



HR DATA ANALYSIS

PSYLIQ DATA ANALYST INTERNSHIP

NEHA DANDEKAR

TASK 1



HR EMPLOYEE ANALYSIS

Department

Human
Resources

Research &
Development

Sales

Gender

Female

Male

Total Employees

4410

Attrition

711

Average Salary

65.00K

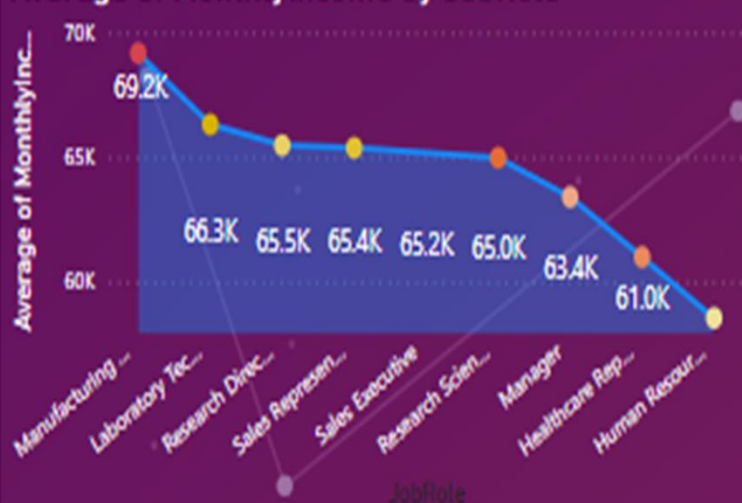
Average Age

36.92

Terminated
Employees

557

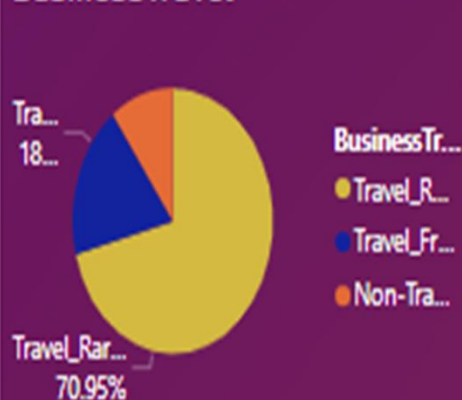
Average of MonthlyIncome by JobRole



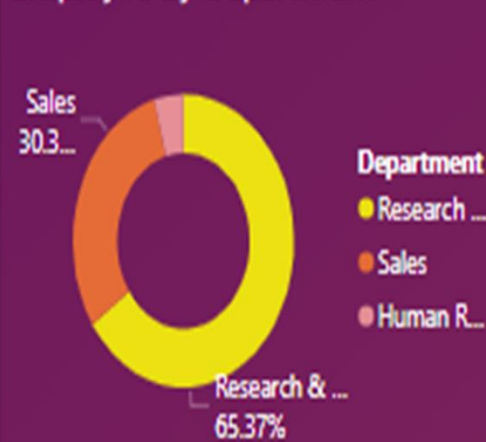
Count of Attrition by YearsAtCompany



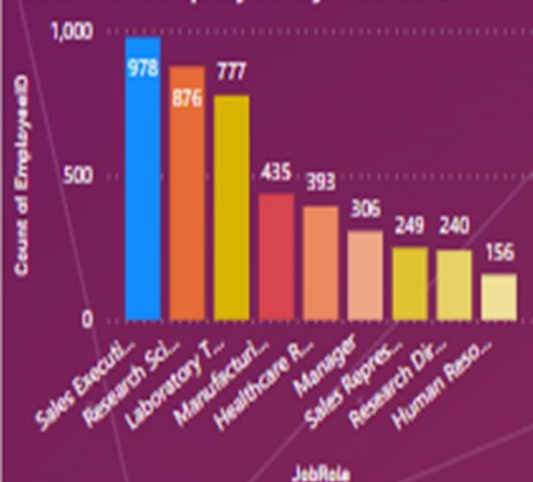
Sum of EmployeeCount by BusinessTravel



Employee by Department



Count of Employee by JobRole



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

	A	B	C	D	E	F	G
1	Age	Attritio	Busine	Depart	Distanc	Educati	Educati
	Sort Smallest to Largest				6	2	Life Scienc
	Sort Largest to Smallest				10	1	Life Scienc
	Sort by Color				17	4	Other
	Clear Filter From "Age"				2	5	Life Scienc
	Filter by Color				10	1	Medical
	Number Filters				8	3	Life Scienc
	Search						
	(Select All)						
	18						
	19						
	20						
	21						
	22						
	23						
	24						
	25						

Custom AutoFilter

Show rows where:

Age

is greater than or equal to 30

☒ And ☐ Or

Use ? to represent any single character
Use * to represent any series of characters

OK Cancel

	A	B	C	D	E	F
1	Age	Attritio	Busine	Depart	Distanc	Edu
