



INTERNSHIP BY PSYLIQ

HR DATA ANALYSIS

PAVAN PANDHARE

Department

All

Gender

All

Age

18

60

JobLevel

1

5

Education

1

5

4410

Total Employee

4.12

Average_year_with_cu...

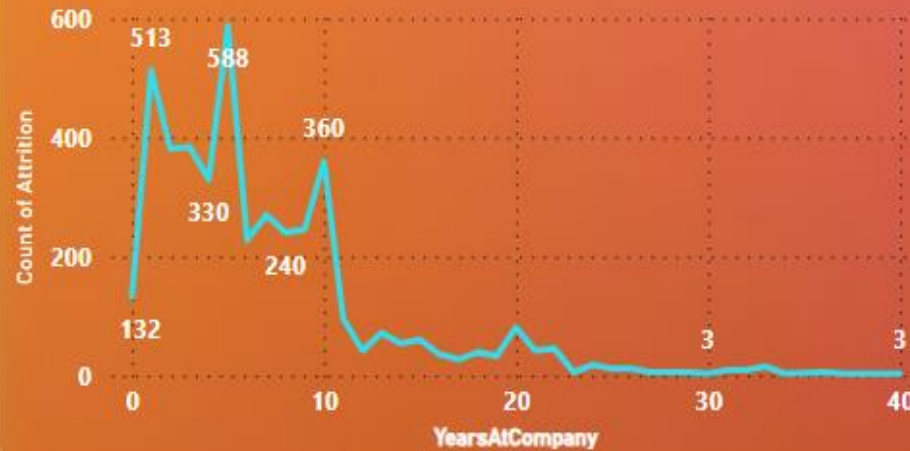
65029.31

Average MonthlyIncome

15.21

Average Salary Hike%

Employee Attrition Over The Years



Average of MonthlyIncome by JobRole

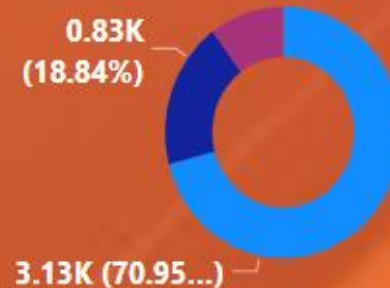


EmployeeID Sum of MonthlyIncome

386	199990
1856	199990
3326	199990
942	199730
2412	199730
3882	199730
1017	199730
Total	286779270

Count of Employee by BusinessTravel

BusinessTravel Traveling Non-Traveling



Count of Employee by MaritalStatus

MaritalStatus Married Single Divorced



- Q1. 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	DistanceFromHome	Education	EducationField	EmployeeCount	Employee	Gender	JobLevel	JobRole	MaritalStatus	MonthlyIncome
	Sort Smallest to Largest			Sales	6	2	Life Sciences	1	1	Female	1	Healthcare	Married	131160
	Sort Largest to Smallest			Research & Dev	10	1	Life Sciences	1	2	Female	1	Research S	Single	41890
	Sort by Color			Research & Dev	17	4	Other	1	3	Male	4	Sales Exec	Married	193280
	Sheet View			Research & Dev	2	5	Life Sciences	1	4	Male	3	Human Re	Married	83210
	Clear Filter From "Age"			Research & Dev	10	1	Medical	1	5	Male	1	Sales Exec	Single	23420
	Filter by Color			Research & Dev	8	3	Life Sciences	1	6	Female	4	Research I	Married	40710
	Number Filters			Research & Dev	11	2	Medical	1	7	Male	2	Sales Exec	Single	58130
	Search			Research & Dev	18	3	Life Sciences	1	8	Male	2	Sales Exec	Married	31430
				Research & Dev	1	3	Life Sciences	1	9	Male	3	Laborator	Married	20440
				Research & Dev	7	4	Medical	1	10	Female	4	Laborator	Divorced	134640
				Research & Dev	17	2	Medical	1	11	Male	2	Laborator	Married	79910
				Research & Dev	28	1	Life Sciences	1	12	Male	1	Laborator	Married	33770
				Research & Dev	14	4	Life Sciences	1	13	Female	1	Sales Exec	Single	55380
				Research & Dev	1	1	Medical	1	14	Male	1	Research S	Married	57620
				Research & Dev	1	3	Life Sciences	1	15	Male	1	Manufactu	Married	25920
				Research & Dev	1	3	Life Sciences	1	16	Male	2	Healthcare	Married	53460
				Research & Dev	3	2	Life Sciences	1	17	Male	1	Laborator	Single	42130
				Research & Dev	1	3	Medical	1	18	Male	2	Sales Exec	Divorced	41270
				Sales	7	4	Life Sciences	1	19	Male	1	Sales Repr	Divorced	24380
				Research & Dev	8	3	Life Sciences	1	20	Female	1	Manager	Divorced	68700
				Research & Dev	1	4	Other	1	21	Male	2	Laborator	Divorced	104470
				Sales	8	4	Life Sciences	1	22	Male	1	Research S	Divorced	96670
24	53	No	Travel_Rarely	Research & Dev	11	4	Life Sciences	1	23	Female	2	Research S	Married	21480
25	42	No	Travel_Rarely	Research & Dev	4	4	Life Sciences	1	24	Male	1	Manufactu	Married	89260
26	29	No	Travel_Frequently	Research & Dev	16	4	Medical	1	25	Male	1	Laborator	Single	65130
27	55	No	Travel_Rarely	Research & Dev	1	4	Other	1	26	Female	1	Research S	Married	67990
28	26	No	Travel_Frequently	Research & Dev	9	3	Life Sciences	1	27	Female	1	Manager	Married	162910

