Data Analyst Internship

TASK 1: HR DATA ANALYSIS

BY

AMAN SHAIKH

Q1. Using Excel, how would you filter the dataset to only show employees aged 30 and above?

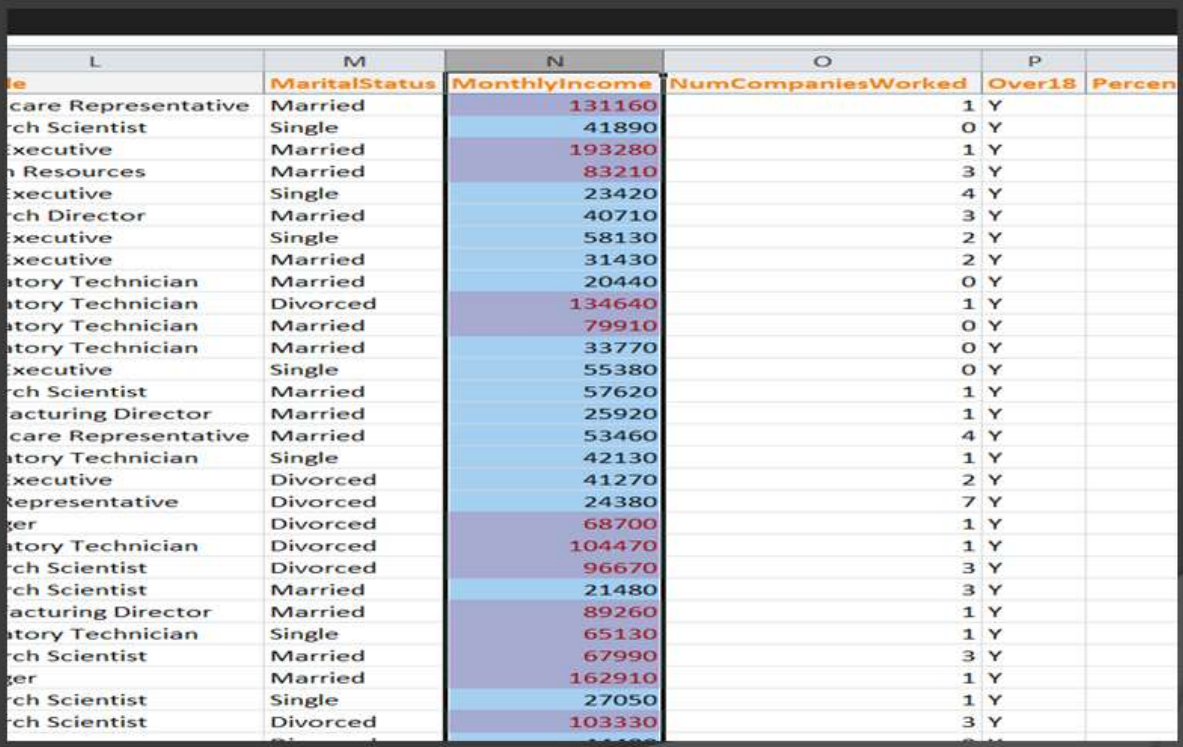
- Open the excel file which contain information about employees.
- Select the age column .
- Apply the shortcut to add filter that is Ctrl+Shift+L
- Click on filter button which will add filter in every column of your dataset.
- Click on the drop down arrow of ' Age' column got to Number Filters and click on greater then equal to option

	A	B	C	D
1	Age	Attritio	Busines	Department
2	51	No	Travel_Rai	Sales
3	31	Yes	Travel_Fre	Research & Developmen
4	32	No	Travel_Fre	Research & Developmen
5	38	No	Non-Trave	Research & Developmen
6	32	No	Travel_Rai	Research & Developmen
7	46	No	Travel_Rai	Research & Developmen
10	31	No	Travel_Rai	Research & Developmen
12	45	No	Travel_Rai	Research & Developmen
13	36	No	Travel_Rai	Research & Developmen
14	55	No	Travel_Rai	Research & Developmen
15	47	Yes	Non-Trave	Research & Developmen
17	37	No	Travel_Rai	Research & Developmen
19	37	No	Non-Trave	Research & Developmen
20	35	No	Travel_Rai	Sales
21	38	No	Travel_Rai	Research & Developmen
23	50	No	Travel_Rai	Sales
24	53	No	Travel_Rai	Research & Developmen
25	42	No	Travel_Rai	Research & Developmen
27	55	No	Travel_Rai	Research & Developmen
29	37	No	Travel_Rai	Sales

Q2. Create a pivot table to summarize the average Monthly Income by Job Role.

