TASK 1

DATA ANALYST INTERNSHIP

PREPARED BY

ANKUSH VERMA

WELCOME MESSAGE

GREETINGS! THIS REPORT MARKS PSYLIQ'S INITIAL DATA ANALYSIS TASK FOR THE DATA ANALYSIS INTERNSHIP. IT AIMS TO DETAIL THE STEPS, METHODS, AND ACTIONS UNDERTAKEN TO ADDRESS ASSESSMENT QUESTIONS AND PROVIDE VALUABLE INSIGHTS.

THIS REPORT IS PRESENTED TO YOU BY ANKUSH VERMA, A FREELANCE DATA ANALYST AND CONTENT WRITER. CURRENTLY PURSUING A MANAGEMENT CAREER WITH A SPECIALIZATION IN FINANCE AND ENTREPRENEURSHIP AT GEETA UNIVERSITY.

FOR CONTACT: ankushverma2807@gmail.com

AGENDA

DATA EXPLORATION

BEFORE WE TACKLE THE QUESTIONS, LET'S FIRST GRASP THE DETAILS OF OUR DATA—HOW IT'S SET UP AND WHAT IT REPRESENTS.

- THE INFORMATION IS SPREAD ACROSS SEVERAL EXCEL FILES.
- IT COMPRISES 4410 ENTRIES WITH 24 CHARACTERISTICS IN A TABLE.
- THIS DATA PERTAINS TO EMPLOYEES IN A SALES COMPANY.
- UPON EXAMINATION, THE DATA SEEMS WELL-ORGANIZED AND PREPARED.

DATA QUESTIONS

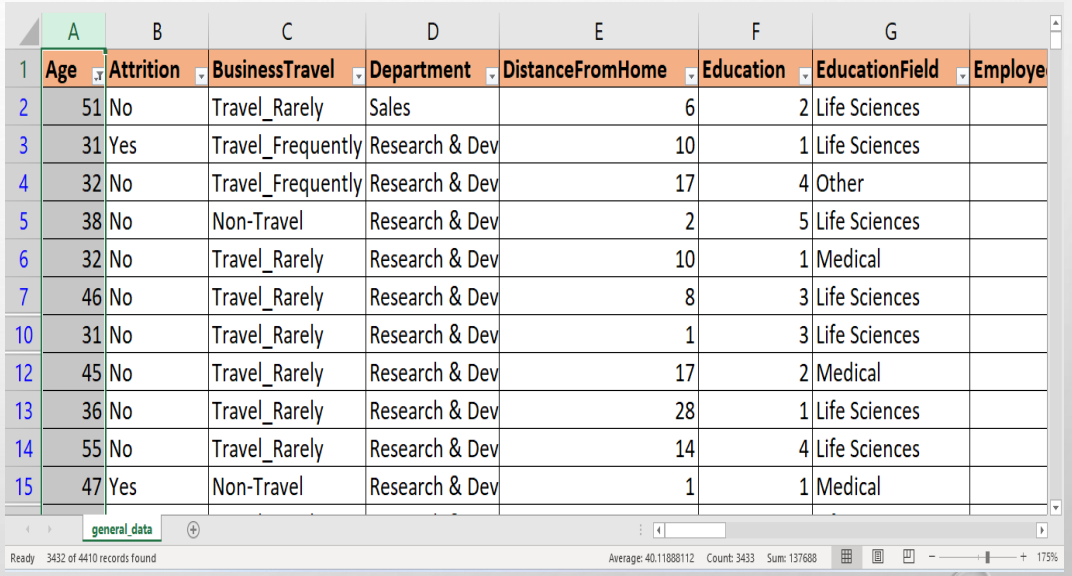
5

Q1. USING EXCEL, HOW WOULD YOU FILTER THE DATASET TO ONLY SHOW EMPLOYEES AGED 30 AND ABOVE?

TO FILTER A DATASET IN EXCEL TO SHOW EMPLOYEES AGED 30 AND ABOVE, FOLLOW THESE STEPS:

1. At the first step, we will transform our dataset into a Table from a Range.
2. Then we will select the age column and use the filter option in the header.
3. We will go to the Number Filtering Option after opening the filter & and sort window.
4. There we will select the “Greater or Equal to” option to get the result.
5. Then in the right text box to the option selected we will write 30 to get the desired result.
6. After clicking “Ok” the row values will be filtered accordingly.

NOW, ONLY THE ROWS WITH EMPLOYEES AGED 30 AND ABOVE WILL BE DISPLAYED IN THE DATASET.



	A	B	C	D	E	F	G	
1	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	Employe
2	51	No	Travel_Rarely	Sales	6	2	Life Sciences	
3	31	Yes	Travel_Frequently	Research & Dev	10	1	Life Sciences	
4	32	No	Travel_Frequently	Research & Dev	17	4	Other	
5	38	No	Non-Travel	Research & Dev	2	5	Life Sciences	
6	32	No	Travel_Rarely	Research & Dev	10	1	Medical	
7	46	No	Travel_Rarely	Research & Dev	8	3	Life Sciences	
10	31	No	Travel_Rarely	Research & Dev	1	3	Life Sciences	
12	45	No	Travel_Rarely	Research & Dev	17	2	Medical	
13	36	No	Travel_Rarely	Research & Dev	28	1	Life Sciences	
14	55	No	Travel_Rarely	Research & Dev	14	4	Life Sciences	
15	47	Yes	Non-Travel	Research & Dev	1	1	Medical	

general_data

Ready 3432 of 4410 records found Average: 40.11888112 Count: 3433 Sum: 137688 175%

DATA QUESTIONS

6

Q2. CREATE A PIVOT TABLE TO SUMMARIZE THE AVERAGE MONTHLY INCOME BY JOB ROLE.