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

Job Roles	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

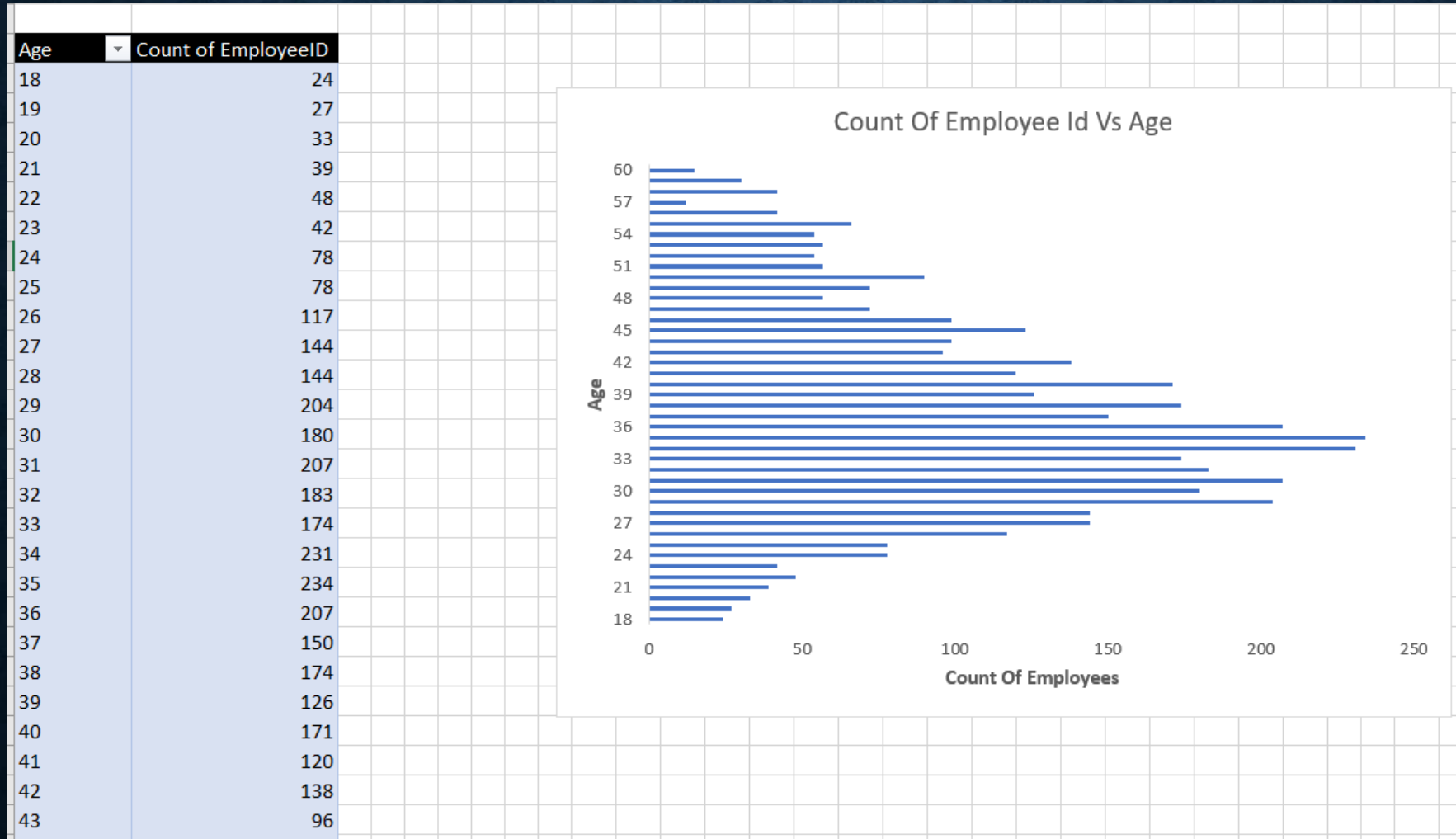
- 3. Apply conditional formatting to highlight employees with Monthly Income above the company's average income.

