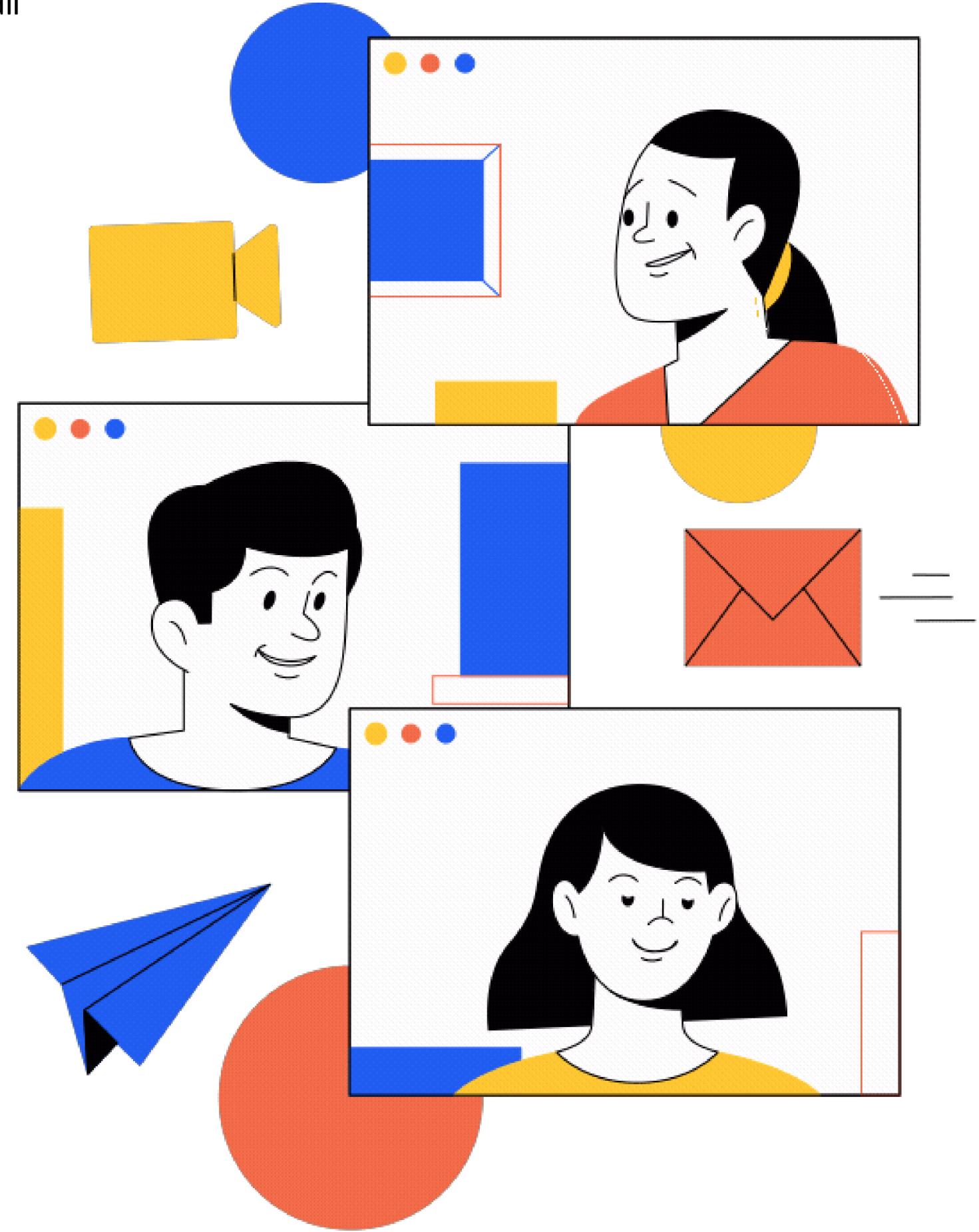


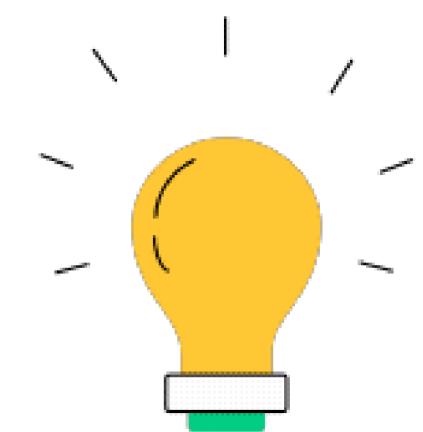
# HR DATA ANALYSIS

Analyzing competitors and their business structures

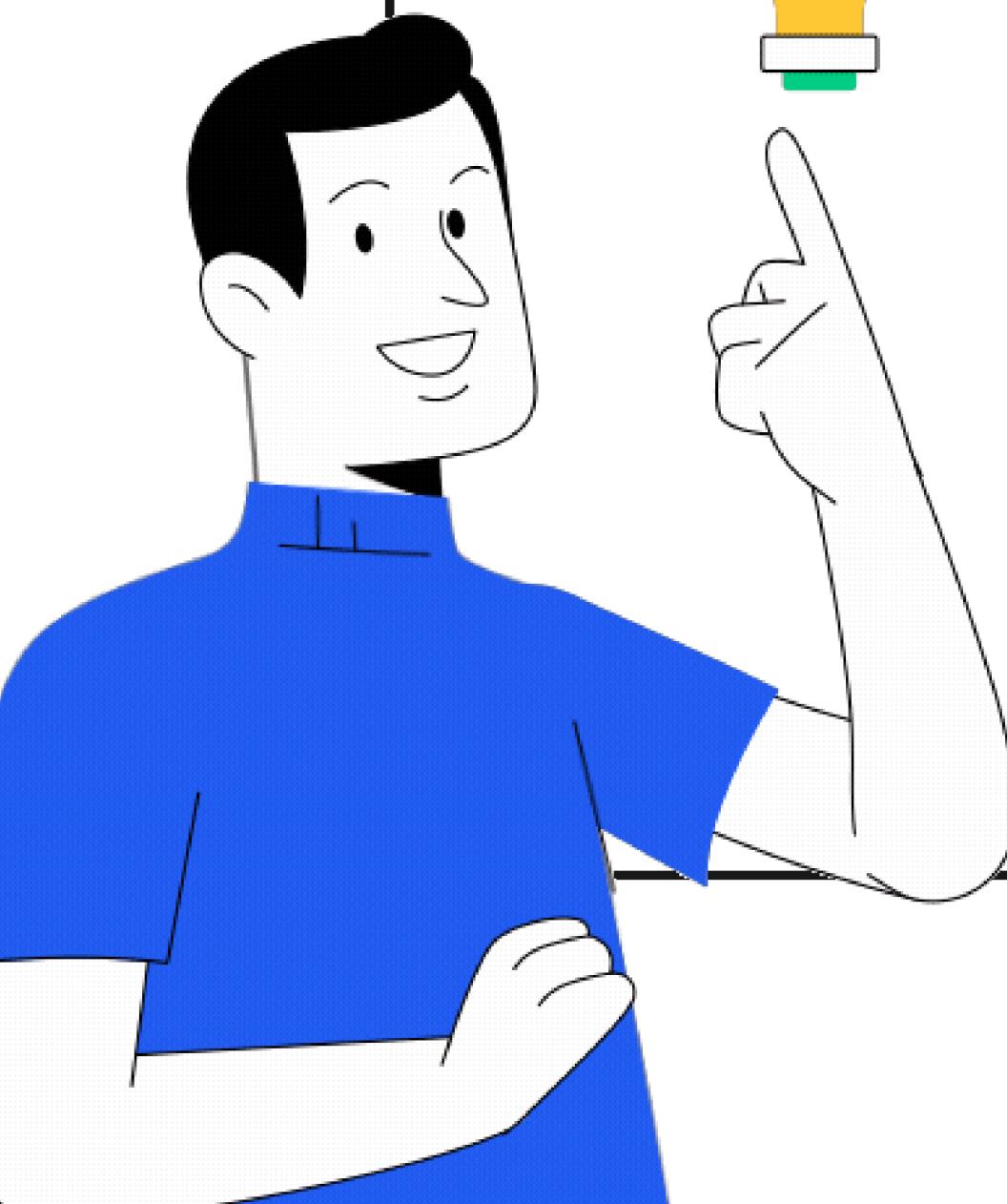
Presented by Rajashree Gavali



# OBJECTIVE



This HR data analysis project aims to leverage Excel and Power BI to gain actionable insights from our employee data. By making trends and patterns, we'll empower decision-makers to optimize talent management strategies, enhance employee engagement, and drive organizational success. Through analysis and visualization, we'll foster a culture of data-driven HR excellence, ensuring our company remains competitive and agile in today's dynamic business environment



# PROJECT TASKS

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

The screenshot shows the Excel ribbon with the 'View' tab selected. A context menu is open over cell A1, with the 'Number Filters' option checked. The sub-menu allows filtering by values greater than or equal to 30. The main table has columns: Age, Attrition, BusinessTravel, Department, DistanceFromHome, Education, EducationField, EmployeeCount, EmployeeID, Gender, JobLevel, and JobRole.

