Q1. Using Excel, how would you filter the dataset to only show employees aged 30 and above?

Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeID	Gender	JobLevel	JobRole	MaritalStatus	MonthlyIncome	NumCompaniesWorked	Over
51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	1	Female	1	Healthcare	Married	131160	1	Y
31	Yes	Travel_Frequently	Research & Dev	10	1	Life Sciences	1	2	Female	1	Research S	Single	41890	0	Y
32	No	Travel_Frequently	Research & Dev	17	4	Other	1	3	Male	4	Sales Execu	Married	193280	1	Y
38	No	Non-Travel	Research & Dev	2	5	Life Sciences	1	4	Male	3	Human Res	Married	83210	3	Y
32	No	Travel_Rarely	Research & Dev	10	1	Medical	1	5	Male	1	Sales Execu	Single	23420	4	Y
46	No	Travel_Rarely	Research & Dev	8	3	Life Sciences	1	6	Female	4	Research D	Married	40710	3	Y
31	No	Travel_Rarely	Research & Dev	1	3	Life Sciences	1	9	Male	3	Laboratory	Married	20440	0	Y
45	No	Travel_Rarely	Research & Dev	17	2	Medical	1	11	Male	2	Laboratory	Married	79910	0	Y
36	No	Travel_Rarely	Research & Dev	28	1	Life Sciences	1	12	Male	1	Laboratory	Married	33770	0	Y
55	No	Travel_Rarely	Research & Dev	14	4	Life Sciences	1	13	Female	1	Sales Execu	Single	55380	0	Y
47	Yes	Non-Travel	Research & Dev	1	1	Medical	1	14	Male	1	Research S	Married	57620	1	Y
37	No	Travel_Rarely	Research & Dev	1	3	Life Sciences	1	16	Male	2	Healthcare	Married	53460	4	Y
37	No	Non-Travel	Research & Dev	1	3	Medical	1	18	Male	2	Sales Execu	Divorced	41270	2	Y
35	No	Travel_Rarely	Sales	7	4	Life Sciences	1	19	Male	1	Sales Repre	Divorced	24380	7	Y
38	No	Travel_Rarely	Research & Dev	8	3	Life Sciences	1	20	Female	1	Manager	Divorced	68700	1	Y
50	No	Travel_Rarely	Sales	8	4	Life Sciences	1	22	Male	1	Research S	Divorced	96670	3	Y
53	No	Travel_Rarely	Research & Dev	11	4	Life Sciences	1	23	Female	2	Research S	Married	21480	3	Y
42	No	Travel_Rarely	Research & Dev	4	4	Life Sciences	1	24	Male	1	Manufactu	Married	89260	1	Y
55	No	Travel_Rarely	Research & Dev	1	4	Other	1	26	Female	1	Research S	Married	67990	3	Y

