



DATA ANALYST INTERNSHIP

HR DATA ANALYSIS

Prepared by - Rehman Khan

INTRODUCTION

Hello and welcome! to MY EXCEL PROJECT REPORT ON HR DATA. In this project, we've utilized various tools such as MS Excel, Pivot Table and POWER BI. The goal of the HR Data Analysis Project is to learn more about understanding and improving predictions through the use of dashboards and technologies like MS EXCEL and POWER BI. This study will help the business make better decisions and meet the expectations of its personnel.

TASK

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	O	P	Q	R	S	T	U
1	Age	Attrition	Business	Department	Distance	Education	Education	Employee	Employee	Gender	JobLevel	JobRole	MaritalS	Monthly	NumCor	Over18	Percent	Standard	StockOp	TotalWc	Training
2	51	No	Travel_Rar	Sales	6	2	Life Science	1	1	Female	1	Healthcare	Married	131160	1	Y	11	8	0	1	6
3	31	Yes	Travel_Fre	Research &	10	1	Life Science	1	2	Female	1	Research S	Single	41890	0	Y	23	8	1	6	3
4	32	No	Travel_Fre	Research &	17	4	Other	1	3	Male	4	Sales Execu	Married	193280	1	Y	15	8	3	5	2
5	38	No	Non-Travel	Research &	2	5	Life Science	1	4	Male	3	Human Res	Married	83210	3	Y	11	8	3	13	5
6	32	No	Travel_Rar	Research &	10	1	Medical	1	5	Male	1	Sales Execu	Single	23420	4	Y	12	8	2	9	2
7	46	No	Travel_Rar	Research &	8	3	Life Science	1	6	Female	4	Research D	Married	40710	3	Y	13	8	0	28	5
10	31	No	Travel_Rar	Research &	1	3	Life Science	1	9	Male	3	Laboratory	Married	20440	0	Y	21	8	0	10	2
12	45	No	Travel_Rar	Research &	17	2	Medical	1	11	Male	2	Laboratory	Married	79910	0	Y	13	8	2	21	2
13	36	No	Travel_Rar	Research &	28	1	Life Science	1	12	Male	1	Laboratory	Married	33770	0	Y	12	8	2	16	2
14	55	No	Travel_Rar	Research &	14	4	Life Science	1	13	Female	1	Sales Execu	Single	55380	0	Y	17	8	0	37	2
15	47	Yes	Non-Travel	Research &	1	1	Medical	1	14	Male	1	Research S	Married	57620	1	Y	11	8	2	10	4
17	37	No	Travel_Rar	Research &	1	3	Life Science	1	16	Male	2	Healthcare	Married	53460	4	Y	11	8	0	7	2
19	37	No	Non-Travel	Research &	1	3	Medical	1	18	Male	2	Sales Execu	Divorced	41270	2	Y	13	8	1	15	2
20	35	No	Travel_Rar	Sales	7	4	Life Science	1	19	Male	1	Sales Repre	Divorced	24380	7	Y	16	8	0	10	5
21	38	No	Travel_Rar	Research &	8	3	Life Science	1	20	Female	1	Manager	Divorced	68700	1	Y	11	8	1	8	5
23	50	No	Travel_Rar	Sales	8	4	Life Science	1	22	Male	1	Research S	Divorced	96670	3	Y	23	8	0	28	2
24	53	No	Travel_Rar	Research &	11	4	Life Science	1	23	Female	2	Research S	Married	21480	3	Y	11	8	0	21	2
25	42	No	Travel_Rar	Research &	4	4	Life Science	1	24	Male	1	Manufactu	Married	89260	1	Y	14	8	0	NA	4
27	55	No	Travel_Rar	Research &	1	4	Other	1	26	Female	1	Research S	Married	67990	3	Y	11	8	0	12	2
29	37	No	Travel_Rar	Sales	5	1	Marketing	1	28	Male	1	Research S	Single	27050	1	Y	11	8	0	17	2
30	44	Yes	Travel_Fre	Research &	1	2	Medical	1	29	Male	2	Research S	Divorced	103330	3	Y	14	8	1	19	2
31	38	No	Travel_Rar	Sales	2	3	Marketing	1	30	Female	1	Manager	Divorced	44480	9	Y	12	8	0	10	3
34	49	No	Travel_Fre	Research &	1	1	Medical	1	33	Female	2	Research S	Single	35910	9	Y	13	8	0	22	2