UPON REVIEWING THE DATA SUMMARY, IT IS EVIDENT THAT THE MANUFACTURING DIRECTOR HAS THE HIGHEST MEAN MONTHLY INCOME.

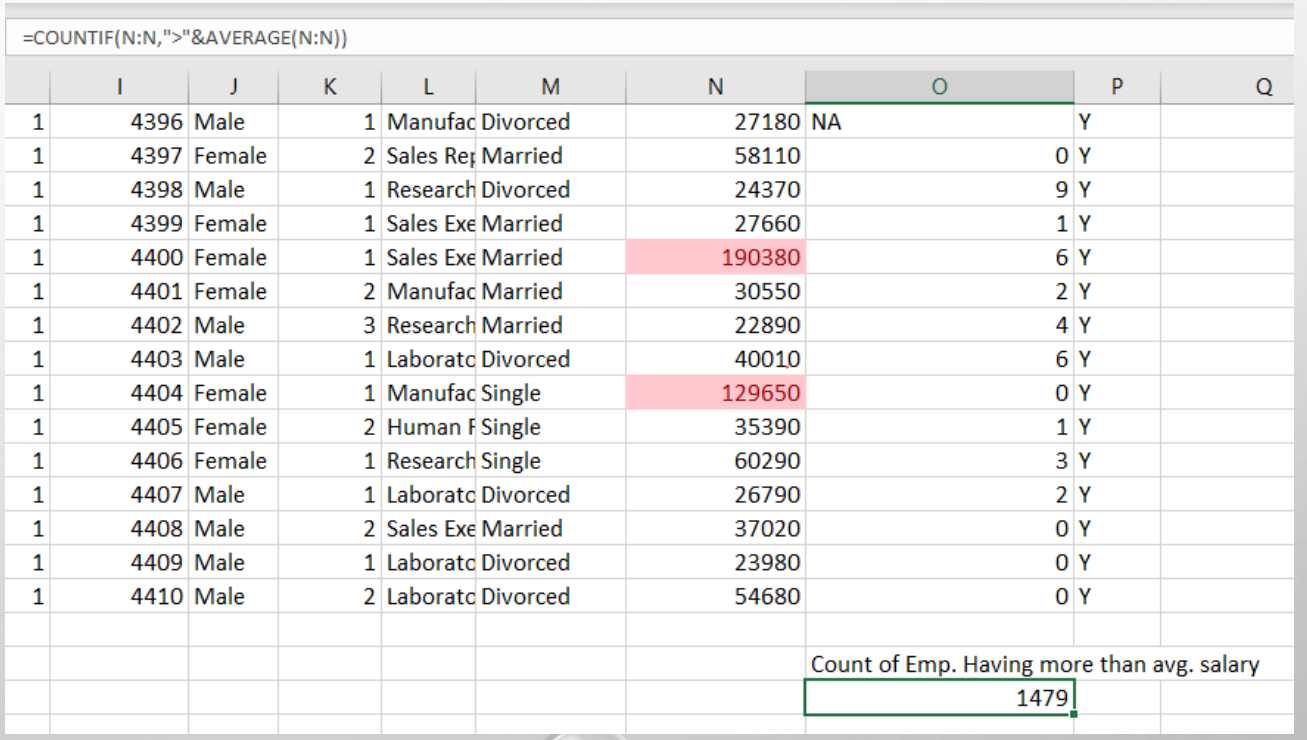
Row Labels	Average of MonthlyIncome
Manufacturing Director	69,183.72
Laboratory Technician	66,314.05
Research Director	65,473.13
Sales Representative	65,370.96
Sales Executive	65,186.69
Research Scientist	64,975.68
Manager	63,395.88
Healthcare Representative	60,983.74
Human Resources	58,528.08
Grand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

DATA QUESTIONS

7

Q3. APPLY CONDITIONAL FORMATTING TO HIGHLIGHT EMPLOYEES WITH MONTHLY INCOME ABOVE THE COMPANY'S AVERAGE INCOME.

Through the implementation of conditional formatting on the dataset, it becomes apparent that 1479 employees receive a monthly income surpassing the company's average earnings.



