1. 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
1	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeID	Gender	JobLevel	JobRole	
2	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	1	Female	1	Healthcare Representative	
3	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1	2	Female	1	Research Scientist	
4	32	No	Travel_Frequently	Research & Development	17	4	Other	1	3	Male	4	Sales Executive	
5	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	4	Male	3	Human Resources	
6	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	5	Male	1	Sales Executive	
7	46	No	Travel_Rarely	Research & Development	8	3	Life Sciences	1	6	Female	1	Research Director	
10	31	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	9	Male	3	Laboratory Technician	
12	45	No	Travel_Rarely	Research & Development	17	2	Medical	1	11	Male	2	Laboratory Technician	
13	36	No	Travel_Rarely	Research & Development	28	1	Life Sciences	1	12	Male	1	Laboratory Technician	
14	55	No	Travel_Rarely	Research & Development	14	4	Life Sciences	1	13	Female	1	Sales Executive	
15	47	Yes	Non-Travel	Research & Development	1	1	Medical	1	14	Male	1	Research Scientist	
17	37	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	16	Male	2	Healthcare Representative	
19	37	No	Non-Travel	Research & Development	1	3	Medical	1	18	Male	1	Sales Executive	
20	35	No	Travel_Rarely	Sales	7	4	Life Sciences	1	19	Male	1	Sales Representative	
21	38	No	Travel_Rarely	Research & Development	8	3	Life Sciences	1	20	Female	1	Manager	
23	50	No	Travel_Rarely	Sales	8	4	Life Sciences	1	22	Male	1	Research Scientist	
24	53	No	Travel_Rarely	Research & Development	11	4	Life Sciences	1	23	Female	2	Research Scientist	
25	42	No	Travel_Rarely	Research & Development	4	4	Life Sciences	1	24	Male	1	Manufacturing Director	
27	55	No	Travel_Rarely	Research & Development	1	4	Other	1	26	Female	1	Research Scientist	
29	37	No	Travel_Rarely	Sales	5	1	Marketing	1	28	Male	1	Research Scientist	
30	44	Yes	Travel_Frequently	Research & Development	1	2	Medical	1	29	Male	2	Research Scientist	
31	38	No	Travel_Rarely	Sales	2	3	Marketing	1	30	Female	1	Manager	

general_data

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

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

Company's Average Income - 65,029

	E	F	G	H	I	J	K	L	M	N	O	P	P
1	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeID	Gender	JobLevel	JobRole	MaritalStatus	MonthlyIncome	NumCompaniesWorked	Over18	
2	6	2	Life Sciences	1	1	Female	1	Healthcare Representative	Married	131160	1	Y	
3	10	1	Life Sciences	1	2	Female	1	Research Scientist	Single	41890	0	Y	
4	17	4	Other	1	3	Male	4	Sales Executive	Married	193280	1	Y	
5	2	5	Life Sciences	1	4	Male	3	Human Resources	Married	83210	3	Y	
6	10	1	Medical	1	5	Male	1	Sales Executive	Single	23420	4	Y	
7	8	3	Life Sciences	1	6	Female	4	Research Director	Married	40710	3	Y	
8	11	2	Medical	1	7	Male	2	Sales Executive	Single	58130	2	Y	
9	18	3	Life Sciences	1	8	Male	2	Sales Executive	Married	31430	2	Y	
10	1	3	Life Sciences	1	9	Male	3	Laboratory Technician	Married	20440	0	Y	
11	7	4	Medical	1	10	Female	4	Laboratory Technician	Divorced	134640	1	Y	
12	17	2	Medical	1	11	Male	1	Laboratory Technician	Married	79910	0	Y	
13	28	1	Life Sciences	1	12	Male	1	Laboratory Technician	Married	33770	0	Y	
14	14	4	Life Sciences	1	13	Female	1	Sales Executive	Single	55380	0	Y	
15	1	1	Medical	1	14	Male	1	Research Scientist	Married	57620	1	Y	
16	1	3	Life Sciences	1	15	Male	1	Manufacturing Director	Married	25920	1	Y	
17	1	3	Life Sciences	1	16	Male	2	Healthcare Representative	Married	53460	4	Y	
18	3	2	Life Sciences	1	17	Male	1	Laboratory Technician	Single	42130	1	Y	
19	1	3	Medical	1	18	Male	1	Sales Executive	Divorced	41270	2	Y	
20	7	4	Life Sciences	1	19	Male	1	Sales Representative	Divorced	24380	7	Y	
21	8	3	Life Sciences	1	20	Female	1	Manager	Divorced	68700	1	Y	
22	1	4	Other	1	21	Male	2	Laboratory Technician	Divorced	104470	1	Y	
23	8	3	Life Sciences	1	22	Male	1	Research Scientist	Divorced	96670	3	Y	