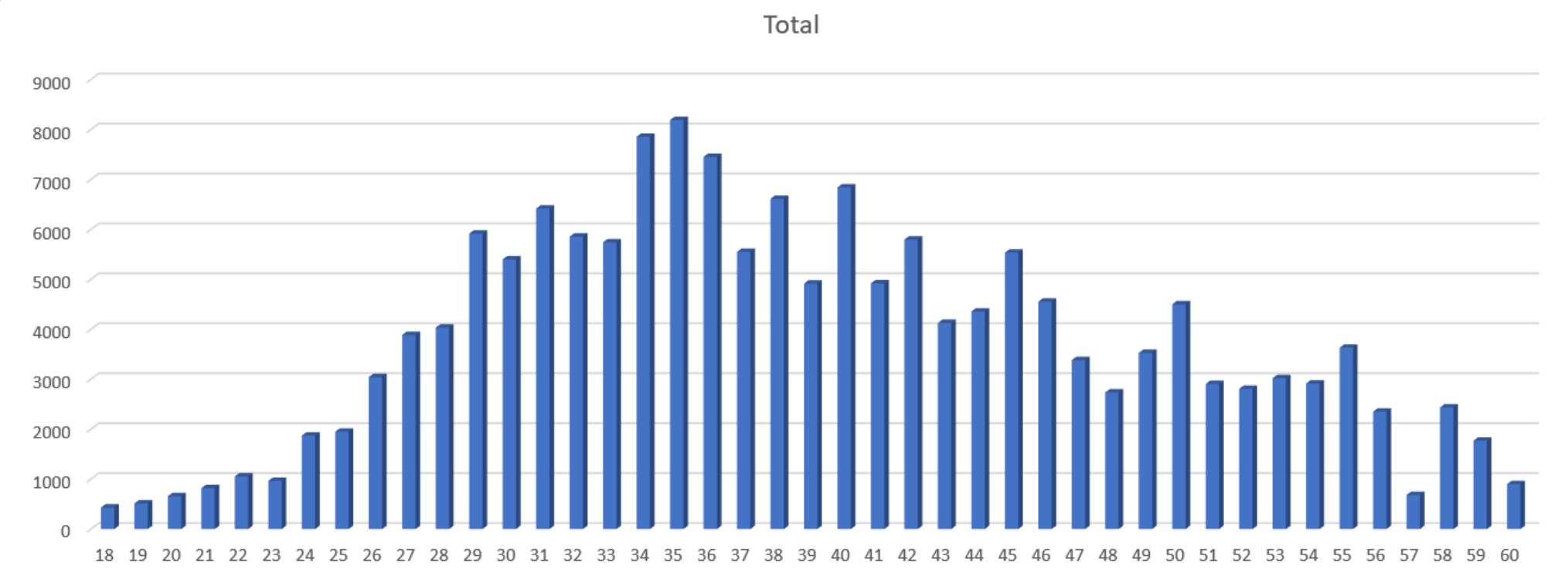
2. Create a pivot table to summarize the average Monthly Income by Job Role.