=COUNTIF(N:N,">"&AVERAGE(N:N))

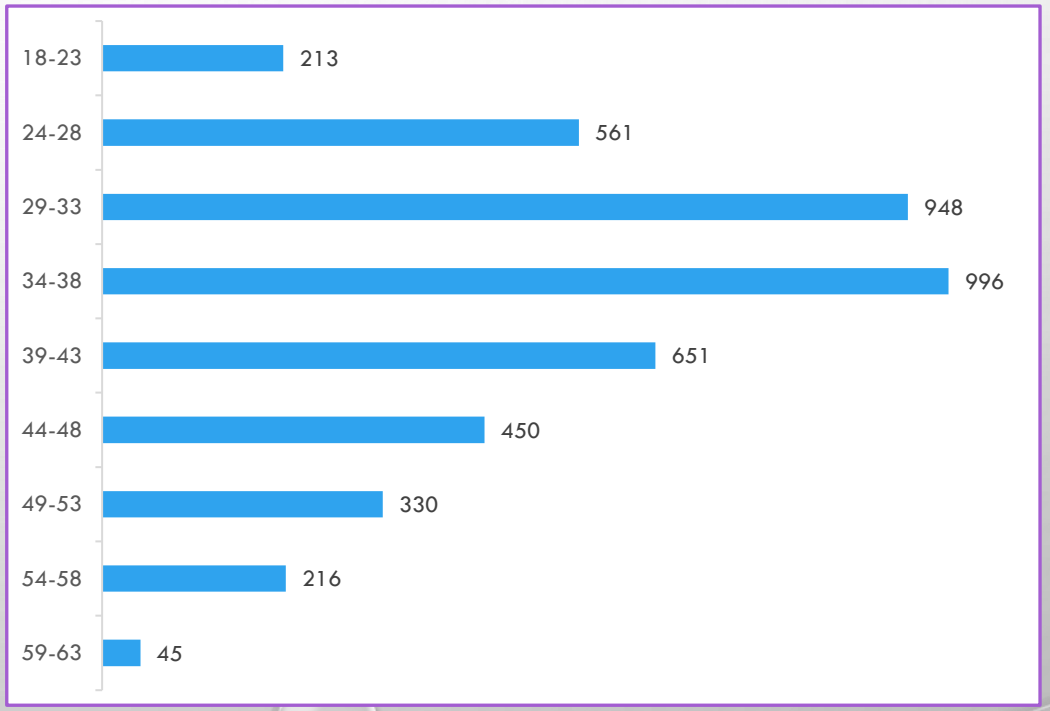
	I	J	K	L	M	N	O	P	Q
1	4396	Male	1	Manufac	Divorced	27180	NA	Y	
1	4397	Female	2	Sales Rep	Married	58110	0	Y	
1	4398	Male	1	Research	Divorced	24370	9	Y	
1	4399	Female	1	Sales Exe	Married	27660	1	Y	
1	4400	Female	1	Sales Exe	Married	190380	6	Y	
1	4401	Female	2	Manufac	Married	30550	2	Y	
1	4402	Male	3	Research	Married	22890	4	Y	
1	4403	Male	1	Laboratc	Divorced	40010	6	Y	
1	4404	Female	1	Manufac	Single	129650	0	Y	
1	4405	Female	2	Human F	Single	35390	1	Y	
1	4406	Female	1	Research	Single	60290	3	Y	
1	4407	Male	1	Laboratc	Divorced	26790	2	Y	
1	4408	Male	2	Sales Exe	Married	37020	0	Y	
1	4409	Male	1	Laboratc	Divorced	23980	0	Y	
1	4410	Male	2	Laboratc	Divorced	54680	0	Y	
							Count of Emp. Having more than avg. salary		
							1479		

DATA QUESTIONS

8

Q4. CREATE A BAR CHART IN EXCEL TO VISUALIZE THE DISTRIBUTION OF EMPLOYEE AGES.

Using a bin width of 5 years, it's evident that more than 80% of our workforce falls within the age range of 23 to 48.



DATA QUESTIONS

Q5. IDENTIFY AND CLEAN ANY MISSING OR INCONSISTENT DATA IN THE "DEPARTMENT" COLUMN.

Inspecting the Department column indicates that the data within the column is consistent, devoid of any discrepancies, and there are no instances of missing data.

DATA QUESTIONS 10

Q6. IN POWER BI, ESTABLISH A RELATIONSHIP BETWEEN THE "EMPLOYEEID" IN THE EMPLOYEE DATA AND THE "EMPLOYEEID" IN THE TIME TRACKING DATA.