Avg monthly income by job role | **general_data** | +

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

Age Distribution of Employees
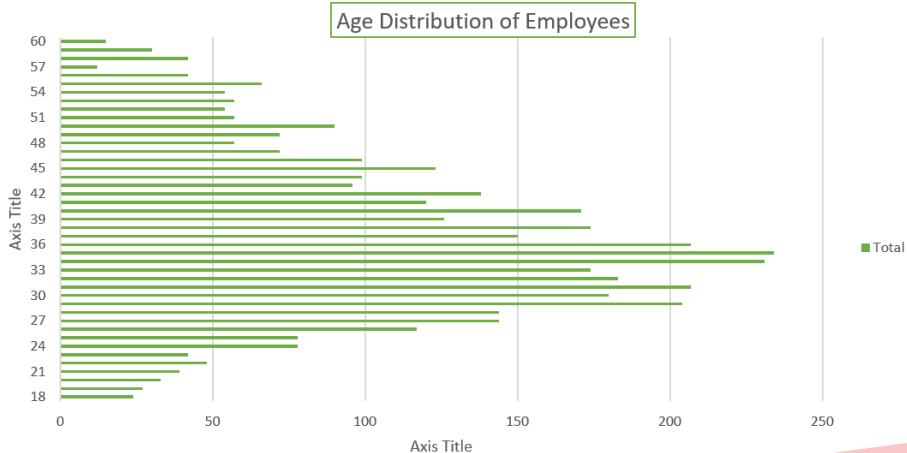
5. Identify and clean any missing or inconsistent data in the "Department" column.

No missing or inconsistent data

6. In Power BI, establish a relationship between the "EmployeeID" in the employee data and the "EmployeeID" in the time tracking data.
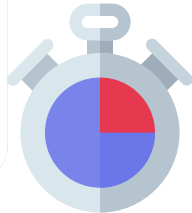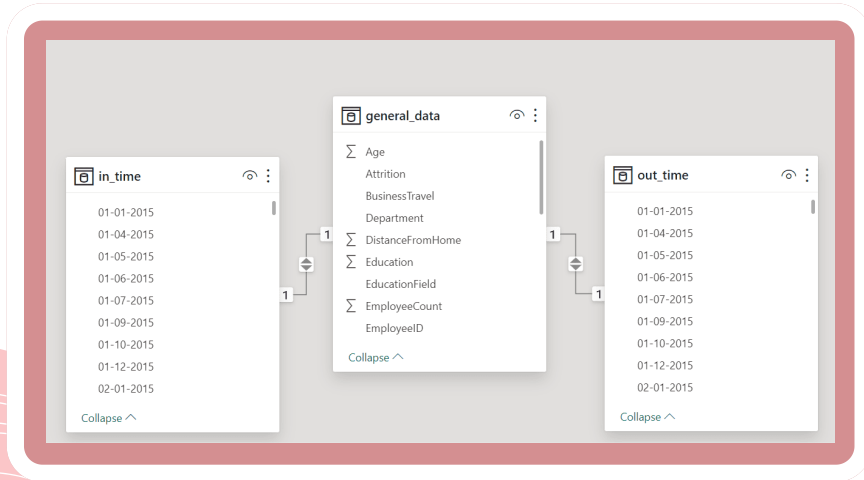
7. Using DAX, create a calculated column that calculates the average years an employee has spent with their current manager.