The screenshot shows the Excel ribbon with the 'Data' tab selected. The 'Sort & Filter' icon is highlighted. The main table has columns: Age, Attrition, BusinessTravel, Department, DistanceFromHome, Education, EducationField, EmployeeCount, EmployeeID, Gender, JobLevel, and JobRole. The status bar at the bottom shows 'Ready' and 'Filter Mode'.

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

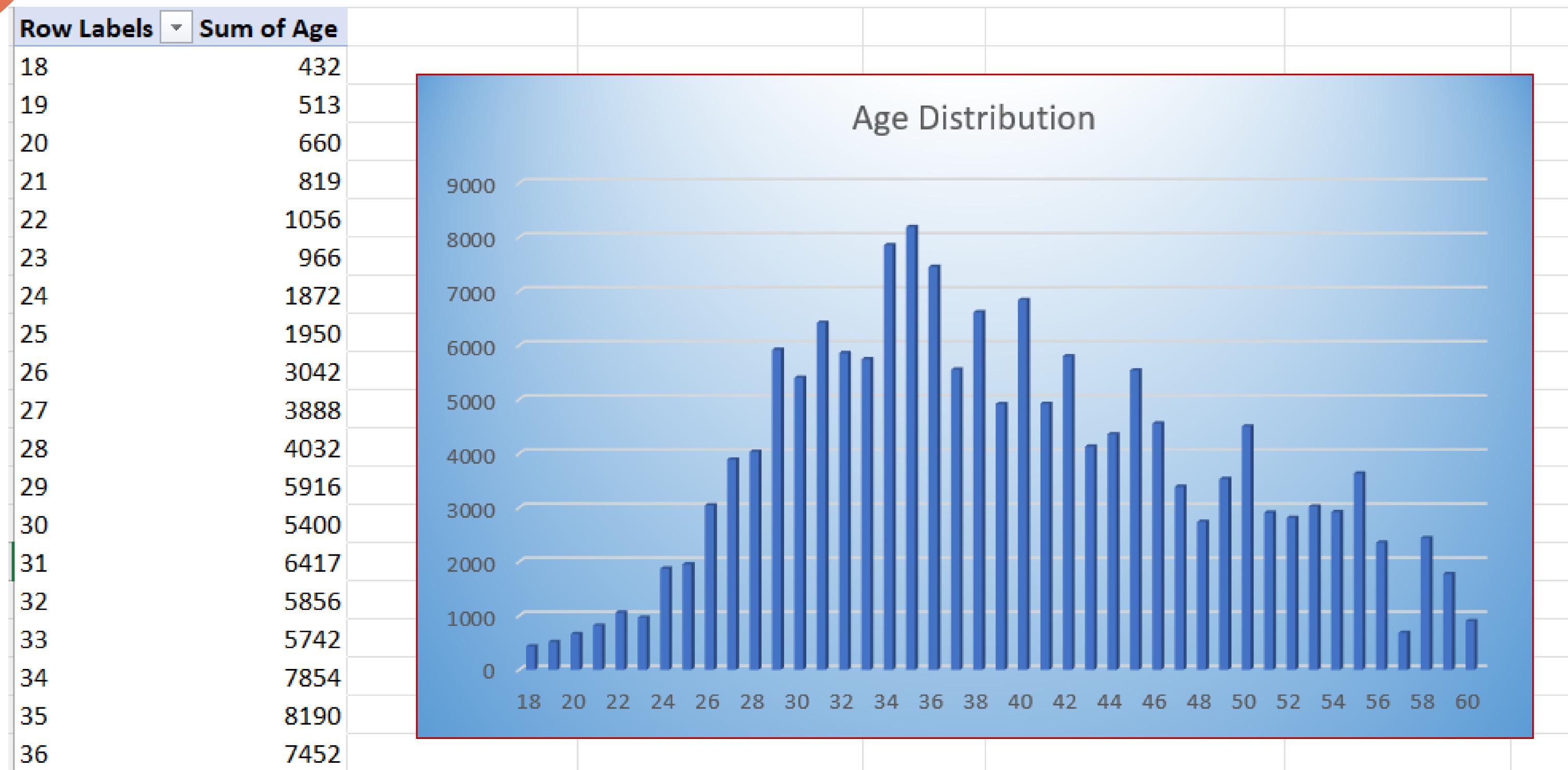
Row Labels	Average of MonthlyIncome
Healthcare Representative	60983.74046
Human Resources	58528.07692
Laboratory Technician	66314.05405
Manager	63395.88235
Manufacturing Director	69183.72414
Research Director	65473.125
Research Scientist	64975.68493
Sales Executive	65186.68712
Sales Representative	65370.96386