Q2. Create a pivot table to summarize the average Monthly Income by Job Role

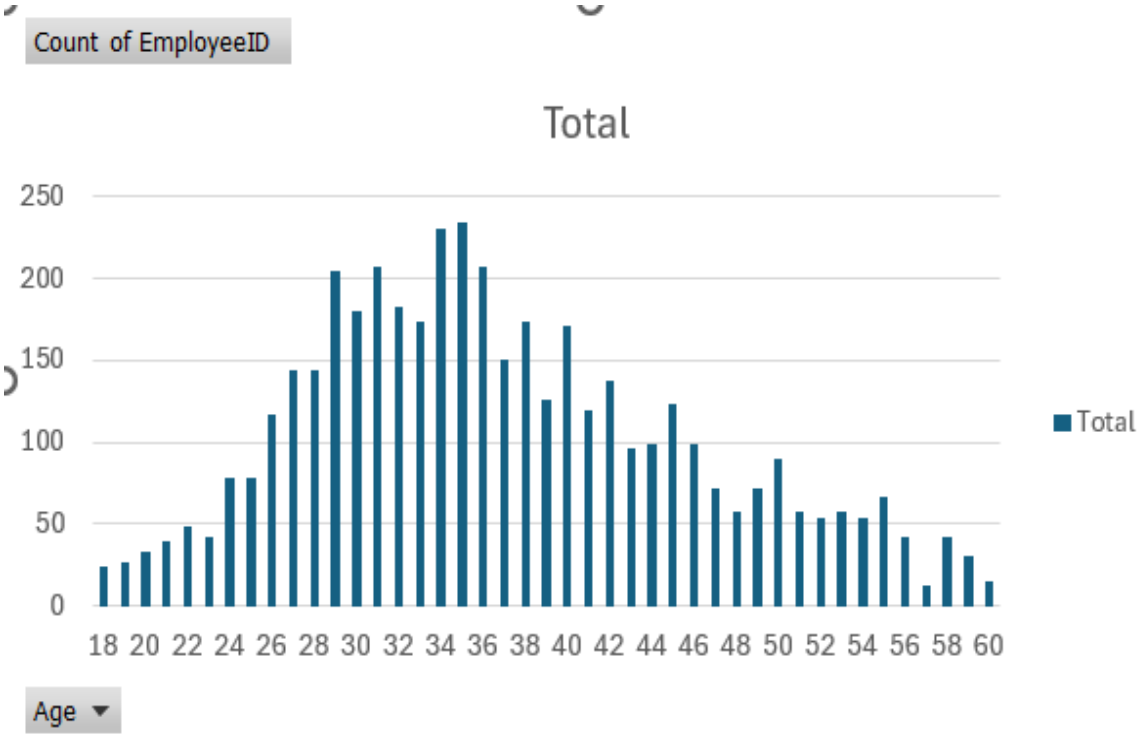
Row Labels	Average of MonthlyIncome
Healthcare Representative	60983.74
Human Resources	58528.08
Laboratory Technician	66314.05
Manager	63395.88
Manufacturing Director	69183.72
Research Director	65473.13
Research Scientist	64975.68
Sales Executive	65186.69
Sales Representative	65370.96
Grand Total	**65029.31**

Q3. Apply conditional formatting to highlight employees with Monthly Income above the company's average income

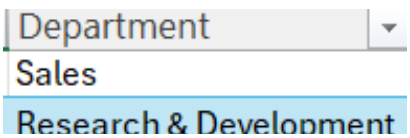
Row	Monthl	Income
1	131160	
2	41890	
3	193280	
4	83210	
5	23420	
6	40710	
7	58130	
8	31430	

8	31450
9	20440
10	134640
11	79910
12	33770
13	55380
14	57620
15	25920
16	53460
17	42130
18	41270
19	24380
20	68700
21	104470
22	96670
23	21480

Q4. Create a bar chart in Excel to visualize the distribution of employee ages.

