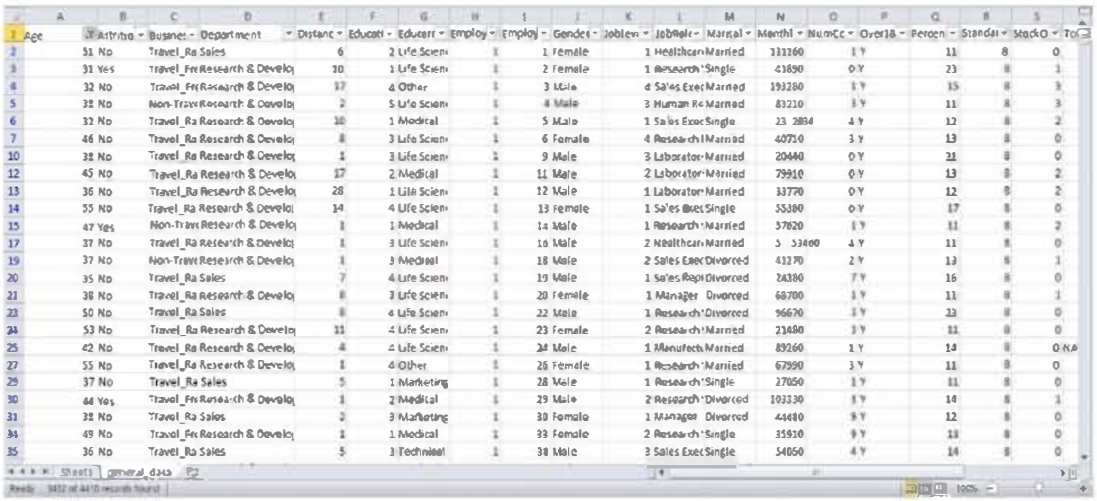
Name : Aditya Srivastava

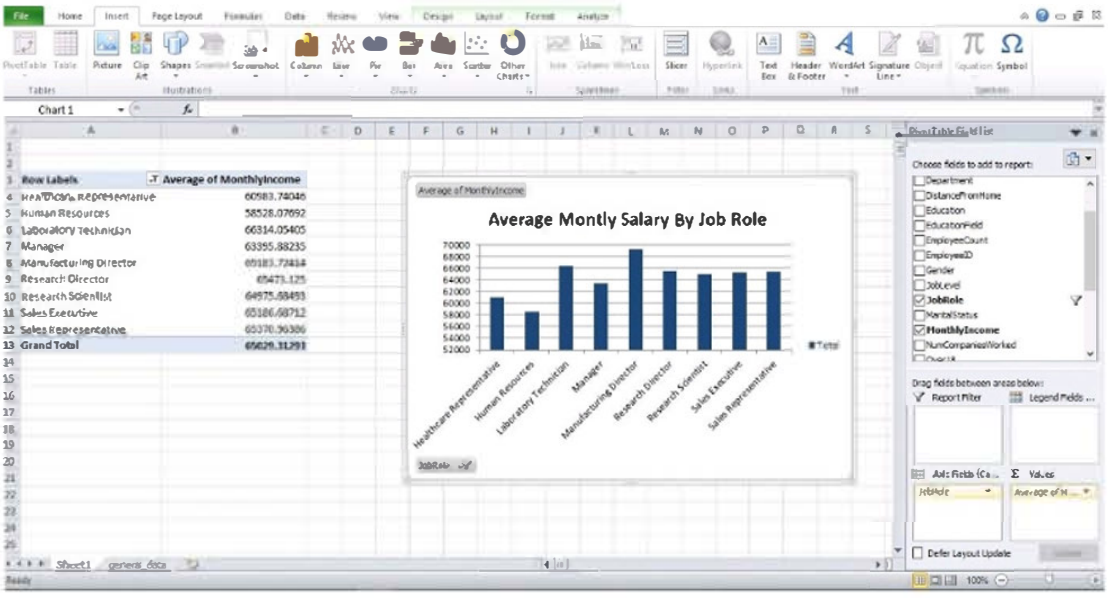
Email : itzadicr7@gmail.com