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

EducationField	EmployeeCount	EmployeeID	Gender	JobLevel	JobRole	MaritalStatus	MonthlyIncome	NumC
2 Life Sciences	1	1	Female	1	Healthcare Representative	Married	131160	1
1 Life Sciences	1	2	Female	1	Research Scientist	Single	41890	0
4 Other	1	3	Male	4	Sales Executive	Married	193280	1
5 Life Sciences	1	4	Male	3	Human Resources	Married	83210	3
1 Medical	1	5	Male	1	Sales Executive	Single	23420	4
3 Life Sciences	1	6	Female	4	Research Director	Married	40710	3
2 Medical	1	7	Male	2	Sales Executive	Single	58130	2
3 Life Sciences	1	8	Male	2	Sales Executive	Married	31430	2
3 Life Sciences	1	9	Male	3	Laboratory Technician	Married	20440	0
4 Medical	1	10	Female	4	Laboratory Technician	Divorced	134640	1
2 Medical	1	11	Male	2	Laboratory Technician	Married	79910	0
1 Life Sciences	1	12	Male	1	Laboratory Technician	Married	33770	0
4 Life Sciences	1	13	Female	1	Sales Executive	Single	55380	0
1 Medical	1	14	Male	1	Research Scientist	Married	57620	1
3 Life Sciences	1	15	Male	1	Manufacturing Director	Married	25920	1
3 Life Sciences	1	16	Male	2	Healthcare Representative	Married	53460	4
2 Life Sciences	1	17	Male	1	Laboratory Technician	Single	42130	1
3 Medical	1	18	Male	2	Sales Executive	Divorced	41270	2
4 Life Sciences	1	19	Male	1	Sales Representative	Divorced	24380	7
3 Life Sciences	1	20	Female	1	Manager	Divorced	68700	1
4 Other	1	21	Male	2	Laboratory Technician	Divorced	104470	1
4 Life Sciences	1	22	Male	1	Research Scientist	Divorced	96670	3
4 Life Sciences	1	23	Female	2	Research Scientist	Married	21480	3
4 Life Sciences	1	24	Male	1	Manufacturing Director	Married	89260	1
4 Medical	1	25	Male	1	Laboratory Technician	Single	65130	1
4 Other	1	26	Female	1	Research Scientist	Married	67990	3

