources	Married	83210	
6	10	1	Medical	1	5	Male	1	Sales Executive	Single	23420	
7	8	3	Life Sciences	1	6	Female	4	Research Director	Married	40710	
8	11	2	Medical	1	7	Male	2	Sales Executive	Single	58130	
9	18	3	Life Sciences	1	8	Male	2	Sales Executive	Married	31430	
10	1	3	Life Sciences	1	9	Male	3	Laboratory Technician	Married	20440	
11	7	4	Medical	1	10	Female	4	Laboratory Technician	Divorced	134640	
12	17	2	Medical	1	11	Male	2	Laboratory Technician	Married	79910	
13	28	1	Life Sciences	1	12	Male	1	Laboratory Technician	Married	33770	
14	14	4	Life Sciences	1	13	Female	1	Sales Executive	Single	55380	
15	1	1	Medical	1	14	Male	1	Research Scientist	Married	57620	
16	1	3	Life Sciences	1	15	Male	1	Manufacturing Director	Married	25920	
17	1	3	Life Sciences	1	16	Male	2	Healthcare Representative	Married	53460	
18	3	2	Life Sciences	1	17	Male	1	Laboratory Technician	Single	42130	
19	1	3	Medical	1	18	Male	2	Sales Executive	Divorced	41270	
20	7	4	Life Sciences	1	19	Male	1	Sales Representative	Divorced	24380	
21	8	3	Life Sciences	1	20	Female	1	Manager	Divorced	68700	
22	1	4	Other	1	21	Male	2	Laboratory Technician	Divorced	104470	
23	8	4	Life Sciences	1	22	Male	1	Research Scientist	Divorced	96670	

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

Age	Count 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

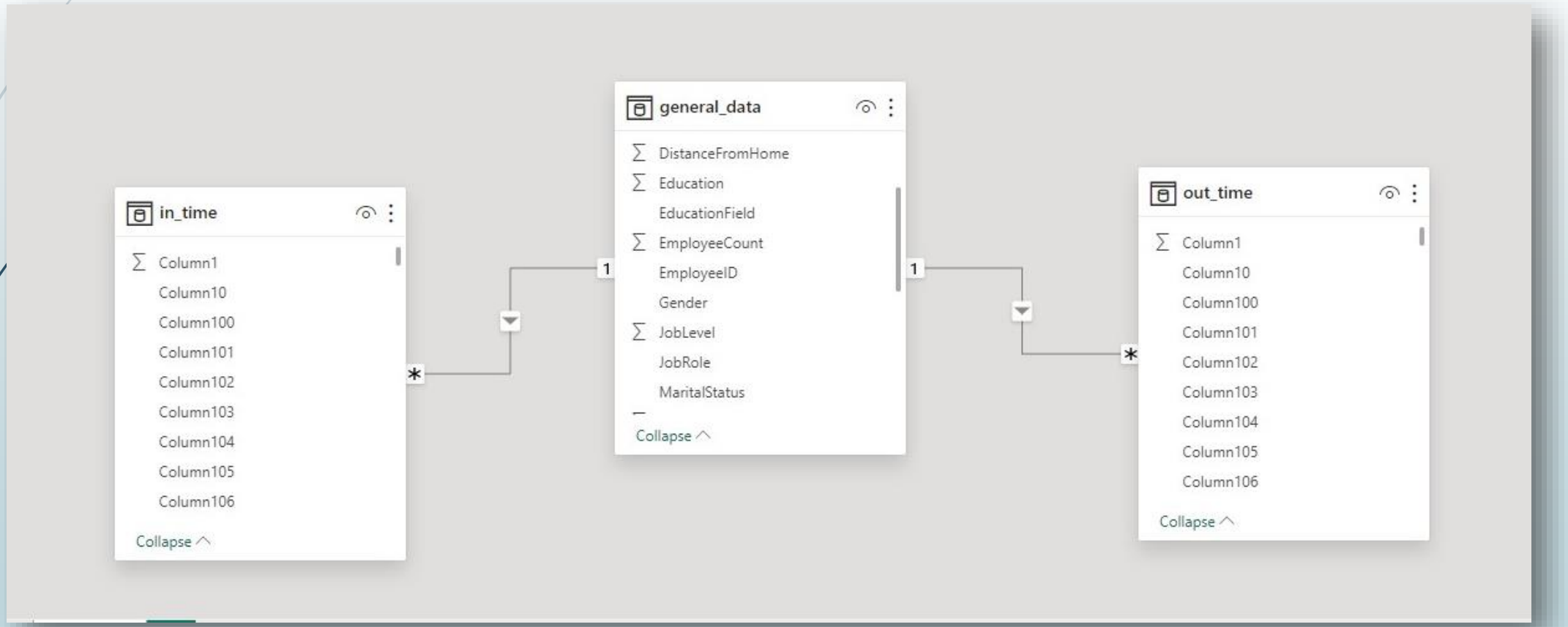


5. Identify and clean any missing or inconsistent data in the "Department" column.



There no 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.



