



HR Data Analysis Assessment

TANYA MITTAL (DATA ANALYST
INTERM)



INTRODUCTION

The data used in this project has been provided by PSYLIQ. It is a Human Resource dataset that covers various aspects of HR management. This dataset contains details of almost 4000+ employees from a large company, which includes information like job role, age, gender, department, salary, performance rating, attrition rate and so on. I will be using Microsoft Excel and Power BI as my tools for data cleaning, data visualization, data modelling, and interpretation. With help of these tools we can come to know the insights and get valuable outcomes.



INSIGHTS



Employee Distribution:

- Total active employees count is 3699, out of which 59.5% are Males.
- 65.7% employees are from Research & Development department.
- 43.3% of the employees belong to age group bracket of 31-40.

Attrition:

- Total Attrition count is 711 employees, at a rate of 16.1.
- Research & Development department has highest Attrition of 64%.
- 177 employees got Attritioned when YearsAtCompany is 1. Rate of attrition is 68.4% for YearsAtCompany is ≤ 5 .
- Rate of attrition is 48.9% for age group bracket of 26-35.
- In terms of JobRole sales executive has highest Attrition of 165 employees.

Rating & Hike%: NOTE: Employee_survey = (Job ,Environment) Satisfaction/Work life balance rating

- Average of Employee_survey Rating is 2.72. • Job role sales representative has lowest Employee_survey rating of 2.64.
- Job role sales executive has highest no of 3+ Employee_survey ratings which is 579.
- 20% is Highest percentage of hike for YearsAtCompany is 30.
- 23.7% of employees has both above-average Monthly Income and above-average Job Satisfaction.

