

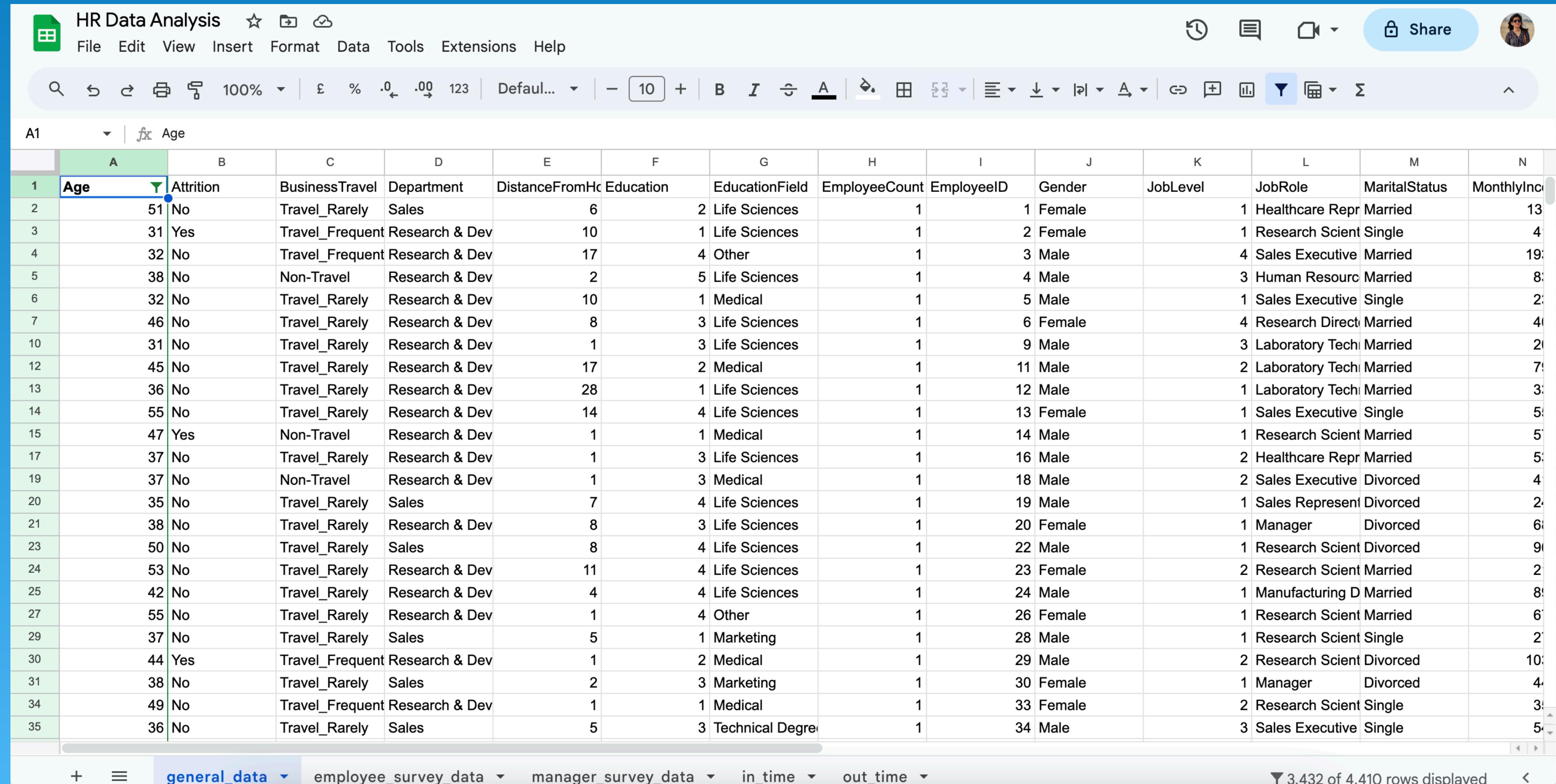
Task-1

HR Data Analysis

Samritha A.R



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



The screenshot shows an Excel spreadsheet titled "HR Data Analysis". The "Age" column is selected, indicated by a blue highlight. The data consists of 3432 rows of employee information, with columns including Age, Attrition, BusinessTravel, Department, DistanceFromHome, Education, EducationField, EmployeeCount, EmployeeID, Gender, JobLevel, JobRole, MaritalStatus, and MonthlyIncome. The "Age" column ranges from 19 to 50+.