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

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

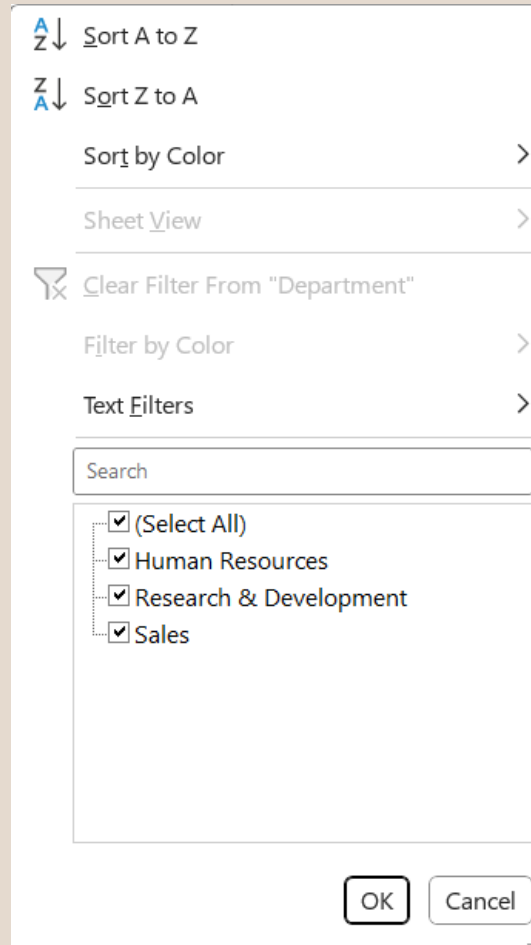
	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	EmployeeCount	EmployeeID	Gender	JobLevel	JobRole	MaritalStatus	MonthlyIncome	NumCompaniesWorked	Over18	PercentSalaryHike	StandardHours	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear
2	1	1	Female	1	Healthcare Representative	Married	131160	1	Y	11	8	0	1	6
3	1	2	Female	1	Research Scientist	Single	41890	0	Y	23	8	1	6	3
4	1	3	Male	4	Sales Executive	Married	193280	1	Y	15	8	3	5	2
5	1	4	Male	3	Human Resources	Married	83210	3	Y	11	8	3	13	5
6	1	5	Male	1	Sales Executive	Single	23420	4	Y	12	8	2	9	2
7	1	6	Female	4	Research Director	Married	40710	3	Y	13	8	0	28	5
8	1	7	Male	2	Sales Executive	Single	58130	2	Y	20	8	1	5	2
9	1	8	Male	2	Sales Executive	Married	31430	2	Y	22	8	3	10	2
10	1	9	Male	3	Laboratory Technician	Married	20440	0	Y	21	8	0	10	2
11	1	10	Female	4	Laboratory Technician	Divorced	134640	1	Y	13	8	1	6	2
12	1	11	Male	2	Laboratory Technician	Married	79910	0	Y	13	8	2	21	2
13	1	12	Male	1	Laboratory Technician	Married	33770	0	Y	12	8	2	16	2
14	1	13	Female	1	Sales Executive	Single	55380	0	Y	17	8	0	37	2
15	1	14	Male	1	Research Scientist	Married	57620	1	Y	11	8	2	10	4
16	1	15	Male	1	Manufacturing Director	Married	25920	1	Y	14	8	0	5	2
17	1	16	Male	2	Healthcare Representative	Married	53460	4	Y	11	8	0	7	2
18	1	17	Male	1	Laboratory Technician	Single	42130	1	Y	12	8	3	3	3
19	1	18	Male	2	Sales Executive	Divorced	41270	2	Y	13	8	1	15	2
20	1	19	Male	1	Sales Representative	Divorced	24380	7	Y	16	8	0	10	5
21	1	20	Female	1	Manager	Divorced	68700	1	Y	11	8	1	8	5
22	1	21	Male	2	Laboratory Technician	Divorced	104470	1	Y	18	8	0	6	3
23	1	22	Male	1	Research Scientist	Divorced	96670	3	Y	23	8	0	28	2
24	1	23	Female	2	Research Scientist	Married	21480	3	Y	11	8	0	21	2
25	1	24	Male	1	Manufacturing Director	Married	89260	1	Y	14	8	0	NA	4
26	1	25	Male	1	Laboratory Technician	Single	65130	1	Y	11	8	1	10	2

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.