EmployeeID	MonthlyIncome	Environment	JobSatisfaction	WorkLifeBalance
1	131160	3	4	2
2	41890	3	2	4
3	193280	2	2	1
4	83210	4	4	3
5	23420	4	1	3
6	40710	3	2	2
7	58130	1	3	1
8	31430	1	2	3
9	20440	2	4	3
10	134640	2	1	3
11	79910	3	4	3
12	33770	NA	4	3
13	55380	4	1	3
14	57620	1	2	2
15	25920	4	4	2
16	53460	3	4	4
17	42130	4	3	4
18	41270	1	4	3
19	24380	2	2	2
20	68700	1	1	3
21	104470	3	2	1
22	96670	1	2	2
23	21480	3	3	2
24	89260	2	3	3
25	65130	2	4	2
26	67990	2	4	3
27	162910	1	1	3
28	27050	4	4	3

employee_survey_data

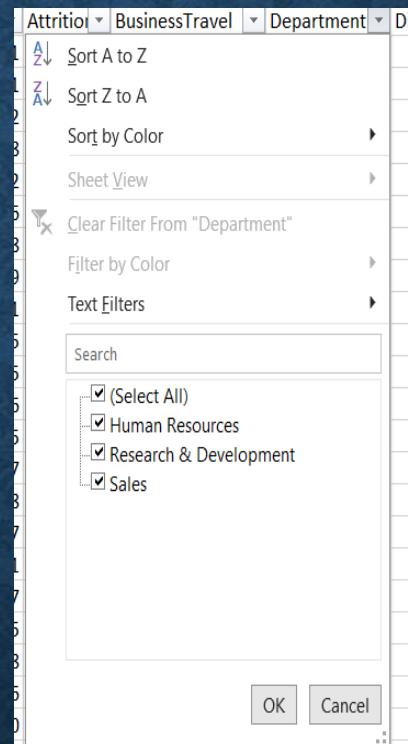


- 4. Create a bar chart in Excel to visualize the distribution of employee ages.

