PSYLIQ

Internship project

NAME: GHODASARA KRIYANSHI
DATA ANALYST USING EXCEL/POWER BI
MAIL: kriyanshighodasara@gmail.com

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

Age	(Multiple Items)		
Count of Age	Column Labels		
Row Labels	Female	Male	Grand Total
Human Resources	57	84	141
Research & Development	933	1299	2232
Sales	417	642	1059
Grand Total	1407	2025	3432

CONCLUSION: In research and development maximum people are of age above 30 as compared to others.

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

Row Labels	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
(blank)	
Grand Total	65029.31293

CONCLUSION: The job role of manufacturing director has the highest average of monthly income and lowest is of research directors.

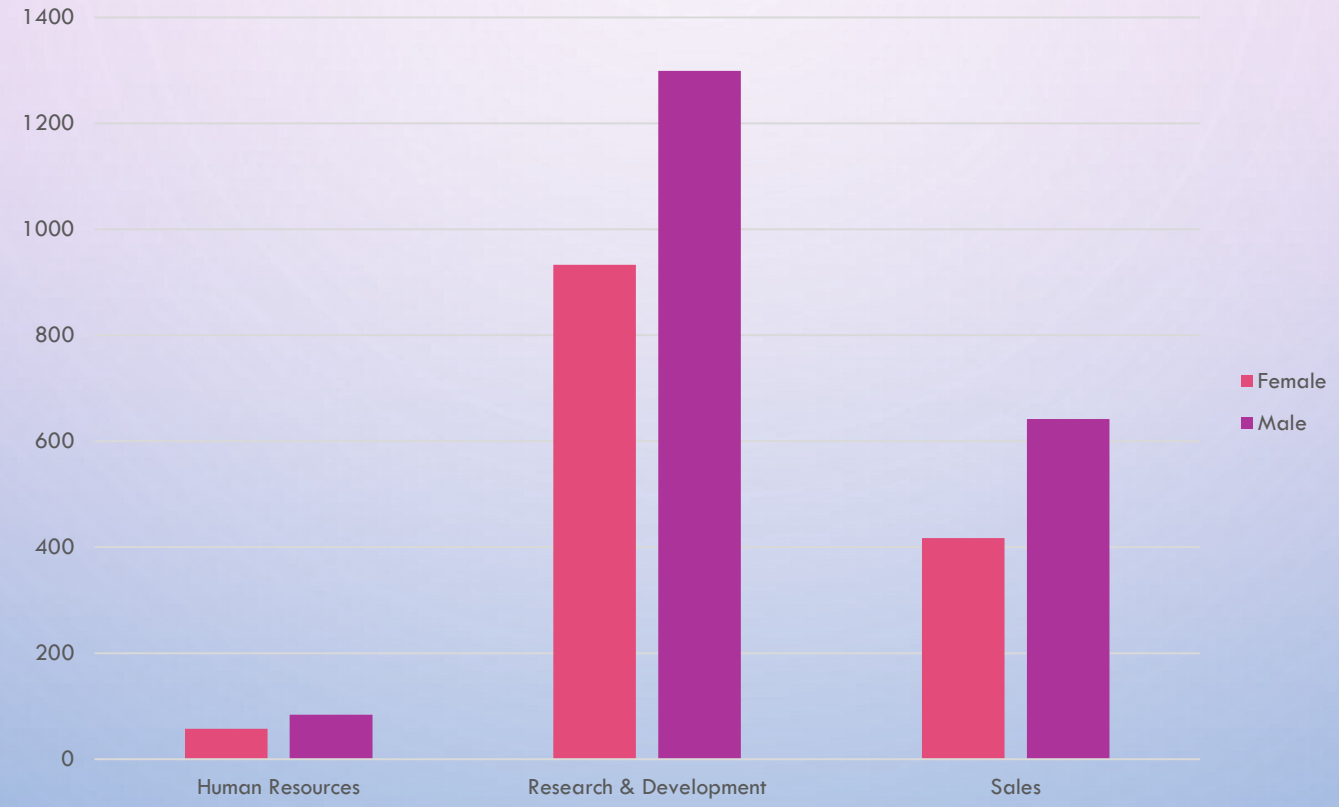
3. APPLY CONDITIONAL FORMATTING TO HIGHLIGHT EMPLOYEES WITH MONTHLY INCOME ABOVE THE COMPANY'S AVERAGE INCOME.