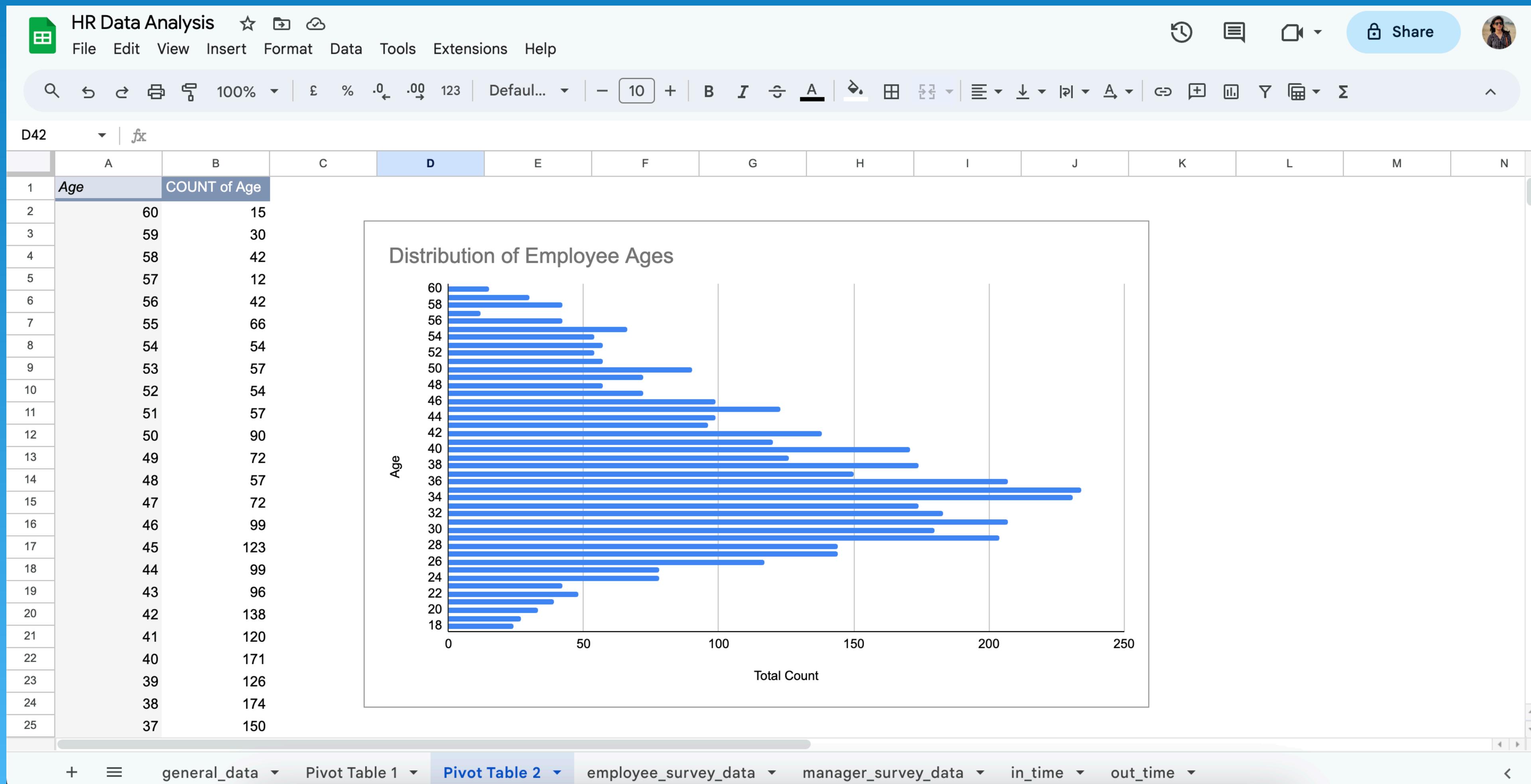
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeID	Gender	JobLevel	JobRole	MaritalStatus	MonthlyIncome
2	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	1	Female	1	Healthcare Repr	Married	13
3	31	Yes	Travel_Frequent	Research & Dev	10	1	Life Sciences	1	2	Female	1	Research Scient	Single	4
4	32	No	Travel_Frequent	Research & Dev	17	4	Other	1	3	Male	4	Sales Executive	Married	19
5	38	No	Non-Travel	Research & Dev	2	5	Life Sciences	1	4	Male	3	Human Resourc	Married	8
6	32	No	Travel_Rarely	Research & Dev	10	1	Medical	1	5	Male	1	Sales Executive	Single	2
7	46	No	Travel_Rarely	Research & Dev	8	3	Life Sciences	1	6	Female	4	Research Direct	Married	4
10	31	No	Travel_Rarely	Research & Dev	1	3	Life Sciences	1	9	Male	3	Laboratory Tech	Married	2
12	45	No	Travel_Rarely	Research & Dev	17	2	Medical	1	11	Male	2	Laboratory Tech	Married	7
13	36	No	Travel_Rarely	Research & Dev	28	1	Life Sciences	1	12	Male	1	Laboratory Tech	Married	3
14	55	No	Travel_Rarely	Research & Dev	14	4	Life Sciences	1	13	Female	1	Sales Executive	Single	5
15	47	Yes	Non-Travel	Research & Dev	1	1	Medical	1	14	Male	1	Research Scient	Married	5
17	37	No	Travel_Rarely	Research & Dev	1	3	Life Sciences	1	16	Male	2	Healthcare Repr	Married	5
19	37	No	Non-Travel	Research & Dev	1	3	Medical	1	18	Male	2	Sales Executive	Divorced	4
20	35	No	Travel_Rarely	Sales	7	4	Life Sciences	1	19	Male	1	Sales Represent	Divorced	2
21	38	No	Travel_Rarely	Research & Dev	8	3	Life Sciences	1	20	Female	1	Manager	Divorced	6
23	50	No	Travel_Rarely	Sales	8	4	Life Sciences	1	22	Male	1	Research Scient	Divorced	9
24	53	No	Travel_Rarely	Research & Dev	11	4	Life Sciences	1	23	Female	2	Research Scient	Married	2
25	42	No	Travel_Rarely	Research & Dev	4	4	Life Sciences	1	24	Male	1	Manufacturing D	Married	8
27	55	No	Travel_Rarely	Research & Dev	1	4	Other	1	26	Female	1	Research Scient	Married	6
29	37	No	Travel_Rarely	Sales	5	1	Marketing	1	28	Male	1	Research Scient	Single	2
30	44	Yes	Travel_Frequent	Research & Dev	1	2	Medical	1	29	Male	2	Research Scient	Divorced	10
31	38	No	Travel_Rarely	Sales	2	3	Marketing	1	30	Female	1	Manager	Divorced	4
34	49	No	Travel_Frequent	Research & Dev	1	1	Medical	1	33	Female	2	Research Scient	Single	3
35	36	No	Travel_Rarely	Sales	5	3	Technical Degree	1	34	Male	3	Sales Executive	Single	5

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

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

HR Data Analysis																					
File		Edit		View		Insert		Format		Data		Tools		Extensions		Help					
Search		Navigation		Zoom		Currency		Number		Text		Font		Style		Table		Filter		Sum	
N1:N4411	fx	MonthlyIncome																			
1	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeID	Gender	JobLevel	JobRole	MaritalStatus	MonthlyIncome	NumCompaniesOver18	PercentSalaryHike	PercentSalaryChange	StandardHours	TotalWorkingYears	YearsAtCompany				
2	Sales	6	2	Life Sciences	1	1	Female	1	Healthcare Representative	Married	131160	1	Y	0.0	40	10	10				
3	Research & Dev	10	1	Life Sciences	1	2	Female	1	Research Scientist	Single	41890	0	Y	0.0	40	10	10				
4	Research & Dev	17	4	Other	1	3	Male	4	Sales Executive	Married	193280	1	Y	0.0	40	10	10				
5	Research & Dev	2	5	Life Sciences	1	4	Male	3	Human Resources	Married	83210	3	Y	0.0	40	10	10				
6	Research & Dev	10	1	Medical	1	5	Male	1	Sales Executive	Single	23420	4	Y	0.0	40	10	10				
7	Research & Dev	8	3	Life Sciences	1	6	Female	4	Research Director	Married	40710	3	Y	0.0	40	10	10				
10	Research & Dev	1	3	Life Sciences	1	9	Male	3	Laboratory Technician	Married	20440	0	Y	0.0	40	10	10				
12	Research & Dev	17	2	Medical	1	11	Male	2	Laboratory Technician	Married	79910	0	Y	0.0	40	10	10				
13	Research & Dev	28	1	Life Sciences	1	12	Male	1	Laboratory Technician	Married	33770	0	Y	0.0	40	10	10				
14	Research & Dev	14	4	Life Sciences	1	13	Female	1	Sales Executive	Single	55380	0	Y	0.0	40	10	10				
15	Research & Dev	1	1	Medical	1	14	Male	1	Research Scientist	Married	57620	1	Y	0.0	40	10	10				
17	Research & Dev	1	3	Life Sciences	1	16	Male	2	Healthcare Representative	Married	53460	4	Y	0.0	40	10	10				
19	Research & Dev	1	3	Medical	1	18	Male	2	Sales Executive	Divorced	41270	2	Y	0.0	40	10	10				
20	Sales	7	4	Life Sciences	1	19	Male	1	Sales Representative	Divorced	24380	7	Y	0.0	40	10	10				
21	Research & Dev	8	3	Life Sciences	1	20	Female	1	Manager	Divorced	68700	1	Y	0.0	40	10	10				
23	Sales	8	4	Life Sciences	1	22	Male	1	Research Scientist	Divorced	96670	3	Y	0.0	40	10	10				
24	Research & Dev	11	4	Life Sciences	1	23	Female	2	Research Scientist	Married	21480	3	Y	0.0	40	10	10				
25	Research & Dev	4	4	Life Sciences	1	24	Male	1	Manufacturing Director	Married	89260	1	Y	0.0	40	10	10				
27	Research & Dev	1	4	Other	1	26	Female	1	Research Scientist	Married	67990	3	Y	0.0	40	10	10				
29	Sales	5	1	Marketing	1	28	Male	1	Research Scientist	Single	27050	1	Y	0.0	40	10	10				
30	Research & Dev	1	2	Medical	1	29	Male	2	Research Scientist	Divorced	103330	3	Y	0.0	40	10	10				
31	Sales	2	3	Marketing	1	30	Female	1	Manager	Divorced	44480	9	Y	0.0	40	10	10				
34	Research & Dev	1	1	Medical	1	33	Female	2	Research Scientist	Single	35910	9	Y	0.0	40	10	10				
35	Sales	5	3	Technical Degree	1	34	Male	3	Sales Executive	Single	54050	4	Y	0.0	40	10	10				