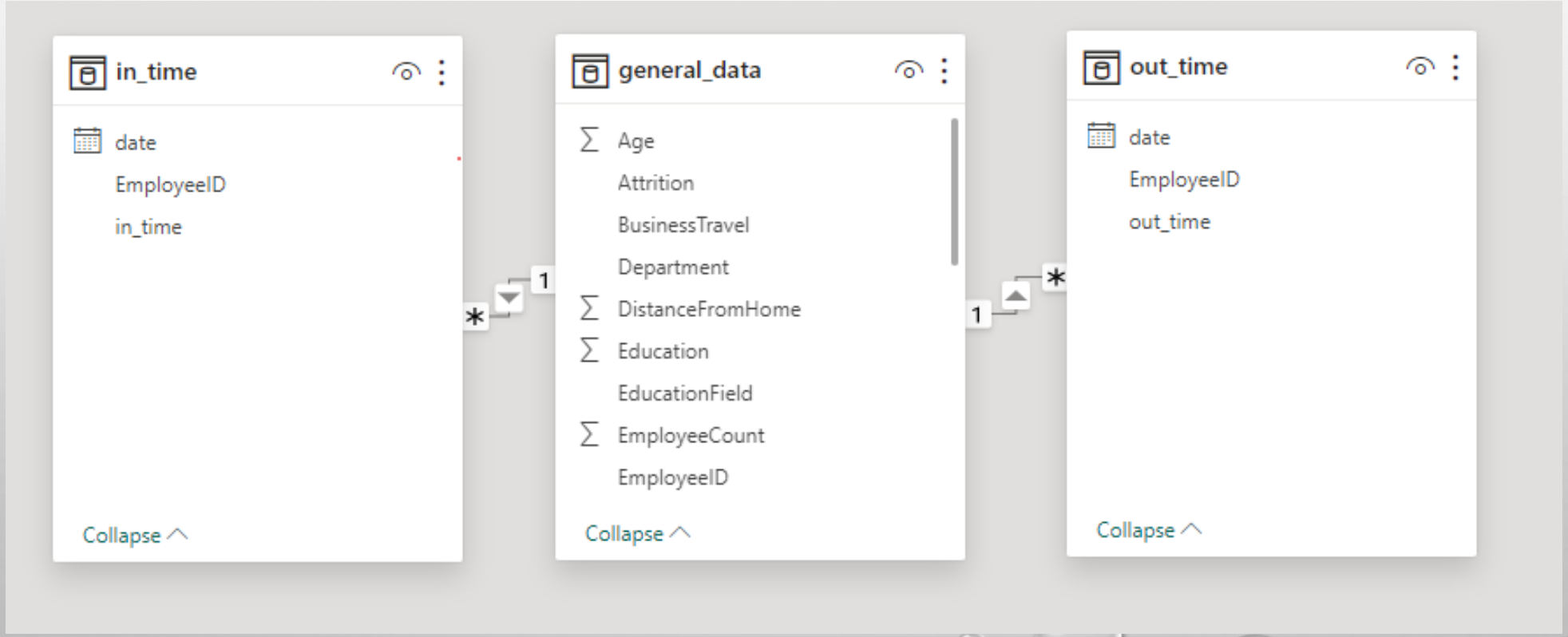
Upon linking the "general_data," "in_time," and "out_time" files in Power BI, we implemented various modifications to the time-tracking data files during the transformation in Power Query.

These modifications are adding headers, pivoting columns, and excluding rows with "NA," presuming them to represent holidays and weekends, to ensure the absence of time-tracking data on those days.

DATA QUESTIONS

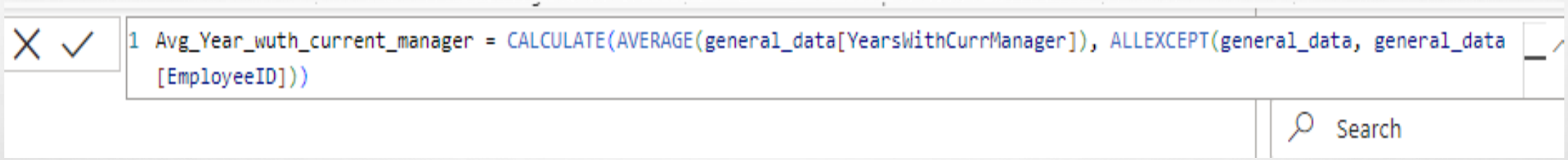
11

Q6. IN POWER BI, ESTABLISH A RELATIONSHIP BETWEEN THE "EMPLOYEEID" IN THE EMPLOYEE DATA AND THE "EMPLOYEEID" IN THE TIME TRACKING DATA.



DATA QUESTIONS

Q7. USING DAX, CREATE A CALCULATED COLUMN THAT CALCULATES THE AVERAGE YEARS AN EMPLOYEE HAS SPENT WITH THEIR CURRENT MANAGER.



To calculate the average year spent with current manager using DAX query, we have used the above given DAX code to answer the given question.

DATA QUESTIONS

Q8. USING EXCEL, CREATE A PIVOT TABLE THAT DISPLAYS THE COUNT OF EMPLOYEES IN EACH MARITAL STATUS CATEGORY, SEGMENTED BY DEPARTMENT.

Here's a quick view of a pivot table that shows how many employees are in each marital status, organized by department.

Marital Status by Department	Employees Count
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
Total	**4410**

DATA QUESTIONS

14

Q9. APPLY CONDITIONAL FORMATTING TO HIGHLIGHT EMPLOYEES WITH BOTH ABOVE-AVERAGE MONTHLY INCOME AND ABOVE-AVERAGE JOB SATISFACTION.

Initially, we used the "VLOOKUP" function to get job satisfaction scores from another Excel file. After applying special formatting to the data, we found out that 984 employees have both higher-than-average monthly income and higher-than-average job satisfaction.