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

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

CONCLUSION: The pink shaded columns represent the id whose monthly income is more than total average of the company.

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



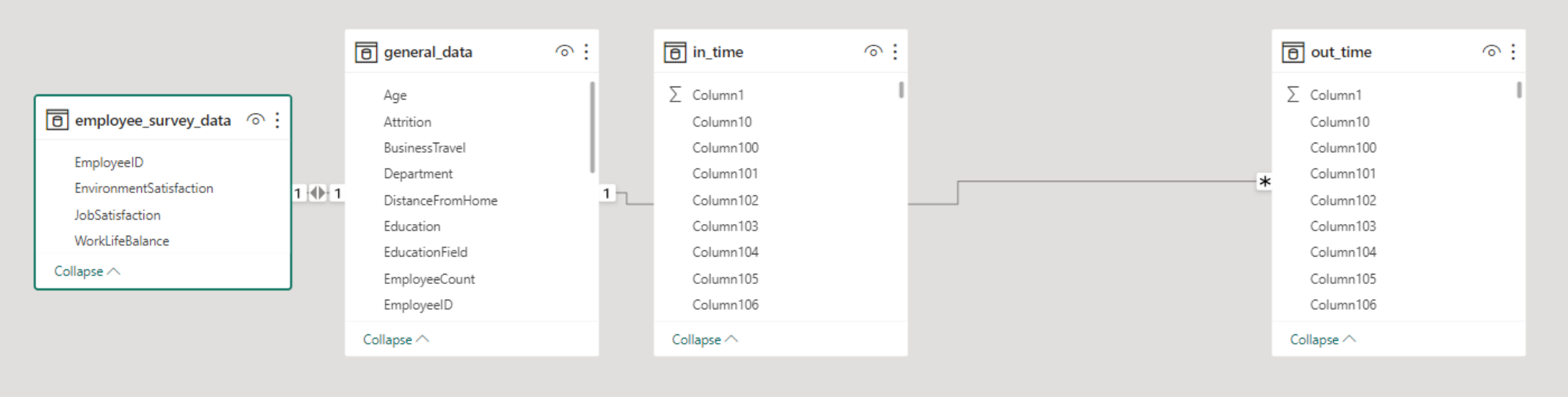
CONCLUSION: In research and development there are more number of males and overall ratio of male is seen more.

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

THERE ARE NO MISSING VALUES IN THE DEPARTMENT COLUMN. HR AND HUMAN RESOURCE ARE SAME SO THEY SHOULD STANDARDIZE FOR CONSISTENCY.

THE STANDARDIZE DATA IS ALWAYS PREFERRED AND USED TO GIVE ANY FURTHER ANALYSIS SO THAT REDUNDANCY IS REMOVED.

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



CONCLUSION: The column employee id is common in all the tables and by that one-one relationship can be observed for in-time and out-time.