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

HR Data Analysis

File Edit View Insert Format Data Tools Extensions Help

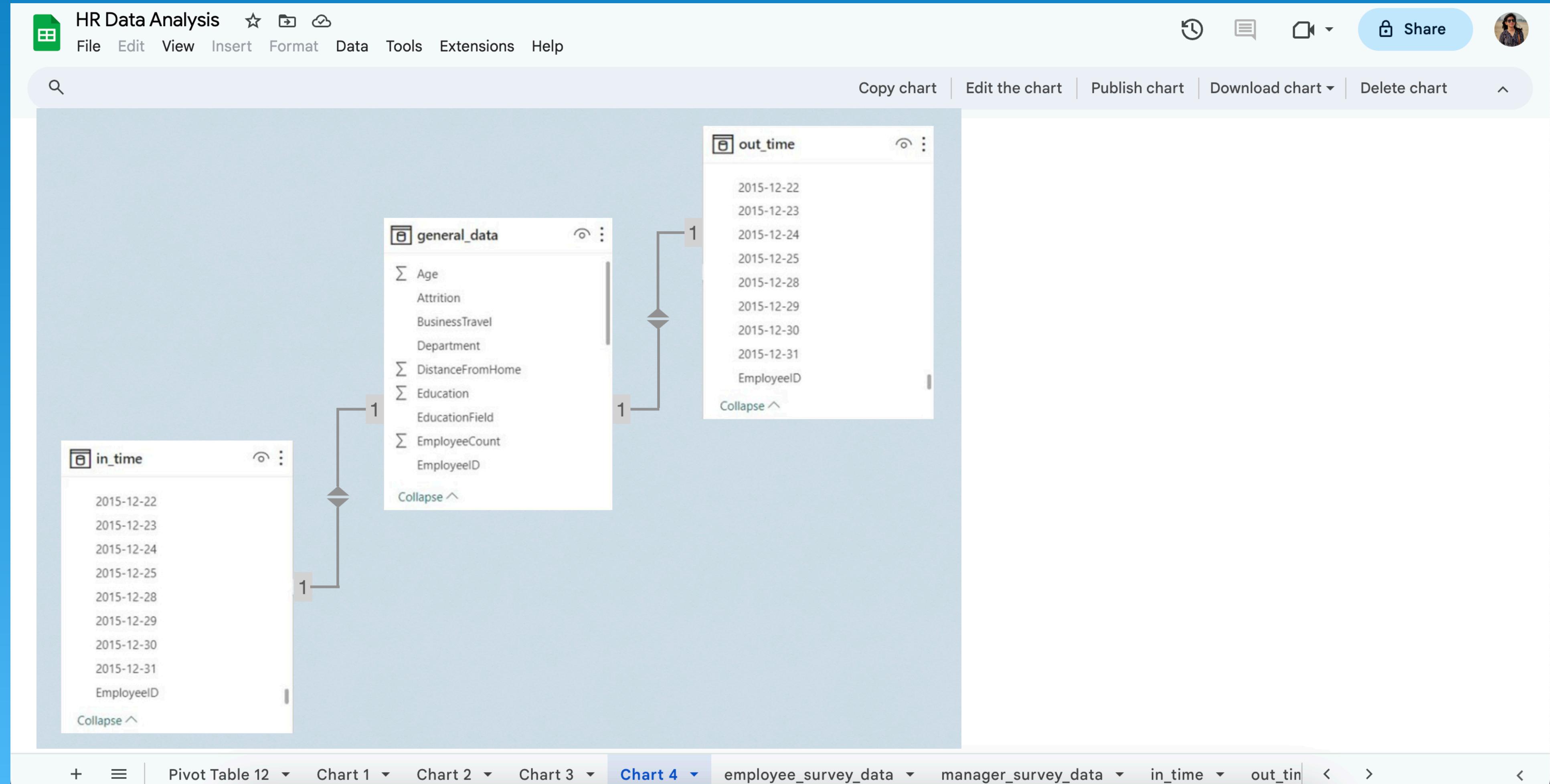
100% 123 Default 10 B I A

D1 Department

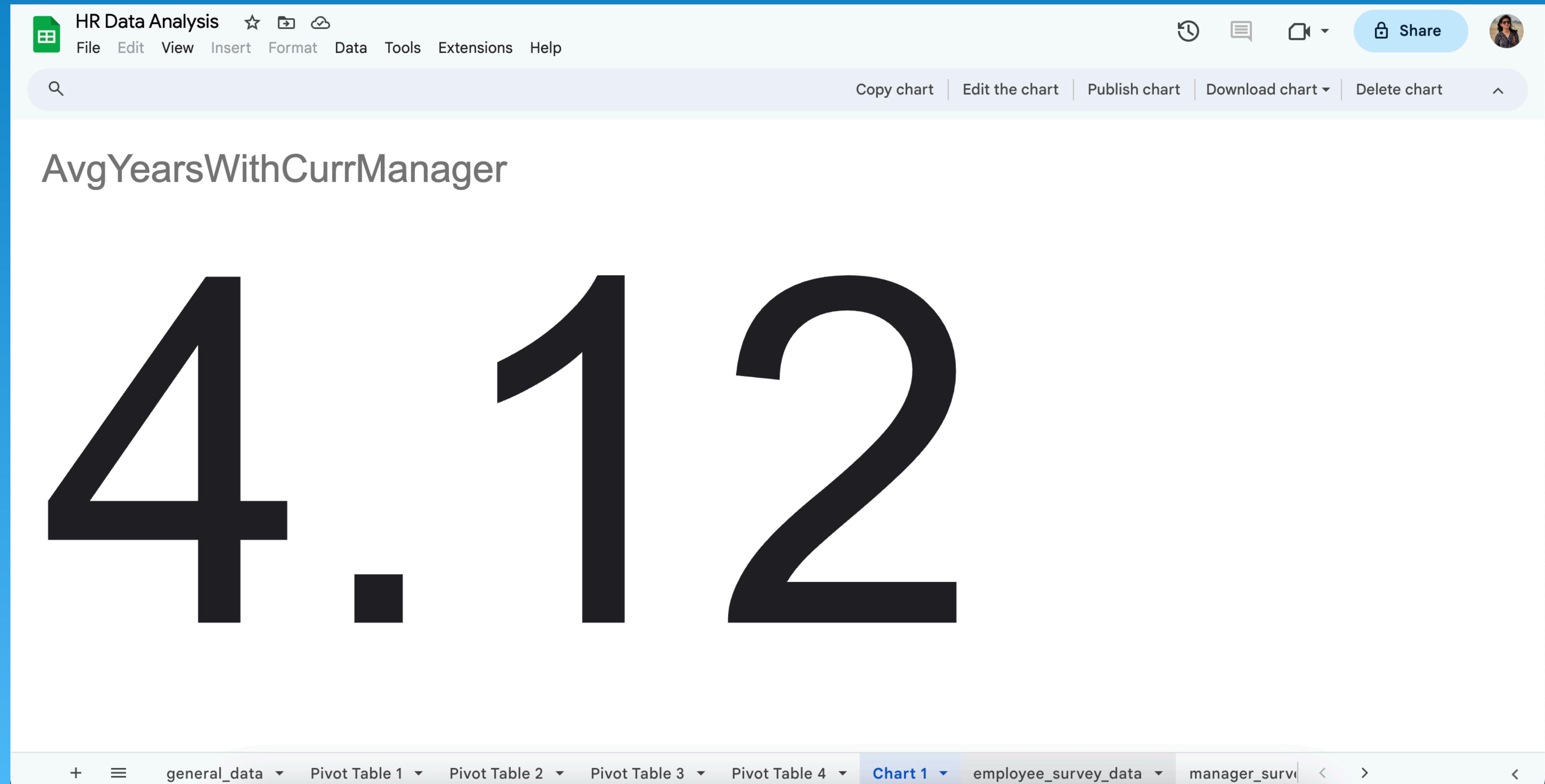
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Age	Attrition	BusinessTravel	Department	DistanceFromHc	Education	EducationField	EmployeeCount	EmployeeID	Gender	JobLevel	JobRole	MaritalStatus	MonthlyIncome
2	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	1	Female	1	Healthcare Rep	Married	13
3	31	Yes	Travel_Frequent	Research & Dev	10	1	Life Sciences	1	2	Female	1	Research Scient	Single	4
4	32	No	Travel_Frequent	Research & Dev	17	4	Other	1	3	Male	4	Sales Executive	Married	19
5	38	No	Non-Travel	Research & Dev	2	5	Life Sciences	1	4	Male	3	Human Resourc	Married	8
6	32	No	Travel_Rarely	Research & Dev	10	1	Medical	1	5	Male	1	Sales Executive	Single	23
7	46	No	Travel_Rarely	Research & Dev	8	3	Life Sciences	1	6	Female	4	Research Directo	Married	40
8	28	Yes	Travel_Rarely	Research & Dev	11	2	Medical	1	7	Male	2	Sales Executive	Single	50
9	29	No	Travel_Rarely	Research & Dev	18	3	Life Sciences	1	8	Male	2	Sales Executive	Married	3
10	31	No	Travel_Rarely	Research & Dev	1	3	Life Sciences	1	9	Male	3	Laboratory Tech	Married	20
11	25	No	Non-Travel	Research & Dev	7	4	Medical	1	10	Female	4	Laboratory Tech	Divorced	13
12	45	No	Travel_Rarely	Research & Dev	17	2	Medical	1	11	Male	2	Laboratory Tech	Married	7