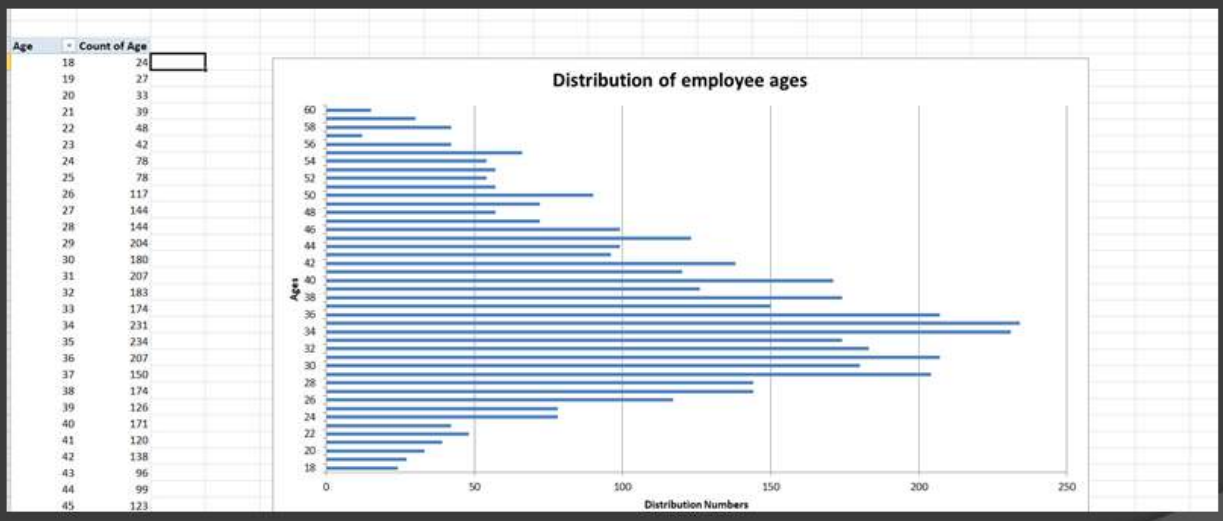
JobRole	Average of MonthlyIncome
Healthcare Representative	60983.74046
Human Resources	58528.07692
Laboratory Technician	66314.05405
Manager	63395.88235
Manufacturing Director	69183.72414
Research Director	65473.125
Research Scientist	64975.68493
Sales Executive	65186.68712
Sales Representative	65370.96386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

Q3. Apply conditional formatting to highlight employees with Monthly Income above the company's average income.



L	M	N	O	P	
le	MaritalStatus	MonthlyIncome	NumCompaniesWorked	Over18	Percen
care Representative	Married	131160	1	Y	
rch Scientist	Single	41890	0	Y	
Executive	Married	193280	1	Y	
n Resources	Married	83210	3	Y	
Executive	Single	23420	4	Y	
rch Director	Married	40710	3	Y	
Executive	Single	58130	2	Y	
Executive	Married	31430	2	Y	
atory Technician	Married	20440	0	Y	
atory Technician	Divorced	134640	1	Y	
atory Technician	Married	79910	0	Y	
atory Technician	Married	33770	0	Y	
Executive	Single	55380	0	Y	
rch Scientist	Married	57620	1	Y	
acturing Director	Married	25920	1	Y	
care Representative	Married	53460	4	Y	
atory Technician	Single	42130	1	Y	
Executive	Divorced	41270	2	Y	
Representative	Divorced	24380	7	Y	
ger	Divorced	68700	1	Y	
atory Technician	Divorced	104470	1	Y	
rch Scientist	Divorced	96670	3	Y	
rch Scientist	Married	21480	3	Y	
acturing Director	Married	89260	1	Y	
atory Technician	Single	65130	1	Y	
rch Scientist	Married	67990	3	Y	
ger	Married	162910	1	Y	
rch Scientist	Single	27050	1	Y	
rch Scientist	Divorced	103330	3	Y	

Q4. Create a bar chart in Excel to visualize the distribution of employee ages.