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

```
1 Average_Year_wih_current_manager = AVERAGE(general_data[YearsWithCurrManager])
```

...tionLevel	TotalWorkingYears	TrainingTimesLastYear	YearsAtCompany	YearsSinceLastPromotion	YearsWithCurrManager	Average_Year_wih_current_manager
0	1	6	1	0	0	4.12
1	1	3	1	0	0	4.12
1	1	2	1	0	0	4.12
0	1	3	1	0	0	4.12
2	1	2	1	0	0	4.12
0	1	3	1	0	0	4.12
0	1	4	1	0	0	4.12
2	1	3	1	0	0	4.12
3	1	2	1	0	0	4.12
2	1	2	1	0	0	4.12
2	1	2	1	0	0	4.12
1	1	2	1	0	0	4.12
1	1	5	1	0	0	4.12
1	1	3	1	0	0	4.12
1	1	3	1	0	0	4.12
0	1	2	1	0	0	4.12
1	1	3	1	0	0	4.12
1	1	5	1	0	0	4.12
0	1	1	1	0	0	4.12
0	1	3	1	0	0	4.12
0	1	3	1	0	0	4.12
0	1	2	1	0	0	4.12

Data

Search

- general_data
  - Age
  - Attrition
  - Average_Year_wih_current_ma...
  - BusinessTravel
  - Department
  - DistanceFromHome
  - Education
  - EducationField
  - EmployeeCount
  - EmployeeID
  - Gender
  - JobLevel
  - JobRole
  - MaritalStatus
  - MonthlyIncome
  - NumCompaniesWorked
  - Over18
  - PercentSalaryHike

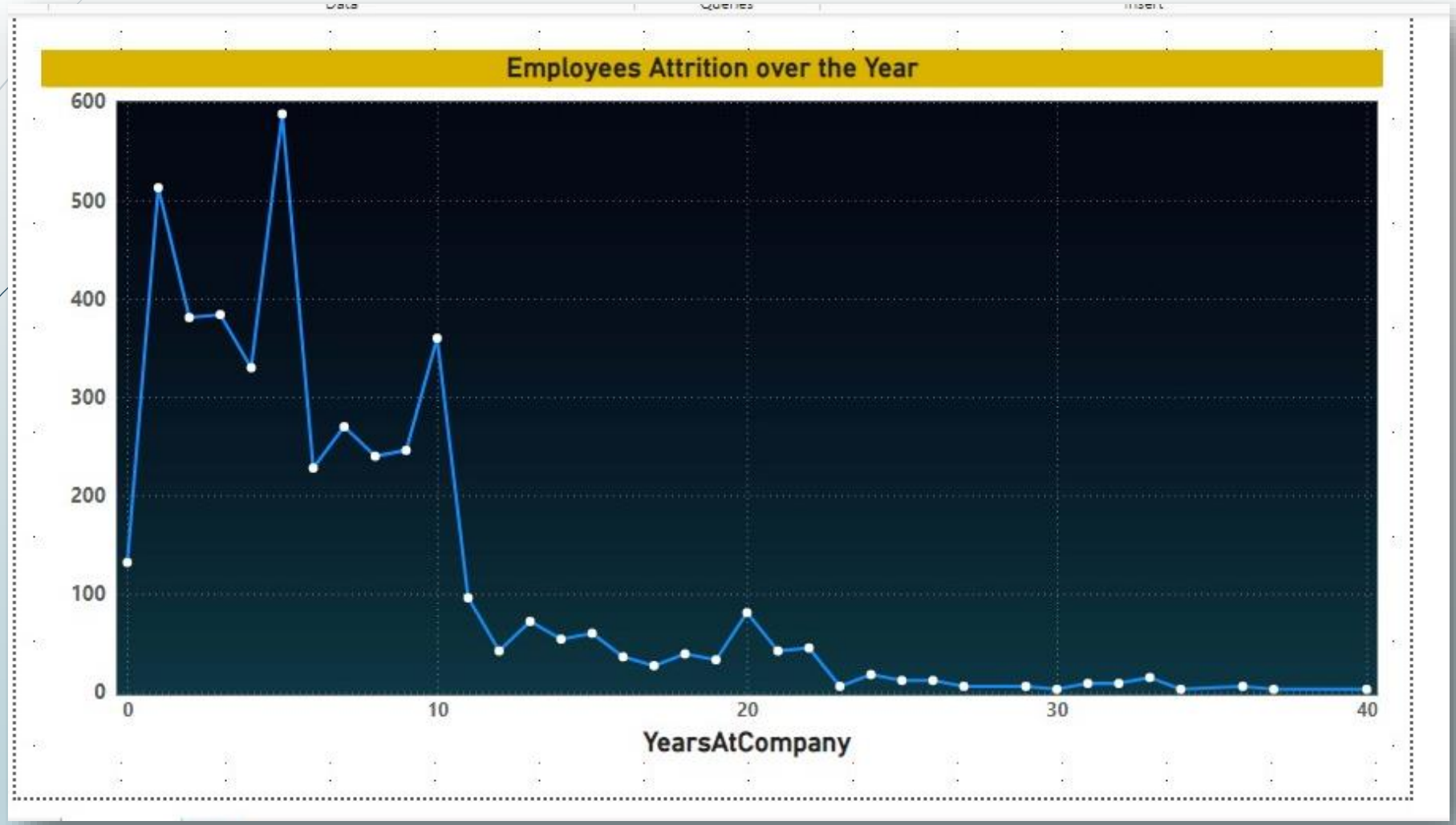
8. Using Excel, create a pivot table that displays the count of employees in each Marital Status category, segmented by Department.