HR DATA ANALYSIS

4410

Count of EmployeeID

65.03K

Average of MonthlyIncome

3.15

Average of PerformanceRating

7.01

Average of YearsAtCompany

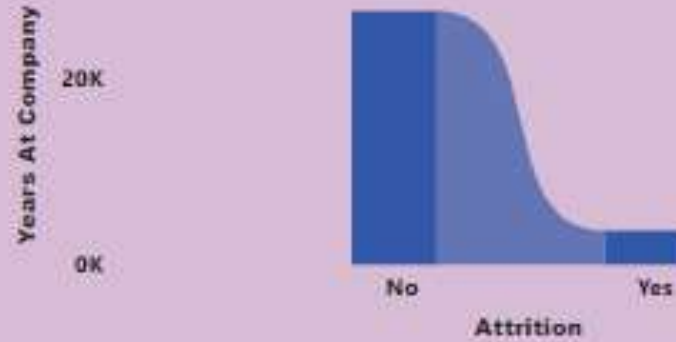
2.91

Average of Education

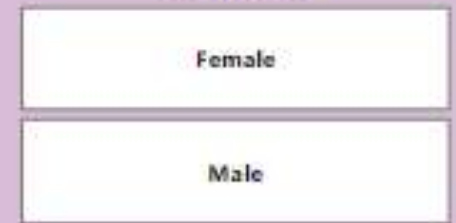
Sum of Age by EducationField



Sum of YearsAtCompany by Attrition



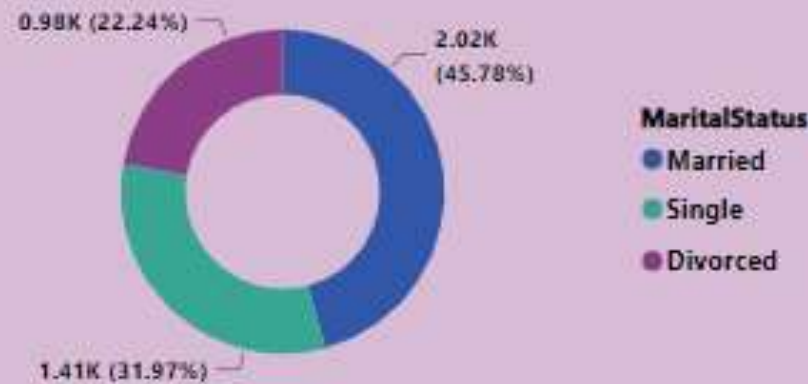
Gender



Department



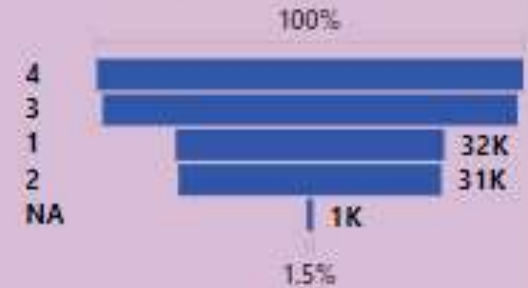
Count of Marital Status by MaritalStatus



Average of MonthlyIncome and PerformanceRating by Department and BusinessTravel



Sum of Age by Job Satisfaction





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

Age	Attrition	BusinessTravel	Department
51	No	Travel_Rarely	Sales
31	Yes	Travel_Frequently	Research & Development
32	No	Travel_Frequently	Research & Development
38	No	Non-Travel	Research & Development
32	No	Travel_Rarely	Research & Development
46	No	Travel_Rarely	Research & Development
31	No	Travel_Rarely	Research & Development
45	No	Travel_Rarely	Research & Development
36	No	Travel_Rarely	Research & Development
55	No	Travel_Rarely	Research & Development
47	Yes	Non-Travel	Research & Development
37	No	Travel_Rarely	Research & Development
37	No	Non-Travel	Research & Development
35	No	Travel_Rarely	Sales
38	No	Travel_Rarely	Research & Development
50	No	Travel_Rarely	Sales
53	No	Travel_Rarely	Research & Development
42	No	Travel_Rarely	Research & Development
55	No	Travel_Rarely	Research & Development
37	No	Travel_Rarely	Sales