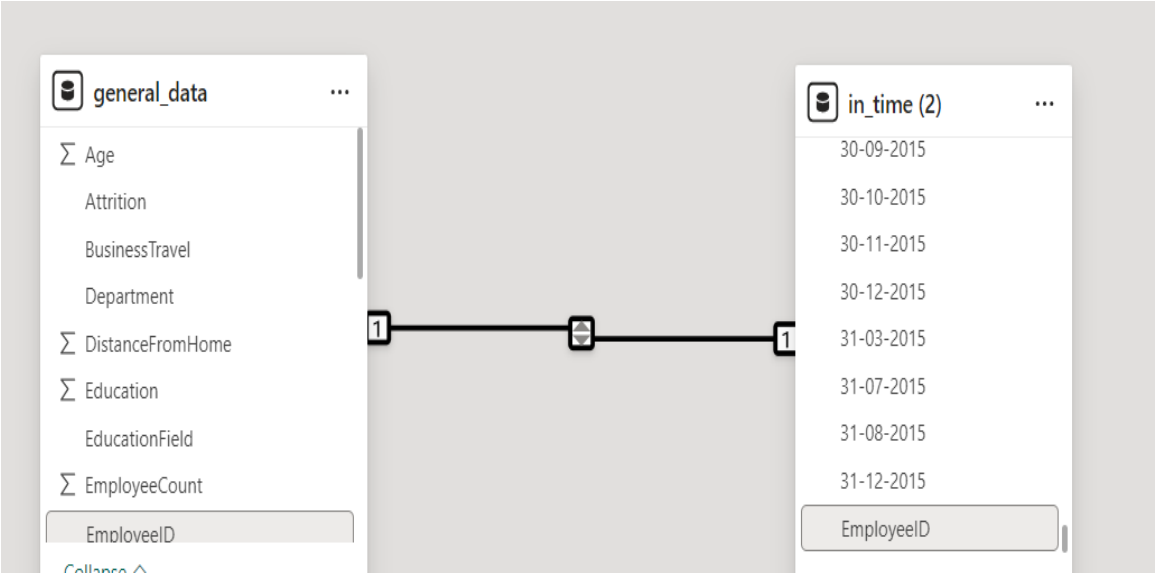


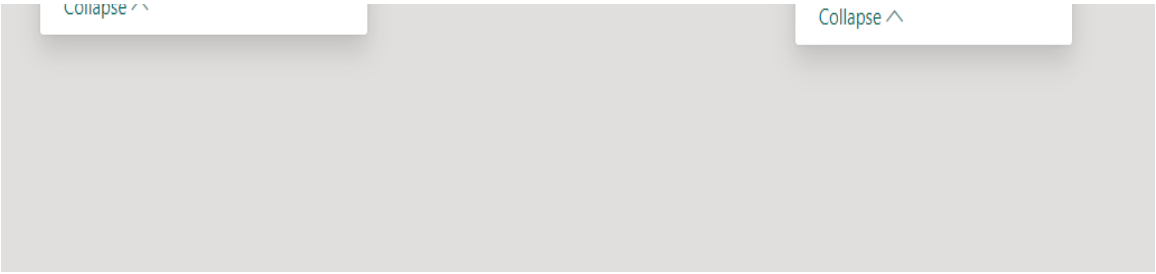
Q5. Identify and clean any missing or inconsistent data in the "Department" column.



Research & Development
Research & Development
Research & Development
Research & Development
Sales
Sales
Sales
Research & Development
Human Resources
Sales
Sales
Sales
Sales
Research & Development
Research & Development
Research & Development
Research & Development
Research & Development
Research & Development

Q6. In Power BI, establish a relationship between the "EmployeeID" in the employee data and the "EmployeeID" in the time tracking data





Q7. Using DAX, create a calculated column that calculates the average years an employee has spent with their current manager.

AverageYearsWithCurrEmployers
9.5
12
7
3
5
7.5
1.5
1
2.5
12
3
2.5
9
5
5.5
1
2.5
2
3
2.5
2
1
2.5

Q8. 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
Human Resources	**189**
Divorced	21
Married	96
Single	72
Research & Development	**2883**
Divorced	621
Married	1350
Single	912
Sales	**1338**
Divorced	339
Married	573
Single	426
Grand Total	**4410**

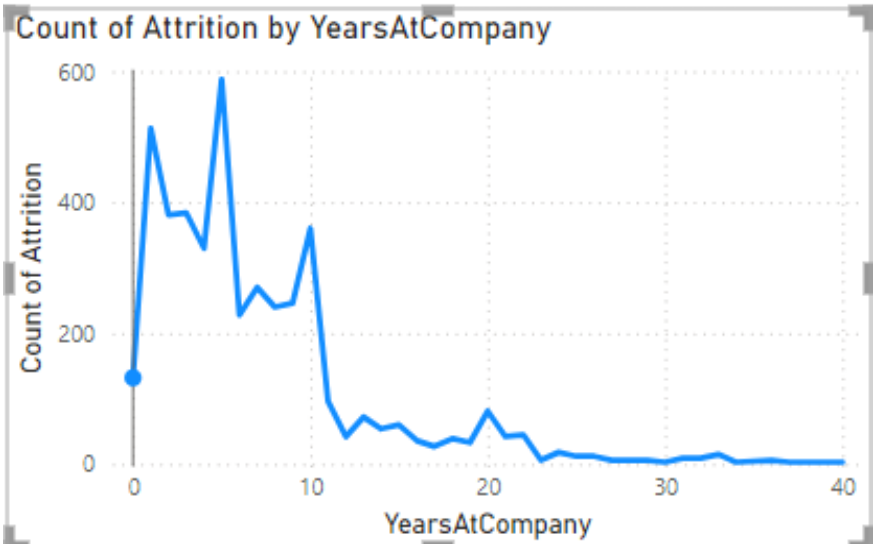
Q9. Apply conditional formatting to highlight employees with both above-average Monthly Income and above-average Job Satisfaction.