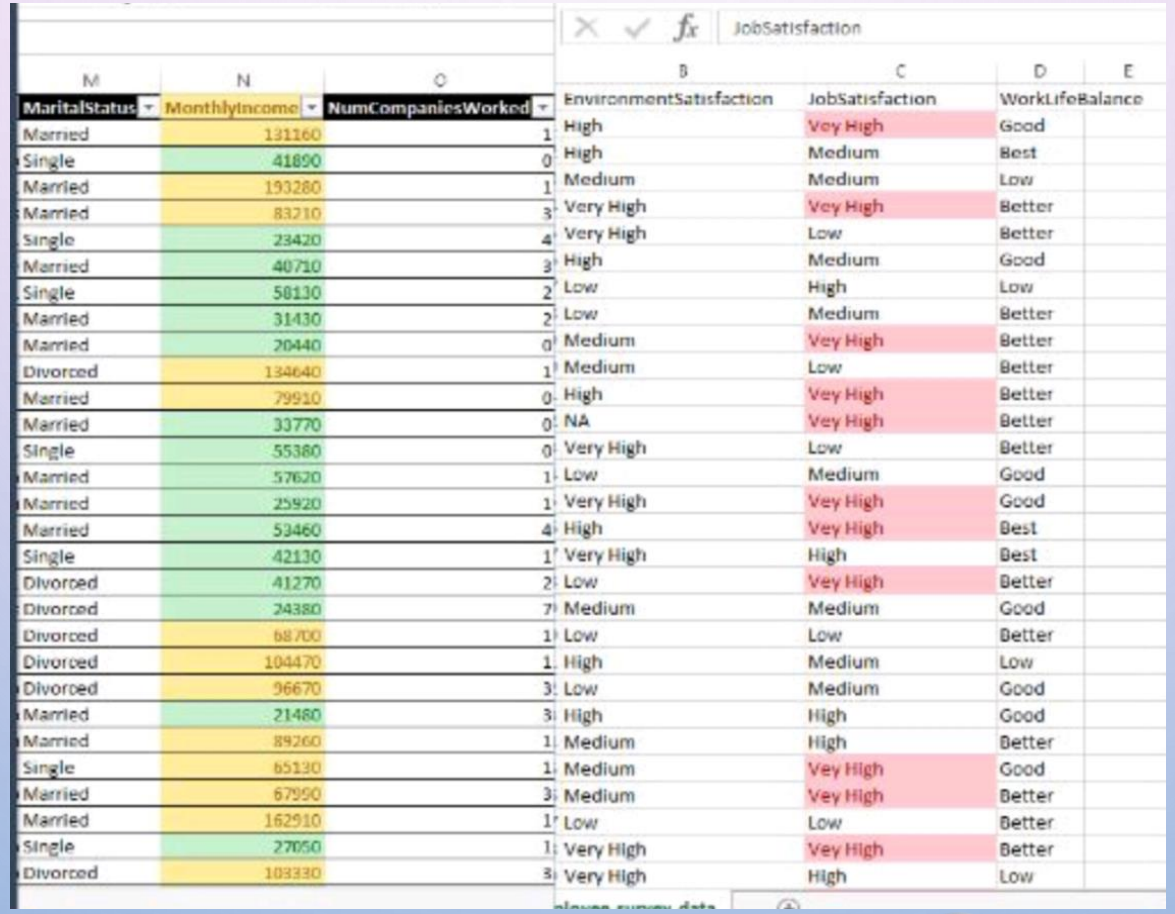
7. USING DAX, CREATE A CALCULATED COLUMN THAT CALCULATES THE AVERAGE YEARS AN EMPLOYEE HAS SPENT WITH THEIR CURRENT MANAGER.

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

Count of MaritalStatus	Column Labels				
Row Labels	Divorced	Married	Single	(blank)	Grand Total
Human Resources	21	96	72		189
Research & Development	621	1350	912		2883
Sales	339	573	426		1338
(blank)					
Grand Total	981	2019	1410		4410

CONCLUSION: Maximum employees are married and minimum are divorced. The researcha nd development has highest married employees.