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

Row Labels	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

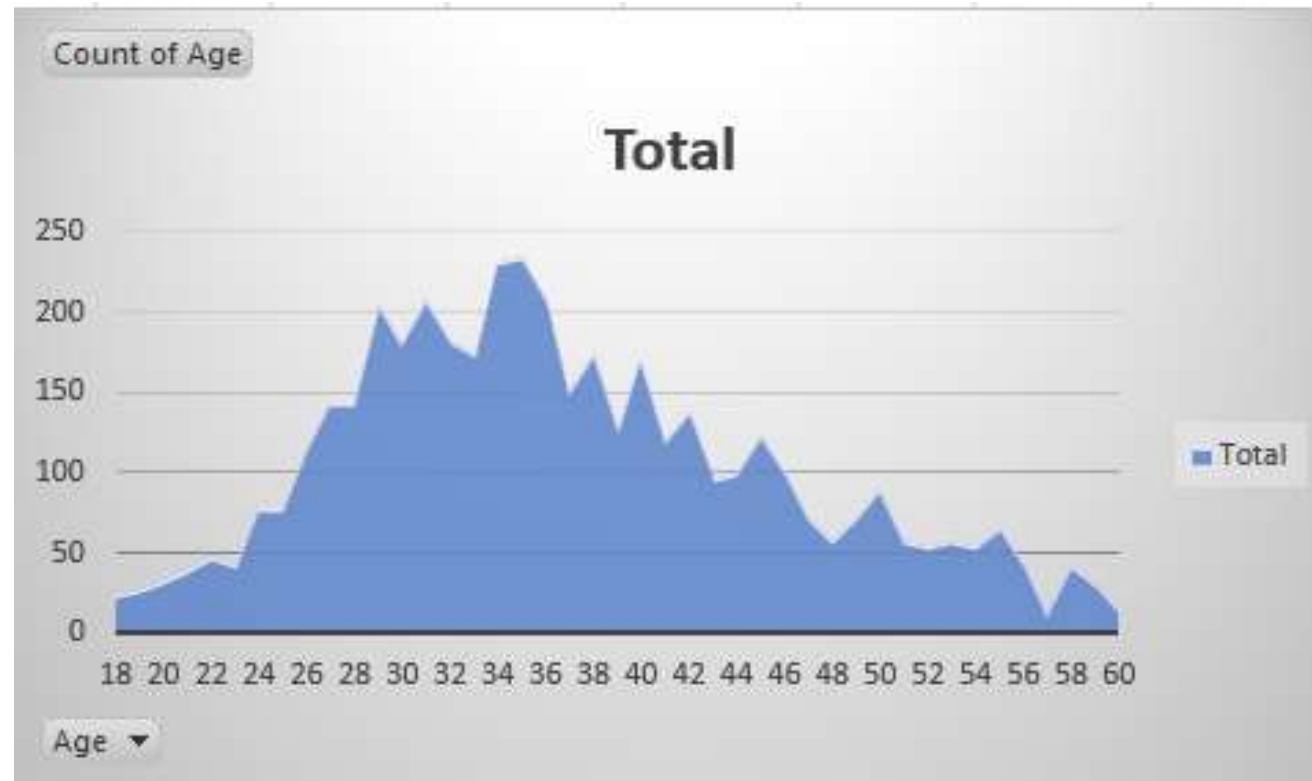


Apply conditional formatting to highlight employees with Monthly Income above the company's average income.

MonthlyIncome
131160
41890
193280
83210
23420
40710
20440
79910
33770
55380
57620
53460
41270
24380
68700
96670
21480
89260
67990
27050
103330



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



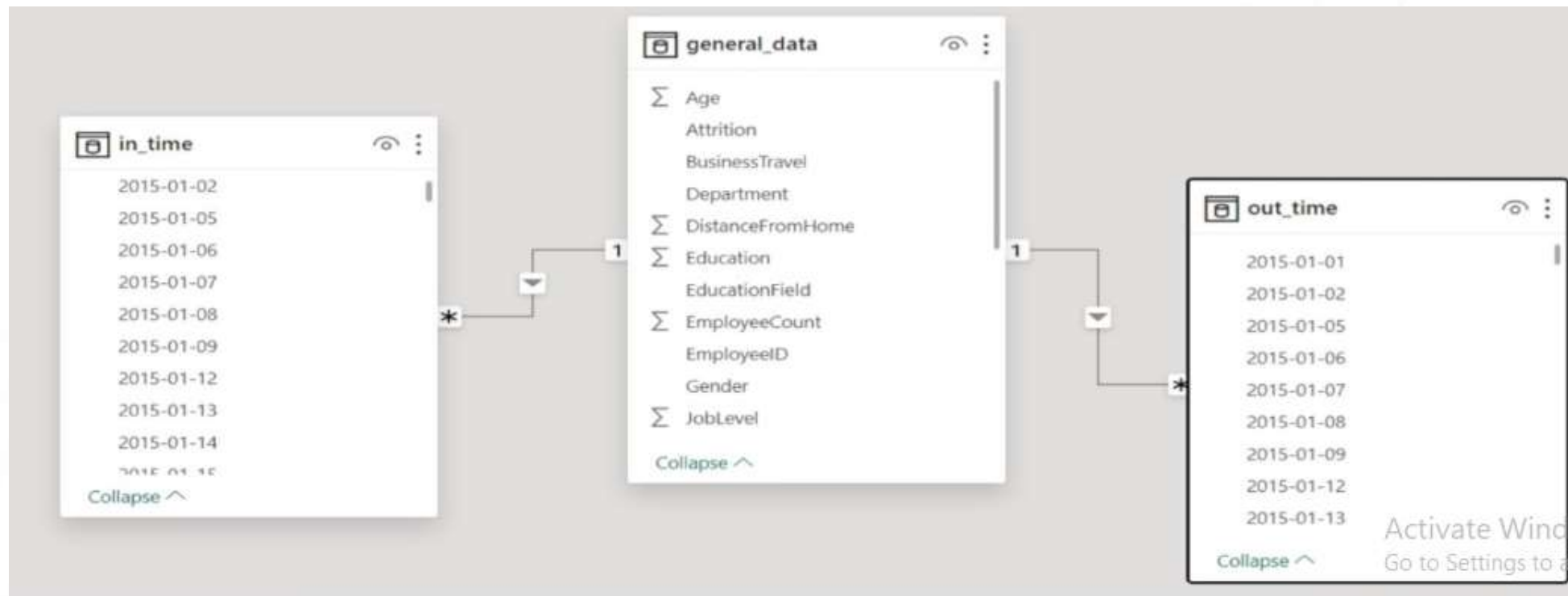


Identify and clean any missing or inconsistent data in the "Department" column.