HR DATA ANALYSIS

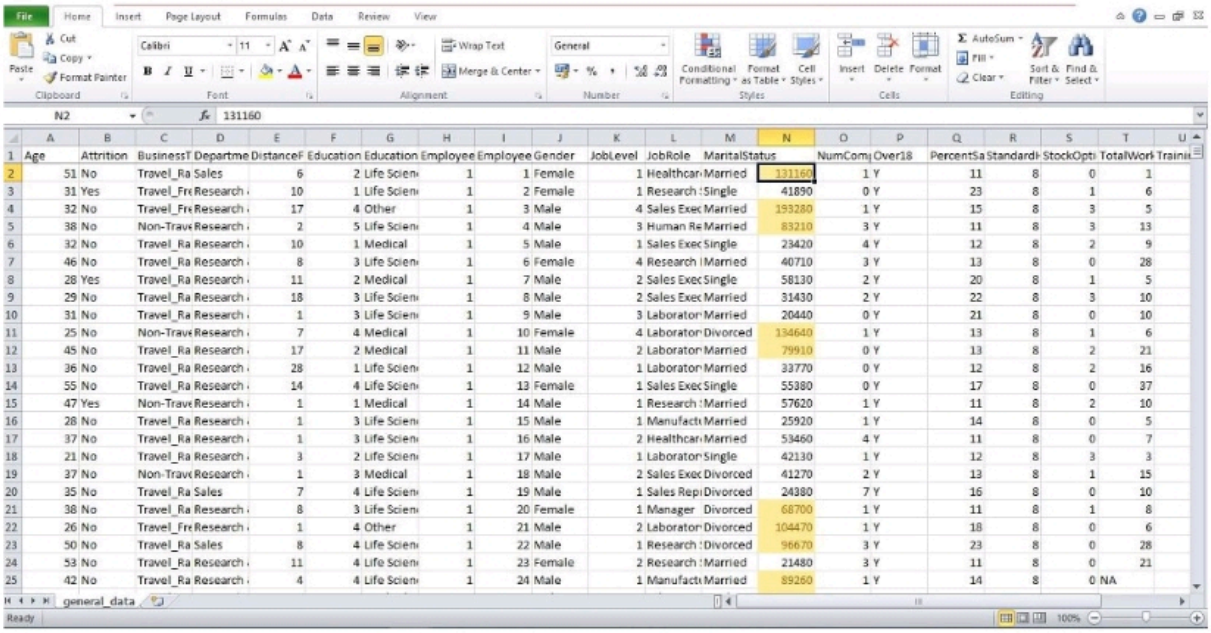
1. Using Excel, how would you filter the dataset to only show employees aged 30 and above?