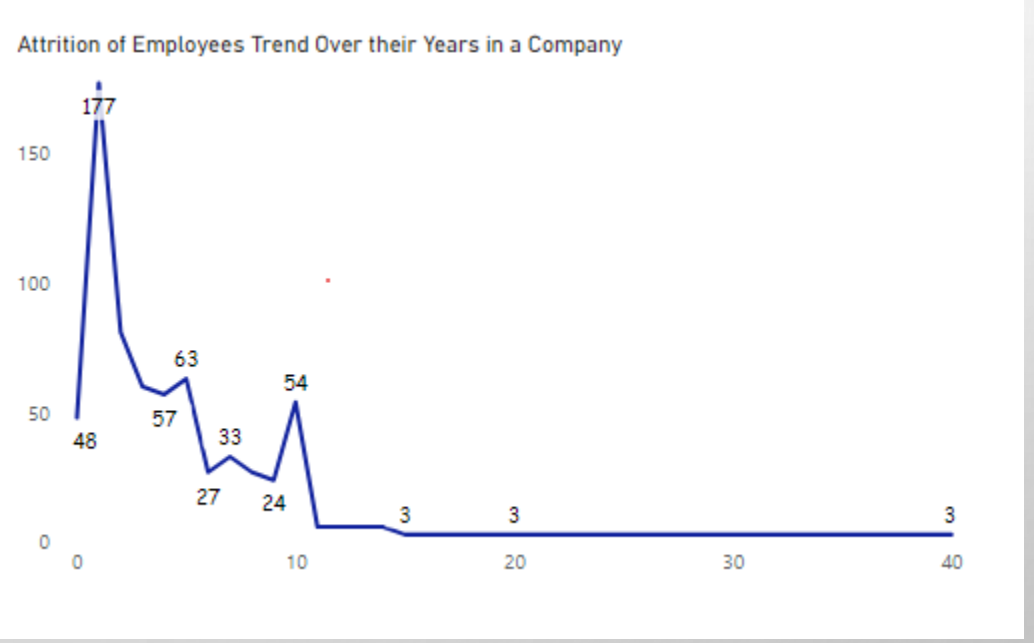
	A	B	C	D	E	F	G	H	I	J
1	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeID	Gender
107	33	Yes	Travel_Rarely	Human Reso	28	2	Human Resourc	1	106	Female
108	37	No	Travel_Rarely	Research & D	5	3	Life Sciences	1	107	Male
109	46	No	Non-Travel	Sales	2	4	Marketing	1	108	Female
110	41	Yes	Travel_Frequen	Research & D	16	1	Medical	1	109	Female
111	50	No	Travel_Rarely	Research & D	9	3	Medical	1	110	Female
112	40	Yes	Travel_Rarely	Research & D	8	3	Medical	1	111	Female
113	31	No	Travel_Rarely	Research & D	1	3	Life Sciences	1	112	Male
114	21	Yes	Travel_Rarely	Human Reso	10	2	Human Resourc	1	113	Male
115	29	No	Travel_Rarely	Research & D	1	3	Life Sciences	1	114	Female
116	35	No	Travel_Rarely	Research & D	29	3	Life Sciences	1	115	Female
117	27	No	Travel_Rarely	Sales	2	3	Life Sciences	1	116	Male
118	28	No	Travel_Rarely	Research & D	2	2	Medical	1	117	Male
119	49	No	Travel_Rarely	Sales	2	3	Technical Degre	1	118	Male
120	51	No	Travel_Rarely	Research & D	1	2	Life Sciences	1	119	Male
121	36	No	Travel_Rarely	Sales	15	2	Life Sciences	1	120	Male
122	34	Yes	Non-Travel	Research & D	7	3	Life Sciences	1	121	Male
123	55	No	Travel_Rarely	Sales	26	3	Marketing	1	122	Male
124	24	No	Travel_Rarely	Research & D	1	3	Life Sciences	1	123	Female
125	30	No	Travel_Rarely	Research & D	3	3	Life Sciences	1	124	Female
126	26	Yes	Travel_Frequen	Sales	14	3	Life Sciences	1	125	Male

DATA QUESTIONS

15

Q10. IN POWER BI, CREATE A LINE CHART THAT VISUALIZES THE TREND OF EMPLOYEE ATTRITION OVER THE YEARS.

Here's a quick look at how employee attrition changes as they spend more time in the company. You can see that when employees stay longer, they tend to be more loyal, and as a result, the rate of leaving (attrition) is lower.