Age	Attrition	BusinessTravel	Department	DistanceFrom
51	No	Tra		
31	Yes	Tra		
32	No	Tra		
38	No	No		
32	No	Tra		
46	No	Tra		
28	Yes	Tra		
29	No	Tra		
31	No	Tra		
25	No	No		
45	No	Tra		
36	No	Tra		
55	No	Tra		
47	Yes	No		
28	No	Tra		
37	No	Tra		
21	No	Tra		
37	No	No		
35	No	Tra		
38	No	Tra		
26	No	Travel_Frequently	Research & Development	
50	No	Travel_Rarely	Sales	




In Power BI, establish a relationship between the "EmployeeID" in the employee data and the "EmployeeID" in the time tracking data.





Using DAX, create a calculated column that calculates the average years an employee has spent with their current manager.

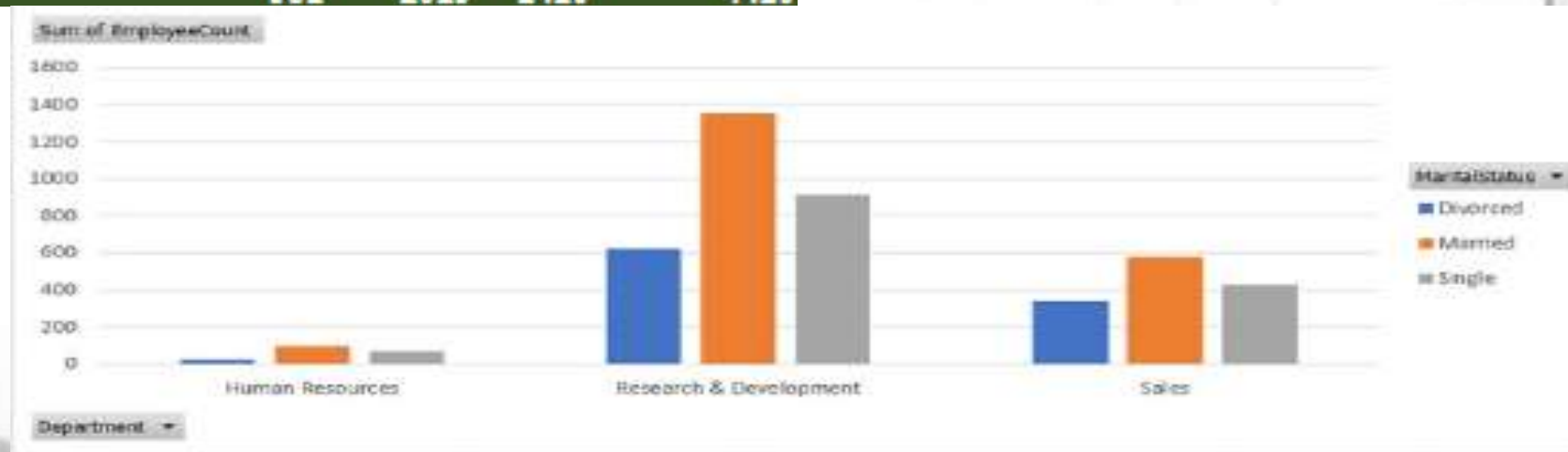
Structure	Formatting	Properties
	<input checked="" type="checkbox"/> <input type="checkbox"/>	<pre>1 Avg_YearsWithCurrMAnager = AVERAGE(general_data[YearsWithCurrManager])</pre>





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			
Marital Status and Department	Divorced	Married	Single	Grand Total
Human Resources	21	96	72	189
Research & Development	621	1350	912	2883
Sales	339	573	426	1338
Grand Total	981	2019	1410	4410