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



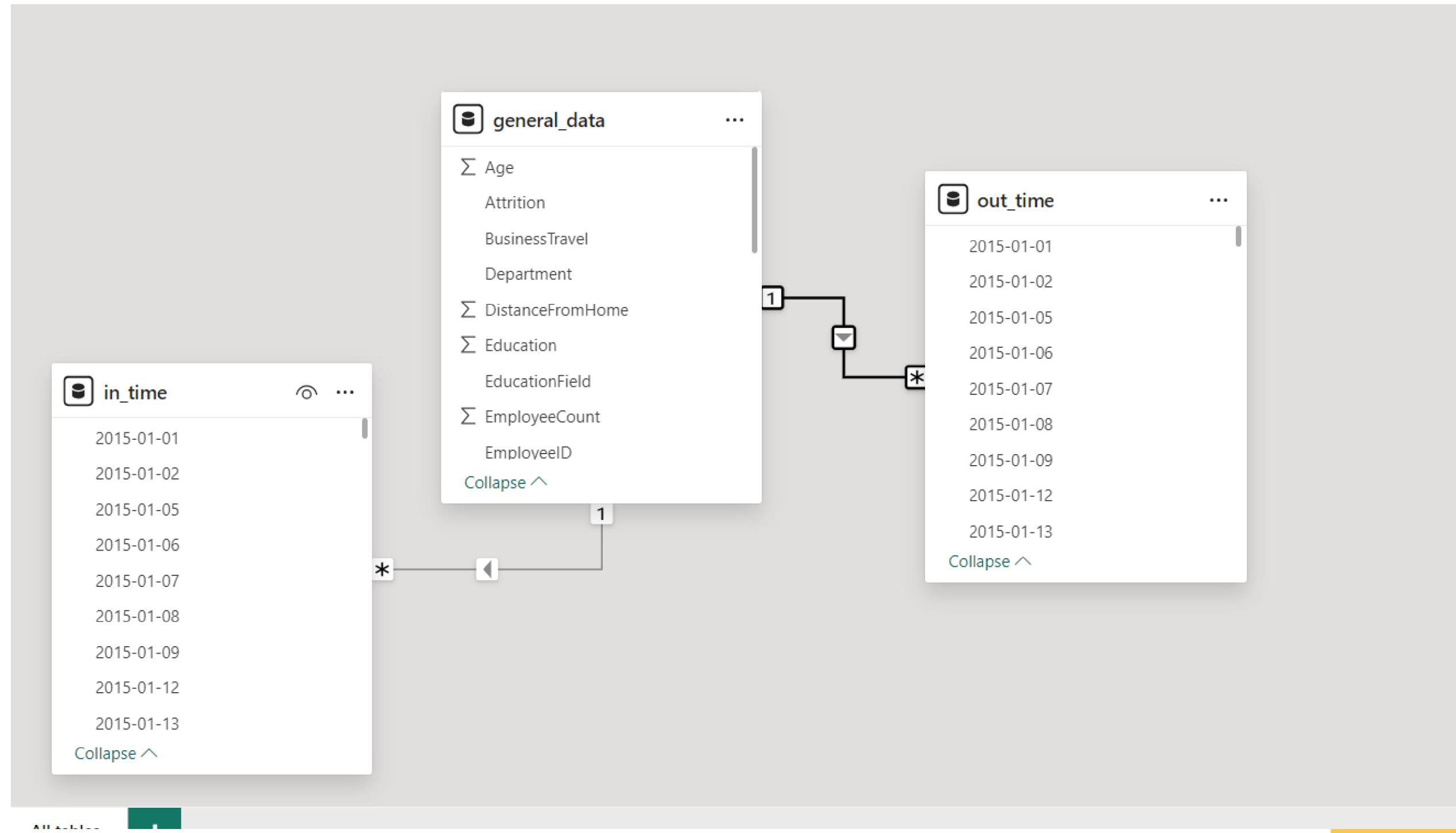
## 5. Identify and clean any missing or inconsistent data in the "Department" column.

The screenshot shows a Microsoft Excel spreadsheet with the 'Data' tab selected in the ribbon. A context menu is open over the 'Department' column header in row 1. The menu includes options for sorting (Sort A to Z, Sort Z to A), filtering by color, and applying text filters. The 'Text Filters' section shows a list of department names: Human Resources, Research & Development, and Sales. The main data area contains 29 rows of employee information, including columns for Age, Attrition, BusinessTravel, Department, DistanceFromHome, Education, EducationField, EmployeeCount, EmployeeID, and Gender.

	A	B	C	D	E	F	G	H	I	J
1	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeID	Gender
2	51	No			6	2	Life Sciences	1	1	Female
3	31	Yes			10	1	Life Sciences	1	2	Female
4	32	No			17	4	Other	1	3	Male
5	38	No			2	5	Life Sciences	1	4	Male
6	32	No			10	1	Medical	1	5	Male
7	46	No			8	3	Life Sciences	1	6	Female
8	28	Yes			11	2	Medical	1	7	Male
9	29	No			18	3	Life Sciences	1	8	Male
10	31	No			1	3	Life Sciences	1	9	Male
11	25	No			7	4	Medical	1	10	Female
12	45	No			17	2	Medical	1	11	Male
13	36	No			28	1	Life Sciences	1	12	Male
14	55	No			14	4	Life Sciences	1	13	Female
15	47	Yes			1	1	Medical	1	14	Male
16	28	No			1	3	Life Sciences	1	15	Male
17	37	No			1	3	Life Sciences	1	16	Male
18	21	No			3	2	Life Sciences	1	17	Male
19	37	No			1	3	Medical	1	18	Male
20	35	No			7	4	Life Sciences	1	19	Male
21	38	No			8	3	Life Sciences	1	20	Female
22	26	No			1	4	Other	1	21	Male
23	50	No			8	4	Life Sciences	1	22	Male
24	53	No			11	4	Life Sciences	1	23	Female
25	42	No			4	4	Life Sciences	1	24	Male
26	29	No	Travel_Frequently	Research & Development	16	4	Medical	1	25	Male
27	55	No	Travel_Rarely	Research & Development	1	4	Other	1	26	Female
28	36	No	Travel_Frequently	Research & Development	2	2	Life Sciences	1	27	Female

