



PSYUIQ

Internship project

NAME: GHODASARA KRIYANSHI

DATA ANALYST USING EXCEL/POWER BI

MAIL: kriyanshigodasara@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.

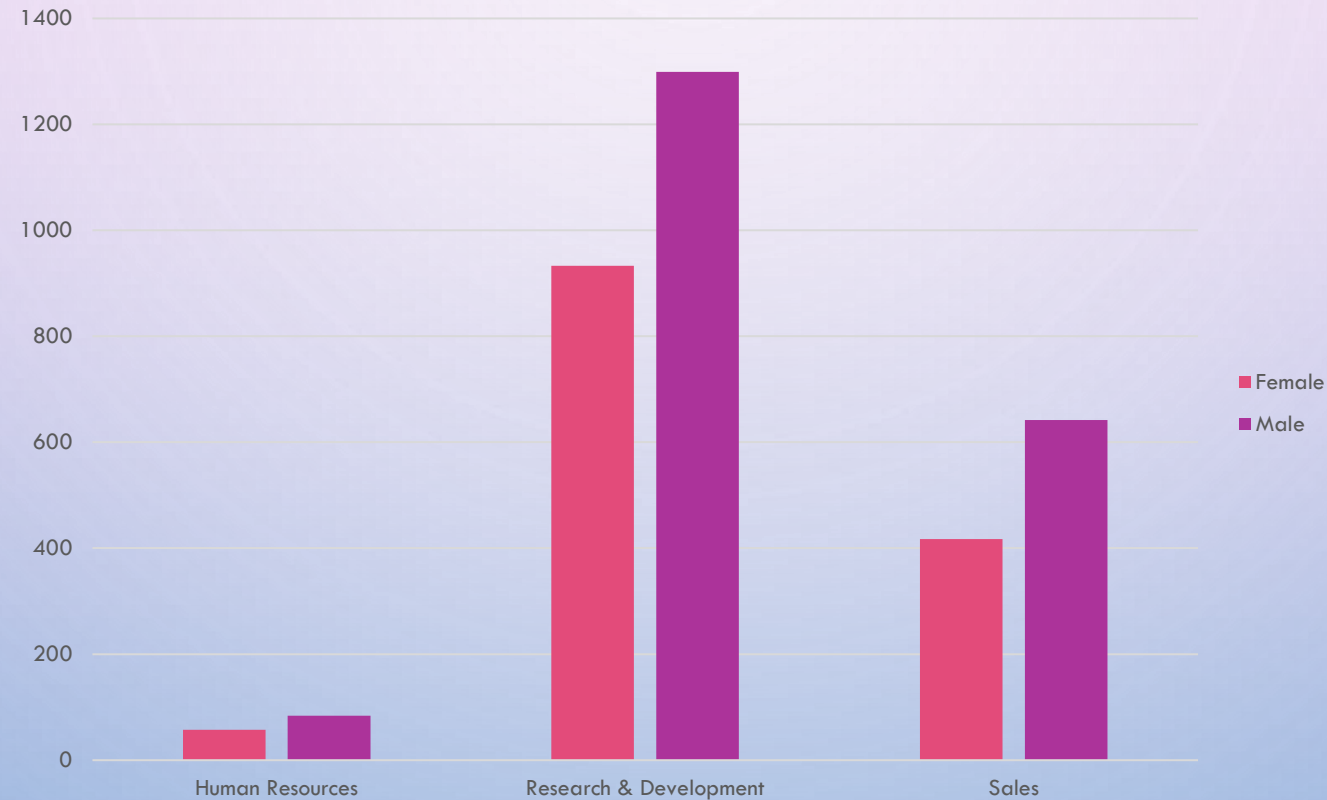
3. APPLY CONDITIONAL FORMATTING TO HIGHLIGHT EMPLOYEES WITH MONTHLY INCOME ABOVE THE COMPANY'S AVERAGE INCOME.

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 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 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.