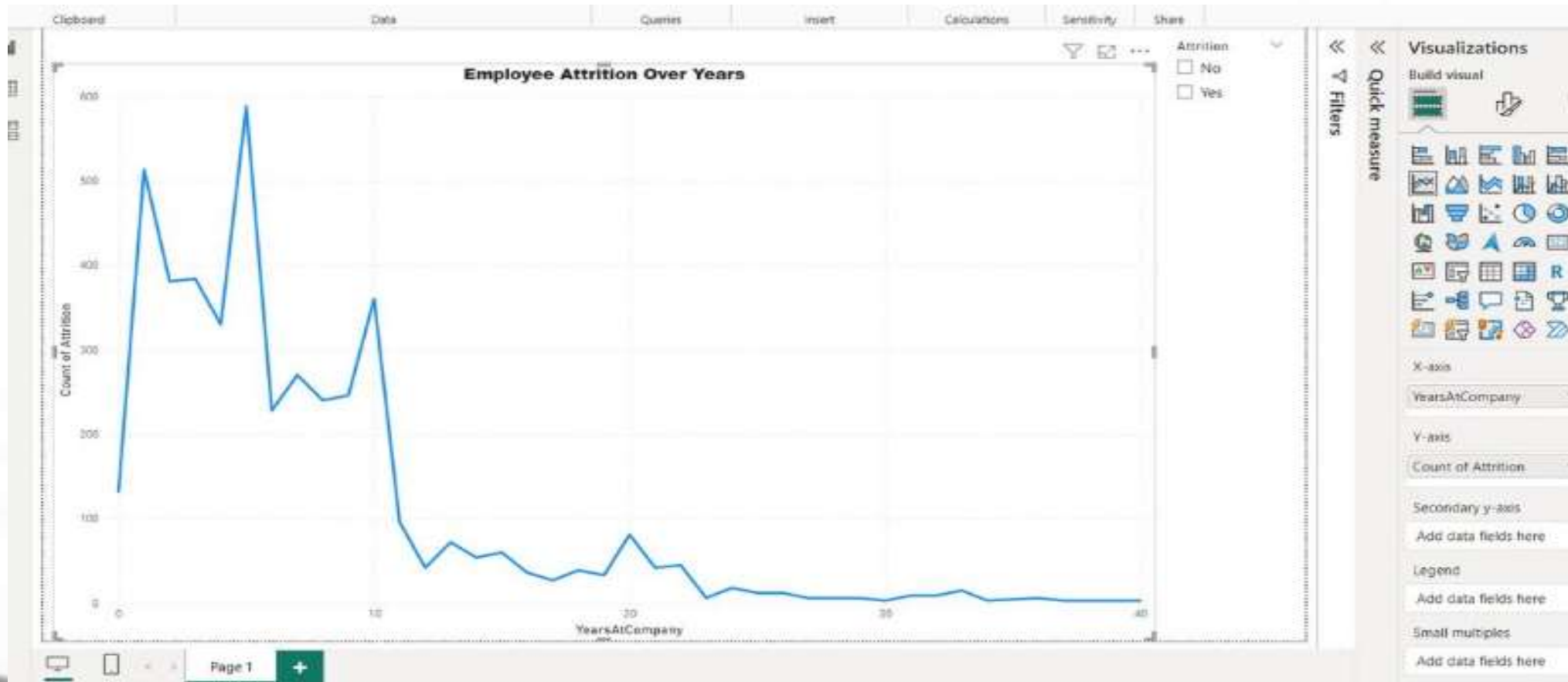


Average of MonthlyIncome	Average of JobSatisfaction
65029.31293	2.728246014

2	2	41890
3	2	193280
4	4	83210
5	1	23420
6	2	40710
7	3	58130
8	2	31430
9	4	20440
10	1	134640
11	4	79910
12	4	33770
13	1	55380
14	2	57620
15	4	25920
16	4	53460
17	3	42130
18	4	41270
19	2	24380
20	1	68700
21	2	104470
22	2	96670
23	3	21480
24	3	89260
25	4	65130
26	4	477000