No missing or inconsistent data in  
Department column

**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.

The screenshot shows the Power BI Data Editor interface. On the left, there's a vertical toolbar with icons for Home, Report, DAX, and Refresh. The main area is divided into three tabs: Structure, Formatting, and Properties. The Structure tab shows a single column named "Average\_Curr\_Man..." with the formula `AVERAGE(general_data[YearsWithCurrManager])`. The Properties tab shows the column is part of the "general\_data" table and is categorized as "Uncategorized". The Formatting tab shows the column is set to "General" format with thousands separator and two decimal places. Below the tabs, there's a large number "4.12" displayed.

Name: Average\_Curr\_Man...  
Home table: general\_data  
Format: General  
Data category: Uncategorized

Structure      Formatting      Properties

1 Average\_Curr\_Manager = AVERAGE(general\_data[YearsWithCurrManager])

Average with Current Manager

4.12

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

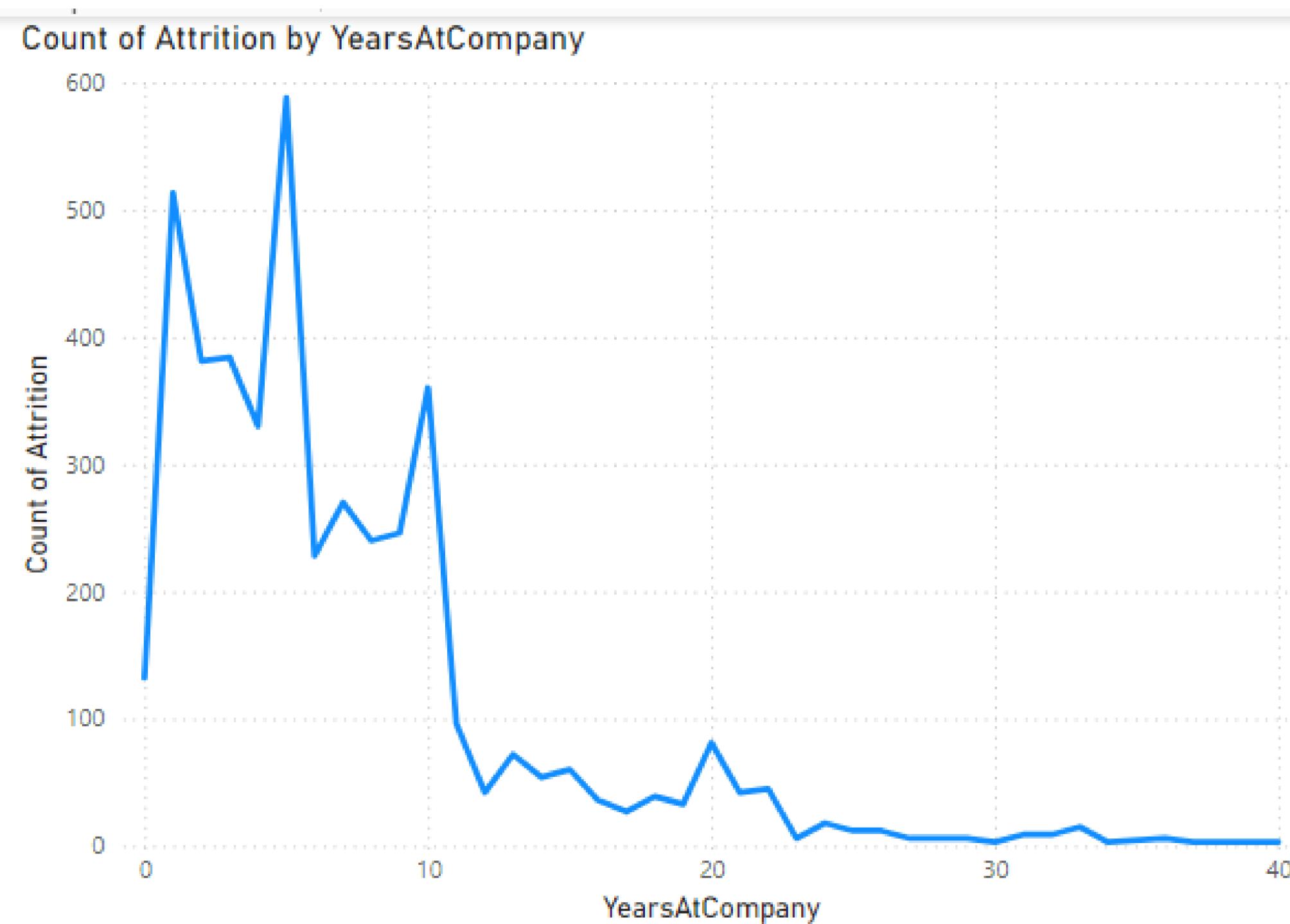
Row Labels	Sum of EmployeeCount
Divorced	981
Human Resources	21
Research & Development	621
Sales	339
Married	2019
Human Resources	96
Research & Development	1350
Sales	573
Single	1410
Human Resources	72
Research & Development	912
Sales	426
<b>Grand Total</b>	<b>4410</b>