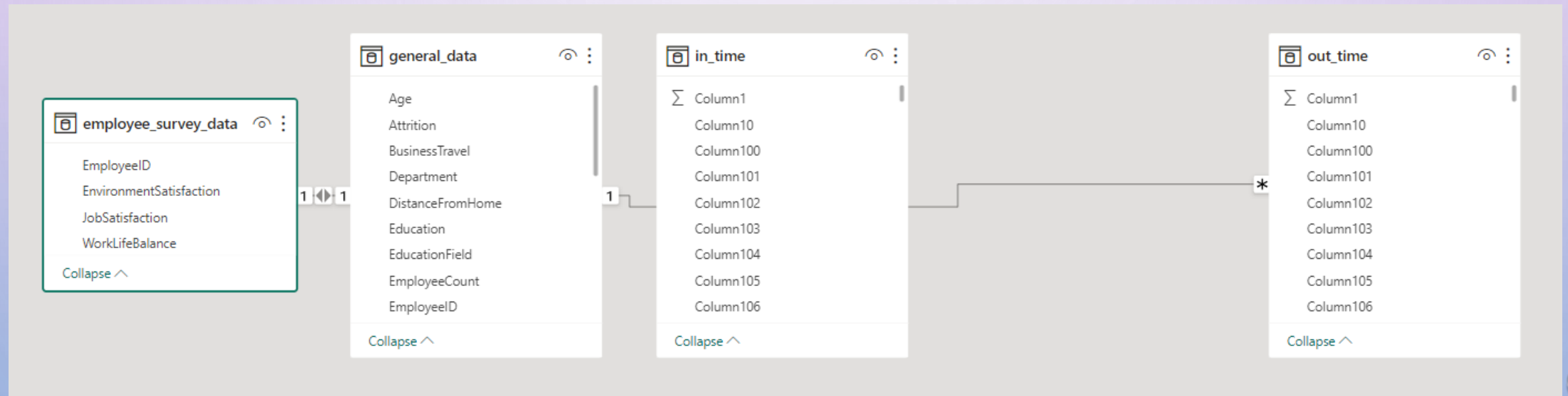
5. IDENTIFY AND CLEAN ANY MISSING OR INCONSISTENT DATA IN THE "DEPARTMENT" COLUMN.

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.

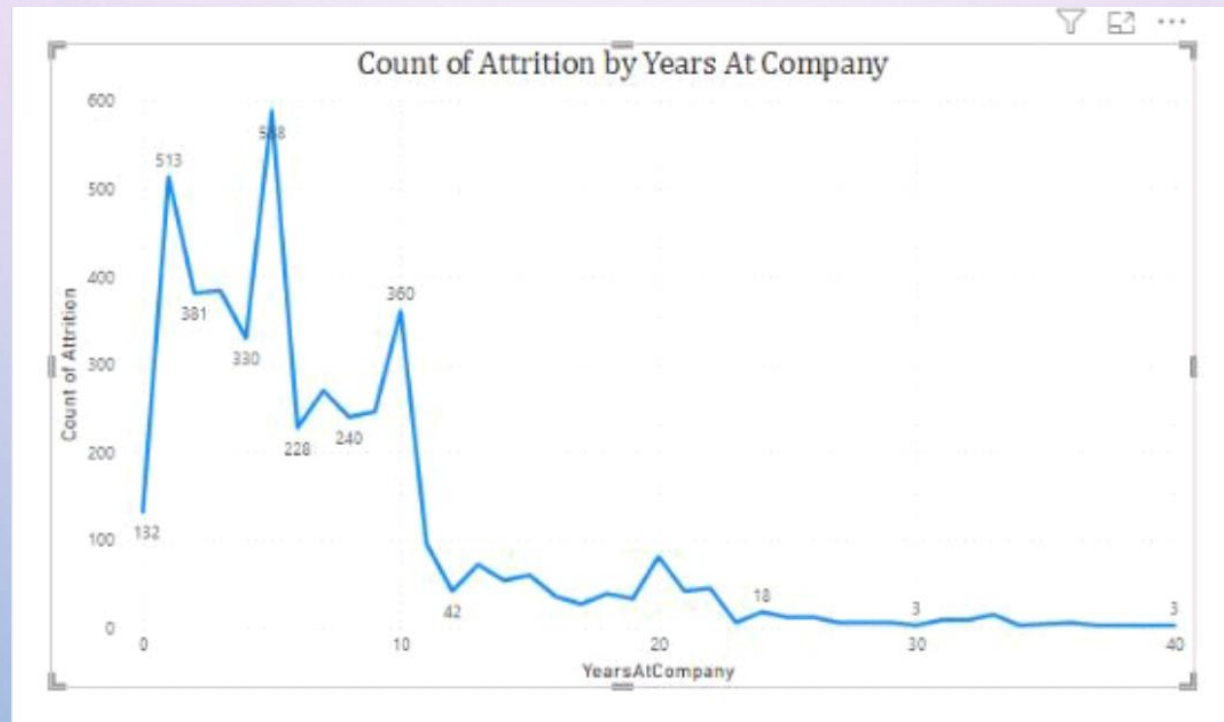
Count of MaritalStatus	Column Labels				
Row Labels	Divorced	Married	Single	(blank)	Grand Total
Human Resources	21	96	72		189
Research & Development	621	1350	912		2883
Sales	339	573	426		1338
(blank)					
Grand Total	981	2019	1410		4410