Formula bar: `1 AvgYearsWithManager = AVERAGE(general_data[YearsWithCurrManager])`

onLevel	TotalWorkingYears	TrainingTimesLastYear	YearsAtCompany	YearsSinceLastPromotion	YearsWithCurrManager	AvgYearsWithManager
0	1		6	1	0	4.12312925170068
1	1		3	1	0	4.12312925170068
1	1		2	1	0	4.12312925170068
0	1		3	1	0	4.12312925170068
2	1		2	1	0	4.12312925170068
0	1		3	1	0	4.12312925170068
0	1		4	1	0	4.12312925170068
2	1		3	1	0	4.12312925170068
3	1		2	1	0	4.12312925170068
2	1		2	1	0	4.12312925170068
2	1		2	1	0	4.12312925170068
1	1		2	1	0	4.12312925170068
1	1		5	1	0	4.12312925170068
1	1		3	1	0	4.12312925170068
1	1		3	1	0	4.12312925170068
0	1		2	1	0	4.12312925170068
1	1		3	1	0	4.12312925170068
1	1		5	1	0	4.12312925170068
0	1		1	1	0	4.12312925170068
0	1		3	1	0	4.12312925170068

Data

Search

- general_data
  - Σ Age
  - Attrition
  - AvgYearsWithManager
  - BusinessTravel
  - Department
  - Σ DistanceFromHome
  - Σ Education
  - EducationField
  - Σ EmployeeCount
  - EmployeeID
  - Gender
  - Σ JobLevel
  - JobRole
 - JobSatisfaction
  - MaritalStatus
  - Σ MonthlyIncome

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

Row Labels ▾	Count of EmployeeID
⊟ **Divorced**	
Human Resources	21
Research & Development	621
Sales	339
⊟ **Married**	
Human Resources	96
Research & Development	1350
Sales	573
⊟ **Single**	
Human Resources	72
Research & Development	912
Sales	426
Grand Total	**4410**

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

	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Education	EducationField	EmployeeCount	Gender	JobLevel	JobRole	MaritalStatus	MonthlyIncome	JobSatisfaction	NumCompaniesWorked	Over18	PercentSalaryHike	
2	2	Life Sciences	1	Female	1	Healthcare Representative	Married	131160	4	1	Y	11	
3	1	Life Sciences	1	Female	1	Research Scientist	Single	41890	2	0	Y	23	
4	4	Other	1	Male	4	Sales Executive	Married	193280	2	1	Y	15	
5	5	Life Sciences	1	Male	3	Human Resources	Married	83210	4	3	Y	11	
6	1	Medical	1	Male	1	Sales Executive	Single	23420	1	4	Y	12	
7	3	Life Sciences	1	Female	4	Research Director	Married	40710	2	3	Y	13	
8	2	Medical	1	Male	2	Sales Executive	Single	58130	3	2	Y	20	
9	3	Life Sciences	1	Male	2	Sales Executive	Married	31430	2	2	Y	22	
10	3	Life Sciences	1	Male	3	Laboratory Technician	Married	20440	4	0	Y	21	
11	4	Medical	1	Female	4	Laboratory Technician	Divorced	134640	1	1	Y	13	
12	2	Medical	1	Male	1	Laboratory Technician	Married	79910	4	0	Y	13	
13	1	Life Sciences	1	Male	1	Laboratory Technician	Married	33770	4	1	Y	12	
14	4	Life Sciences	1	Female	1	Sales Executive	Single	55380	1	0	Y	17	
15	1	Medical	1	Male	1	Research Scientist	Married	57620	2	1	Y	11	
16	3	Life Sciences	1	Male	1	Manufacturing Director	Married	25920	4	1	Y	14	
17	3	Life Sciences	1	Male	2	Healthcare Representative	Married	53460	4	4	Y	11	
18	3	Life Sciences	1	Male	1	Laboratory Technician	Single	42130	3	1	Y	12	
19	3	Medical	1	Male	2	Sales Executive	Divorced	41270	4	2	Y	13	
20	4	Life Sciences	1	Male	1	Sales Representative	Divorced	24380	2	7	Y	16	
21	3	Life Sciences	1	Female	1	Manager	Divorced	68700	1	1	Y	11	
22	4	Other	1	Male	2	Laboratory Technician	Divorced	104470	2	1	Y	18	
23	4	Life Sciences	1	Male	1	Research Scientist	Divorced	96670	2	3	Y	23	

... | Age distribution of employees | Marital status over department | **general_data** |