13	36	No	Travel_Rarely	Research & Dev	28	1	Life Sciences	1	12	Male	1	Laboratory Tech	Married	3
14	55	No	Travel_Rarely	Research & Dev	14	4	Life Sciences	1	13	Female	1	Sales Executive	Single	5
15	47	Yes	Non-Travel	Research & Dev	1	1	Medical	1	14	Male	1	Research Scient	Married	5
16	28	No	Travel_Rarely	Research & Dev	1	3	Life Sciences	1	15	Male	1	Manufacturing D	Married	2
17	37	No	Travel_Rarely	Research & Dev	1	3	Life Sciences	1	16	Male	2	Healthcare Rep	Married	5
18	21	No	Travel_Rarely	Research & Dev	3	2	Life Sciences	1	17	Male	1	Laboratory Tech	Single	4
19	37	No	Non-Travel	Research & Dev	1	3	Medical	1	18	Male	2	Sales Executive	Divorced	4
20	35	No	Travel_Rarely	Sales	7	4	Life Sciences	1	19	Male	1	Sales Represent	Divorced	2
21	38	No	Travel_Rarely	Research & Dev	8	3	Life Sciences	1	20	Female	1	Manager	Divorced	6
22	26	No	Travel_Frequent	Research & Dev	1	4	Other	1	21	Male	2	Laboratory Tech	Divorced	10
23	50	No	Travel_Rarely	Sales	8	4	Life Sciences	1	22	Male	1	Research Scient	Divorced	9
24	53	No	Travel_Rarely	Research & Dev	11	4	Life Sciences	1	23	Female	2	Research Scient	Married	2
25	42	No	Travel_Rarely	Research & Dev	4	4	Life Sciences	1	24	Male	1	Manufacturing D	Married	8

No missing 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

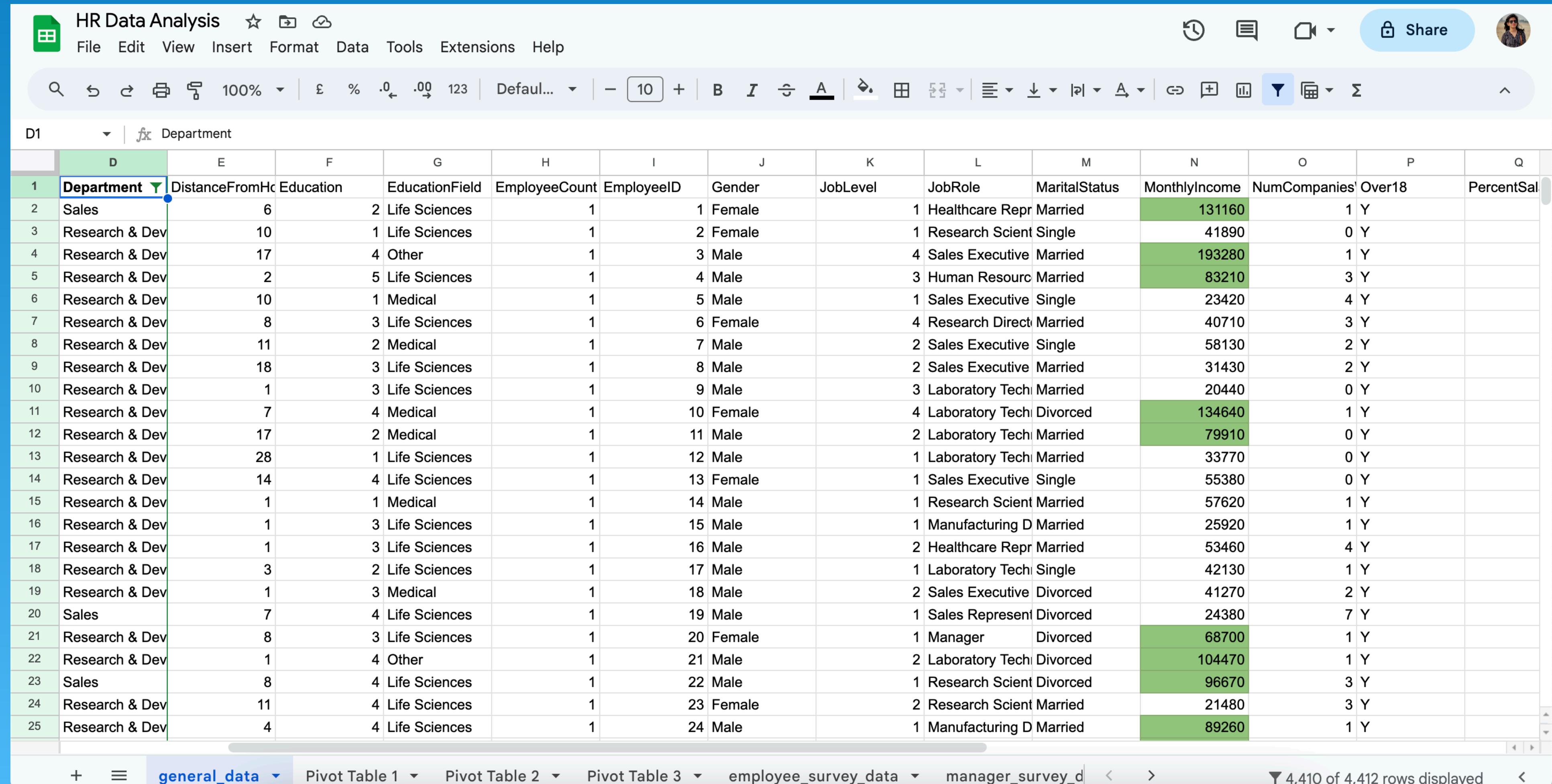


7. Using DAX, create a calculated column that calculates the average years an employee has spent with their current manager