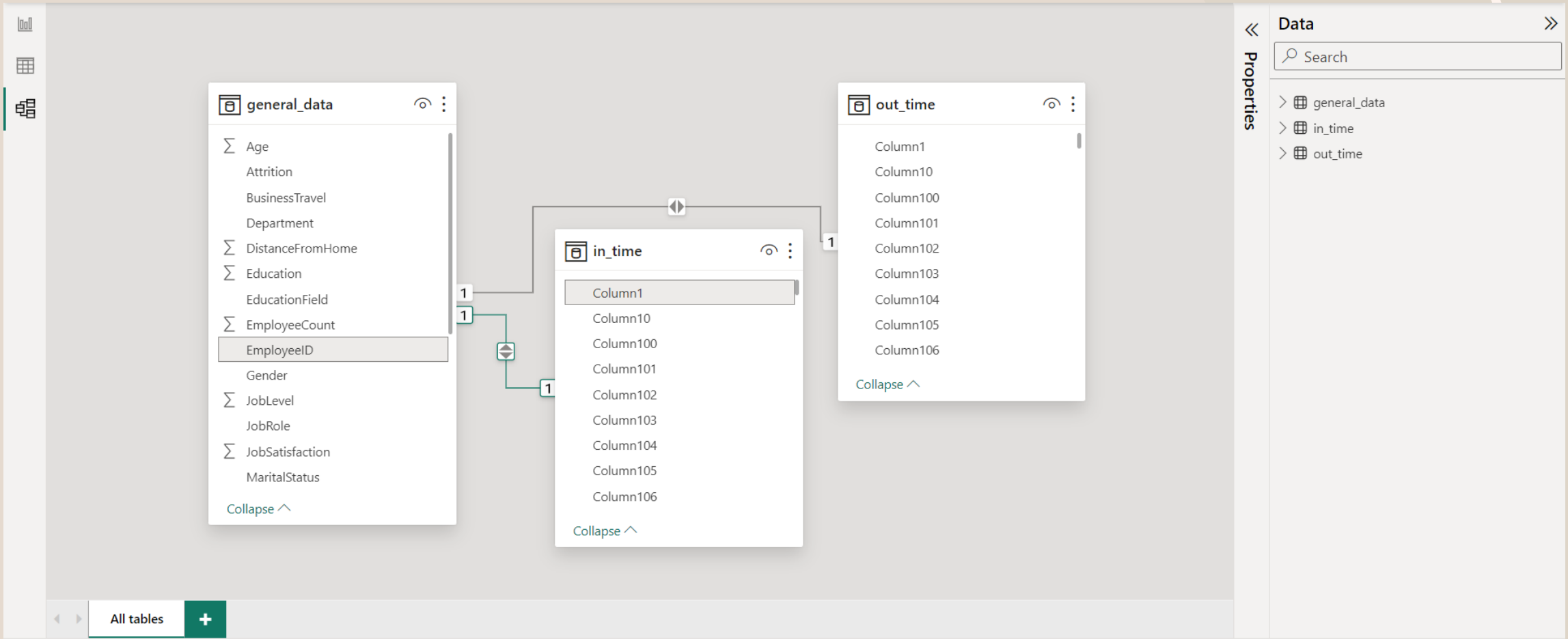
Neither missing nor inconsistent data in the departments column.