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

MonthlyIncome
131160
41890
193280
83210
23420
40710
58130
31430
20440
134640
79910
33770
55380
57620
25920
53460
42130
41270
24380
68700
104470

JobSatisfaction
4
2
2
4
1
2
3
2
4
1
4
4
1
2
4
4
3
4
2
1
2
2
3
4
3
4
4
1
4
3
4

**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.**

**To create a star schema for any data, you can follow these steps:**

- 1. Identify the fact table:** The fact table contains the main metrics or measurements that you want to analyse. This table typically contains numerical values and is surrounded by dimension tables.
- 2. 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.
- 3. 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.
- 4. Normalize the dimension tables:** Dimension tables should be normalized to reduce redundancy and improve data consistency.

## Benefits of using a star schema include:

- 1. Simplified queries:** Star schemas are optimised for querying and reporting, making it easier to retrieve specific information from the database.
- 2. Improved query performance:** Star schemas are designed for faster query performance as they involve fewer joins compared to other schema designs.
- 3. 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.
- 4. Scalability:** Star schemas are scalable and can accommodate large volumes of data while maintaining query performance.
- 5. Flexibility:** Star schemas allow for adding new dimensions or attributes easily without affecting existing data structures. Overall, star schemas are widely used in data warehousing and analytics environments due to their simplicity, efficiency, and effectiveness in organizing and analyzing data.

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

```
1 Rolling 3-Month Avg Monthly Income =  
2 VAR CurrentIndex = 'general_data'[Index1]  
3 RETURN  
4     AVERAGEX(  
5         FILTER(  
6             'general_data',  
7             'general_data'[Index] <= CurrentIndex &&  
8             'general_data'[Index] >= CurrentIndex - 2  
9         ),  
10        'general_data'[MonthlyIncome]  
11    )
```

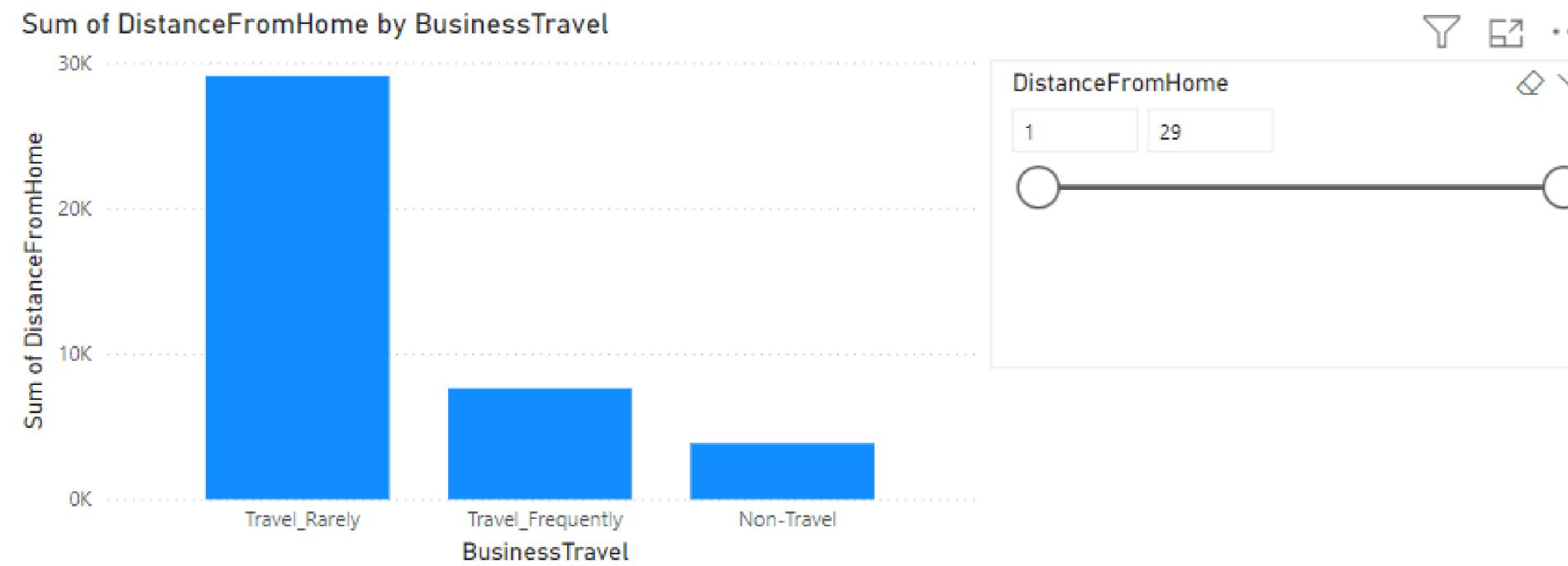
# 13. Create a hierarchy in Power BI that allows users to drill down from Department to Job Role to further narrow their analysis.