	A	B	C	D	E
2					
3	**Departments**	**Marital Status**			
4		**Divorced**	**Married**	**Single**	**Grand Total**
5	Human Resources	21	96	72	189
6	Research & Development	621	1350	912	2883
7	Sales	339	573	426	1338
8	**Grand Total**	**981**	**2019**	**1410**	**4410**

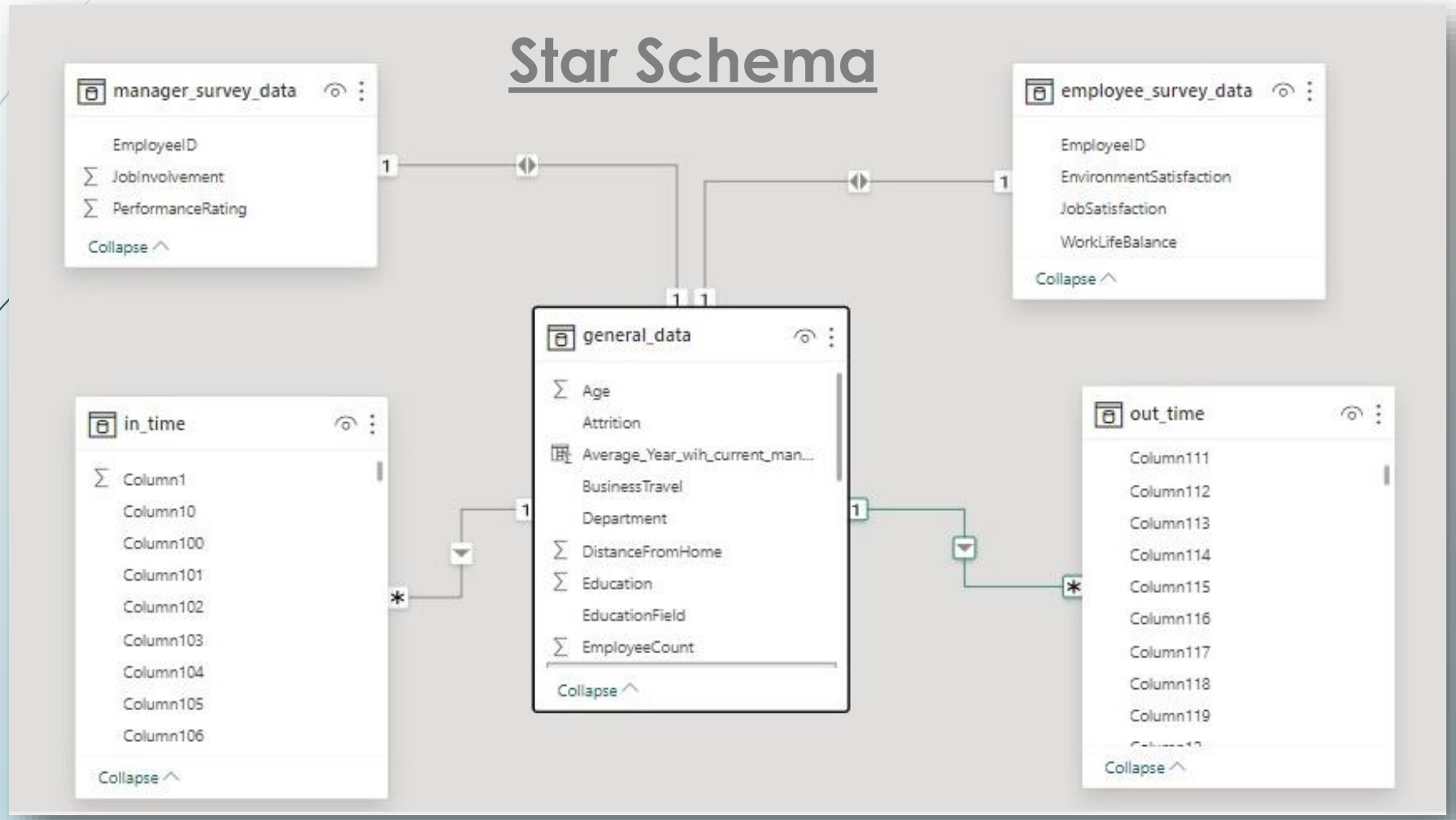
9. Apply conditional formatting to highlight employees with both above-average Monthly Income and above-average Job Satisfaction.