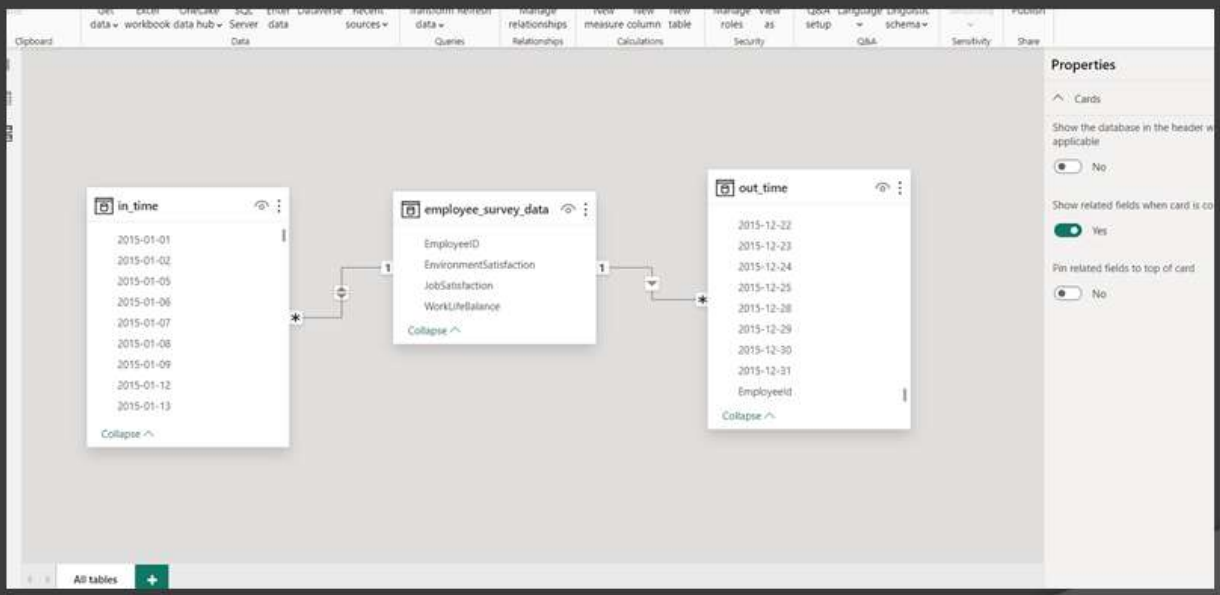


Age	Count of Age
18	24
19	27
20	33
21	39
22	48
23	42
24	78
25	78
26	117
27	144
28	144
29	204
30	180
31	207
32	183
33	174
34	231
35	234
36	207
37	150
38	174
39	126
40	171
41	120
42	138
43	96
44	99
45	123