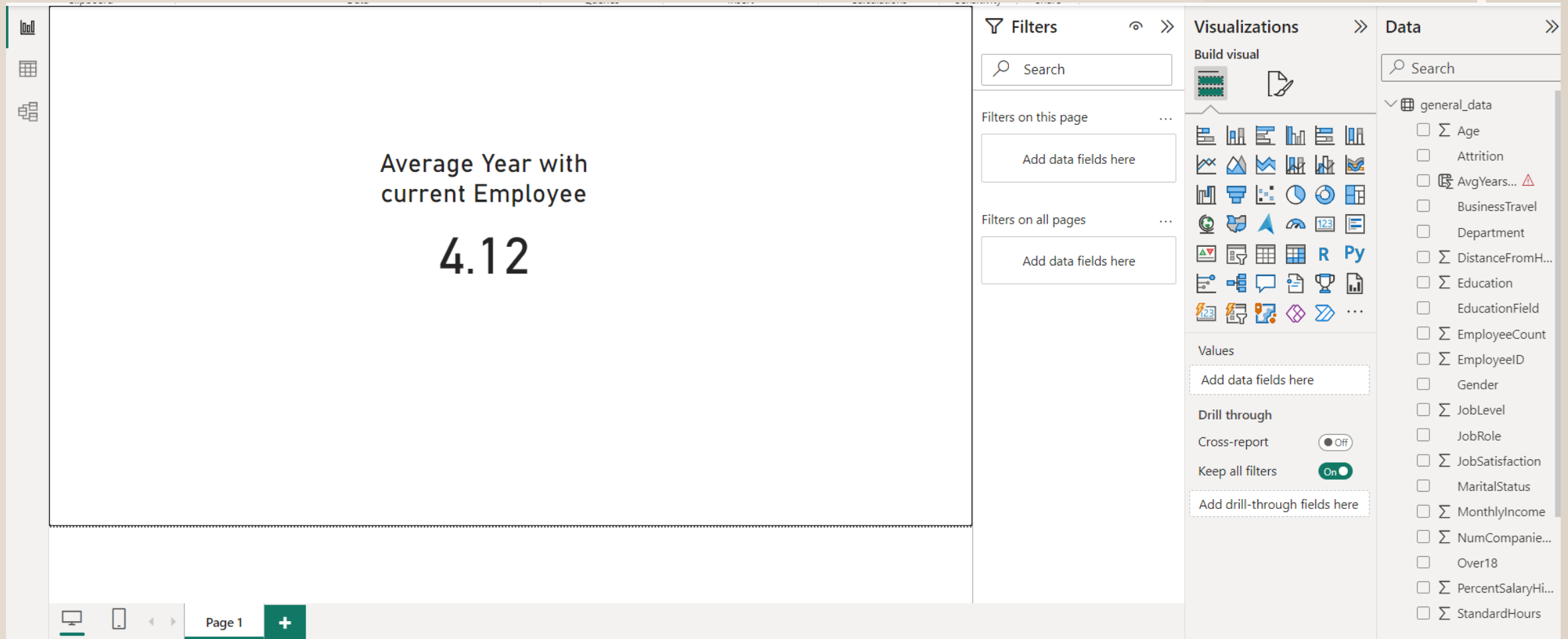
The image shows the Excel Filter menu for the "Department" column. The menu includes options for sorting (A to Z, Z to A), sorting by color, and sheet view. It also shows the "Clear Filter From 'Department'" option, which is currently selected. Below this, there are options for "Filter by Color" and "Text Filters". A search bar is present, and a list of departments is shown with checkboxes: (Select All), Human Resources, Research & Development, and Sales. All checkboxes are checked. At the bottom, there are "OK" and "Cancel" buttons.

- ☒ (Select All)
- ☒ Human Resources
- ☒ Research & Development
- ☒ Sales

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.