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

Trend of Employee Attrition over Years
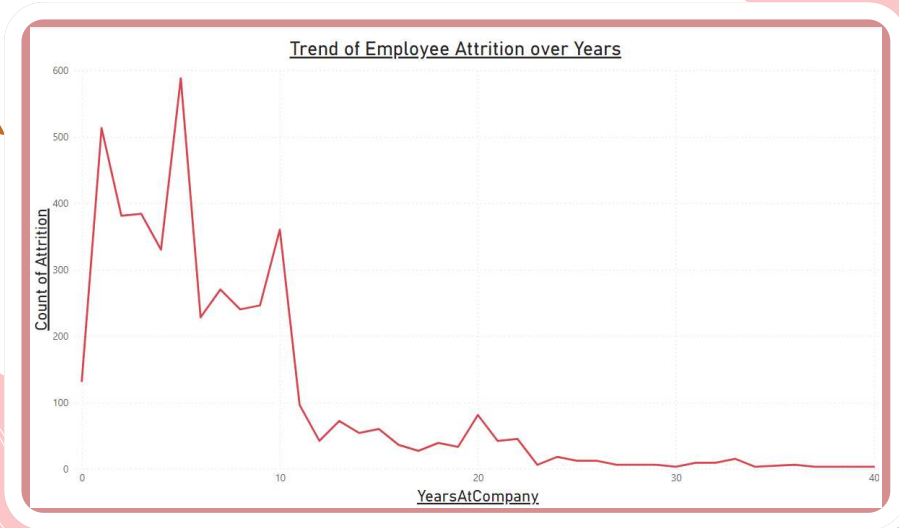
11. Describe how you would create a star schema for this dataset, explaining the benefits of doing so.

To create a star schema out of this dataset we would need 'Fact' and 'Dimensions' tables separated with appropriate columns a nd foreign keys; the general steps should be

1)To mormalizethe table and divide it into more tables i.e. we may create 'Employee Attrition Fact Table', 'Satisfaction Fact Table', 'Job Performance Fact Table' which would be the center of the star schema.

2)Furthermore, we need Dimension tables as well consisting of 'Employee Dimension Table', 'Satisfaction Dimension Table', 'Jo b Performance Dimension Table', 'In-Out Time Dimension Table'.

3)All these dimension tables would contain a primary key ' EmployeeID' which is connected to the foreign key 'Employee ID' of the Fact tables.

4)This ensures that the data is correctly related, and Power BI can perform accurate analysis and reporting.

The benefits of creating a star schema for this dataset would be:

A)Simplicity and Understandability for technical as well as non -technical users.
B)The separation of dimensions and facts allows for efficient querying, especially for aggregations and reporting.
C)The star schema is scalable and can handle large datasets and Star schemas are flexible and can adapt to changing reporting requirements.
D)Data redundancy is minimized in a star schema. Also, security can be managed more efficiently in a star schema since access controls can be implemented at the dimension and fact table levels

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

```
Rolling 3-Month Average =
CALCULATE(
    AVERAGE('general_data'[MonthlyIncome]),
    FILTER(
        ALL('DateTable'[Value]),
        'DateTable'[Value] >= MAX('DateTable'[Value]) - 90 &&
        'DateTable'[Value] <= MAX('DateTable'[Value])
    )
)
```
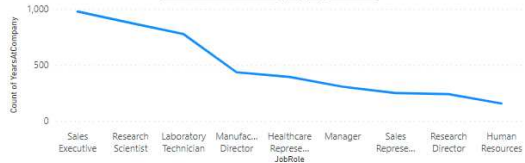
EmployeeID	Rolling 3-Month Average
1	131160
2	41890
3	193280
4	83210
5	23420
6	40710
7	58130
8	31430
9	20440
10	134640
11	79910
12	33770
13	55380

13. Create a hierarchy in Power BI that allows users to drill down from Department to Job Role to further narrow their analysis.
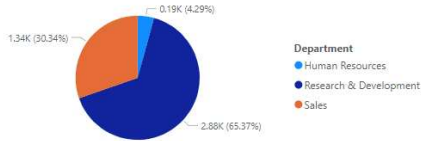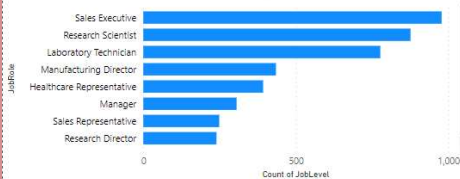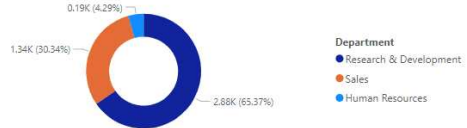
14. How can you set up parameterized queries in Power BI to allow users to filter data based 2 of 2 on the Distance from Home column?