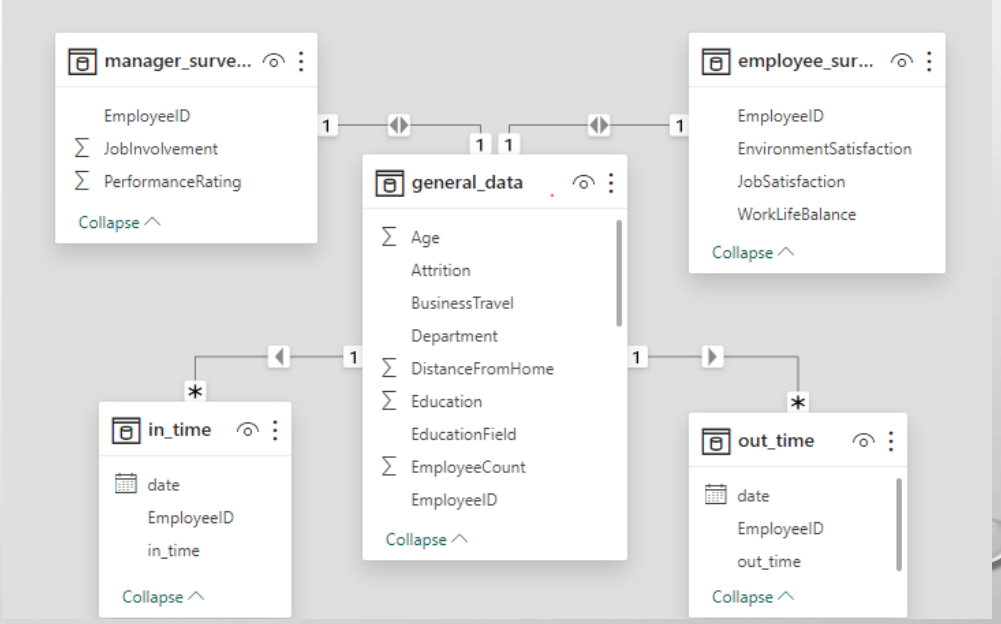


DATA QUESTIONS

Q11. DESCRIBE HOW YOU WOULD CREATE A STAR SCHEMA FOR THIS DATASET, EXPLAINING THE BENEFITS OF DOING SO.

Think of a star schema like organizing information in a simple way. There's a main table (fact table) with numbers and measurements, like EmployeeID and MonthlyIncome. Then, there are other tables (dimension tables) with details like EducationField and BusinessTravel.

This setup makes things easier, works faster, and plays well with different systems.

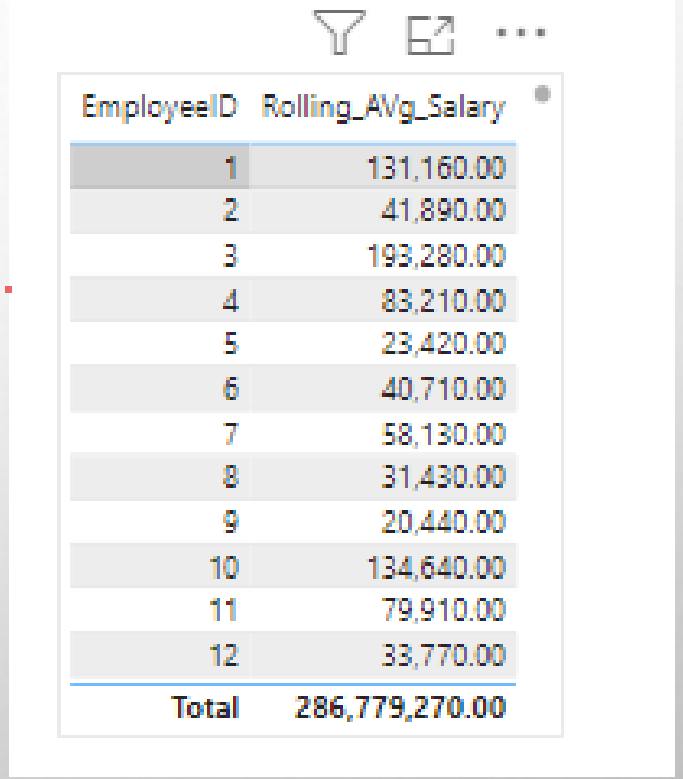


DATA QUESTIONS

17

Q12. USING DAX, CALCULATE THE ROLLING 3-MONTH AVERAGE OF MONTHLY INCOME FOR EACH EMPLOYEE.