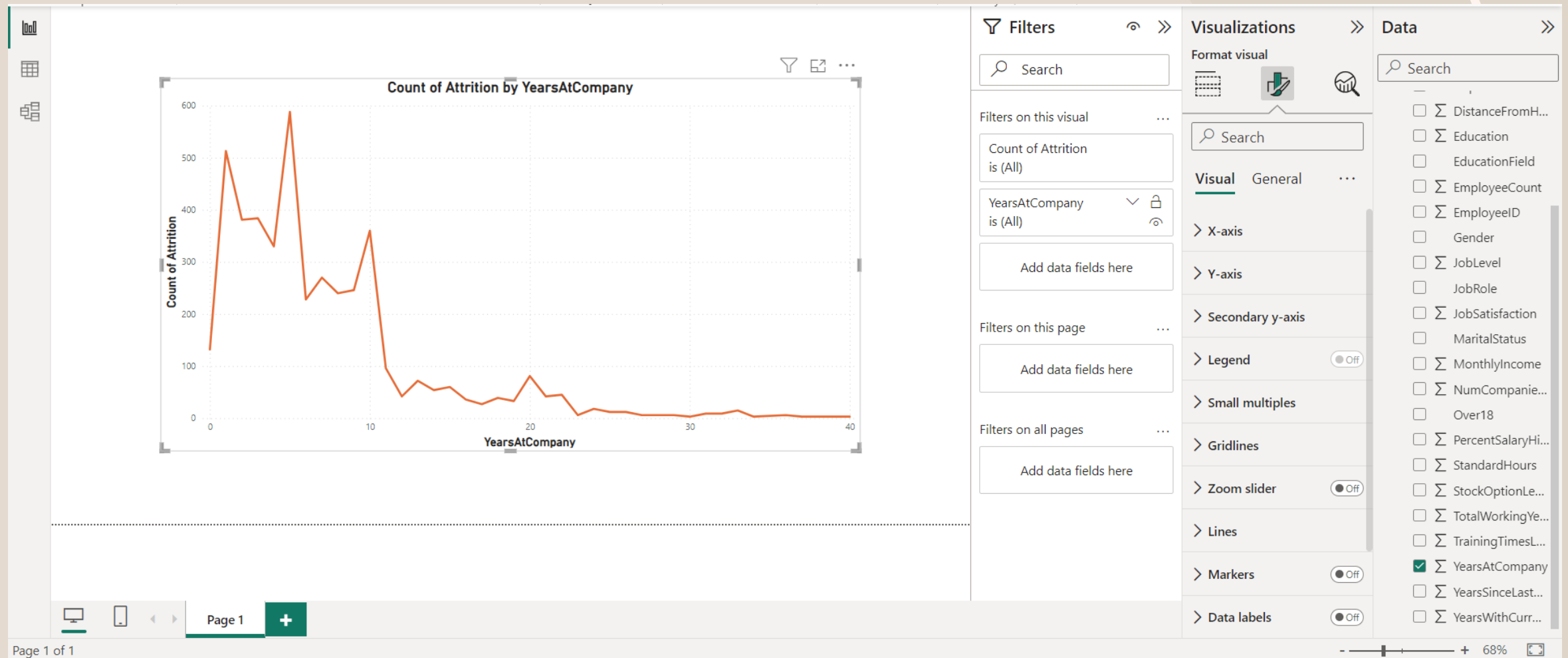
8. Using Excel, create a pivot table that displays the count of employees in each Marital Status category, segmented by Department.

Row Labels	Count of Department
Divorced	981
Human Resources	21
Research & Development	621
Sales	339
Married	2019
Human Resources	96
Research & Development	1350
Sales	573
Single	1410
Human Resources	72
Research & Development	912
Sales	426
Grand Total	4410