MonthlyIncome	JobSatisfa
131160	4
41890	2
193280	2
83210	4
23420	1
40710	2
58130	3
31430	2
20440	4
134640	1
79910	4
33770	4
55380	1
	2

57620
25920
53460
42130
41270
24380
68700

4
4
3
4
2
1
2
2
3

Q10. In Power BI, create a line chart that visualizes the trend of Employee Attrition over the years.



Q11. Describe how you would create a star schema for this dataset, explaining the benefits of doing so.

1. Identify the Fact Table:

The **Fact Table** should hold the primary quantitative metrics you want to analyze. In this case, **YearsAtCompany, and TotalWorkingYears.**

the primary fact could be Employee Attrition and other metrics such as Monthly Income,

Fact Table: **Employee Fact**

Columns:

EmployeeID (Foreign Key)

Monthly Income

Years At Company

Total Working Years
Years Since Last Promotion
Years With Curr Manager
Attrition (Yes/No)
TerminationType
TerminationDescription
Stock Option Level
Percent Salary Hike
Job Involvement (from Sheet 3)
Performance Rating (from Sheet 3)
Environment Satisfaction (from Sheet 4)
Job Satisfaction (from Sheet 4)
Work-Life Balance (from Sheet 4)

2. Create Dimension Tables: Dimension tables will hold categorical data that describes the facts in the fact table. These tables are designed to avoid redundancy and provide context for the metrics in the fact table.

Dimension Table 1: Employee Dimension:

EmployeeID (Primary Key)
Age
Gender
Marital Status
Num Companies Worked
Business Travel
Department
Job Role
Job Level
Education
Education Field

Dimension Table 2: Time Dimension (if your dataset includes dates or you want to add date-based analysis, you would create a date/time table).

DateKey (Primary Key)
Full Date
Year
Month
Day
Quarter

Dimension Table 3: Geography Dimension:

Distance From Home
Location (if available, otherwise derive from business-related data).

Dimension Table 4: Job Role and Department Dimension:

JobRoleID (Primary Key)
Job Role
Job Level
Department
Stock Option Level
Over18

3. Linking Dimension Tables with Fact Table:

All the dimension tables will be linked to the fact table via **foreign keys** (e.g., EmployeeID in the Fact Table will link to the EmployeeID in the Employee Dimension table).
In Power BI or a database management system, you would create **one-to-many relationships** between the fact table and each dimension table, based on the corresponding keys.

4. Star Schema Structure:

Fact Table: Employee Fact (Central Table)

Foreign keys: EmployeeID, DateKey, etc.
Quantitative metrics: Monthly Income, Attrition, Years At Company, etc.

Dimension Table 1: Employee Dimension

Descriptive attributes related to the employee: Age, Gender, Job Role, etc.

Dimension Table 2: Time Dimension (if relevant)

Attributes for date-based analysis: Year, Month, Quarter, etc.

Dimension Table 3: Job Role and Department Dimension

Descriptive attributes related to the job and department: Job Role, Department, etc.

Dimension Table 4: Geography Dimension

Descriptive attributes related to location: Distance From Home.

Example Star Schema Design:

Fact Table: Employee Fact

EmployeeID (FK)
Attrition
Monthly Income
Years At Company
Total Working Years
Job Involvement
Performance Rating
Environment Satisfaction
Job Satisfaction
Work-Life Balance
Termination Type
Termination Description

Dimension Table 1: Employee Dimension

EmployeeID (PK)
Age
Gender
Marital Status
Education
Job Role
Job Level
Department
Num Companies Worked

Dimension Table 2: Job Role and Department Dimension

JobRoleID (PK)
Job Role
Job Level
Department
Stock Option Level

Dimension Table 3: Geography Dimension

Distance From Home

Dimension Table 4: Time Dimension (Optional for Date Analytics)

DateKey (PK)
Year
Month
Day

Benefits of a Star Schema:

1. Simplifies Queries:

A star schema makes queries more straightforward by organizing data into a clear structure of fact and dimension tables. Analysts can run queries by simply joining the fact table to dimension tables.