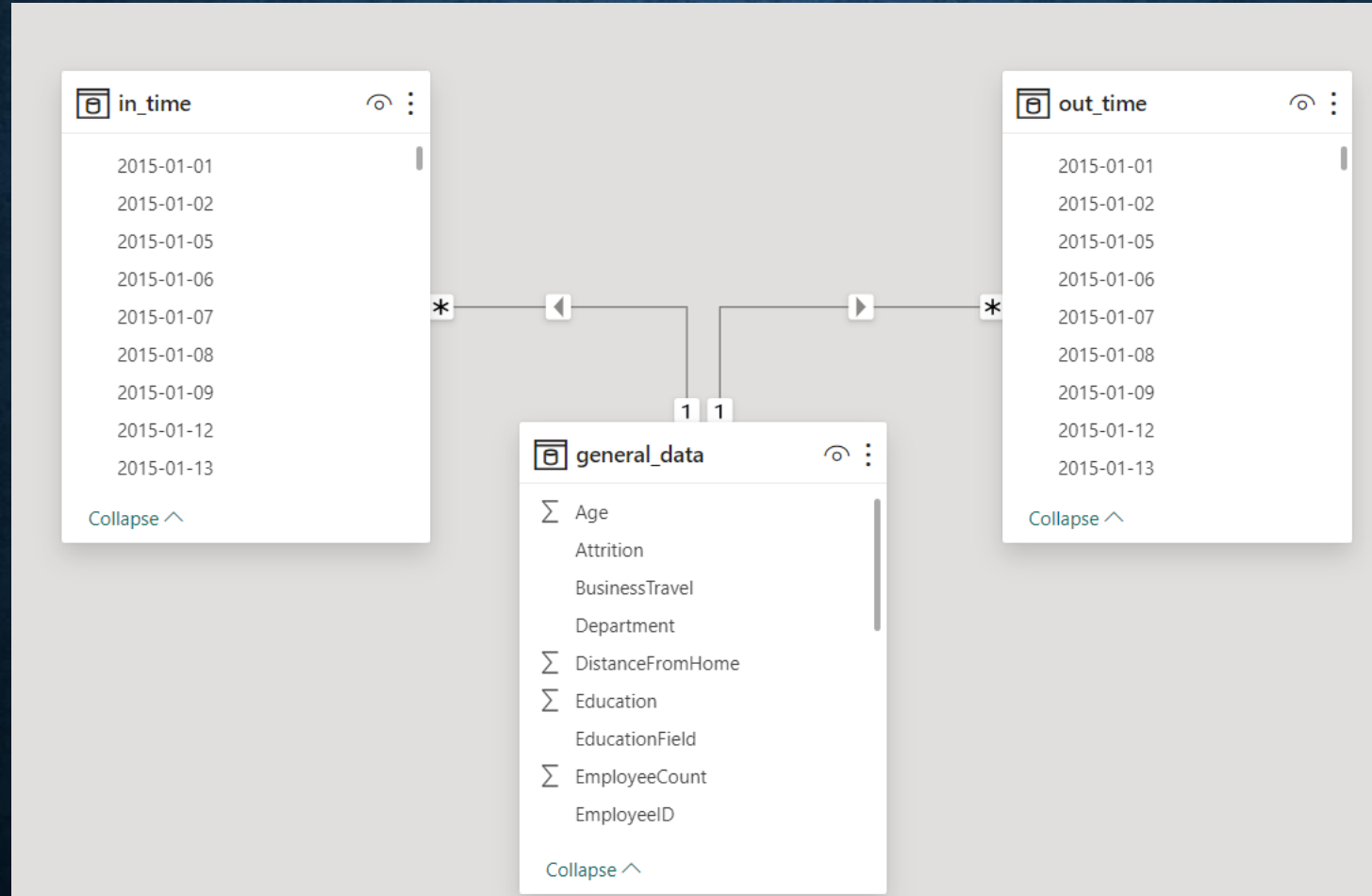


- 5. Identify and clean any missing or inconsistent data in the "Department" column.

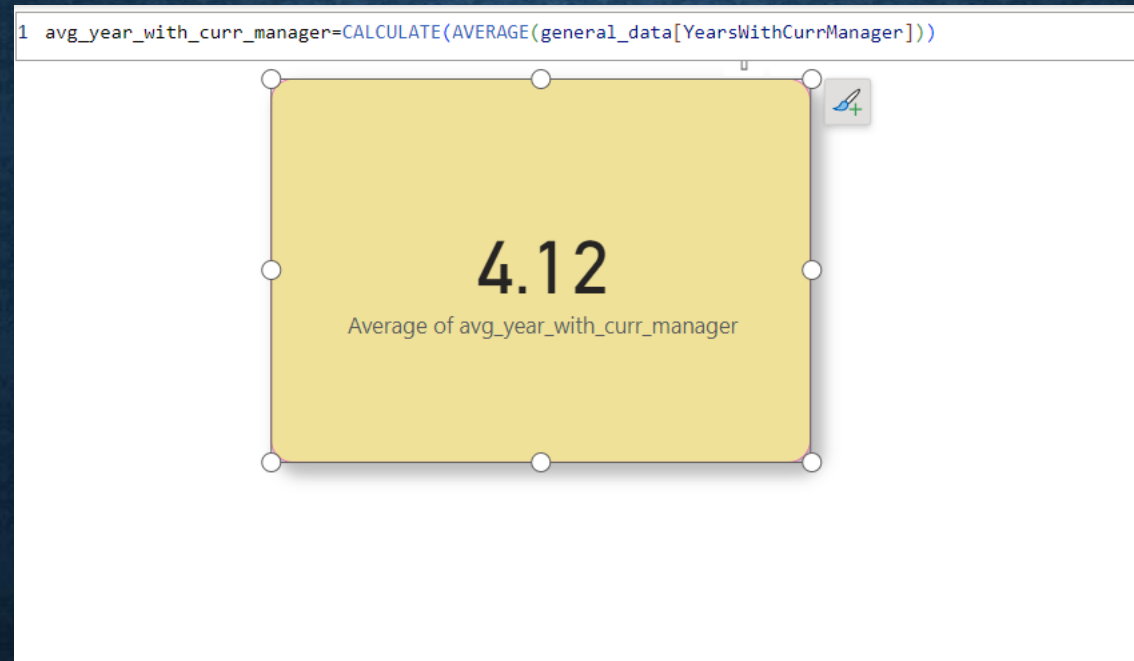
By checking the Department column, it seems that the column data is consistent, clean and doesn't have any missing data.