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

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	EmployeeID	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Gender	JobLevel	JobRole	MaritalStatus	MonthlyIncome	JobSatisfaction	NumCompaniesWorked	Over18	Pe
2	1	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	Female	1	Healthcare Representative	Married	131160	4		1 Y	
3	2	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1	Female	1	Research Scientist	Single	41890	2		0 Y	
4	3	32	No	Travel_Frequently	Research & Development	17	4	Other	1	Male	4	Sales Executive	Married	193280	2		1 Y	
5	4	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	Male	3	Human Resources	Married	83210	4		3 Y	
6	5	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	Male	1	Sales Executive	Single	23420	1		4 Y	
7	6	46	No	Travel_Rarely	Research & Development	8	3	Life Sciences	1	Female	4	Research Director	Married	40710	2		3 Y	
8	7	28	Yes	Travel_Rarely	Research & Development	11	2	Medical	1	Male	2	Sales Executive	Single	58130	3		2 Y	
9	8	29	No	Travel_Rarely	Research & Development	18	3	Life Sciences	1	Male	2	Sales Executive	Married	31430	2		2 Y	
10	9	31	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	Male	3	Laboratory Technician	Married	20440	4		0 Y	
11	10	25	No	Non-Travel	Research & Development	7	4	Medical	1	Female	4	Laboratory Technician	Divorced	134640	1		1 Y	
12	11	45	No	Travel_Rarely	Research & Development	17	2	Medical	1	Male	2	Laboratory Technician	Married	79910	4		0 Y	
13	12	36	No	Travel_Rarely	Research & Development	28	1	Life Sciences	1	Male	1	Laboratory Technician	Married	33770	4		0 Y	
14	13	55	No	Travel_Rarely	Research & Development	14	4	Life Sciences	1	Female	1	Sales Executive	Single	55380	1		0 Y	
15	14	47	Yes	Non-Travel	Research & Development	1	1	Medical	1	Male	1	Research Scientist	Married	57620	2		1 Y	
16	15	28	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	Male	1	Manufacturing Director	Married	25920	4		1 Y	
17	16	37	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	Male	2	Healthcare Representative	Married	53460	4		4 Y	
18	17	21	No	Travel_Rarely	Research & Development	3	2	Life Sciences	1	Male	1	Laboratory Technician	Single	42130	3		1 Y	
19	18	37	No	Non-Travel	Research & Development	1	3	Medical	1	Male	2	Sales Executive	Divorced	41270	4		2 Y	
20	19	35	No	Travel_Rarely	Sales	7	4	Life Sciences	1	Male	1	Sales Representative	Divorced	24380	2		7 Y	
21	20	38	No	Travel_Rarely	Research & Development	8	3	Life Sciences	1	Female	1	Manager	Divorced	68700	1		1 Y	
22	21	26	No	Travel_Frequently	Research & Development	1	4	Other	1	Male	2	Laboratory Technician	Divorced	104470	2		1 Y	
23	22	50	No	Travel_Rarely	Sales	8	4	Life Sciences	1	Male	1	Research Scientist	Divorced	96670	2		3 Y	
24	23	53	No	Travel_Rarely	Research & Development	11	4	Life Sciences	1	Female	2	Research Scientist	Married	21480	3		3 Y	
25	24	42	No	Travel_Rarely	Research & Development	4	4	Life Sciences	1	Male	1	Manufacturing Director	Married	89260	3		1 Y	
26	25	29	No	Travel_Frequently	Research & Development	16	4	Medical	1	Male	1	Laboratory Technician	Single	65130	4		1 Y	
27	26	55	No	Travel_Rarely	Research & Development	1	4	Other	1	Female	1	Research Scientist	Married	67990	4		3 Y	
28	27	26	No	Travel_Frequently	Research & Development	9	3	Life Sciences	1	Female	1	Manager	Married	162910	1		1 Y	
29	28	37	No	Travel_Rarely	Sales	5	1	Marketing	1	Male	1	Research Scientist	Single	27050	4		1 Y	
30	29	44	Yes	Travel_Frequently	Research & Development	1	2	Medical	1	Male	2	Research Scientist	Divorced	103330	3		3 Y	
31	30	38	No	Travel_Rarely	Sales	2	3	Marketing	1	Female	1	Manager	Divorced	44480	4		9 Y	
32	31	26	Yes	Travel_Rarely	Research & Development	4	3	Medical	1	Male	3	Research Scientist	Divorced	68540	2		2 Y	

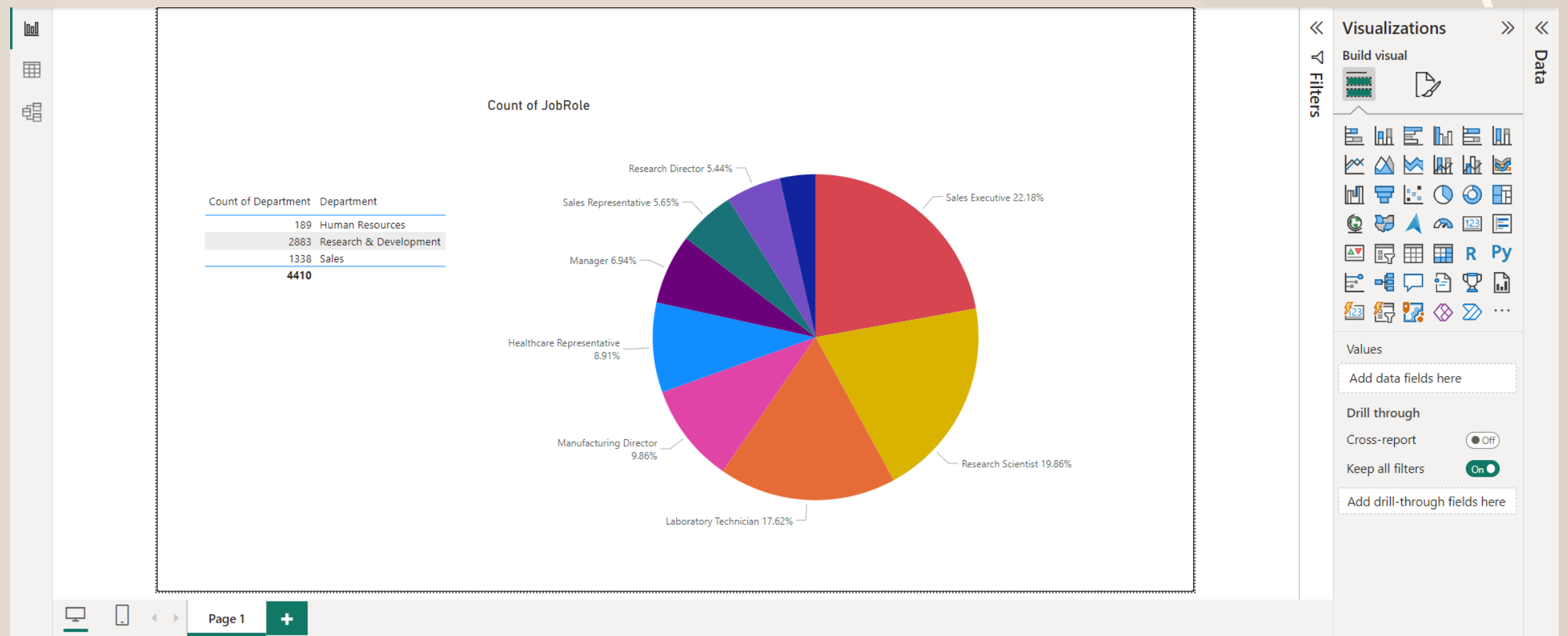
10. In Power BI, create a line chart that visualizes the trend of Employee Attrition over the years.