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

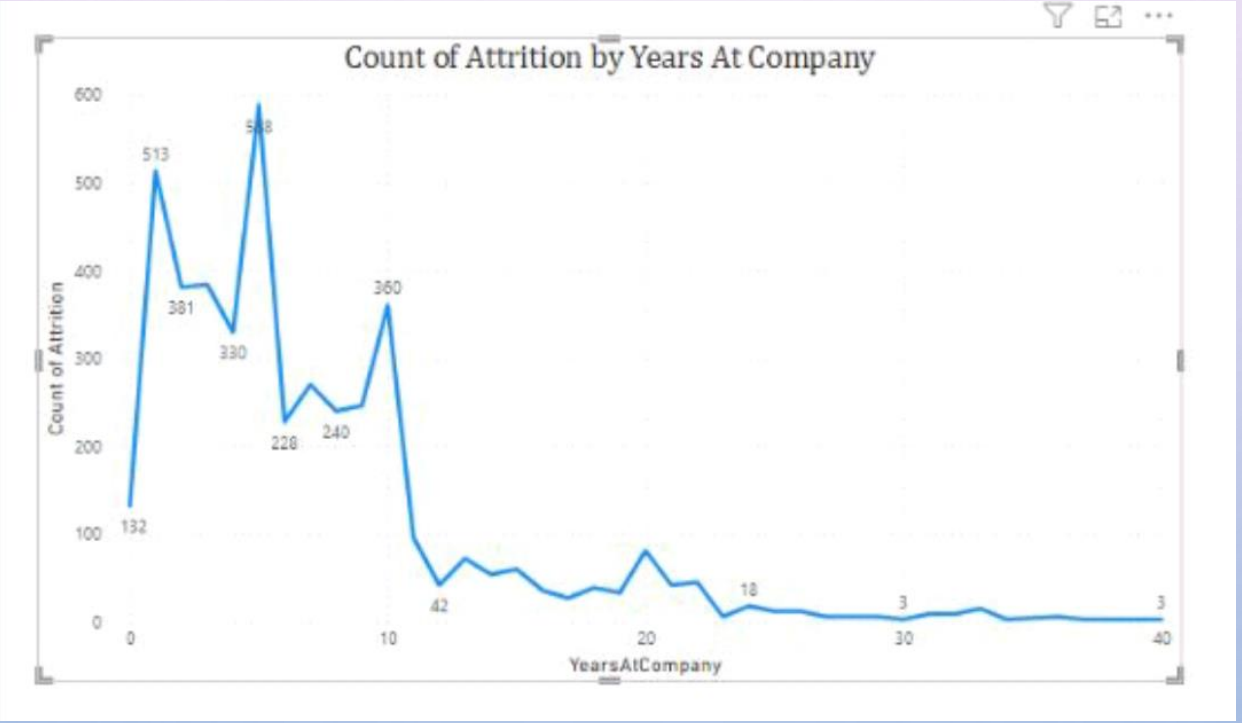


M	N	O
MaritalStatus	MonthlyIncome	NumCompaniesWorked
Married	131160	1
Single	41890	0
Married	193280	1
Married	83210	3
Single	23420	4
Married	40710	3
Single	58130	2
Married	31430	2
Married	20440	0
Divorced	134640	1
Married	79910	0
Married	33770	0
Single	55380	0
Married	57620	1
Married	25920	1
Married	53460	4
Single	42130	1
Divorced	41270	2
Divorced	24380	7
Divorced	68700	1
Divorced	104470	1
Divorced	96670	3
Married	21480	3
Married	89260	1
Single	65130	1
Married	67990	3
Married	162910	1
Single	27050	1
Divorced	103330	3

JobSatisfaction

B	C	D	E
EnvironmentSatisfaction	JobSatisfaction	WorkLifeBalance	
High	Very High	Good	
High	Medium	Best	
Medium	Medium	Low	
Very High	Very High	Better	
Very High	Low	Better	
High	Medium	Good	
Low	High	Low	
Low	Medium	Better	
Medium	Very High	Better	
Medium	Low	Better	
High	Very High	Better	
NA	Very High	Better	
Very High	Low	Better	
Low	Medium	Good	
Very High	Very High	Good	
High	Very High	Best	
Very High	High	Best	
Low	Very High	Better	
Medium	Medium	Good	
Low	Low	Better	
High	Medium	Low	
Low	Medium	Good	
High	High	Good	
Medium	High	Better	
Medium	Very High	Good	
Medium	Very High	Better	
Low	Low	Better	
Very High	Very High	Better	
Very High	High	Low	

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.

. I WOULD CREATE IT IN SQL, ASSUMING THERE ARE TWO TABLES EMPLOYEE TABLE AND DEPARTMENT TABLE.

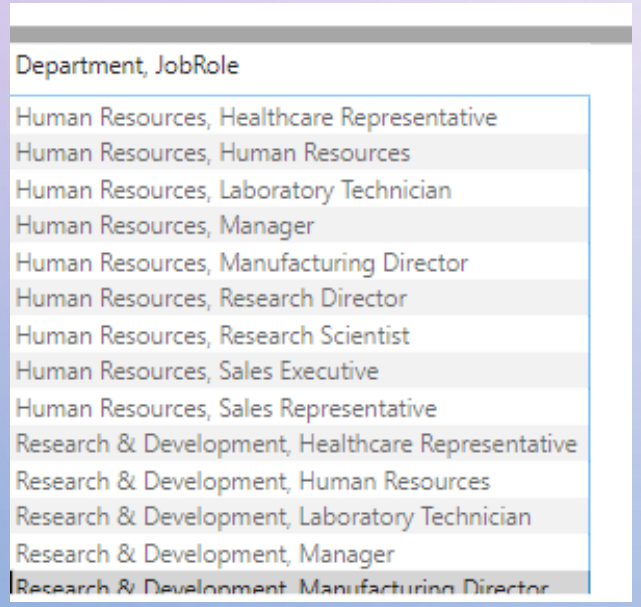
BENEFITS OF STAR SCHEMA:

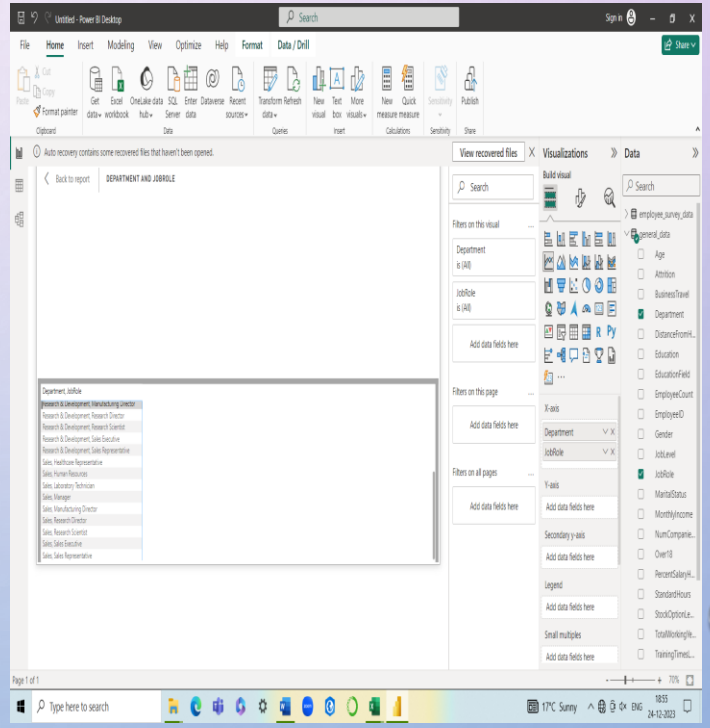
- SIMPLICITY AND UNDERSTANDABILITY (USER-FRIENDLY FOR TECHNICAL AND NON-TECHNICAL USERS)
- PERFORMANCE (QUERY PERFORMANCE IS BETTER IN STAR SCHEMA TO NORMALIZE THE SCHEMA)
- FLEXIBILITY (TO ADAPT AND EXTEND NEW DIMENSIONS OR FACTS)
- AGGREGATION (STORED AT DIMENSION LEVEL, IMPROVING QUERY PERFORMANCE)
- SEPARATIONS OF CONCERNS (CLEAR SEPARATION BETWEEN DIMENSION AND FACT TABLES SIMPLIFIES MAINTENANCE AND UPDATES)
- SCALABILITY (PERFORM GOOD EVEN WHEN DATA INCREASE)
- TOOL COMPATIBILITY (IS COMPATIBLE WITH MANY BI TOOLS AND REPORTING SYSTEM)