- 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.



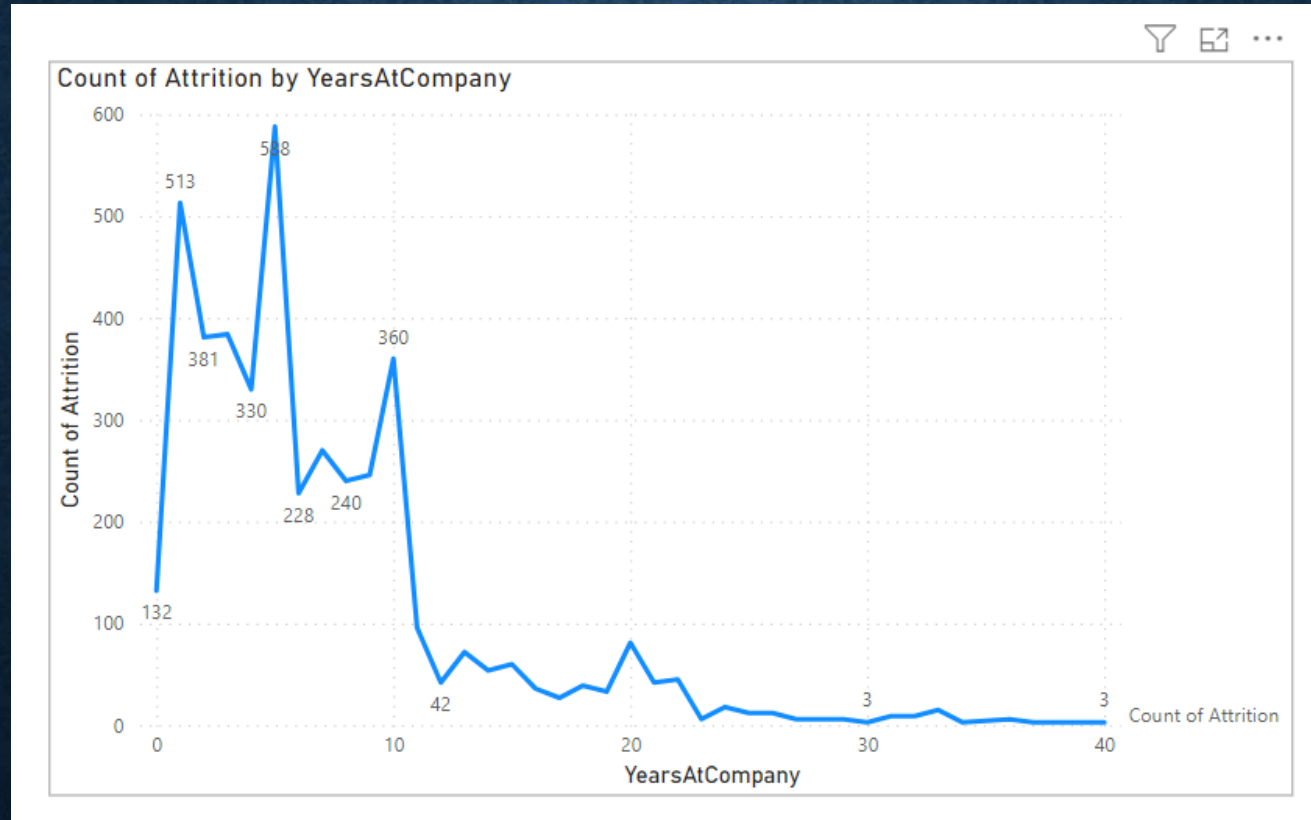
- 8. Using Excel, create a pivot table that displays the count of employees in each Marital Status category, segmented by Department

Count of EmployeeID					
Column Labels					
Row Labels		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