2. Improved Performance:

Star schemas reduce the number of joins needed in queries, which leads to faster query performance, especially in large datasets.

3. Better Organization:
Data is neatly organized, with facts (e.g., income, tenure) stored in the central table and descriptive context stored in dimension tables. This improves data clarity and usability.

4. Scalability:
Adding new dimensions (e.g., job satisfaction or geography) is easy because they can be linked to the existing fact table without affecting the core structure.

5. Flexibility in Reporting:

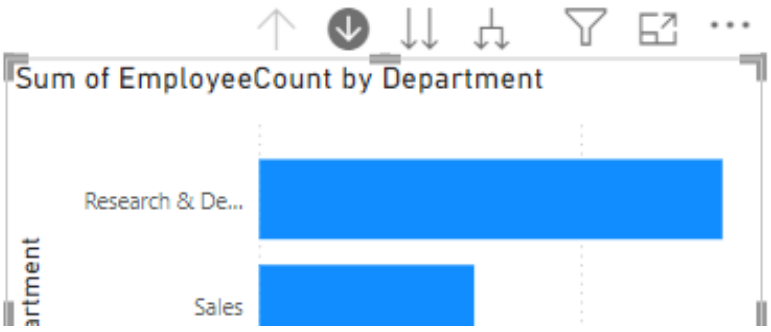
You can slice and dice data easily across different dimensions, such as analyzing attrition by department, income by job role, or performance by years with the manager.

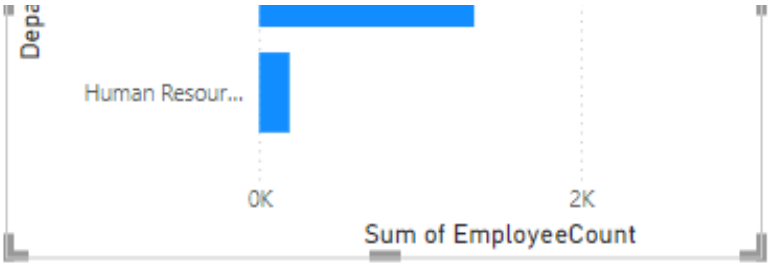
In this schema, you have a central fact table linked to several dimension tables that describe different attributes of the employee and their roles, making reporting, aggregation, and analysis intuitive and efficient.

Q12. Using DAX, calculate the rolling 3-month average of Monthly Income for each employee.

Rolling3MonthAvgMonthlyIncome
54680
54680
54680
54680
54680
54680
54680
54680
54680
54680
54680
54680

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





14Q. 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 up parameterized queries in Power BI for filtering data based on the **Distance from Home** column, follow these steps:

1. Create a Parameter:

Go to the **Modeling** tab and select **New Parameter**.
Define the parameter (e.g., DistanceFromHomeParameter) with a suitable data type, range, and default value.

2. Use the Parameter in Queries:

Open the **Power Query Editor**.
Select the relevant table and apply a filter on the **Distance from Home** column.
Set the filter to use the parameter by referencing it in the filter condition (e.g., "is less than or equal to DistanceFromHomeParameter").