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

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



The screenshot shows a Google Sheets spreadsheet titled "HR Data Analysis". The main table has columns labeled D through Q. The "Department" column (D) contains values like Sales, Research & Dev, and Other. The "MonthlyIncome" column (N) contains numerical values such as 131160, 41890, 193280, etc. The "JobSatisfaction" column (Q) contains values like Y, 0 Y, 1 Y, etc. A green highlight is applied to several rows where both the monthly income and job satisfaction are above average. The status bar at the bottom indicates "4,410 of 4,412 rows displayed".

1	Department	DistanceFromHq	Education	EducationField	EmployeeCount	EmployeeID	Gender	JobLevel	JobRole	MaritalStatus	MonthlyIncome	NumCompaniesOver18	PercentSal
2	Sales	6	2 Life Sciences	Life Sciences	1	1	Female	1	Healthcare Rep	Married	131160	1 Y	
3	Research & Dev	10	1 Life Sciences	Life Sciences	1	2	Female	1	Research Scient	Single	41890	0 Y	
4	Research & Dev	17	4 Other	Other	1	3	Male	4	Sales Executive	Married	193280	1 Y	
5	Research & Dev	2	5 Life Sciences	Life Sciences	1	4	Male	3	Human Resourc	Married	83210	3 Y	
6	Research & Dev	10	1 Medical	Medical	1	5	Male	1	Sales Executive	Single	23420	4 Y	
7	Research & Dev	8	3 Life Sciences	Life Sciences	1	6	Female	4	Research Direct	Married	40710	3 Y	
8	Research & Dev	11	2 Medical	Medical	1	7	Male	2	Sales Executive	Single	58130	2 Y	
9	Research & Dev	18	3 Life Sciences	Life Sciences	1	8	Male	2	Sales Executive	Married	31430	2 Y	
10	Research & Dev	1	3 Life Sciences	Life Sciences	1	9	Male	3	Laboratory Tech	Married	20440	0 Y	
11	Research & Dev	7	4 Medical	Medical	1	10	Female	4	Laboratory Tech	Divorced	134640	1 Y	
12	Research & Dev	17	2 Medical	Medical	1	11	Male	2	Laboratory Tech	Married	79910	0 Y	
13	Research & Dev	28	1 Life Sciences	Life Sciences	1	12	Male	1	Laboratory Tech	Married	33770	0 Y	
14	Research & Dev	14	4 Life Sciences	Life Sciences	1	13	Female	1	Sales Executive	Single	55380	0 Y	
15	Research & Dev	1	1 Medical	Medical	1	14	Male	1	Research Scient	Married	57620	1 Y	
16	Research & Dev	1	3 Life Sciences	Life Sciences	1	15	Male	1	Manufacturing D	Married	25920	1 Y	
17	Research & Dev	1	3 Life Sciences	Life Sciences	1	16	Male	2	Healthcare Rep	Married	53460	4 Y	
18	Research & Dev	3	2 Life Sciences	Life Sciences	1	17	Male	1	Laboratory Tech	Single	42130	1 Y	
19	Research & Dev	1	3 Medical	Medical	1	18	Male	2	Sales Executive	Divorced	41270	2 Y	
20	Sales	7	4 Life Sciences	Life Sciences	1	19	Male	1	Sales Represent	Divorced	24380	7 Y	
21	Research & Dev	8	3 Life Sciences	Life Sciences	1	20	Female	1	Manager	Divorced	68700	1 Y	
22	Research & Dev	1	4 Other	Other	1	21	Male	2	Laboratory Tech	Divorced	104470	1 Y	
23	Sales	8	4 Life Sciences	Life Sciences	1	22	Male	1	Research Scient	Divorced	96670	3 Y	
24	Research & Dev	11	4 Life Sciences	Life Sciences	1	23	Female	2	Research Scient	Married	21480	3 Y	
25	Research & Dev	4	4 Life Sciences	Life Sciences	1	24	Male	1	Manufacturing D	Married	89260	1 Y	

Employees with above-average Monthly Income

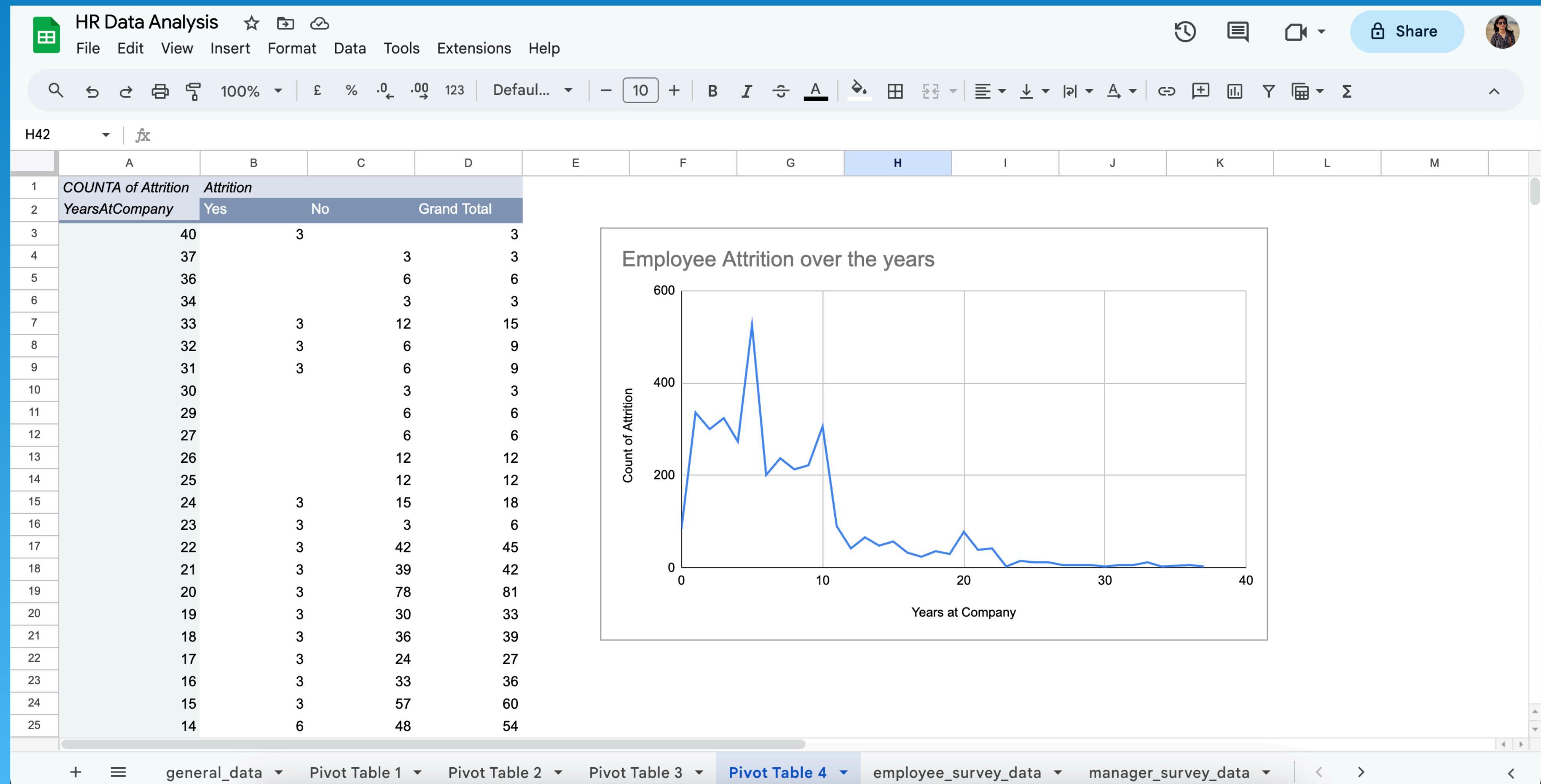
9. Apply conditional formatting to highlight employees with both above-average Monthly Income and above-average Job Satisfaction