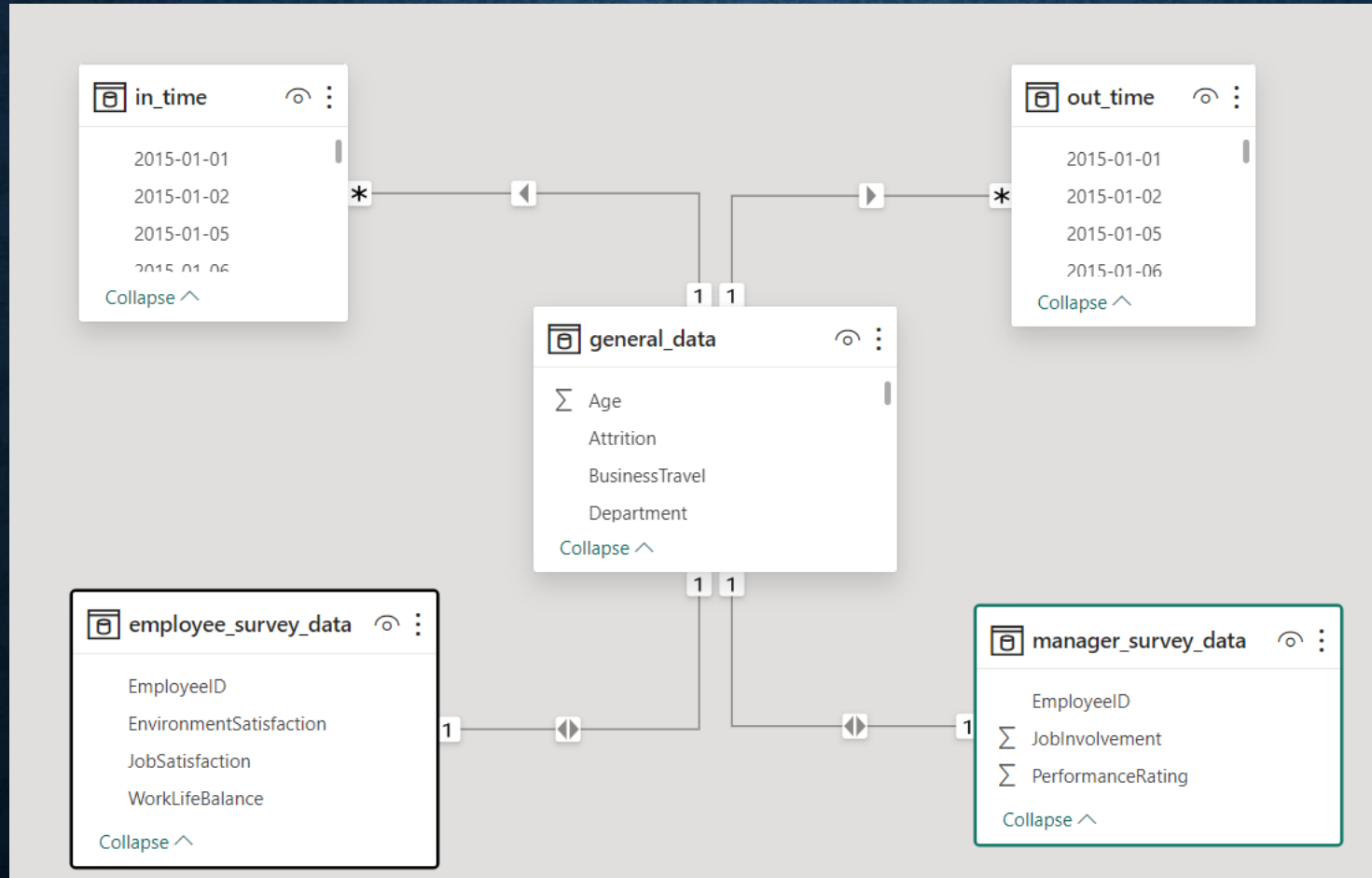
- 9. Apply conditional formatting to highlight employees with both above average Monthly Income and above-average Job Satisfaction.

A	B	C
EmployeeID	MonthlyIncome	JobSatisfaction
1	131160	4
2	41890	2
3	193280	2
4	83210	4
5	23420	1
6	40710	2
7	58130	3
8	31430	2
9	20440	4
10	134640	1
11	79910	4
12	33770	4
13	55380	1
14	57620	2
15	25920	4
16	53460	4
17	42130	3
18	41270	4
19	24380	2
20	68700	1
21	104470	2
22	96670	2
23	21480	3
24	89260	3
25	65130	4
26	67990	4
27	162910	1
28	27050	4