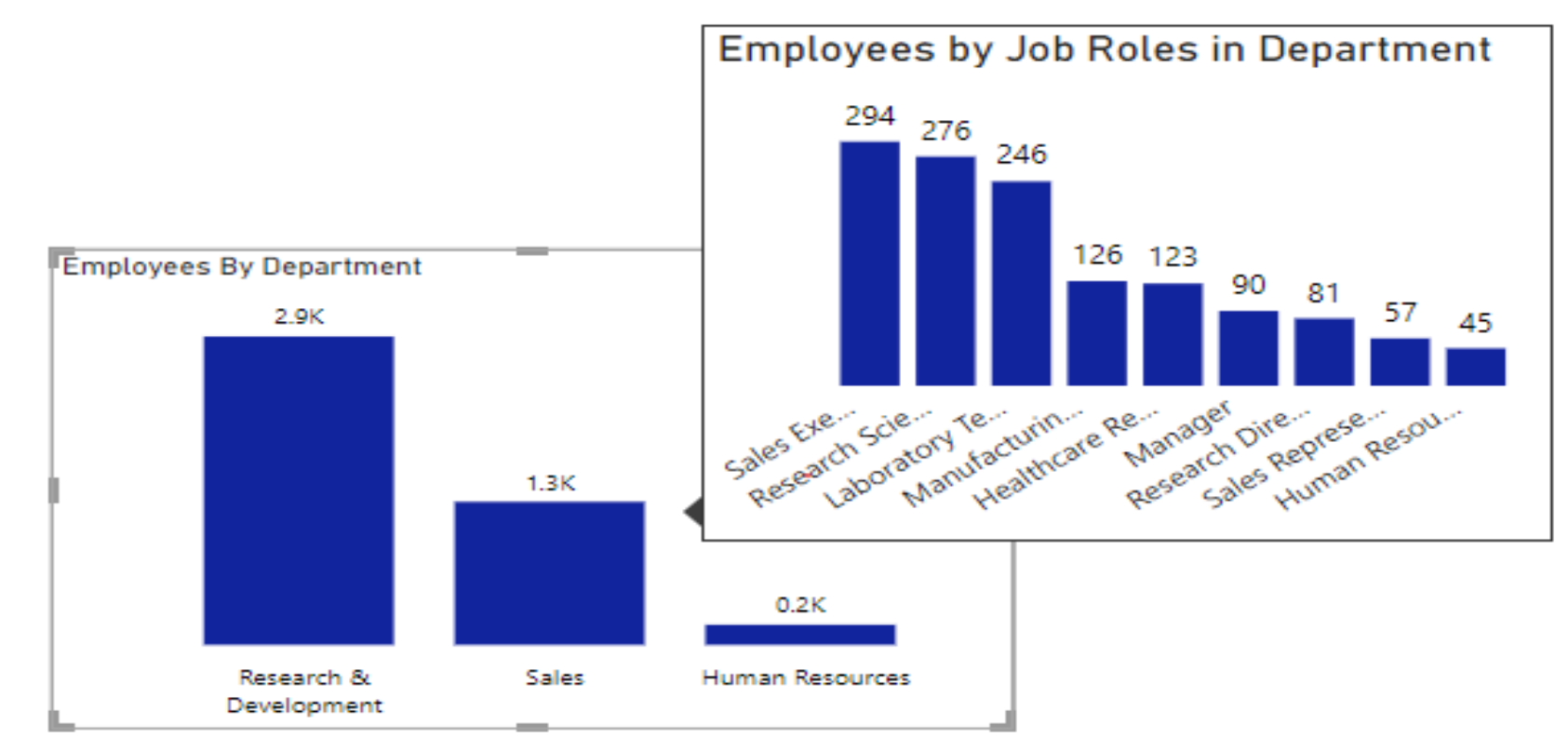
In Power BI, the DAX formula for the rolling 3-month average of Monthly Income per employee considers a dynamic window, smoothing variations and providing insights into income trends over time. This valuable metric helps analyze individual employees' financial patterns for informed decision-making in areas like compensation and financial performance.



EmployeeID	Rolling_AVg_Salary
1	131,160.00
2	41,890.00
3	193,280.00
4	83,210.00
5	23,420.00
6	40,710.00
7	58,130.00
8	31,430.00
9	20,440.00
10	134,640.00
11	79,910.00
12	33,770.00
Total	**286,779,270.00**

DATA QUESTIONS

Q13. CREATE A HIERARCHY IN POWER BI THAT ALLOWS USERS TO DRILL DOWN FROM DEPARTMENT TO JOB ROLE TO FURTHER NARROW THEIR ANALYSIS.



DATA QUESTIONS

Q14. HOW CAN YOU SET UP PARAMETERIZED QUERIES IN POWER BI TO ALLOW USERS TO FILTER DATA BASED ON THE DISTANCE FROM HOME COLUMN?

In Power BI, we can set up parameterized queries by creating a parameter in the Power Query Editor. Define the parameter using the "Manage Parameters" option and then reference it in our query, allowing users to dynamically filter data based on the Distance from the Home column, providing a more interactive and personalized analysis experience.

DATA QUESTIONS

20

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.

By configuring the pivot table with the Job Level filter accurately, we can present the monthly earnings for employees at job levels 3 and higher.

JobLevel	(Multiple Items)
Department	**Sum of MonthlyIncome**
Human Resources	3,259,140
Research & Development	53,502,900
Sales	22,974,330
Total	**79,736,370**

DATA QUESTIONS

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.

	Normal Case	10% Increase Case
Average Monthly Salary	65,014	
Average Salary Hike (%)	15.21	25.21
Average Income After Hike	74,903	81,404

If the Average Salary Hike increased by 10% then the Average Salary will have to increase by 6,501.

DATA QUESTIONS

Q17. VERIFY IF THE DATA ADHERES TO A PREDEFINED SCHEMA. WHAT ACTIONS WOULD YOU TAKE IF YOU FIND INCONSISTENCIES?

The HR department datasets provided are structured in CSV format, adhering to a predefined schema that delineates the organization, relationships, and attributes within the data. To ensure conformity to this schema, I'd employ a systematic verification process. Initially, I'd utilize data profiling and schema validation tools to compare the actual structure with the predefined schema. Any identified inconsistencies would prompt a meticulous review. I'd document these inconsistencies, communicate findings with stakeholders, and initiate corrective actions. This could involve data cleansing, modification of validation rules, and potential schema updates to align with valid variations. Continuous monitoring and quality assurance measures would be established to maintain adherence to the schema and sustain data integrity.

DATA INSIGHTS

- **DATA OVERVIEW:**
- EMPLOYEES: 2646 MALES, 1764 FEMALES.
- TOTAL MONTHLY INCOME: $286,779,270.
- AVERAGE MONTHLY INCOME: $65,029.

- **WORKFORCE STATISTICS:**
- AVERAGE ATTRITION RATE: 16.1%.
- AVERAGE SALARY HIKE: 15.21%.
- MOST EMPLOYEES RARELY TRAVEL TO WORK.