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

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

			JobSatisfaction			
M	N	O	B	C	D	E
MaritalStatus	MonthlyIncome	NumCompaniesWorked	EnvironmentSatisfaction	JobSatisfaction	WorkLifeBalance	
Married	131160	1	High	Very High	Good	
Single	41890	0	High	Medium	Best	
Married	193280	1	Medium	Medium	Low	
Married	83210	3	Very High	Very High	Better	
Single	23420	4	Very High	Low	Better	
Married	40710	3	High	Medium	Good	
Single	58130	2	Low	High	Low	
Married	31430	2	Low	Medium	Better	
Married	20440	0	Medium	Very High	Better	
Divorced	134640	1	Medium	Low	Better	
Married	79910	0	High	Very High	Better	
Married	33770	0	NA	Very High	Better	
Single	55380	0	Very High	Low	Better	
Married	57620	1	Low	Medium	Good	
Married	25920	1	Very High	Very High	Good	
Married	53460	4	High	Very High	Best	
Single	42130	1	Very High	High	Best	
Divorced	41270	2	Low	Very High	Better	
Divorced	24380	7	Medium	Medium	Good	
Divorced	68700	1	Low	Low	Better	
Divorced	104470	1	High	Medium	Low	
Divorced	96670	3	Low	Medium	Good	
Married	21480	3	High	High	Good	
Married	89260	1	Medium	High	Better	
Single	65130	1	Medium	Very High	Good	
Married	67950	3	Medium	Very High	Better	
Married	162910	1	Low	Low	Better	
Single	27050	1	Very High	Very High	Better	
Divorced	103330	3	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)