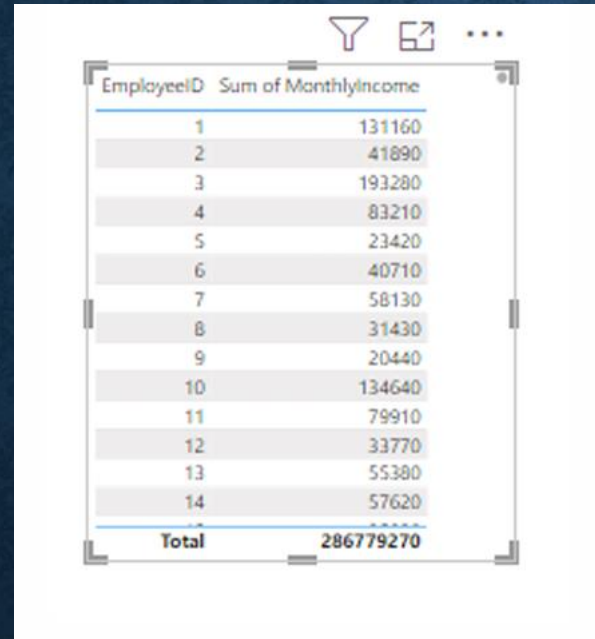
- 10. In Power BI, create a line chart that visualizes the trend of Employee Attrition over the years.



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



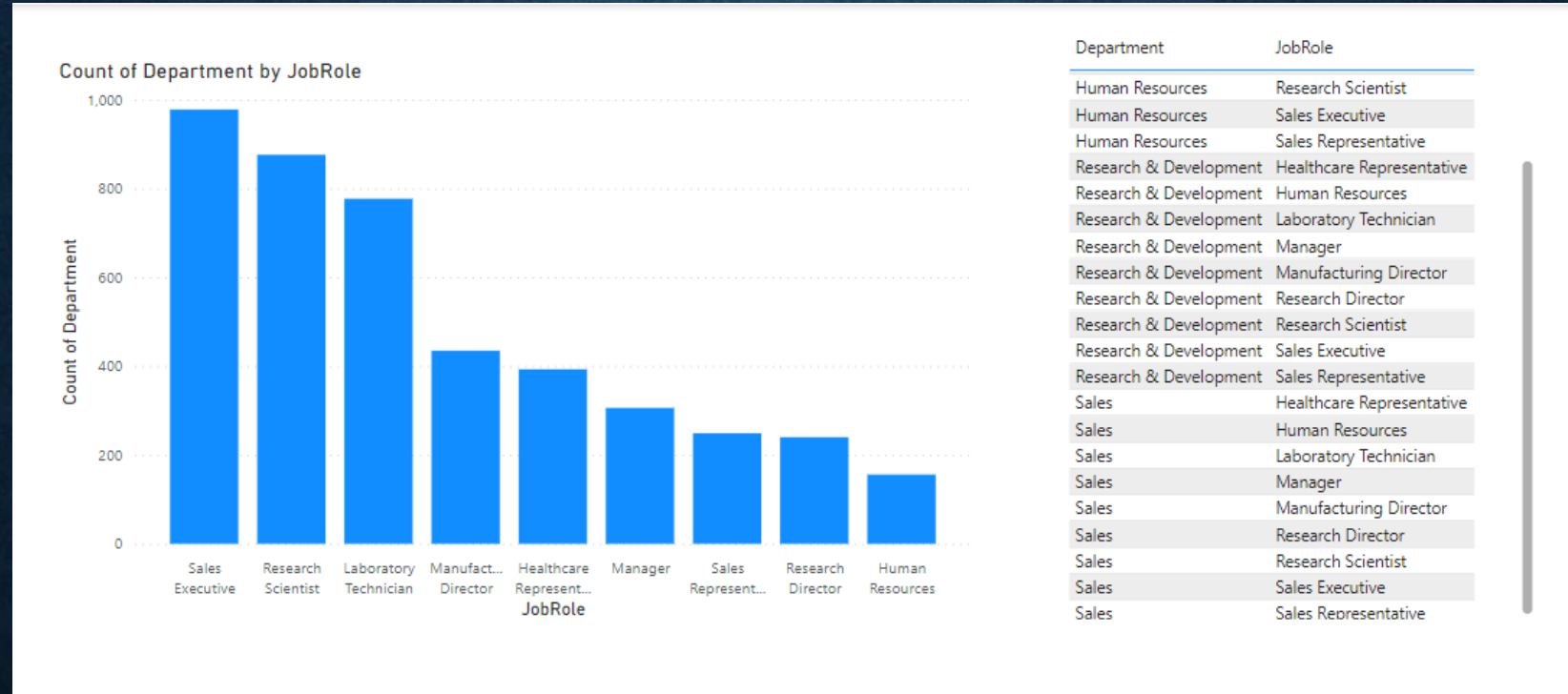
- 12. Using DAX, calculate the rolling 3-month average of Monthly Income for each employee.