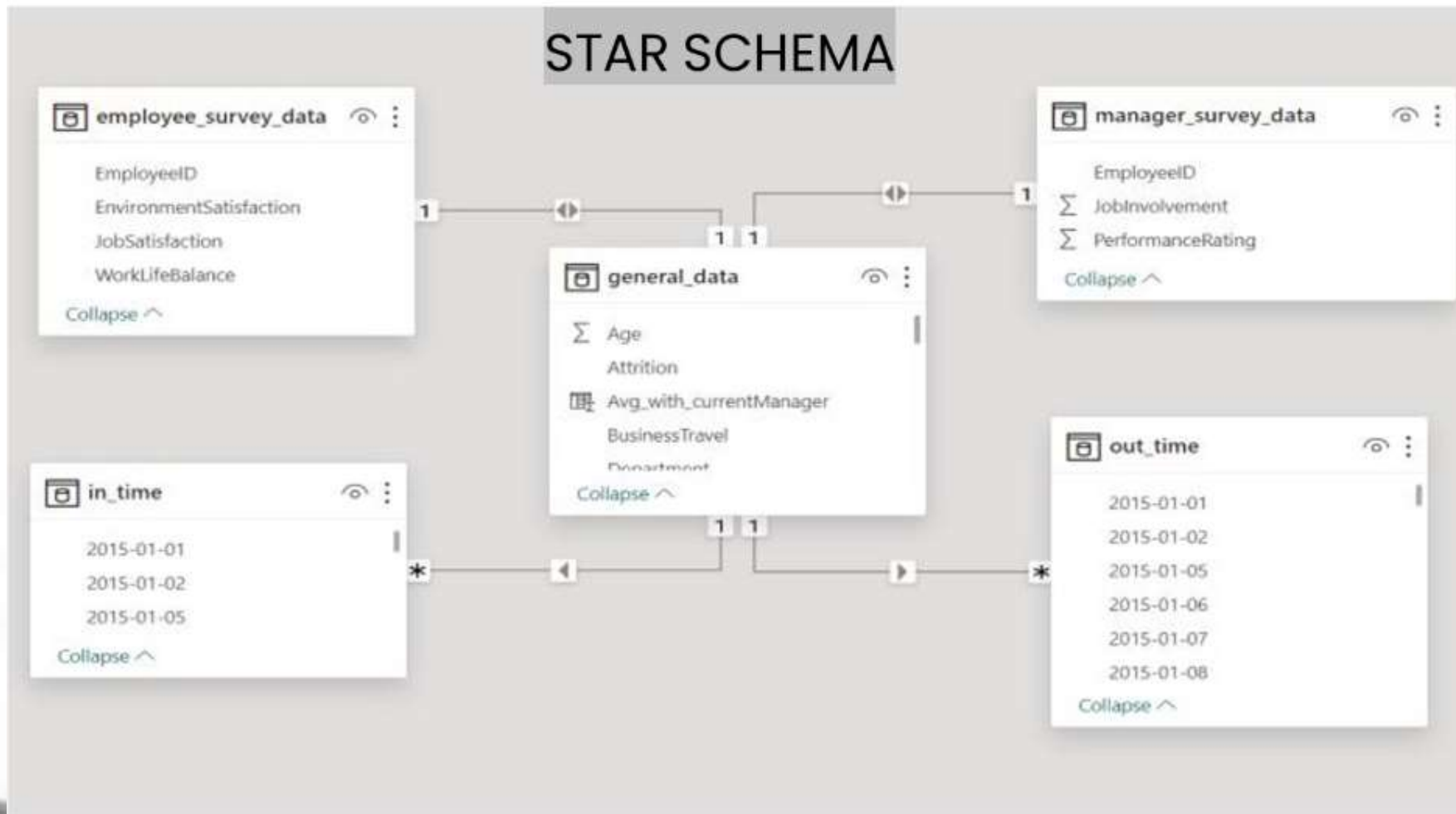


In Power BI, create a line chart that visualizes the trend of Employee Attrition over the years.