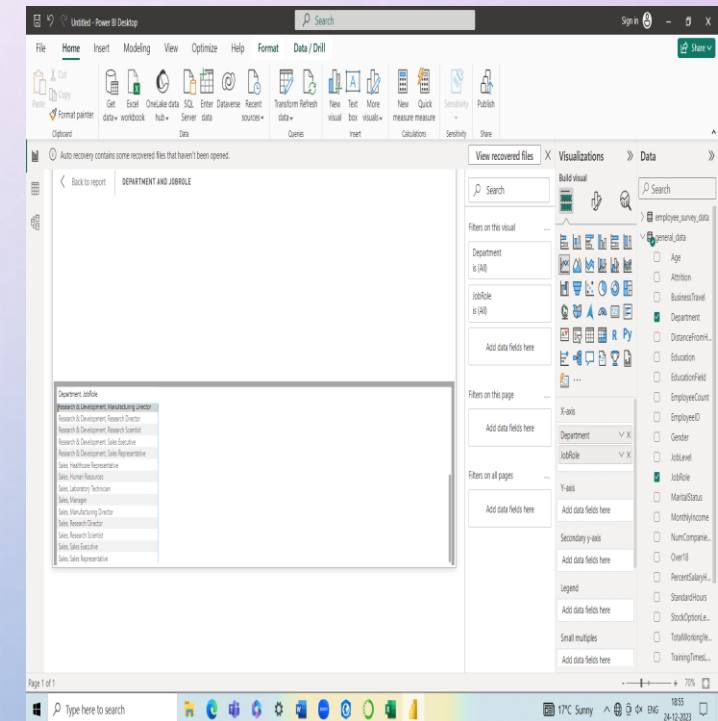
12. USING DAX, CALCULATE THE ROLLING 3-MONTH AVERAGE OF MONTHLY INCOME FOR EACH EMPLOYEE.

```
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.

Department, JobRole
Human Resources, Healthcare Representative
Human Resources, Human Resources
Human Resources, Laboratory Technician
Human Resources, Manager
Human Resources, Manufacturing Director
Human Resources, Research Director
Human Resources, Research Scientist
Human Resources, Sales Executive
Human Resources, Sales Representative
Research & Development, Healthcare Representative
Research & Development, Human Resources
Research & Development, Laboratory Technician
Research & Development, Manager
Research & Development, Manufacturing Director

Department, JobRole
Research & Development, Manufacturing Director
Research & Development, Research Director
Research & Development, Research Scientist
Research & Development, Sales Executive
Research & Development, Sales Representative
Sales, Healthcare Representative
Sales, Human Resources
Sales, Laboratory Technician
Sales, Manager
Sales, Manufacturing Director
Sales, Research Director
Sales, Research Scientist
Sales, Sales Executive
Sales, Sales Representative



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

USE THE 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.

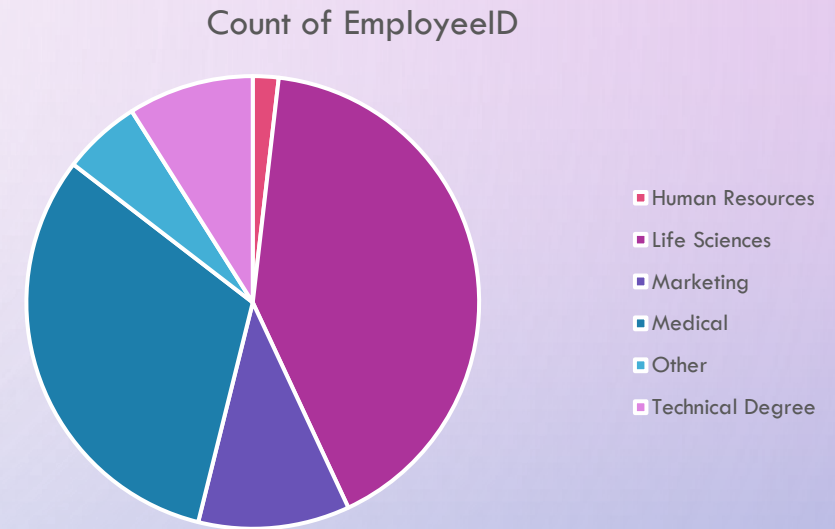
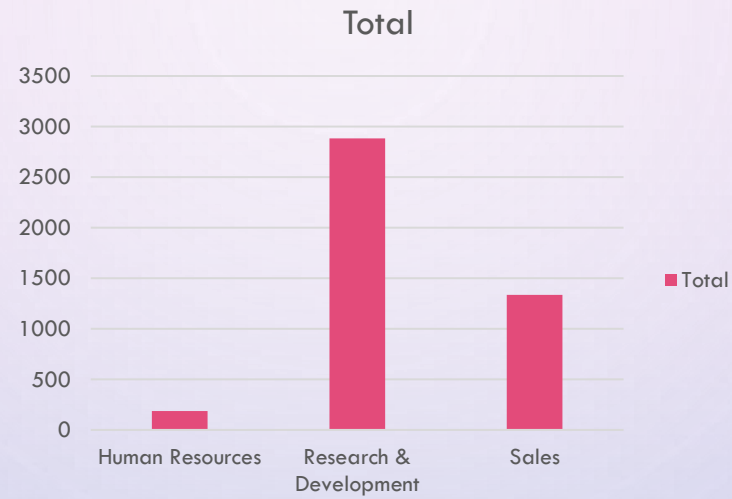
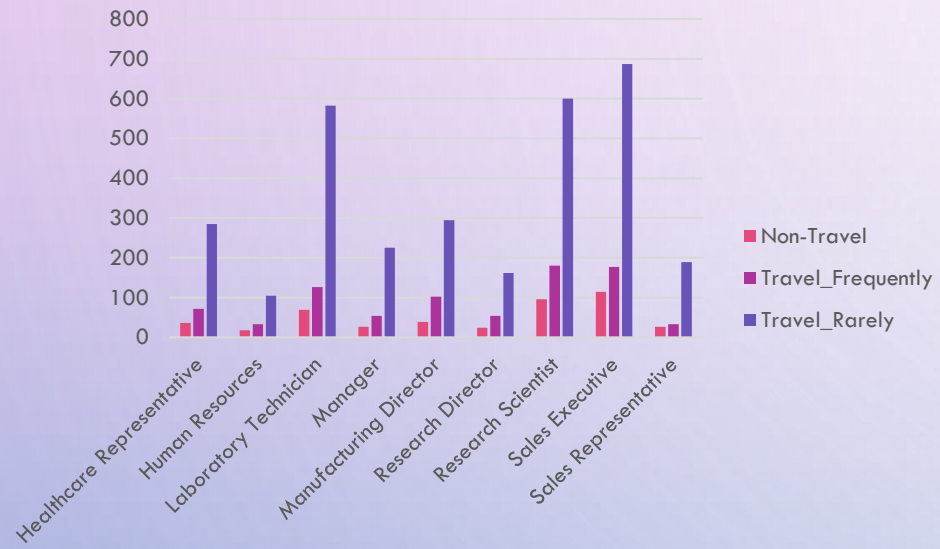
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	Gender	JobLevel	JobRole	MaritalStatus	CurrentSalary	MonthlyIncome	NumComp	Over18	PercentSalaryHike	StandardHours	StockOptions	TotalWorkingYears	TrainingTimesCompleted	YearsAtCompany	