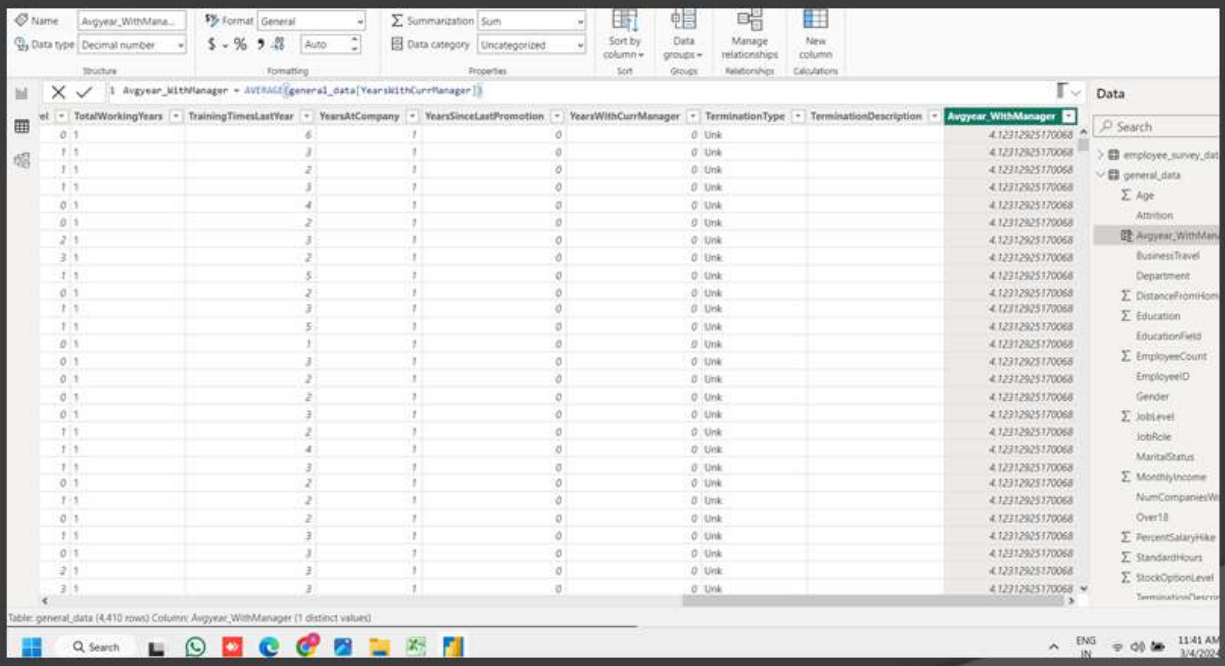
Q5. Identify and clean any missing or inconsistent data in the "Department" column.

No missing or inconsistent data in the Department column

Q6. In Power BI, establish a relationship between the "EmployeeID" in the employee data and the "EmployeeID" in the time tracking data.



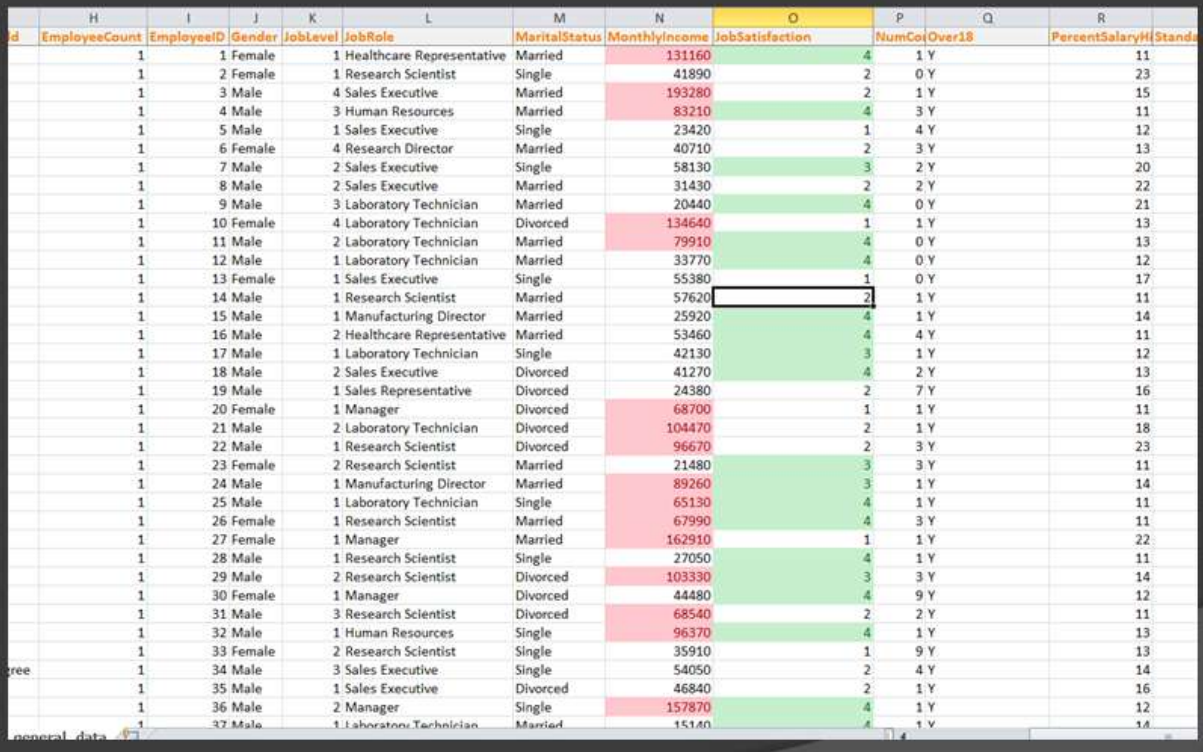
Q7. Using DAX, create a calculated column that calculates the average years an employee has spent with their current manager.