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

```
ROLLING 3-MONTH AVG=
CALCULATE(AVERAGE('YOUR TABLENAME'[MONTHLY INCOME]),
        FILTER(ALL('YOUR TABLE NAME'),
                'YOUR TABLENAME'[EMPLOYEEID]=
EARLIER('YOUR TABLENAME'[EMPLOYEEID])&&
                'YOUR TABLENAME'[DATE]<=
EARLIER('YOUR TABLENAME'[DATE])&&
                'YOUR TABLENAME'[DATE]>
DATEADD(EARLIER('YOUR TABLENAME'[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 2 OF 2 ON THE DISTANCE FROM HOME COLUMN?

- CREATE A PARAMETER

TO CREATE NEW PARAMETER, SET THE DATA TYPE AS DECIMAL NUMBERS.

- MODIFY THE QUERY

LOCATE THE QUERY (DISTANCE FROM HOME), APPLY THE FILTER AND ADD NEW QUERY, REPLACE THE ACTUAL DISTANCE WITH THE PARAMETER NAME

- USE THE PARAMETER IN THE VISUALIZATION

IN REPORT, USE THE FILTERED DATA, ADD A FILTER FOR THE 'DISTANCE FROM HOME' UNDER ADVANCE FILTERS (LESS THAN OR EQUAL TO)

- TEST AND REFINE

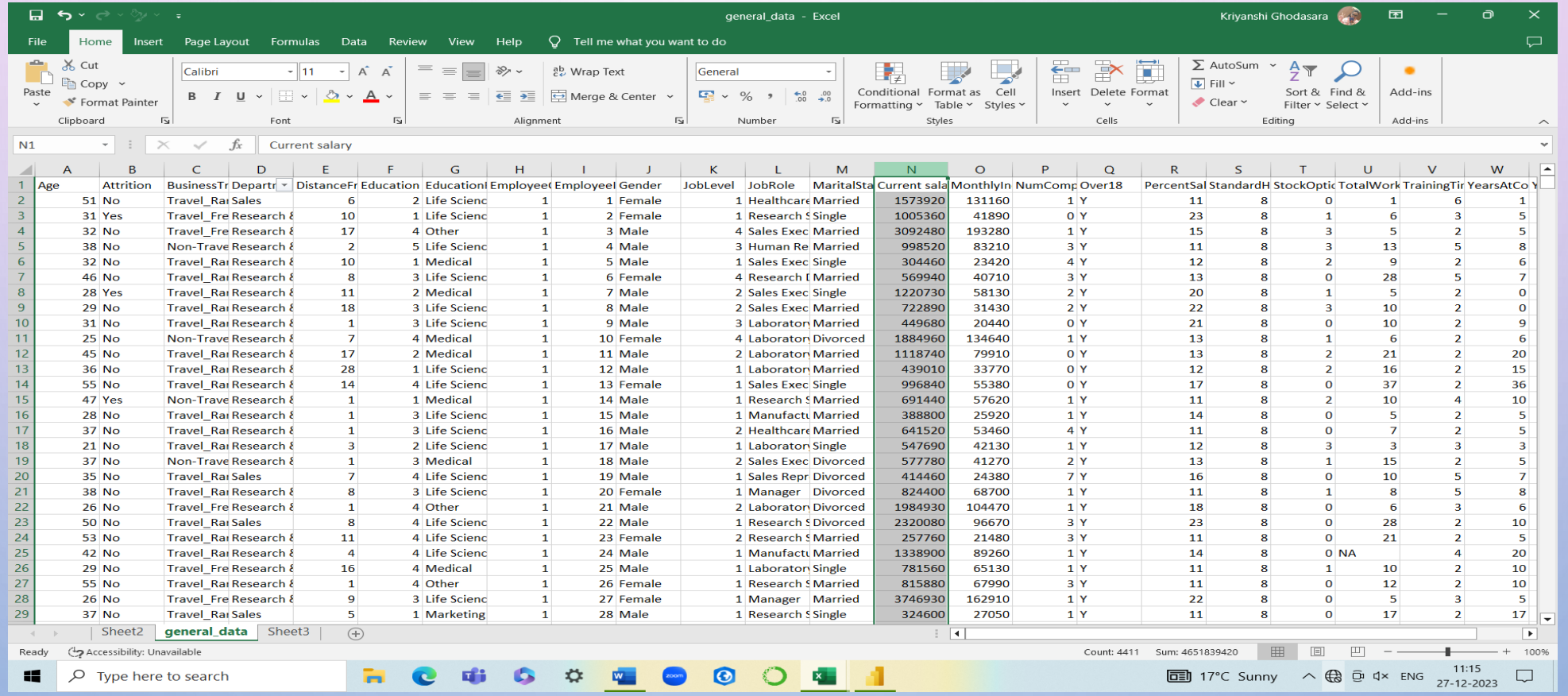
US ETHE FILTER PANE TO TEST THE VALUES AND UPDATE ACCORDINGLY