- **DEPARTMENTAL DISTRIBUTION:**
- DEPARTMENTS: R&D, SALES, HR.
- 65% OF EMPLOYEES WORK IN RESEARCH & DEVELOPMENT.

- **EDUCATIONAL BACKGROUND:**
- 39% HAVE A BACHELOR'S DEGREE.
- 27% HAVE A MASTER'S DEGREE.

- **MARITAL STATUS:**
- 45% OF EMPLOYEES ARE MARRIED.
- 31% ARE SINGLE.

DATA INSIGHTS

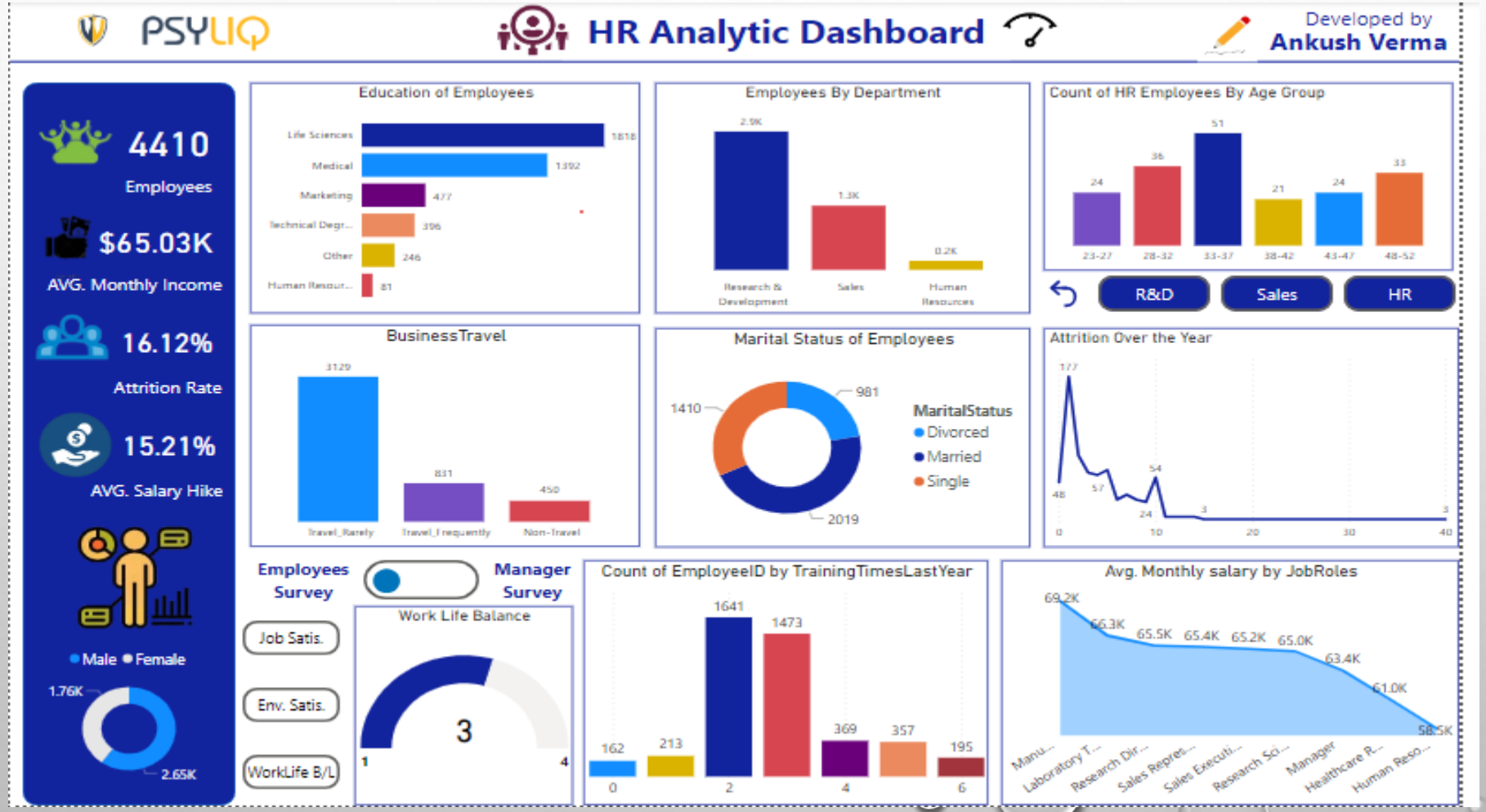
- **EMPLOYEES:**
- AVERAGE JOB & ENVIRONMENT SATISFACTION: 2.72.
- AVERAGE WORK-LIFE BALANCE: 2.75.

- **MANAGERS:**
- AVERAGE JOB INVOLVEMENT: 2.73.
- AVERAGE PERFORMANCE RATING: 3.19.

- **1 = BAD**
- **2 = GOOD**
- **3 = BETTER**
- **4 = BEST**

DASHBOARD

FROM THIS LINK, THE DASHBOARD CAN BE ACCESSED AND VIEWED



DATA ANALYST INTERNSHIP

PSYLIQ

THANK YOU

PREPARED BY

ANKUSH VERMA