1	51	No	Travel_Rai	Sales	6	2	Life Scienc	1	1	Female	1	Healthcare	Married	1573920	131160	1	Y	11	8	0	1	6	1	
2	31	Yes	Travel_Fre	Research &	10	1	Life Scienc	1	2	Female	1	Research &	Single	1005360	41890	0	Y	23	8	1	6	3	5	
3	32	No	Travel_Fre	Research &	17	4	Other	1	3	Male	4	Sales Exec	Married	3092480	193280	1	Y	15	8	3	5	2	5	
4	38	No	Non-Travel	Research &	2	5	Life Scienc	1	4	Male	3	Human Re	Married	998520	83210	3	Y	11	8	3	13	5	8	
5	32	No	Travel_Rai	Research &	10	1	Medical	1	5	Male	1	Sales Exec	Single	304460	23420	4	Y	12	8	2	9	2	6	
6	46	No	Travel_Rai	Research &	8	3	Life Scienc	1	6	Female	4	Research &	Married	569940	40710	3	Y	13	8	0	28	5	7	
7	28	Yes	Travel_Rai	Research &	11	2	Medical	1	7	Male	2	Sales Exec	Single	1220730	58130	2	Y	20	8	1	5	2	0	
8	29	No	Travel_Rai	Research &	18	3	Life Scienc	1	8	Male	2	Sales Exec	Married	722890	31430	2	Y	22	8	3	10	2	0	
9	31	No	Travel_Rai	Research &	1	3	Life Scienc	1	9	Male	3	Laborator	Married	449680	20440	0	Y	21	8	0	10	2	9	
10	25	No	Non-Travel	Research &	7	4	Medical	1	10	Female	4	Laborator	Divorced	1884960	134640	1	Y	13	8	1	6	2	6	
11	45	No	Travel_Rai	Research &	17	2	Medical	1	11	Male	2	Laborator	Married	1118740	79910	0	Y	13	8	2	21	2	20	
12	36	No	Travel_Rai	Research &	28	1	Life Scienc	1	12	Male	1	Laborator	Married	439010	33770	0	Y	12	8	2	16	2	15	
13	55	No	Travel_Rai	Research &	14	4	Life Scienc	1	13	Female	1	Sales Exec	Single	996840	55380	0	Y	17	8	0	37	2	36	
14	47	Yes	Non-Travel	Research &	1	1	Medical	1	14	Male	1	Research &	Married	691440	57620	1	Y	11	8	2	10	4	10	
15	28	No	Travel_Rai	Research &	1	3	Life Scienc	1	15	Male	1	Manufactu	Married	388800	25920	1	Y	14	8	0	5	2	5	