3. Add a Slicer:

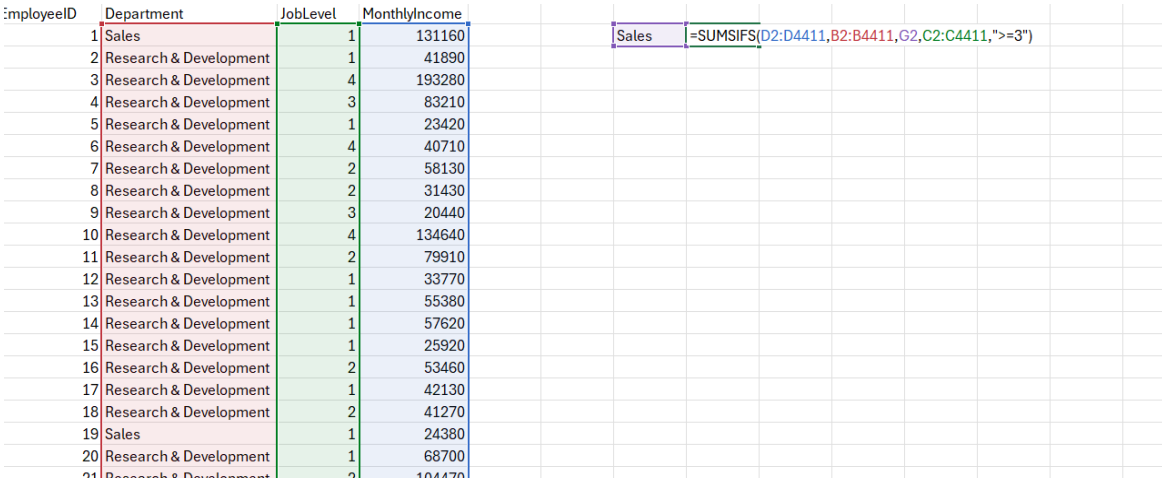
In the report view, add a slicer visual and drag the parameter into it to allow users to select their desired distance.

4. Test the Slicer:

Change the slicer value to see the visuals update dynamically based on the selected distance.

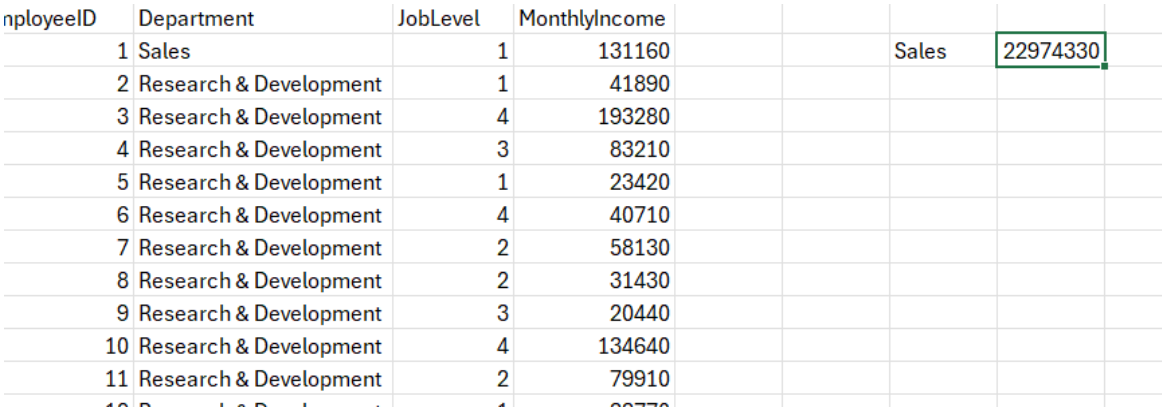
This setup allows users to filter data interactively based on the **Distance from Home** parameter.

Q15. In Excel, calculate the total Monthly Income for each Department, considering only the employees with a Job Level greater than or equal to 3.



EmployeeID	Department	JobLevel	MonthlyIncome
1	Sales	1	131160
2	Research & Development	1	41890
3	Research & Development	4	193280
4	Research & Development	3	83210
5	Research & Development	1	23420
6	Research & Development	4	40710
7	Research & Development	2	58130
8	Research & Development	2	31430
9	Research & Development	3	20440
10	Research & Development	4	134640
11	Research & Development	2	79910
12	Research & Development	1	33770
13	Research & Development	1	55380
14	Research & Development	1	57620
15	Research & Development	1	25920
16	Research & Development	2	53460
17	Research & Development	1	42130
18	Research & Development	2	41270
19	Sales	1	24380
20	Research & Development	1	68700

Sales =SUMSIFS(D2:D4411,B2:B4411,G2,C2:C4411,">=3")



nployeeID	Department	JobLevel	MonthlyIncome
1	Sales	1	131160
2	Research & Development	1	41890
3	Research & Development	4	193280
4	Research & Development	3	83210
5	Research & Development	1	23420
6	Research & Development	4	40710
7	Research & Development	2	58130
8	Research & Development	2	31430
9	Research & Development	3	20440
10	Research & Development	4	134640
11	Research & Development	2	79910

Sales 22974330

Q16. 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.