Q8. Using Excel, create a pivot table that displays the count of employees in each Marital Status category, segmented by Department.

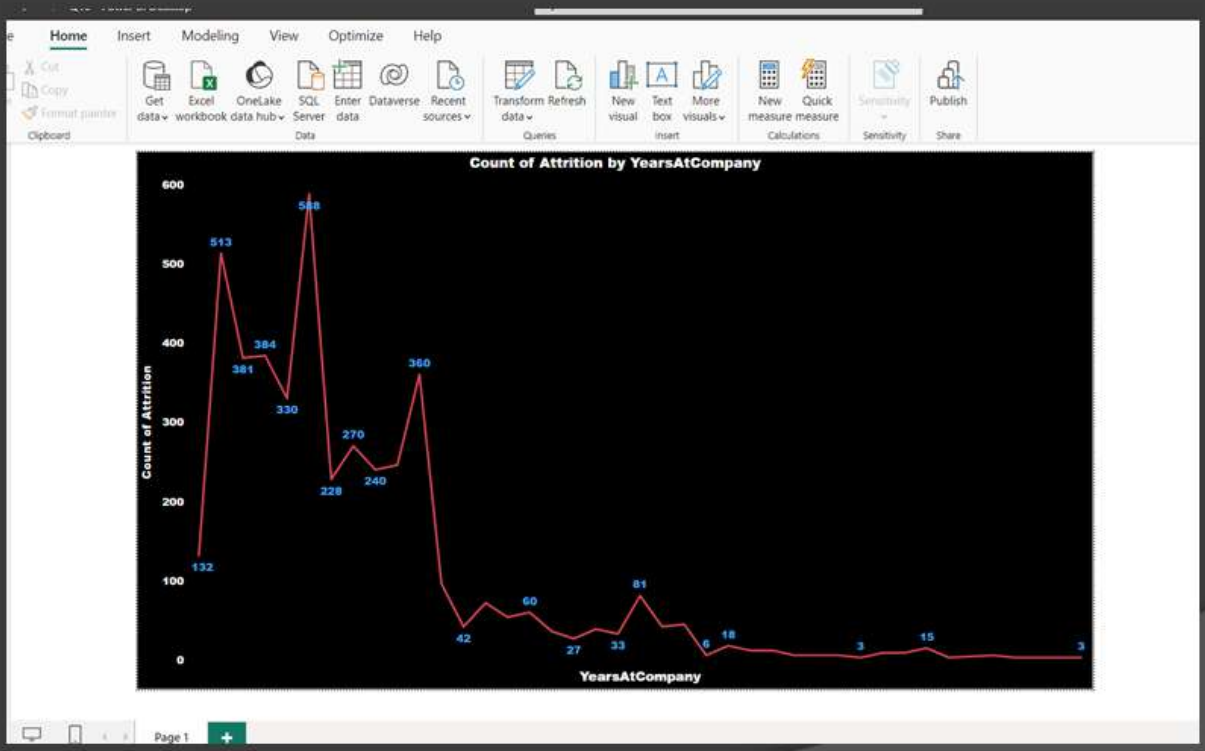
MaritalStatus	Department	Count of EmployeeCount
Divorced		**981**
	Human Resources	21
	Research & Development	621
	Sales	339
Married		**2019**
	Human Resources	96
	Research & Development	1350
	Sales	573
Single		**1410**
	Human Resources	72
	Research & Development	912
	Sales	426
Grand Total		**4410**

Q9. Apply conditional formatting to highlight employees with both above-average Monthly Income and above-average Job Satisfaction.