1)Parameterized queries can be set up using Power Query, to allow users to filter data based on the Distance from Home column using parameters.

2)Load your data, create a new parameter in the Power Query Editor for Distance Parameter, and set its datatype to decimal/whole number.

3)In the Power Query Editor, locate the query that loads your data and add a filter step to filter the data based on the parameter. = Table.SelectRows(YourPreviousStep, each [DistanceFromHome] <= DistanceParameter)

4)Click "Close & Apply" in the Home tab to apply the changes.

5)In your Power BI report, create a slicer visual or any other method for users to input the parameter value and manually update the filter condition in your visual to reference the parameter.

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.

Step 1: Set up your data.
 - In one column, list the current monthly salaries of your employees.
 - In another column, enter the current percent salary hike for each employee.
 - Create a third column to calculate the current monthly salary hike amount using the formula =Current Salary * Current Percent Hike/100.
 - Finally, add a fourth column to calculate the new monthly income with the increased salary hike by adding the Current Salary and Current Monthly Salary Hike columns.

Step 2: Apply the 10% increase.
 - In a separate cell, type the value 10 or the desired percentage increase.
 - In the Current Percent Hike column for each employee, use the formula =Original Percent Hike + Increase Percent (e.g., =A2 + B1, where A2 is the original percent hike and B1 is the value in the increase cell).

Step 3: Analyze the impact.
 - Observe the changes in the Current Monthly Salary Hike and New Monthly Income columns. These will show the increase in amount and overall income for each employee due to the 10% hike.

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

1) 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.

2) As per current status of the data, there is a need to reorder EmployeeID column, changing data type of 'TotalWorkingHours' column, flling NA values and blank values from general_data, employee_survey_data, manager_survey_data.
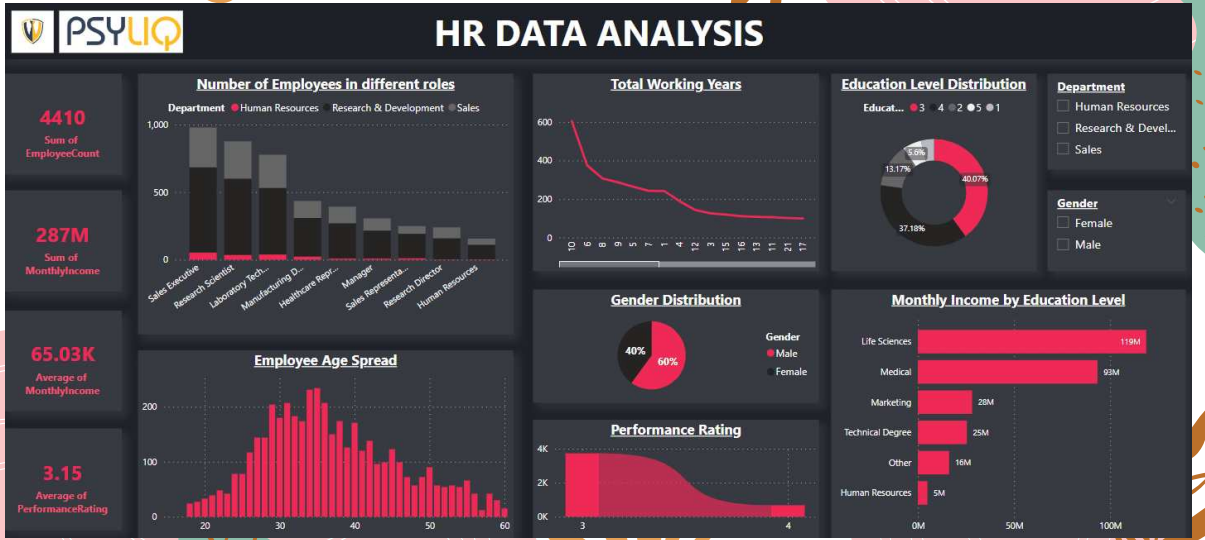
3) There is 'EmployeeID' label missing in the 'in-time' and 'out-time' data.

4) 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.

5) The dataset would then conform to the predefined schema following the resolution of all identified inconsistencies and the implementation of necessary data quality measures.