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

A data warehouse schema that consists of a center fact table surrounded by dimension tables is known as a star schema. It entails organizing data into a central fact table that is linked to dimension tables. This may include a fact table with measurements like Employee ID, Monthly Income, and Job Satisfaction for an employee dataset, as well as dimension tables for employees, departments, and time. Improved query performance, simplicity, scalability, and analysis flexibility make it easier to comprehend and maintain.

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



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

Parametrized queries can be created in POWERBI using POWER QUERY, By filtering data based on the DISTANCE FROM HOME COLUMN utilizing many criteria.

- Step 1: Create a Parameter
- Step 2: Modify Query to Use the Parameter
- Step 3: Use the Parameter in Visualizations
- Step 4: Test the Report

14. 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		Column Labels			
Row Labels		Human Resources	Research & Development	Sales	Grand Total
3		1648500	28117740	11792400	41558640
4		754800	15277290	8753070	24785160
5		855840	10107870	2428860	13392570
Grand Total		3259140	53502900	22974330	79736370

15. 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.

	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1																					
2																					
3																					
4	144276	131160	41890	193280	83210																
5	0%	131160	41890	193280	83210																
6	5%	137718	43985%	202944	87371%																
7	10%	144276	46079%	212608	91531%																
8	15%	150834	48174%	222272	95692%																
9																					
10																					
11																					
12																					
13																					
14																					
15																					
16																					
17																					
18																					
19																					
20																					
21																					
22																					
23																					
24																					

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

1. Review the Predefined Schema:

Examine the predefined schema documentation or specifications to understand the expected structure, data types, and relationships.

2. Data Types and Constraints:

Check if the data types of each column match the expected types in the schema.

Verify that constraints such as unique keys and foreign keys are honored.

3. Missing or Extra Columns:

Confirm that all required fields in the schema are present in the dataset.

Identify and handle any extra columns that are not part of the predefined schema.

4. Null Values:

Examine null or missing values. Decide whether these nulls are acceptable or if they need to be addressed.

5. Relationships:

If there are relationships between tables, ensure that foreign key constraints are satisfied. Verify that the relationships align with the schema.

6. Duplicate Records:

Check for duplicate records. Identify and handle any duplicate entries that violate uniqueness constraints.

7. Outliers and Anomalies:

Examine numerical values for outliers that might be errors. Use statistical methods to identify values that fall outside expected ranges.

8. Date and Time Consistency:

For date and time fields, check for consistency in formatting and validate that the dates fall within a reasonable range.



Thank You

Rehman Khan

ark.khan0123@gmail.com

<https://www.linkedin.com/in/khanrehman/>