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
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?

To set Parameterized in PowerBI:

1. Open Query Editor: Access Power BI's Query Editor.
2. Create Parameters: Make two parameters for Minimum and Maximum Distance values.
3. Filter Data: Set up a filter for the Distance from Home column, using the parameters for the range.
4. Apply Changes: Confirm the filter settings and apply them.
5. Use Parameters: Now adjust the parameters to dynamically filter the data based on Distance from Home

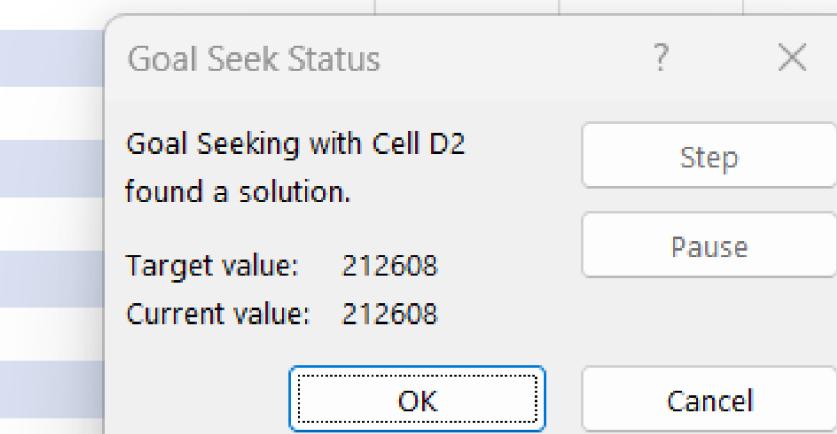


15. In Excel, calculate the total Monthly Income for each Department, considering only the employees with a Job Level greater than or equal to 3.

Sum of MonthlyIncome	Column Labels	3	4	5	Grand Total
Row Labels					
Human Resources		1648500	754800	855840	3259140
Research & Development		28117740	15277290	10107870	53502900
Sales		11792400	8753070	2428860	22974330
<b>Grand Total</b>		<b>41558640</b>	<b>24785160</b>	<b>13392570</b>	<b>79736370</b>

## 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
	MonthlyIncome	10% Added	10% MonthlyIncome			
	193280	19328	212608			
	83210					
	40710					
	20440					
	134640					
	68540					
	54050					
	29560					
	69620					
	63840					
	99070					
	193920					
	24390					
	73140					
	39020					
	26620					
	28560					
	41970					
	47880					
	38860					
	50330					
	25870					
	81030					
	39780					
	25440					
	22260					



## **17. 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. Data Cleaning: Perform data cleaning operations to correct any inconsistencies, such as removing duplicates, filling in missing values, or transforming data to match the expected format.
2. Data Transformation: Use Power Query in Power BI to transform the data to align with the predefined schema.
3. Data Validation: Implement data validation checks to ensure that new data added to the dataset adheres to the predefined schema.
4. Error Handling: Set up error handling mechanisms to flag or handle data inconsistencies automatically.

# THANK YOU!