2	51	No	Travel_Ra	Sales	6	
3	31	Yes	Travel_Fre	Research	10	
4	32	No	Travel_Fre	Research	17	
5	38	No	Non-Travel	Research	2	
6	32	No	Travel_Ra	Research	10	
7	46	No	Travel_Ra	Research	8	
10	31	No	Travel_Ra	Research	1	
12	45	No	Travel_Ra	Research	17	
13	36	No	Travel_Ra	Research	28	
14	55	No	Travel_Ra	Research	14	
15	47	Yes	Non-Travel	Research	1	
17	37	No	Travel_Ra	Research	1	
19	37	No	Non-Travel	Research	1	
20	35	No	Travel_Ra	Sales	7	

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

Filters	Columns
Rows	Values
JobRole ▼	Average of MonthlyInc... ▼

Job Role	Average of MonthlyIncome
Healthcare Representative	\$ 60,983.74
Human Resources	\$ 58,528.08
Laboratory Technician	\$ 66,314.05
Manager	\$ 63,395.88
Manufacturing Director	\$ 69,183.72
Research Director	\$ 65,473.13
Research Scientist	\$ 64,975.68
Sales Executive	\$ 65,186.69
Sales Representative	\$ 65,370.96
Grand Total	\$ 65,029.31

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

New Formatting Rule ? X

Select a Rule Type:

- ▶ Format all cells based on their values
- ▶ Format only cells that contain
- ▶ Format only top or bottom ranked values
- ▶ Format only values that are above or below average
- ▶ Format only unique or duplicate values
- ▶ Use a formula to determine which cells to format

Edit the Rule Description:

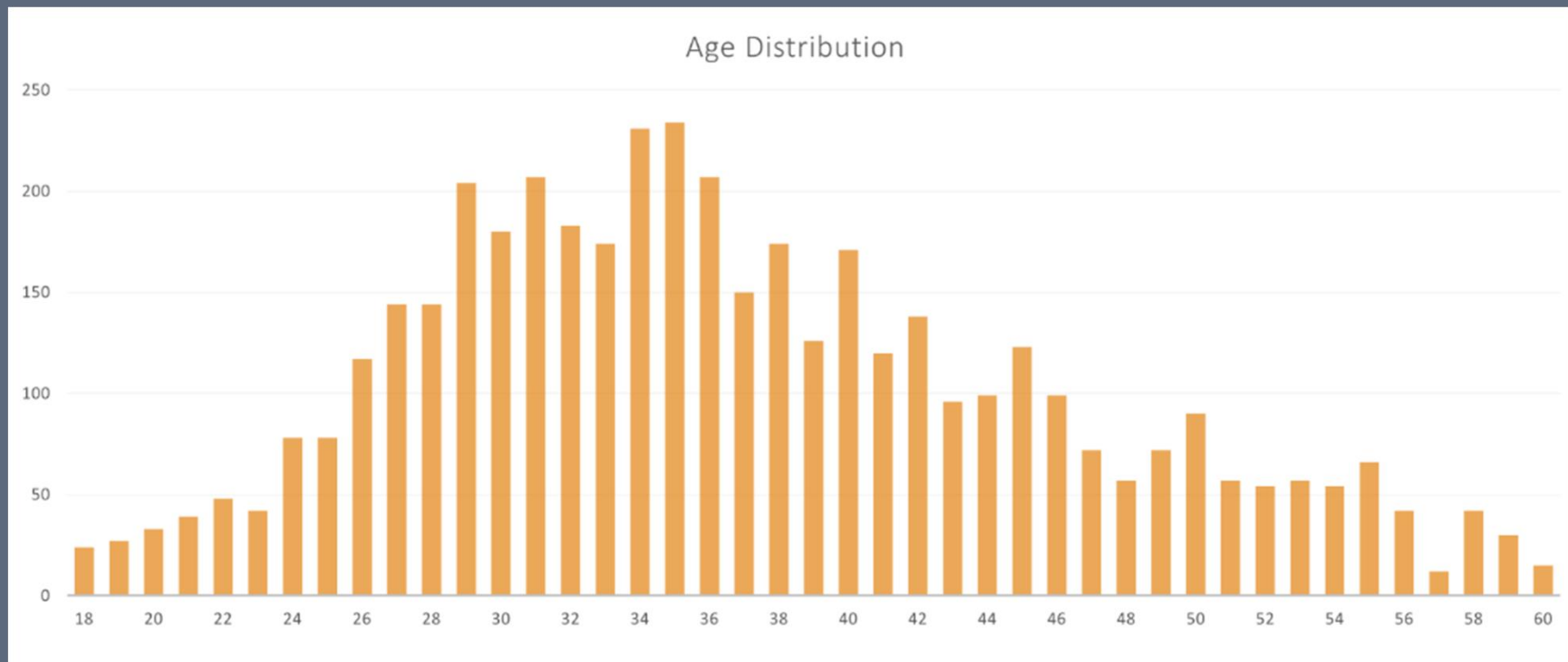
Format values that are:

above the average for the selected range

Preview:

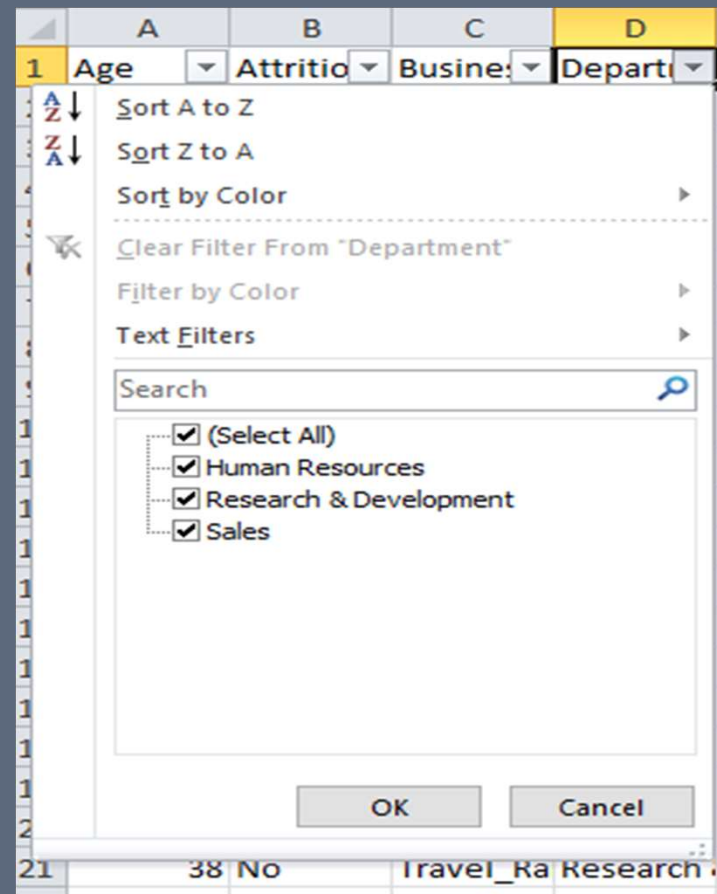
	H	I	J	K	L	M	N	O
	Employee	Employee	Gender	JobLevel	JobRole	MaritalSta	MonthlyIr	Numo
1	1	1	Female	1	Healthcar	Married	131160	
2	1	2	Female	1	Research	Single	41890	
3	1	3	Male	4	Sales Exec	Married	193280	
4	1	4	Male	3	Human Re	Married	83210	
5	1	5	Male	1	Sales Exec	Single	23420	
6	1	6	Female	4	Research	Married	40710	
7	1	7	Male	2	Sales Exec	Single	58130	
8	1	8	Male	2	Sales Exec	Married	31430	
9	1	9	Male	3	Laborator	Married	20440	
10	1	10	Female	4	Laborator	Divorced	134640	
11	1	11	Male	2	Laborator	Married	79910	
12	1	12	Male	1	Laborator	Married	33770	
13	1	13	Female	1	Sales Exec	Single	55380	
14	1	14	Male	1	Research	Married	57620	
15	1	15	Male	1	Manufact	Married	25920	
16	1	16	Male	2	Healthcar	Married	53460	
17	1	17	Male	1	Laborator	Single	42130	