H	I	J	K	L	M	N	O	P	Q	R
EmployeeCount	EmployeeID	Gender	JobLevel	JobRole	MaritalStatus	MonthlyIncome	JobSatisfaction	NumCo	Over18	PercentSalaryH
1	1	Female	1	Healthcare Representative	Married	131160	4	1	Y	11
1	2	Female	1	Research Scientist	Single	41890	2	0	Y	23
1	3	Male	4	Sales Executive	Married	193280	2	1	Y	15
1	4	Male	3	Human Resources	Married	83210	4	3	Y	11
1	5	Male	1	Sales Executive	Single	23420	1	4	Y	12
1	6	Female	4	Research Director	Married	40710	2	3	Y	13
1	7	Male	2	Sales Executive	Single	58130	3	2	Y	20
1	8	Male	2	Sales Executive	Married	31430	2	2	Y	22
1	9	Male	3	Laboratory Technician	Married	20440	4	0	Y	21
1	10	Female	4	Laboratory Technician	Divorced	134640	1	1	Y	13
1	11	Male	2	Laboratory Technician	Married	79910	4	0	Y	13
1	12	Male	1	Laboratory Technician	Married	33770	4	0	Y	12
1	13	Female	1	Sales Executive	Single	55380	1	0	Y	17
1	14	Male	1	Research Scientist	Married	57620	2	1	Y	11
1	15	Male	1	Manufacturing Director	Married	25920	4	1	Y	14
1	16	Male	2	Healthcare Representative	Married	53460	4	4	Y	11
1	17	Male	1	Laboratory Technician	Single	42130	3	1	Y	12
1	18	Male	2	Sales Executive	Divorced	41270	4	2	Y	13
1	19	Male	1	Sales Representative	Divorced	24380	2	7	Y	16
1	20	Female	1	Manager	Divorced	68700	1	1	Y	11
1	21	Male	2	Laboratory Technician	Divorced	104470	2	1	Y	18
1	22	Male	1	Research Scientist	Divorced	96670	2	3	Y	23
1	23	Female	2	Research Scientist	Married	21480	3	3	Y	11
1	24	Male	1	Manufacturing Director	Married	89260	3	1	Y	14
1	25	Male	1	Laboratory Technician	Single	65130	4	1	Y	11
1	26	Female	1	Research Scientist	Married	67990	4	3	Y	11
1	27	Female	1	Manager	Married	162910	1	1	Y	22
1	28	Male	1	Research Scientist	Single	27050	4	1	Y	11
1	29	Male	2	Research Scientist	Divorced	103330	3	3	Y	14
1	30	Female	1	Manager	Divorced	44480	4	9	Y	12
1	31	Male	3	Research Scientist	Divorced	68540	2	2	Y	11
1	32	Male	1	Human Resources	Single	96370	4	1	Y	13
1	33	Female	2	Research Scientist	Single	35910	1	9	Y	13
1	34	Male	3	Sales Executive	Single	54050	2	4	Y	14
1	35	Male	1	Sales Executive	Divorced	46840	2	1	Y	16
1	36	Male	2	Manager	Single	157870	4	1	Y	12
1	37	Male	1	Laboratory Technician	Married	15140	4	1	Y	14

Q10. In Power BI, create a line chart that visualizes the trend of Employee Attrition over the years.



Q11. Describe how you would create a star schema for this dataset, explaining the benefits of doing so.

- We can create **Star Schema** using following steps:

1. Import data
2. Go to Model view
3. Create Measures
4. Visualization

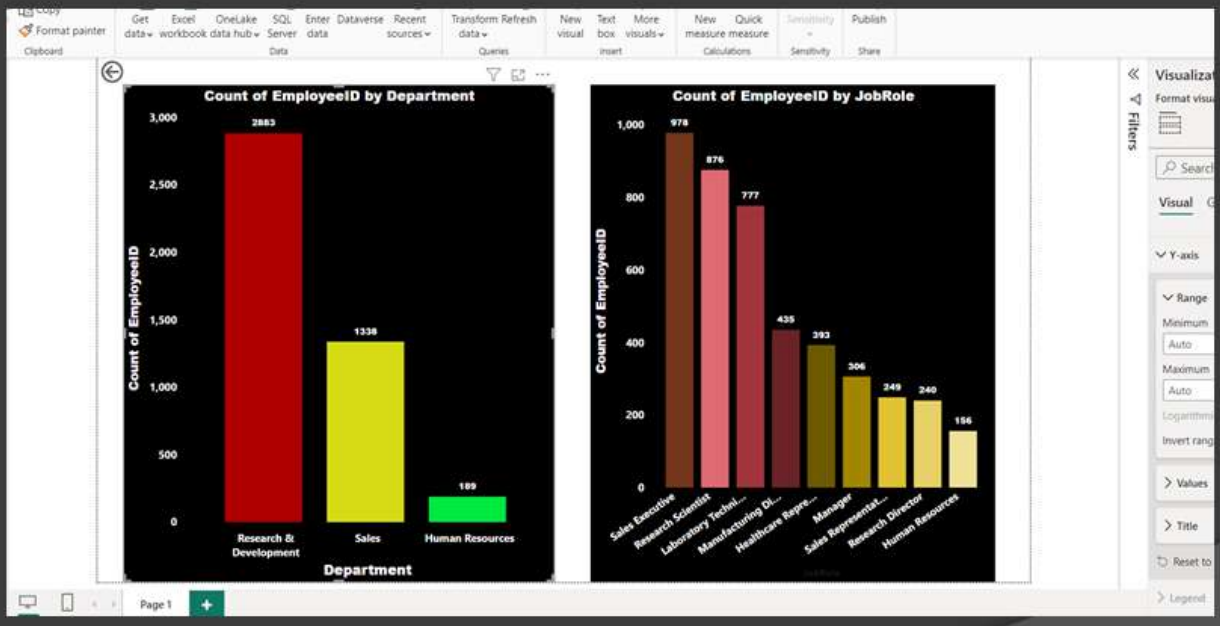
- Benefits:

1. Simple and easy to understand: The star like structure make it easy to visualize the relationships between data points.
2. Query Performance: Queries can be executed efficiently due to the denormalized nature of the fact table and well defined relationship.
3. Scalability: Star Schema can handle large dataset effectively.
4. Flexibility: its easy to add new dimensions or measures without significantly altering the existing structure.

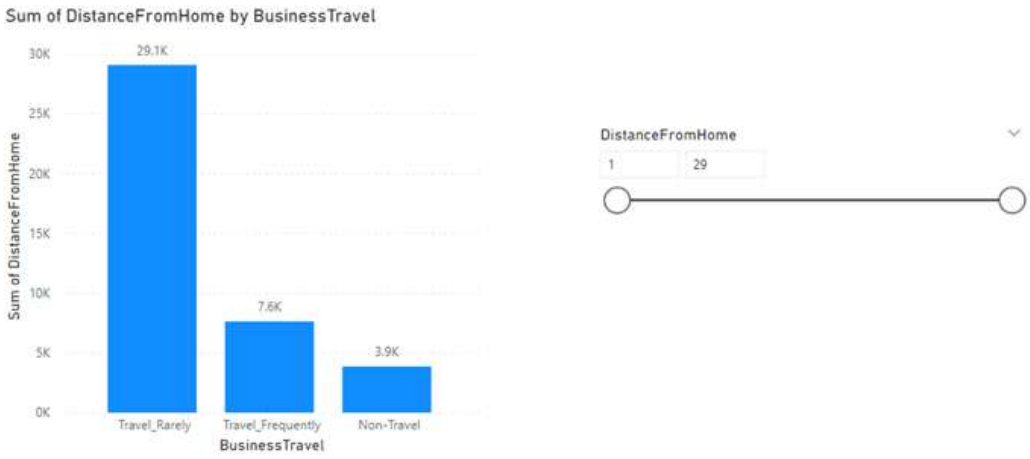
Q12. Using DAX, calculate the rolling 3-month average of Monthly Income for each employee.

```
Rolling 3-Month Avg Monthly Income =
VAR CurrentEmployeeID = 'Employee'[EmployeeID]
RETURN
    AVERAGEX(
    FILTER(
       ALL('Date),
       'Date'[Date] <= EARLIER('Date]) &&
       'Date'[Date] > DATEADD(EARLIER('Date'[Date]),-3,MONTH)
     ),
     CALCULATE(
     SUM('Sales'[MonthlyIncome]),
     'Employee'[EmployeeID] = CurrentEmployeeID
    )
  )
```

Q13. Create a hierarchy in Power BI that allows users to drill down from Department to Job Role to further narrow their analysis.