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



A star schema is a specialized data model commonly used in relational data warehouses. The schema is named after its resemblance to a star, where the fact table resides at the center, and dimension tables surround it like points in the star.



To create astar schema for any data, you can follow these steps:

- **Identify the fact table:** The fact table contains the main metrics or measurements that you want to analyze. This table typically contains numerical values and is surrounded by dimension tables.
- **Identify the dimension tables:** Dimension tables contain descriptive attributes related to the data in the fact table. These attributes provide context to the measurements in the fact table.
- **Create relationships between the fact table and dimension tables:** The fact table is connected to dimension tables through foreign keys. Each dimension table will have a primary key that is referenced as a foreign key in the fact table.
- **Normalize the dimension tables:** Dimension tables should be normalized to reduce redundancy and improve data consistency.



Benefits of using a star schema include:

- **Simplified queries:** Star schemas are optimized for querying and reporting, making it easier to retrieve specific information from the database.
- **Performance Improvement:** Star schemas are designed for faster query performance as they involve fewer joins compared to other schema designs. It optimizes query performance by minimizing redundant data.
- **Easier data analysis:** With clear relationships between the fact and dimension tables, it is easier to analyze data and gain insights from the information stored in the database.
- **Scalability:** It can handle large data volumes and maintain query performance while easily merging additional dimensions.
- **Flexibility:** Star schemas are efficient and widely used in data warehousing and analytics. They allow for easy addition of new dimensions without affecting existing structures.



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

```
Rolling 3-MonthAverage  
=CALCULATE(AVERAGE('general  
data'[MonthlyIncome]),DATESINPERIOD  
(('general data'[Date],LASTDATE('general  
data'[DATE]), -3, MONTH)
```

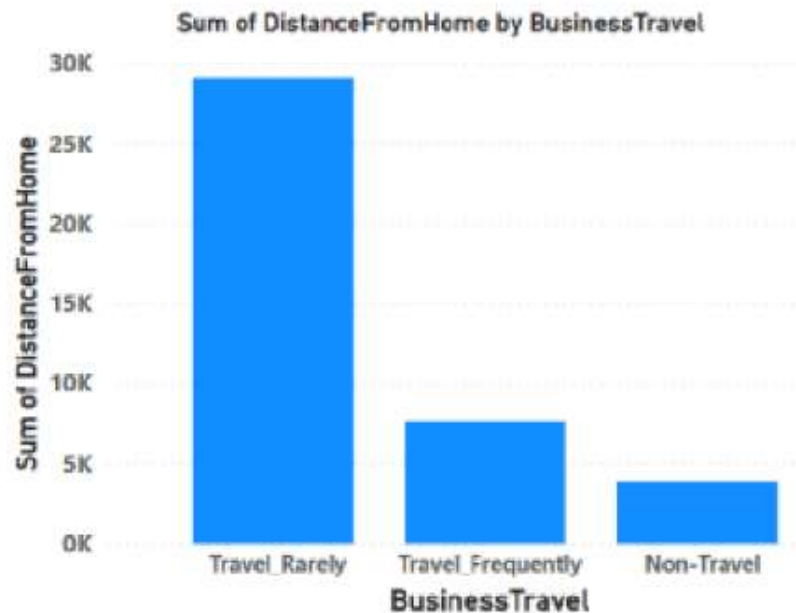


Create a hierarchy in Power BI that allows users to drill down from Department to Job Role to further narrow their analysis.

Department	Sum of MonthlyIncome
Human Resources	10943940
Healthcare Representative	831240
Human Resources	60420
Laboratory Technician	3031530
Manager	788430
Manufacturing Director	1336560
Research Director	223380
Research Scientist	1899960
Sales Executive	2433360
Sales Representative	339060
Research & Development	193702890
Healthcare Representative	15498330
Human Resources	6942660
Laboratory Technician	33204210
Manager	14055810
Manufacturing Director	20416080
Research Director	10586790
Research Scientist	38381580
Sales Executive	42613140
Sales Representative	12004290
Sales	82132440
Healthcare Representative	7637040
Human Resources	2127300
Laboratory Technician	15290280
Manager	4554900
Manufacturing Director	8342280
Research Director	4903380
Research Scientist	16637160
Sales Executive	18706080
Sales Representative	3934070
Total	286779270



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?