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

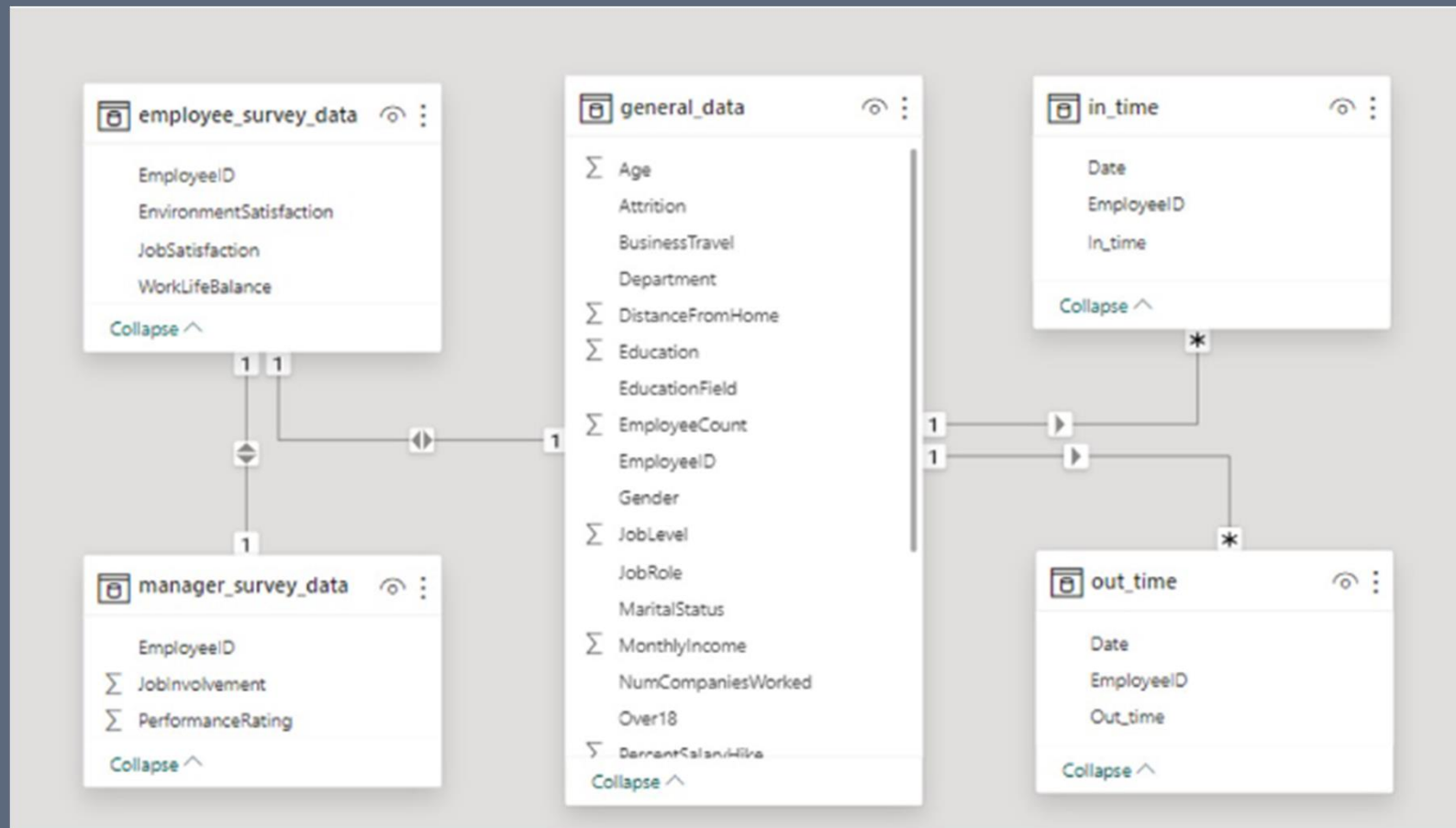


5. Identify and clean any missing or inconsistent data in the "Department" column.

- There are No missing values 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

```
Average_yearswcurrmanager = AVERAGE(general_data[YearsWithCurrManager])
```

4.12

Average_yearswcurrmanager

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

Row Labels	Sum 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

PivotTable Field List

Choose fields to add to report:

- ☐ EducationField
- ☒ EmployeeCount
- ☐ EmployeeID
- ☐ Gender
- ☐ JobLevel
- ☐ JobRole
- ☒ MaritalStatus

Drag fields between areas below:

Report Filter

Column Labels

Row Labels

- MaritalStatus
- Department

Values

- Sum of Employee...

☐ Defer Layout Update

Update

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

New Formatting Rule ? X

Select a Rule Type:

- ▶ Format all cells based on their values
- ▶ Format only cells that contain
- ▶ Format only top or bottom ranked values
- ▶ **Format only values that are above or below average**
- ▶ Format only unique or duplicate values
- ▶ Use a formula to determine which cells to format

Edit the Rule Description:

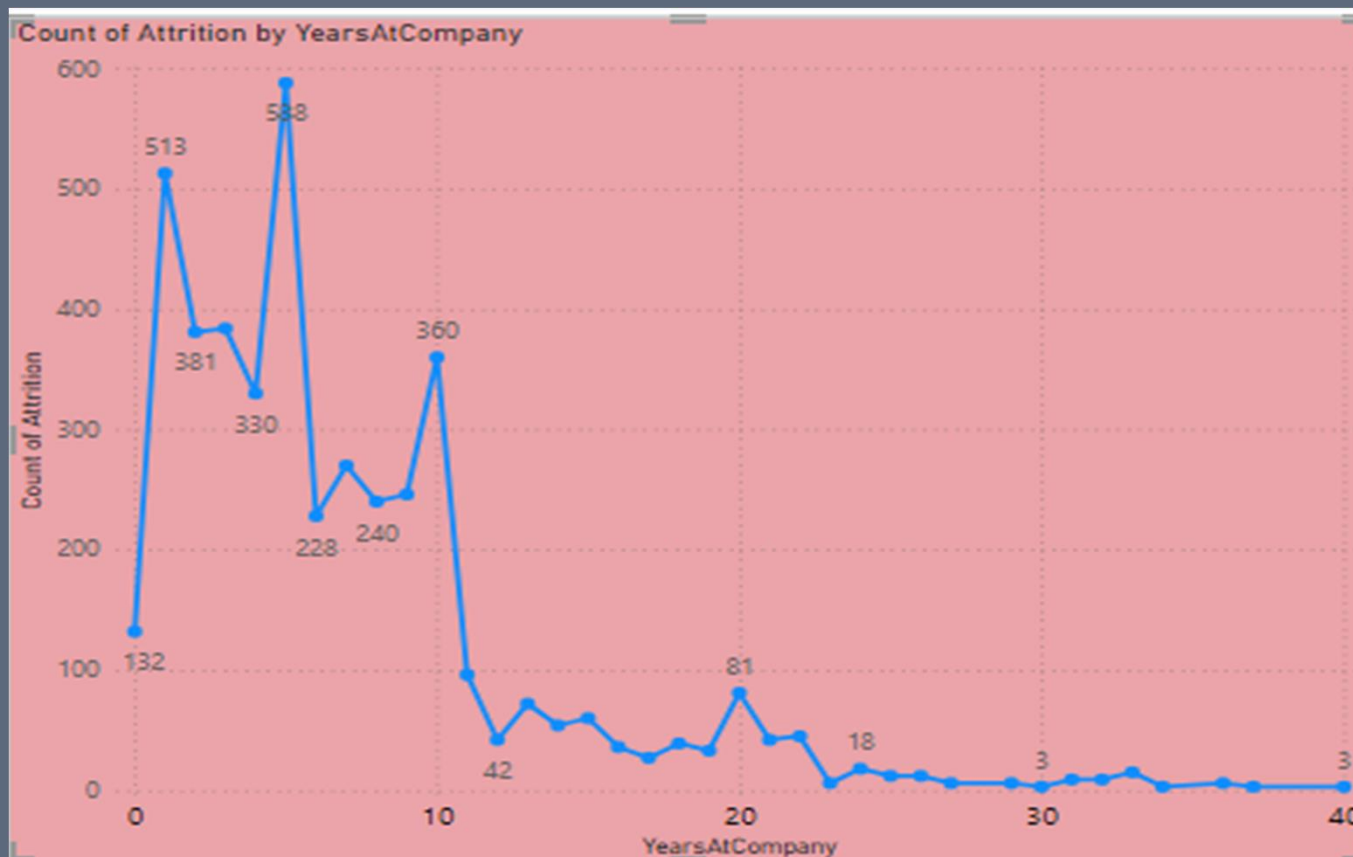
Format values that are:

above the average for the selected range

Preview:

M	N	O	P
ritalSta	MonthlyIr	NumCom	Over18
ried	131160	1	Y
gle	41890	0	Y
ried	193280	1	Y
ried	83210	3	Y
gle	23420	4	Y
ried	40710	3	Y
gle	58130	2	Y
ried	31430	2	Y
ried	20440	0	Y
orced	134640	1	Y
ried	79910	0	Y
ried	33770	0	Y
gle	55380	0	Y
ried	57620	1	Y
ried	25920	1	Y
ried	53460	4	Y
gle	42130	1	Y
orced	41270	2	Y
orced	24380	7	Y
	68780	1	Y