The screenshot shows a Google Sheets spreadsheet titled "HR Data Analysis". The data is organized into columns A through F. Columns A, B, C, D, and E contain numerical values ranging from 1 to 4. Column F is empty. The rows are numbered from 1 to 25. Rows 2 through 25 have a green background, while row 1 has a light blue background. This visual cue highlights the rows where both monthly income and job satisfaction are above average.

	A	B	C	D	E	F
1	EmployeeID	Environment	JobSatisfact	WorkLifeBal		
2		1	3	4	2	
3		2	3	2	4	
4		3	2	2	1	
5		4	4	4	3	
6		5	4	1	3	
7		6	3	2	2	
8		7	1	3	1	
9		8	1	2	3	
10		9	2	4	3	
11		10	2	1	3	
12		11	3	4	3	
13		12	NA	4	3	
14		13	4	1	3	
15		14	1	2	2	
16		15	4	4	2	
17		16	3	4	4	
18		17	4	3	4	
19		18	1	4	3	
20		19	2	2	2	
21		20	1	1	3	
22		21	3	2	1	
23		22	1	2	2	
24		23	3	3	2	
25		24	2	3	3	

Employees with above-average Job Satisfaction

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

A star schema is a type of database schema used in data warehousing and business intelligence. It consists of one or more fact tables that store quantitative performance data and multiple dimension tables that store descriptive data related to the dimensions of the business.

Steps to Create a Star Schema:

Step 1: Identify Fact Table

- Determine the fact table that contains the primary quantitative measures (e.g., Attrition, MonthlyIncome, PercentSalaryHike, TotalWorkingYears)

Step 2: Identify Dimensions

- Identify dimensions related to the business process (e.g., EmployeeID, Gender, Department, JobRole)

Step 3: Define Primary Keys

- Assign primary keys to each dimension table and the fact table.

Step 4: Establish Relationships

- Connect the primary key in the fact table to the corresponding foreign keys in the dimension tables.

Step 5: Create Indexes

- Index the primary key columns in the dimension tables to improve query performance.

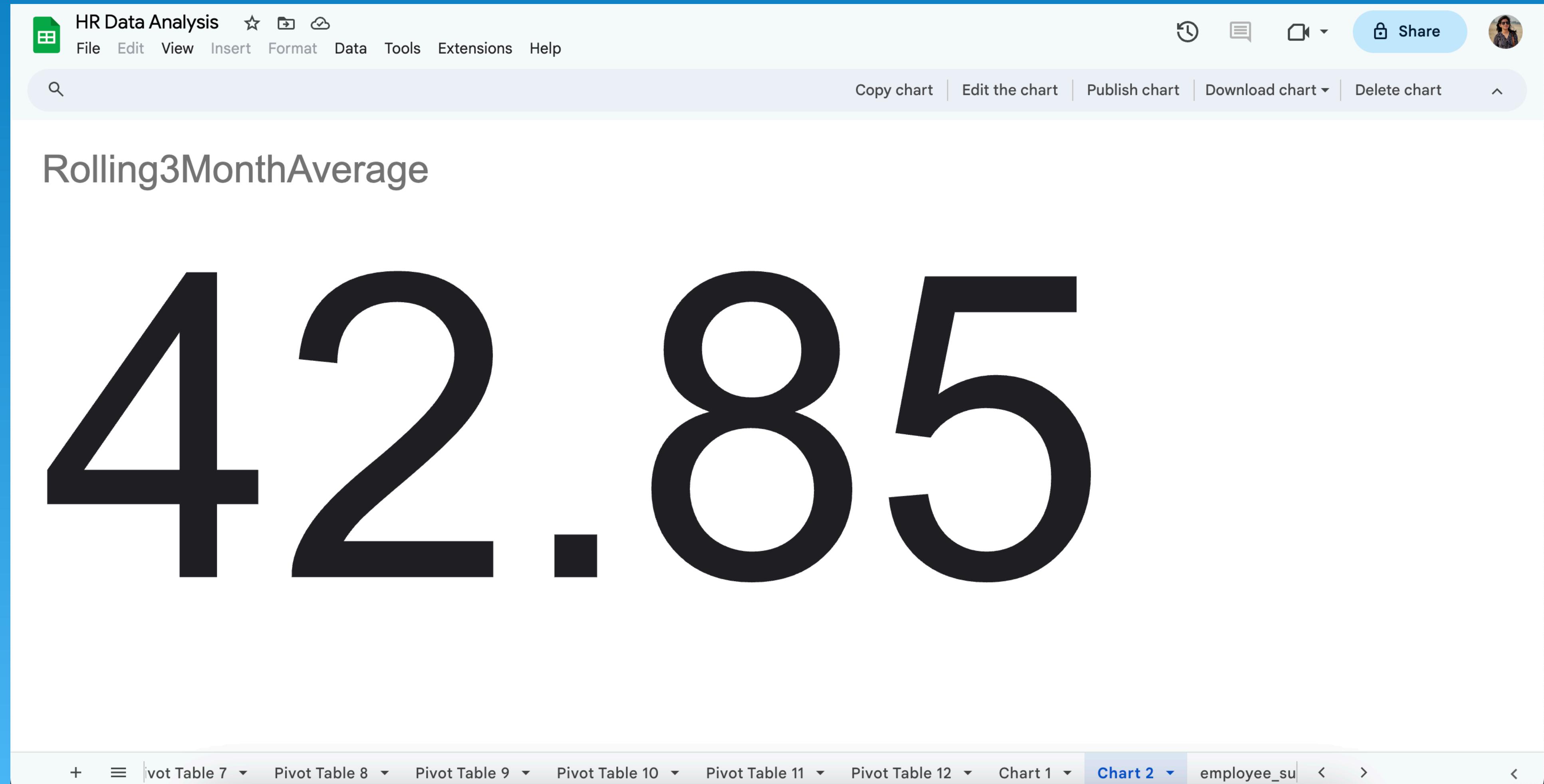
Step 6: Normalize Dimension Tables

- Normalize dimension tables to reduce redundancy and improve data consistency.