16	37	No	Travel_Rai	Research &	1	3	Life Scienc	1	16	Male	2	Healthcare	Married	641520	53460	4	Y	11	8	0	7	2	5	
17	21	No	Travel_Rai	Research &	3	2	Life Scienc	1	17	Male	1	Laborator	Single	547690	42130	1	Y	12	8	3	3	3	3	
18	37	No	Non-Travel	Research &	1	3	Medical	1	18	Male	2	Sales Exec	Divorced	577780	41270	2	Y	13	8	1	15	2	5	
19	35	No	Travel_Rai	Sales	7	4	Life Scienc	1	19	Male	1	Sales Repr	Divorced	414460	24380	7	Y	16	8	0	10	5	7	
20	38	No	Travel_Rai	Research &	8	3	Life Scienc	1	20	Female	1	Manager	Divorced	824400	68700	1	Y	11	8	1	8	5	8	
21	26	No	Travel_Fre	Research &	1	4	Other	1	21	Male	2	Laborator	Divorced	1984930	104470	1	Y	18	8	0	6	3	6	
22	50	No	Travel_Rai	Sales	8	4	Life Scienc	1	22	Male	1	Research &	Divorced	2320080	96670	3	Y	23	8	0	28	2	10	
23	53	No	Travel_Rai	Research &	11	4	Life Scienc	1	23	Female	2	Research &	Married	257760	21480	3	Y	11	8	0	21	2	5	
24	42	No	Travel_Rai	Research &	4	4	Life Scienc	1	24	Male	1	Manufactu	Married	1338900	89260	1	Y	14	8	0	NA	4	20	
25	29	No	Travel_Fre	Research &	16	4	Medical	1	25	Male	1	Laborator	Single	781560	65130	1	Y	11	8	1	10	2	10	
26	55	No	Travel_Rai	Research &	1	4	Other	1	26	Female	1	Research &	Married	815880	67990	3	Y	11	8	0	12	2	10	
27	26	No	Travel_Fre	Research &	9	3	Life Scienc	1	27	Female	1	Manager	Married	3746930	162910	1	Y	22	8	0	5	3	5	
28	37	No	Travel_Rai	Sales	5	1	Marketing	1	28	Male	1	Research &	Single	324600	27050	1	Y	11	8	0	17	2	17	

CONCLUSION: What-if analysis can be performed by using the formula $\text{monthly income} \times (1 + \text{percent salary hike})$

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