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



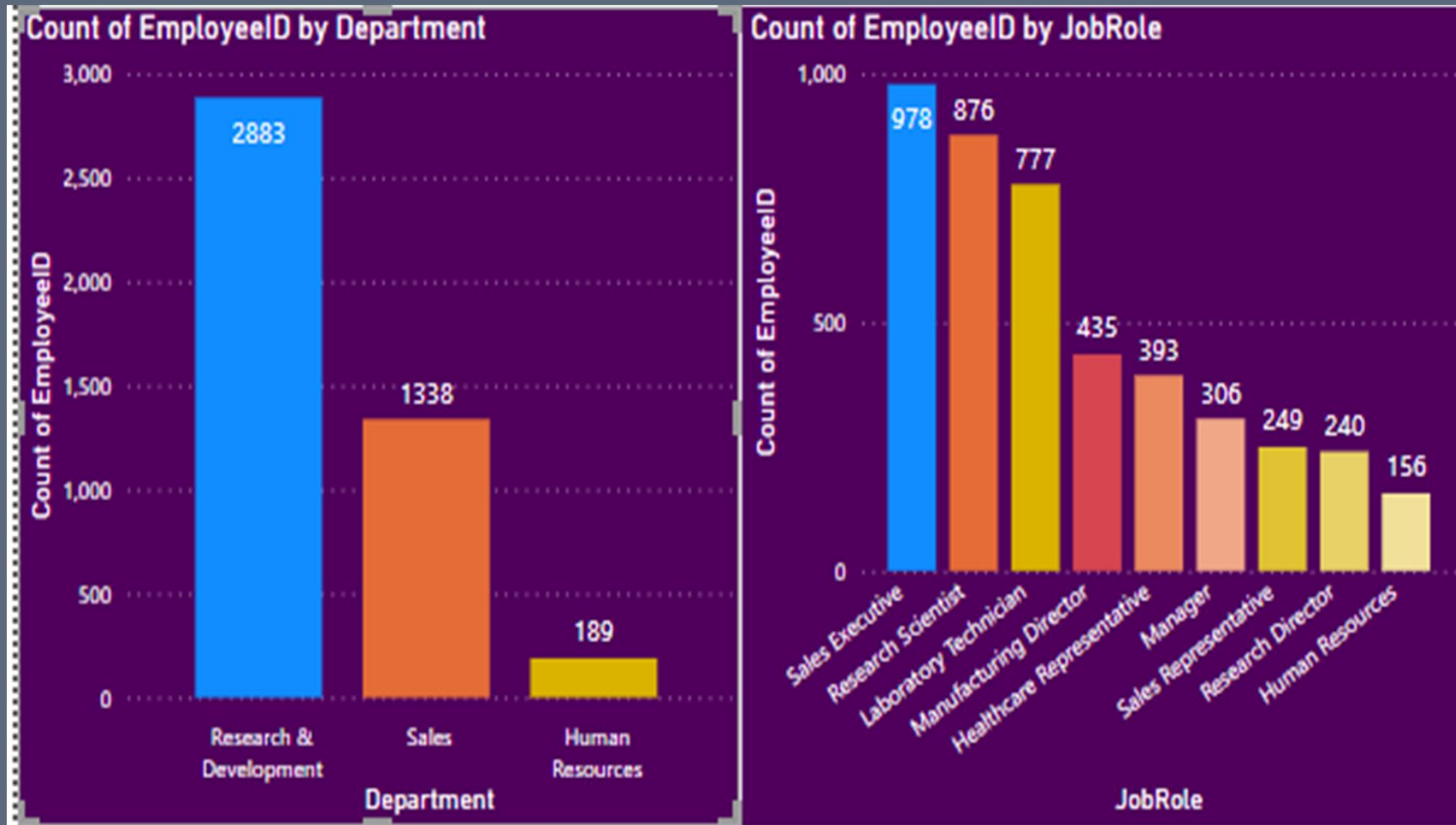
11. Describe how you would create a star schema for this dataset, explaining the benefits of doing so.

- Steps to Create star schema for this dataset :
 - Import all the tables
 - Load all the table.
 - Go to the Model view section
 - Take General_data as central query.
 - Establish relationship with employee_survey_data, in_time, out_time,manager_survey.
- Benefits :
 - Simple and easy-to understand structure
 - Better Performance for analytical queries
 - Fewer joins to access data