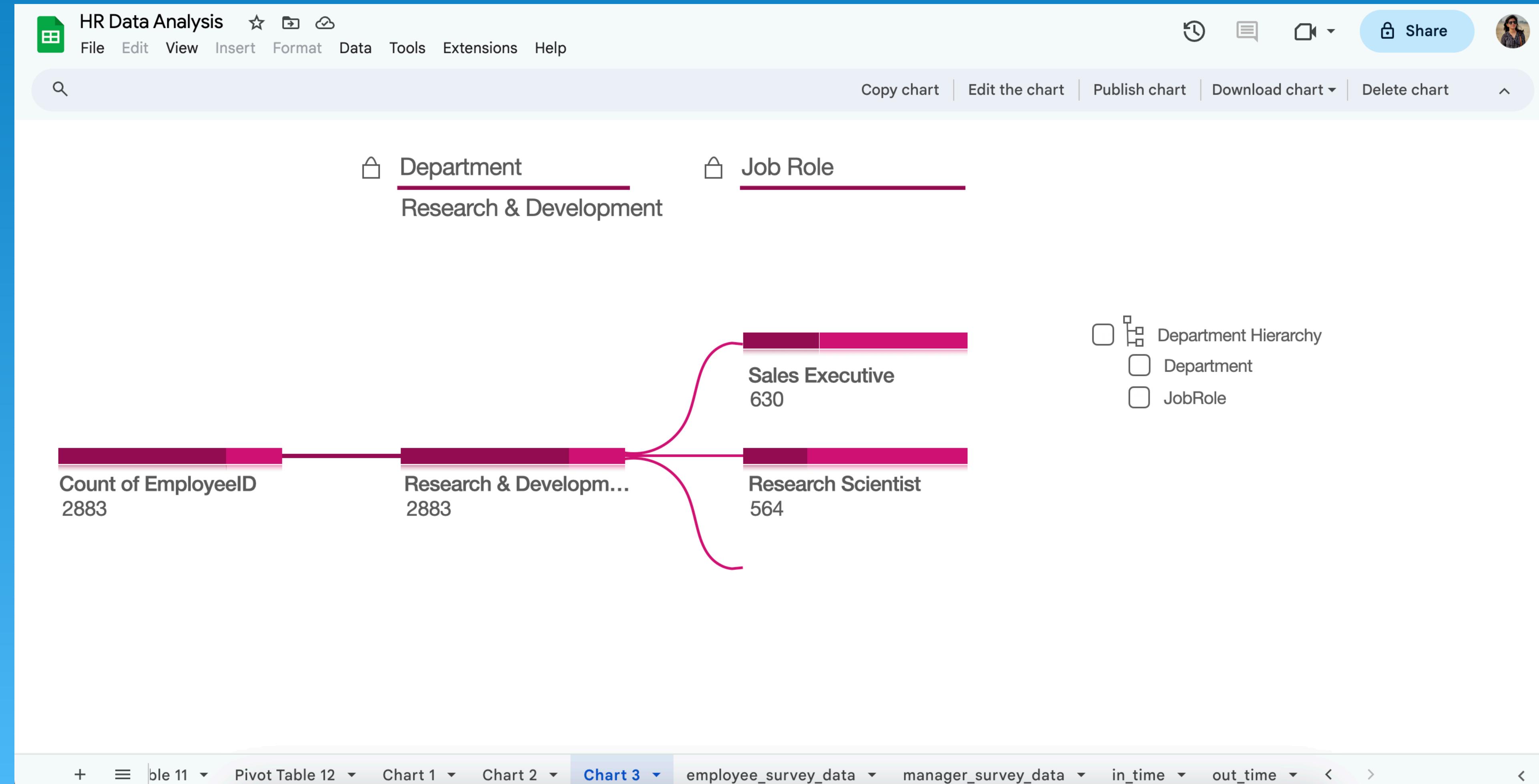
Benefits of Star Schema:

- Simplicity and Readability
- Query Performance
- Flexibility
- Ease of Maintenance
- Scalability
- Facilitates Business Intelligence (BI) Tools
- Optimized for OLAP
- Enhanced Data Consistency

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



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 on the Distance from Home column

In Power BI, parameterized queries can be set up using Power Query, the data preparation and transformation tool integrated into Power BI. Parameterized queries allow users to filter data dynamically based on user input.

Steps to set up a parameterized query:

Step 1: Load the Data into Power BI

- Open Power BI Desktop, Go to "Home" and click on "Get Data".
- Select your data source and load the data into Power BI.

Step 2: Open Power Query Editor

- In Power BI Desktop, go to "Home" and click on "Transform Data." This opens the Power Query Editor.

Step 3: Clean and Transform the Data

- Locate the "DistanceFromHome" column in the Power Query Editor.
- Clean and transform the data as needed. You may want to handle null values, data type conversions, or any other cleaning operations.

Step 4: Parameterize the Query

- In the Power Query Editor, go to the "Home" tab. Click on "Manage Parameters."
- In the Manage Parameters window, click on "New."
- Enter a name for the parameter (e.g., DistanceParam) and choose the data type (e.g., Decimal Number). Set a default value for the parameter.
- Click "OK" to create the parameter.

Step 5: Apply the Parameterized Filter

- Go back to the Power Query Editor. In the "Home" tab, select the "Advanced Editor".
- Replace the existing formula with the following, adjusting column names and parameter names as needed
- Ensure that [DistanceFromHome] matches the actual column name in your dataset.

Step 6: Load the Parameterized Data

- Click "Done" to close the Advanced Editor.
- Click "Close & Apply" in the Home tab to apply the changes and load the parameterized data.

Step 7: Use the Parameter in Report

- Create your report as usual, and add a slicer visual. Use the parameter you created (e.g., DistanceParam) as the slicer field.
- Users can now interactively filter data based on the Distance from Home column using the slicer.

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

The screenshot shows an Excel spreadsheet titled "HR Data Analysis". The Pivot Table is located in the range A1:L6. The table structure is as follows:

	A	B	C	D	E	F	G	H	I	J	K	L
1	SUM of MonthlyIncome	JobLevel										
2	Department		3	4	5	Grand Total						
3	Human Resources		1648500	754800	855840	3259140						
4	Research & Development		28117740	15277290	10107870	53502900						
5	Sales		11792400	8753070	2428860	22974330						
6	Grand Total		41558640	24785160	13392570	79736370						

The Pivot Table is part of a larger sheet containing other data and charts. The ribbon at the top shows tabs for "general_data", "Pivot Table 1", "Pivot Table 2", "Pivot Table 3", "Pivot Table 4", "Pivot Table 5" (which is selected), "Chart 1", and "employee_survey_data". The status bar at the bottom indicates the formula bar is active.

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

Performing a What-If analysis in Excel involves using a data table to see how changing one or more input values (in this case, a 10% increase in Percent Salary Hike) impacts a calculated result (Monthly Income).

Steps to conduct What-If Analysis:

Having a dataset in Excel with columns like "Percent Salary Hike" and "Monthly Income".