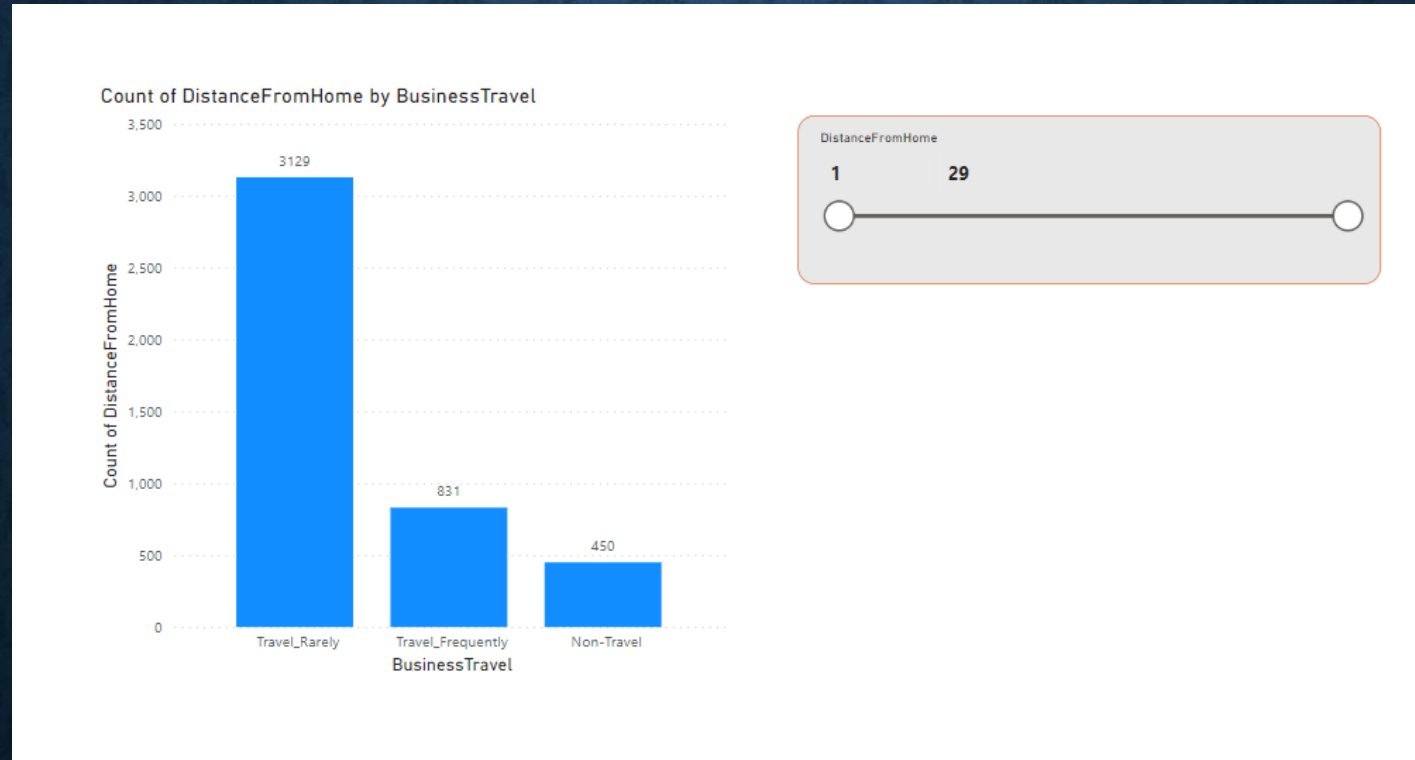
A screenshot of a Power BI table. The table has two columns: 'EmployeeID' and 'Sum of Monthly Income'. It contains 14 rows of data, with a 'Total' row at the bottom. The table is displayed in a standard Power BI view with a filter icon, a refresh icon, and a menu icon at the top. The table is bordered by a thin grey line, and the 'Total' row is highlighted in a darker shade.

EmployeeID	Sum of Monthly Income
1	131160
2	41890
3	193280
4	83210
5	23420
6	40710
7	58130
8	31430
9	20440
10	134640
11	79910
12	33770
13	55380
14	57620
Total	286779270

- 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?



- 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	Column Labels			
Row Labels	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.

N	O
MonthlyIncome	10% hike monthaly incom
131160	144276
41890	46079
193280	212608
83210	91531
23420	25762
40710	44781
58130	63943
31430	34573
20440	22484
134640	148104
79910	87901
33770	37147
55380	60918
57620	63382
25920	28512
53460	58806
42130	46343
41270	45397
24380	26818
68700	75570
104470	114917
96670	106337
21480	23628
89260	98186
65130	71643
67990	74789
162910	179201
27050	29755

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.
- 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.