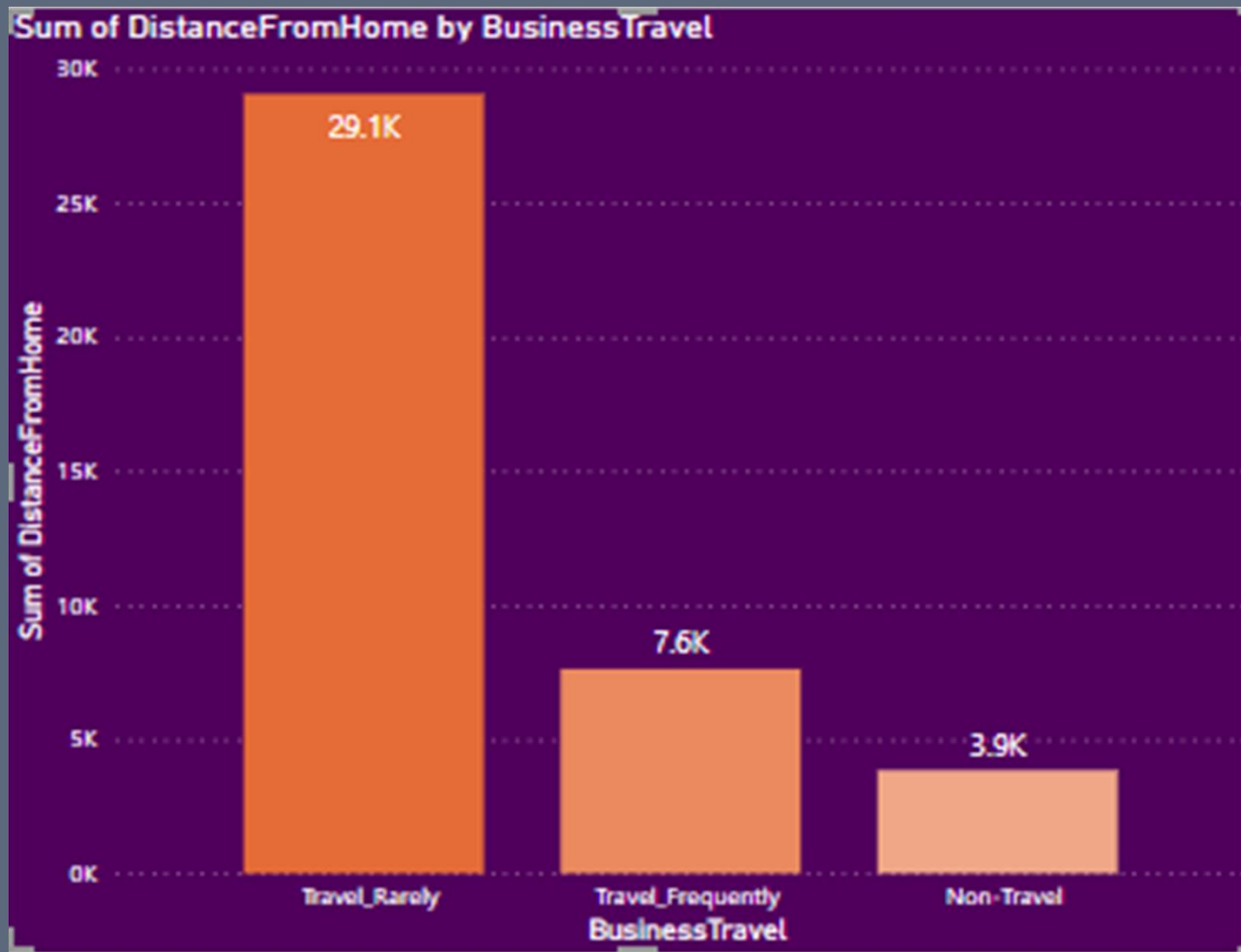
12. Using DAX, calculate the rolling 3-month average of Monthly Income for each employee.

- Rolling 3-Month Average =
- CALCULATE(
• AVERAGE(general_data'[Monthly Income]),
• DATESINPERIOD(
• 'general_data'[Date],
• LASTDATE('general_data'[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.



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



DistanceFromHome

1 29

0 100

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.

```
=SUMIFS(Gral[ [ MonthlyIncome ] ], Gral[Department], [@Department], Gral[JobLevel], ">=" & 3)
```

Department	Monthly Income
Human Resources	\$ 3,259,140
Research & Development	\$ 53,502,900
Sales	\$ 22,974,330

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

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

Calculate the average monthly income.

Multiply the average monthly income by 1 + the 10% of salary hike.

% Salary hike		Avg Monthly Incc		Income after salary hike
10%	\$	65,029	\$	71,532

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

To verify if the data adheres to a predefined schema we need to first understand the business rules to make sure that our relations, tables, columns and data types are the correct ones.

In the case I find inconsistencies, I would correct some of them manually, and others using a software process to clean my data , communicate my findings with the interested parts and actualize the schema if necessary.