16A.1: Set Up Your Data

Start by preparing your data in Excel. You need a table with three

EmployeeID
Monthly Income
Percent Salary Hike

For example, you could have the following data:

Employee 1: Monthly Income of 4000 and a Percent Salary Hike of

Employee 2: Monthly Income of 5000 and a Percent Salary Hike of
Employee 3: Monthly Income of 6000 and a Percent Salary Hike of

2: Calculate New Monthly Income

Next, we will calculate the new Monthly Income after increasing the Percent Salary Hike by 10%.

1. Add a New Column: Create a new column next to your data and name it **New Monthly Income**.

2. Enter the Formula: In the first cell of this new column, type a formula to calculate the new income. The formula should take the Monthly Income and add 10% of the current Percent Salary Hike to it.

3. Copy the Formula: Click on the small square at the bottom right of the cell and drag it down to apply this formula to all employees.

3: Create a Summary Table

Now, let's create a summary table to show both the original and new

1. Set Up a New Table: Create a new table that lists each EmployeeID, their original Monthly Income, and the newly calculated

4: Set Up a Data Table for What-If Analysis

1. Create a New Section: Set up a small table to explore different increases in the Percent Salary Hike, like 10%, 20%, and 30%.

2. Enter Increases: In the first column, write down the different percentage increases you want to analyze.

3. Link to Your Calculation: In the first row of the New Monthly Income column, enter a formula that calculates the new Monthly

4. Create the Data Table: Select the whole table you just created. Go to the Data tab, click on What-If Analysis, and choose Data Table. For the Column Input Cell, select the cell that contains the original

5: Check the Results

Once you set everything up, Excel will automatically calculate and fill in the New Monthly Income based on the percentage increases you specified. This will let you see how different increases in Percent

Q17. Verify if the data adheres to a predefined schema. What actions would you take if you find inconsistencies

17A.When working with data, it's super important to make sure it follows a specific structure or schema. This means checking that all the data matches what we expect it to look like. Here's how I'd go

Define the Schema: First, I'd outline what I expect from the data.
Column Names: I'd list all the columns I want to see.
Data Types: I'd specify the type of data for each column, like
Constraints: If certain fields are required or need to be unique, I'd

Validate the Data: Next, I'd check the actual data against this
Check Column Names: I'd make sure all the column names match what I defined and look out for any typos or missing ones.
Verify Data Types: It's crucial that each column has the correct data type. For instance, I'd check that date fields are in a valid date
Assess Constraints: Here, I'd look for any missing values in required fields, check for duplicates where they shouldn't be, and make sure all the data falls within the expected ranges.

Identify Inconsistencies: To find any issues, I'd use data profiling tools like Power BI or Excel. If the dataset isn't too large, I might even do a manual review to catch things that tools might miss.

Take Action on Inconsistencies:
Fix Errors: If I find any typos or format issues, I'd correct them. For missing values, I could fill them in with averages or appropriate placeholders. If there are duplicates, I'd remove those to ensure
Communicate with Stakeholders: If I come across major inconsistencies or if fixing the data requires input from others, I'd reach out to the relevant people to clarify and resolve those issues.
Document Changes: It's also important to keep a record of what changes I made. This helps maintain data integrity and provides a

Implement Data Quality Checks: After making corrections, I'd set up processes to regularly check the data against the schema in the
Conducting regular audits of the data.

Automating checks with scripts to ensure that any new data entries comply with the schema.

Re-validate: Finally, once I've made corrections, I'd re-check the data to confirm that it now adheres to the predefined schema. It's like a final check to ensure everything's good to go!

By following these steps, I can effectively verify that the data meets the required standards and take necessary actions to fix any inconsistencies I find. It's all about keeping our data clean and reliable!