Step 1: Set Up Your Data

- Data:
Have a column with various values for Percent Salary Hike and a column with the corresponding Monthly Income values.
- Calculate New Monthly Income:
In an empty cell, calculate the new Monthly Income based on a 10% increase. The original Monthly Income is in cell N1, you can use the formula:
 $=N1 * (1 + 0.10)$

Step 2: Set Up Data Table

- Enter Input Values:
In a nearby empty area of your worksheet, create a column for different values of Percent Salary Hike. For example, you might have values like 5%, 10%, 15%, etc.
- Create Data Table:
Highlight the range of your input values and the cell where you calculated the new Monthly Income (including headers).
"Data" tab -> "What-If Analysis" -> "Data Table."
- Enter Row Input Cell:
In the Data Table dialog box, enter the reference to the cell containing the original Percent Salary Hike. The original Percent Salary Hike is in cell Q1, enter \$A\$1 as the Row Input Cell.
- Click OK
Click "OK" to create the data table.

Step 3: Review the What-If Analysis

- Review the Data Table:
The data table will populate with different Monthly Income values based on the Percent Salary Hike values you entered.
- Interpret the Results:
Examine how changes in Percent Salary Hike impact the resulting Monthly Income. This gives you insights into the potential impact of a 10% increase in Percent Salary Hike on Monthly Income.

Step 4: Visualize the Results (Optional)

- Create a Chart:
You can create a chart to visualize the impact. For example, create a line chart with Percent Salary Hike on the x-axis and Monthly Income on the y-axis.
- Analyze the Chart:
The chart allows for a visual representation of the What-If analysis, making it easier to interpret the impact of different Percent Salary Hike values on Monthly Income.

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

Verifying if data adheres to a predefined schema involves the process of checking whether the actual data in a dataset aligns with the expected structure and rules outlined in the predefined schema. As of the current status of the data, there are specific actions that need to be taken to ensure alignment with the predefined schema:

1. Reordering 'EmployeeID' Column:

The 'EmployeeID' column needs to be reordered to adhere to the expected sequence outlined in the predefined schema. This ensures consistency and facilitates better data organization.

2. Changing Data Type of 'TotalWorking Hours' Column:

The data type of the 'TotalWorking Hours' column should be modified to match the specified data type in the predefined schema. This adjustment ensures data uniformity and accurate representation of working hours.

3. Handling Missing and Blank Values:

Missing values and blank entries in the "general_data," "employee_survey_data," and "manager_survey_data" need to be addressed. Appropriate methods, such as imputation or filling with default values, should be applied to ensure completeness and reliability in the dataset.

4. Adding 'EmployeeID' Label in 'In-Time' and 'Out-Time' Data:

The 'EmployeeID' label is missing in the 'in-time' and 'out-time' data. It is crucial to include this label to maintain consistency across different datasets and enable proper linkage between data sources.

5. Data Profiling and Validation:

Rigorous data profiling and validation processes should be implemented to identify and rectify any inconsistencies, anomalies, or discrepancies in the dataset. This involves scrutinizing the data against the predefined schema to ensure accurate representation and adherence to expected standards.

6. Addressing Inconsistencies in Data Files:

Any inconsistencies found in the data files should be carefully addressed. This may include correcting errors, standardizing formats, and resolving discrepancies to bring the data in line with the predefined schema.

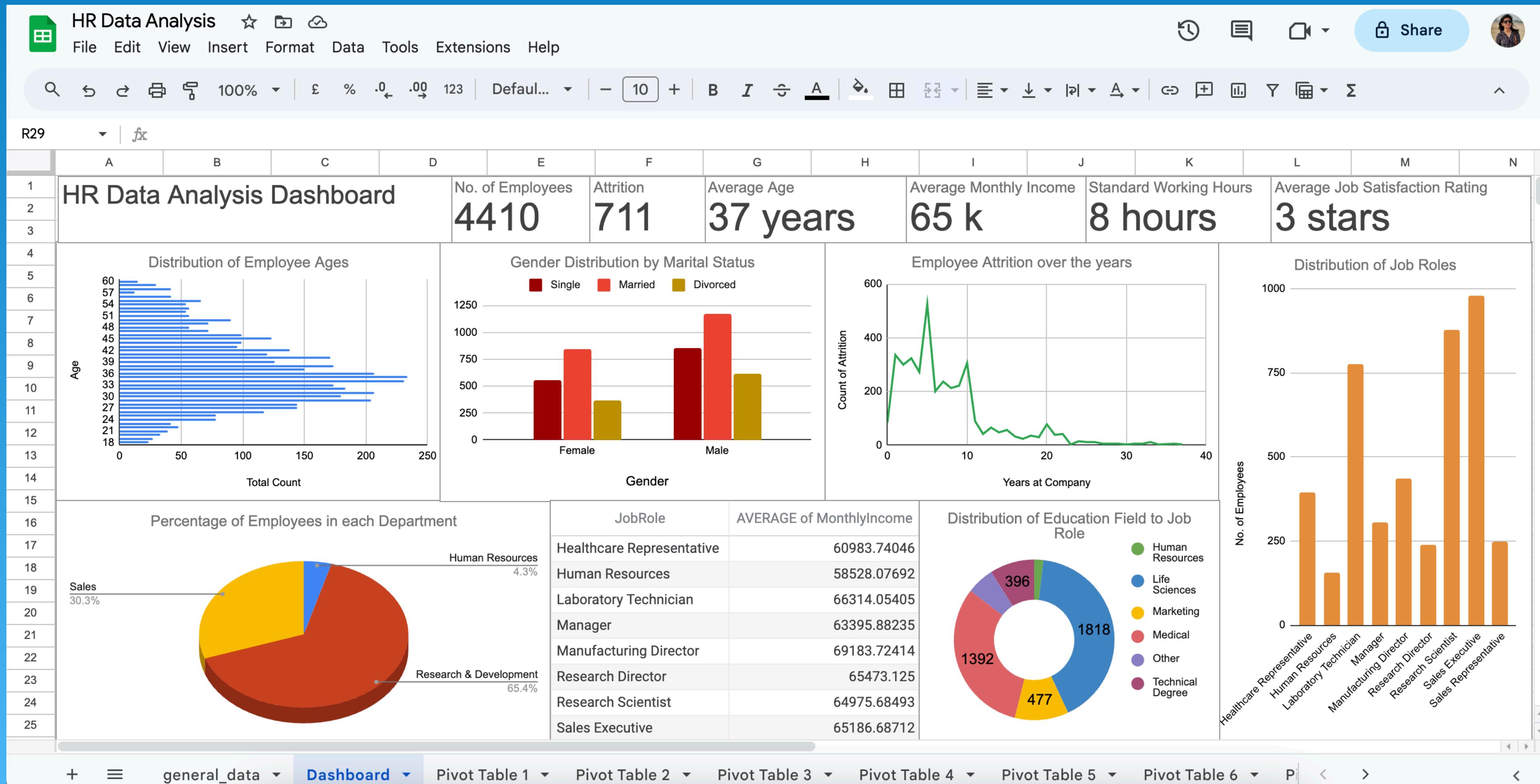
7. Implementing Data Quality Measures:

Robust data quality measures need to be implemented to enhance the overall quality of the dataset. This involves establishing and enforcing standards, ensuring accuracy, and promoting reliability in the data.

8. Conforming to the Predefined Schema:

The dataset will conform to the predefined schema once the above-mentioned inconsistencies are resolved, and the recommended data quality measures are implemented. This alignment is essential for maintaining the integrity and reliability of the data.

HR Data Analysis Dashboard



Thank you