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.

JobLevel	(Multiple Items)
Row Labels	Sum of MonthlyIncome
Human Resources	3259140
Research & Development	53502900
Sales	22974330
Grand Total	79736370

CONCLUSION: Monthly income amongst the department is of Research and development.

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.

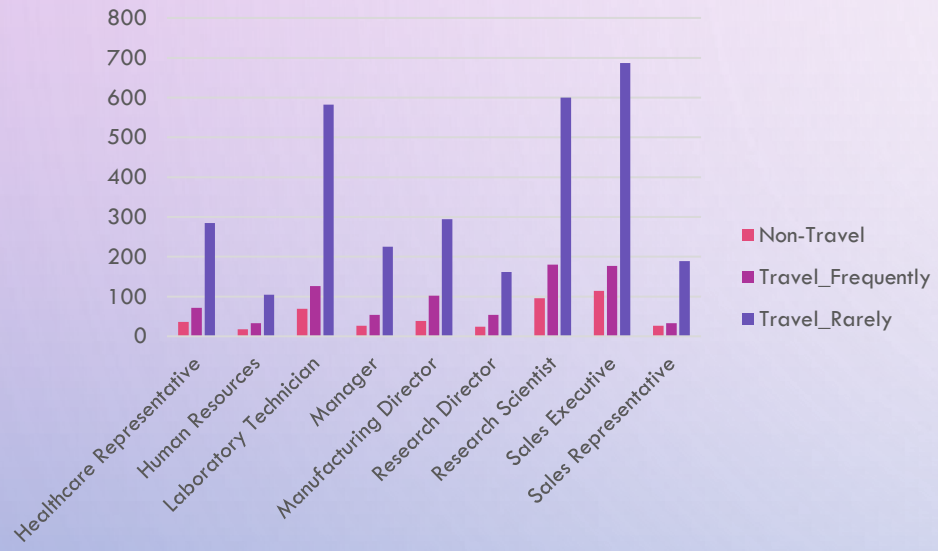


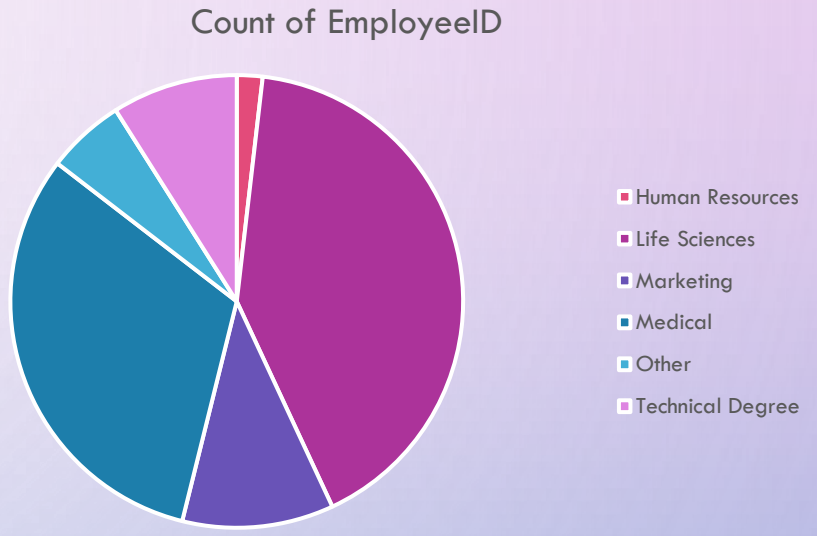
CONCLUSION: What-if analysis can be performed by using the formula monthly income(1+ percent salary hike)

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

THE GIVEN DATA ADHERES TO PREDEFINED SCHEMA. THE ACTION THAT NEEDS TO BE TAKEN TO FIND INCONSISTENCIES (DATA ENTRY ERRORS, SYSTEM ISSUES) AND RECTIFY THEM. STANDARDIZE OR CLEAN DATA VALUES TO ALIGN WITH THE SCHEMA. TO COMMUNICATE WITH RELEVANT PERSON TO DISCUSS AND RESOLVE THE SIGNIFICANT SCHEMA DEVIATIONS. VALIDATIONS NEED TO BE PERFORM ESPECIALLY, WHEN DATA IS INTRODUCED TO MAINTAIN THE QUALITY.

EXTRA GRAPHS





THANK YOU