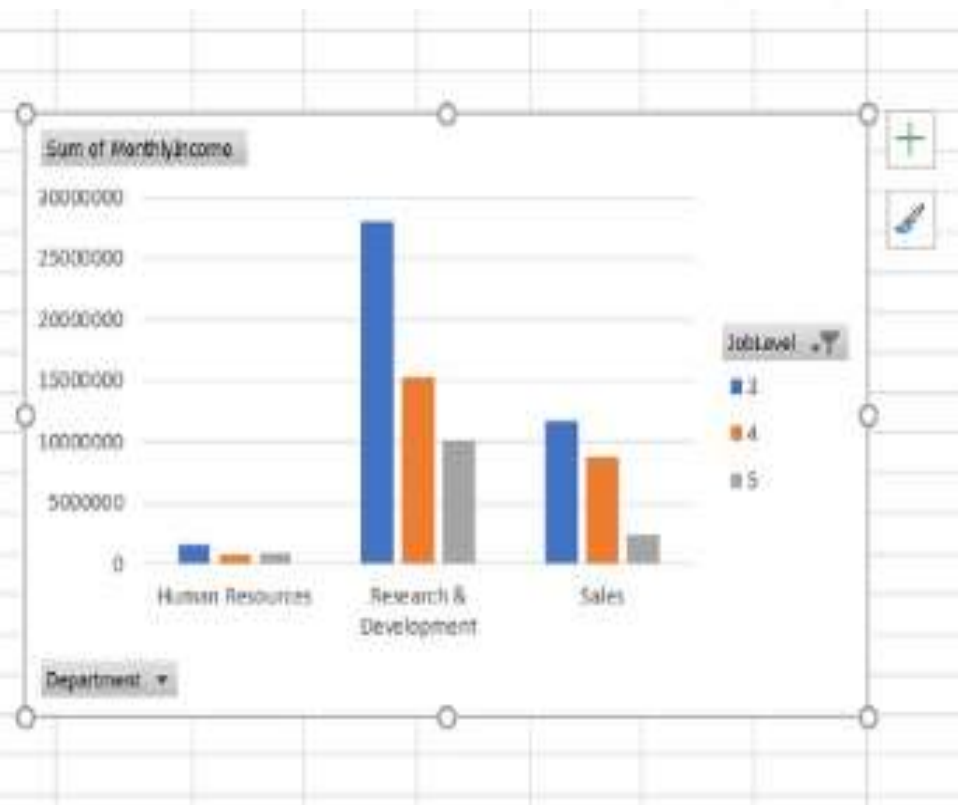
DistanceFromHome





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





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 Scenario	After 10% Hike
Average Monthly Income	65029.31293	
Average Salary Hike	15.20952381	25
Average Monthly Income after Salary Hike	1054094.196	1690762.136

Scenario Summary		
	Current Values:	New
Changing Cells:		
\$J\$7	15.20952381	25
Result Cells:		
\$J\$9	1054094.196	1690762.136
\$D\$15	15.20952381	25

Notes: Current Values column represents values of changing cells at time Scenario Summary Report was created. Changing cells for each scenario are highlighted in gray.



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

- **If we find inconsistencies in the data that do not adhere to the predefined schema, here are some actions we can take:**
 1. Review the data to find any fields, values or formats that deviate from the predefined schema.
 2. Document the inconsistencies, noting the specific fields or records that don't adhere to the specific schema.
 3. Understand the reasons behind the inconsistencies. It could be data entry errors, system issues or changes in source systems.
 4. Set up validation rules to prevent future inconsistencies. This ensures that new data entering the system conforms to the defined schema.
 5. Implement ongoing monitoring processes to catch and rectify inconsistencies early on, maintaining data quality over time.
 6. Update documentation to reflect any changes made to the schema or validation rules.
 7. Establish a continuous improvement process to regularly review and enhance data quality measures.



THANKYOU