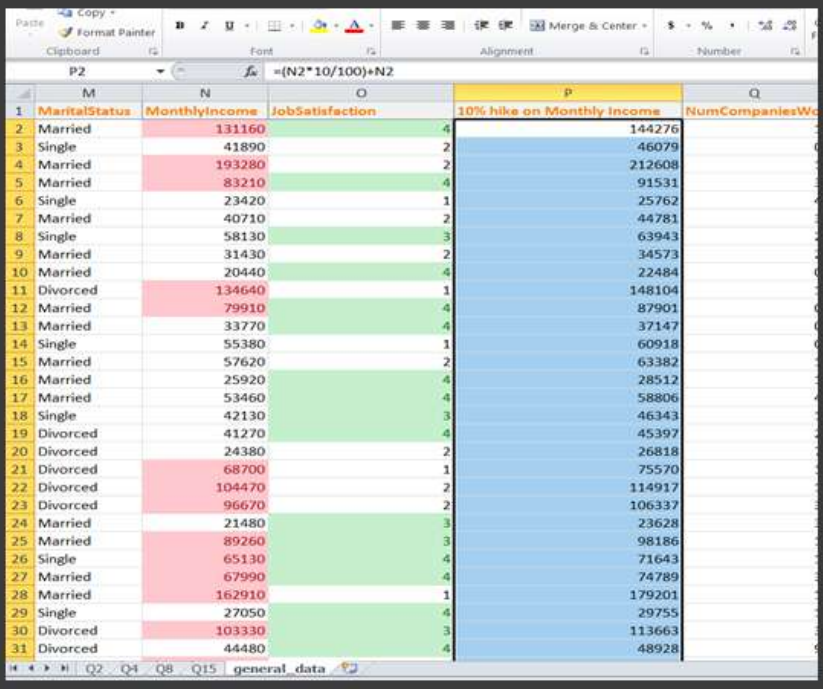
Q14. How can you set up parameterized queries in Power BI to allow users to filter data based on the Distance from Home column?



Q15. 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	Department			
JobLevel	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**

Q16. 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.



Paste | Copy | Format Painter | Clipboard | Font | Merge & Center | Alignment | Number

P2 =(N2*10/100)+N2

	M	N	O	P	Q
1	MaritalStatus	MonthlyIncome	JobSatisfaction	10% hike on Monthly Income	NumCompaniesWo
2	Married	131160	4	144276	
3	Single	41890	2	46079	
4	Married	193280	2	212608	
5	Married	83210	4	91531	
6	Single	23420	1	25762	
7	Married	40710	2	44781	
8	Single	58130	3	63943	
9	Married	31430	2	34573	
10	Married	20440	4	22484	
11	Divorced	134640	1	148104	
12	Married	79910	4	87901	
13	Married	33770	4	37147	
14	Single	55380	1	60918	
15	Married	57620	2	63382	
16	Married	25920	4	28512	
17	Married	53460	4	58806	
18	Single	42130	3	46343	
19	Divorced	41270	4	45397	
20	Divorced	24380	2	26818	
21	Divorced	68700	1	75570	
22	Divorced	104470	2	114917	
23	Divorced	96670	2	106337	
24	Married	21480	3	23628	
25	Married	89260	3	98186	
26	Single	65130	4	71643	
27	Married	67990	4	74789	
28	Married	162910	1	179201	
29	Single	27050	4	29755	
30	Divorced	103330	3	113663	
31	Divorced	44480	4	48928	

Q2 Q4 Q8 Q15 general_data

Q17. Verify if the data adheres to a predefined schema. What actions would you take if you find inconsistencies?

Check the data against the established schema. Compare the restrictions, types, and structure. Find discrepancies such as duplicates, wrong kinds, or missing values. Put quality control procedures into action. Record conclusions and actions. Engage in dialogue with intrested parties. Repeat to ensure continous data integrity.