	M	N	O	
	JobSatisfaction	MaritalStatus	MonthlyIncome	NumCo
ative	4	Married	131160	
	2	Single	41890	
	2	Married	193280	
	4	Married	83210	
	1	Single	23420	
	2	Married	40710	
	3	Single	58130	
	2	Married	31430	
	4	Married	20440	
	1	Divorced	134640	
	4	Married	79910	
	4	Married	33770	
	1	Single	55380	
	2	Married	57620	
r	4	Married	25920	
ative	4	Married	53460	
	3	Single	42130	
	4	Divorced	41270	
	2	Divorced	24380	
	1	Divorced	68700	
	2	Divorced	104470	
	2	Divorced	96670	
	3	Married	21480	
r	3	Married	89260	

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.



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.

```
RollingAvg =
CALCULATE(AVERAGE('Employee'[MonthlyIncome]),
DATESINPERIOD( 'Employee'[Date],
LASTDATE('Employee'[Date]), -3, MONTH ) )
```

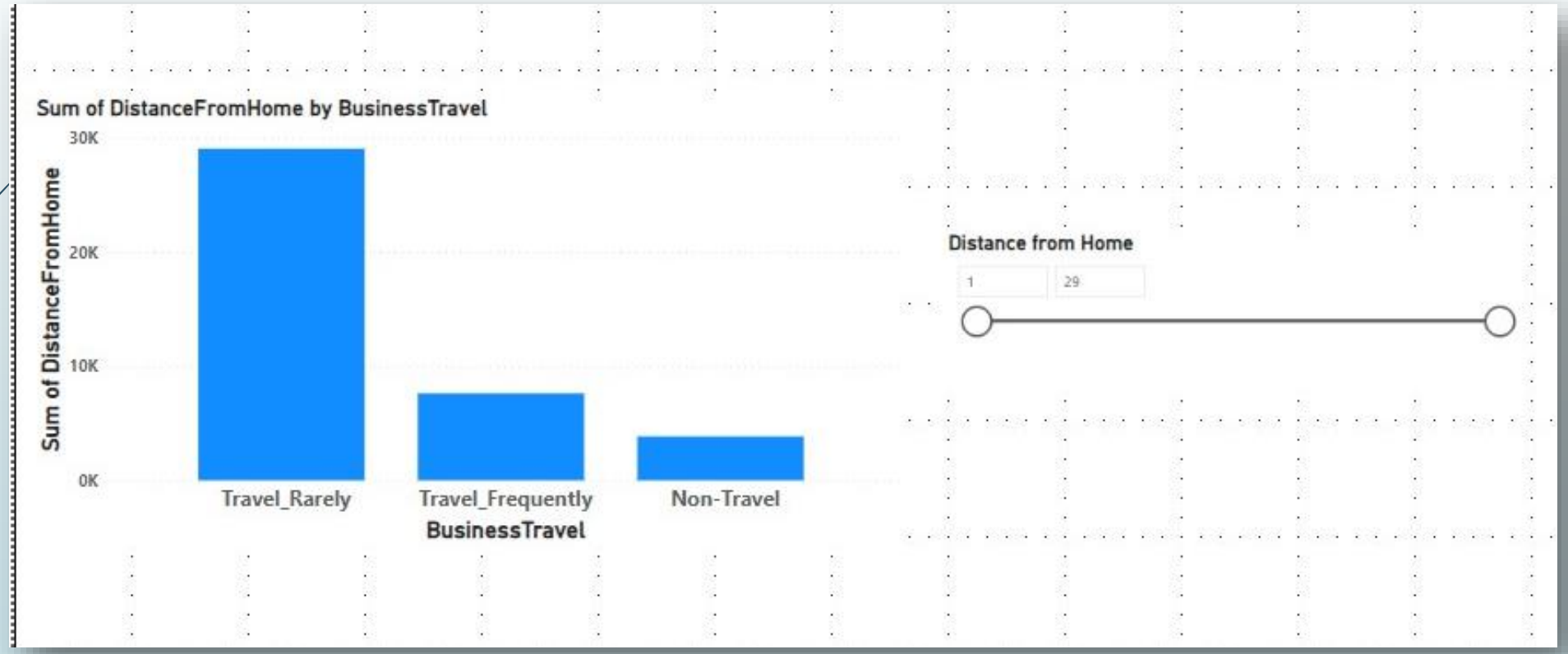
13. Create a hierarchy in Power BI that allows users to drill down from Department to Job Role to further narrow their analysis

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.

Department	Count of EmployeeID
⊞ Human Resources	189
⊞ Sales	1338
⊞ Research & Development	2883

Department	Count of EmployeeID
⊟ **Human Resources**	
Human Resources	3
Research Director	3
Healthcare Representative	9
Manager	9
Sales Representative	12
Manufacturing Director	24
Research Scientist	36
Laboratory Technician	39
Sales Executive	54
⊞ **Sales**	**1338**
⊞ **Research & Development**	**2883**

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

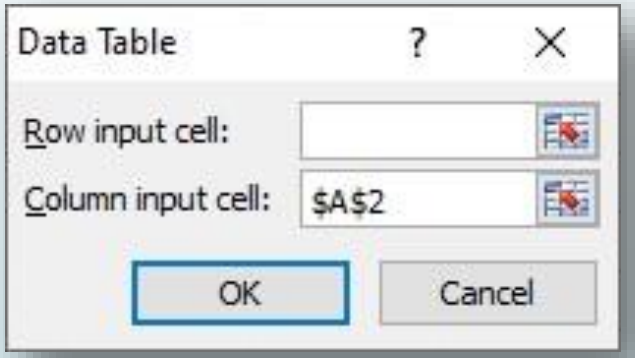


15. In Excel, calculate the total Monthly Income for each Department, considering only the employees with a Job Level greater than or equal to 3.

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.

	A	B
1	MonthlyIncome	Percent Salary Hike
2	131160	=A2*(1+10%)
3	41890	
4	193280	
5	83210	
6	23420	
7	40710	
8	58130	
9	31430	
10	20440	
11	134640	
12	79910	
13	33770	
14	55380	
15	57620	
16	25920	
17	53460	
18	42130	
19	41270	
20	24380	
21	68700	
22	104470	
23	96670	
24	21480	
25	89260	



	A	B
1	MonthlyIncome	Percent Salary Hike
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
23	96670	106337
24	21480	23628
25	89260	98186

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

HR DATA ANALYTICS DASHBOARD [POWER BI]

Data Insights and Summary

- Data Overview In Our workforce, there are 60% males and 40% females.
- The average monthly income for employees is 65.029K .
- The workforce experiences an average attrition rate of 16.1% coupled with an average salary hike of 15.21%. Moreover, majority of employees needs to travel rarely.
- There are 3 departments named Research & Development, Sales and HR with almost 65% of employees belonging to R&D employee.
- We have highest number of employees in Sales Executive.

THANK YOU!

Vyas Kumar
7717706832
vyaskv123@gmail.com
www.linkedin.com/in/vyas-kumar-969388220