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

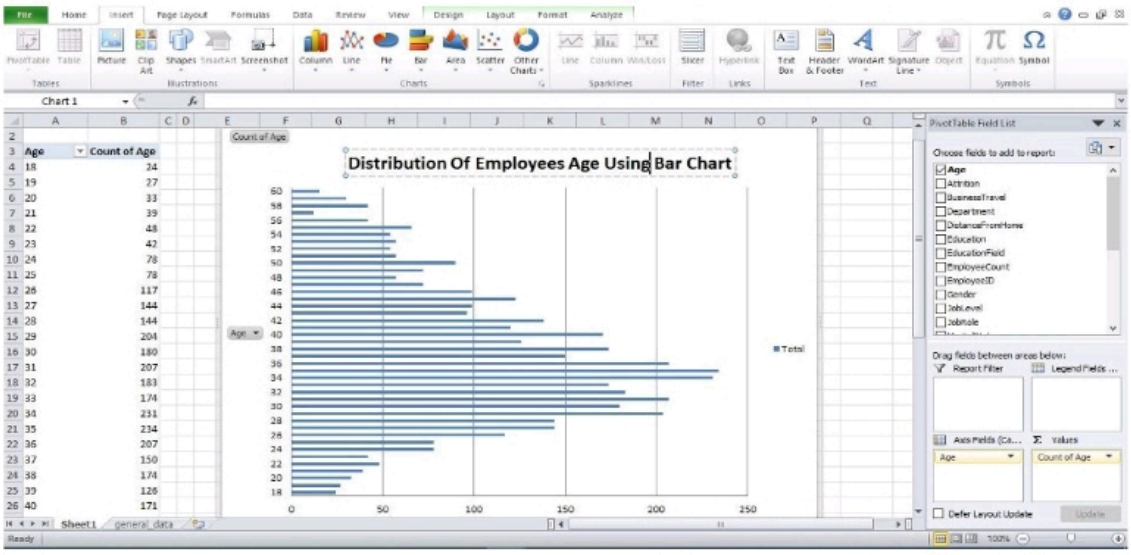


3. Apply conditional formatting to highlight employees with Monthly Income above the company's average income.



Age	Attrition	BusinessT	Departme	DistanceF	Education	Education	Employee	Employee	Gender	JobLevel	JobRole	MaritalStatus		NumCom	Over18	PercentSa	StandardH	StockOpti	TotalWork	Trainin
51	No	Travel_Ra	Sales	6	2	Life Scien	1	1	Female	1	Healthcar	Married	131160	1	Y	11	8	0	1	
31	Yes	Travel_Fr	Research	10	1	Life Scien	1	2	Female	1	Research	Single	41890	0	Y	23	8	1	6	
32	No	Travel_Fr	Research	17	4	Other	1	3	Male	4	Sales Exec	Married	193280	1	Y	15	8	3	5	
38	No	Non-Trave	Research	2	5	Life Scien	1	4	Male	3	Human Re	Married	83210	3	Y	11	8	3	13	
32	No	Travel_Ra	Research	10	1	Medical	1	5	Male	1	Sales Exec	Single	23420	4	Y	12	8	2	9	
46	No	Travel_Ra	Research	8	3	Life Scien	1	6	Female	4	Research	Married	40710	3	Y	13	8	0	28	
28	Yes	Travel_Ra	Research	11	2	Medical	1	7	Male	2	Sales Exec	Single	58130	2	Y	20	8	1	5	
29	No	Travel_Ra	Research	18	3	Life Scien	1	8	Male	2	Sales Exec	Married	31430	2	Y	22	8	3	10	
31	No	Travel_Ra	Research	1	3	Life Scien	1	9	Male	3	Laborator	Married	20440	0	Y	21	8	0	10	
25	No	Non-Trave	Research	7	4	Medical	1	10	Female	4	Laborator	Divorced	134640	1	Y	13	8	1	6	
45	No	Travel_Ra	Research	17	2	Medical	1	11	Male	2	Laborator	Married	79910	0	Y	13	8	2	21	
36	No	Travel_Ra	Research	28	1	Life Scien	1	12	Male	1	Laborator	Married	33770	0	Y	12	8	2	16	
55	No	Travel_Ra	Research	14	4	Life Scien	1	13	Female	1	Sales Exec	Single	55380	0	Y	17	8	0	37	
47	Yes	Non-Trave	Research	1	1	Medical	1	14	Male	1	Research	Married	57620	1	Y	11	8	2	10	
28	No	Travel_Ra	Research	1	3	Life Scien	1	15	Male	1	Manufactu	Married	25920	1	Y	14	8	0	5	
37	No	Travel_Ra	Research	1	3	Life Scien	1	16	Male	2	Healthcar	Married	53460	4	Y	11	8	0	7	
21	No	Travel_Ra	Research	3	2	Life Scien	1	17	Male	1	Laborator	Single	42130	1	Y	12	8	3	3	
37	No	Non-Trave	Research	1	3	Medical	1	18	Male	2	Sales Exec	Divorced	41270	2	Y	13	8	1	15	
35	No	Travel_Ra	Sales	7	4	Life Scien	1	19	Male	1	Sales Rep	Divorced	24380	7	Y	16	8	0	10	
38	No	Travel_Ra	Research	8	3	Life Scien	1	20	Female	1	Manager	Divorced	68700	1	Y	11	8	1	8	
26	No	Travel_Fr	Research	1	4	Other	1	21	Male	2	Laborator	Divorced	104470	1	Y	18	8	0	6	
50	No	Travel_Ra	Sales	8	4	Life Scien	1	22	Male	1	Research	Divorced	96670	3	Y	23	8	0	28	
53	No	Travel_Ra	Research	11	4	Life Scien	1	23	Female	2	Research	Married	21480	3	Y	11	8	0	21	
42	No	Travel_Ra	Research	4	4	Life Scien	1	24	Male	1	Manufactu	Married	89260	1	Y	14	8	0	NA	

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

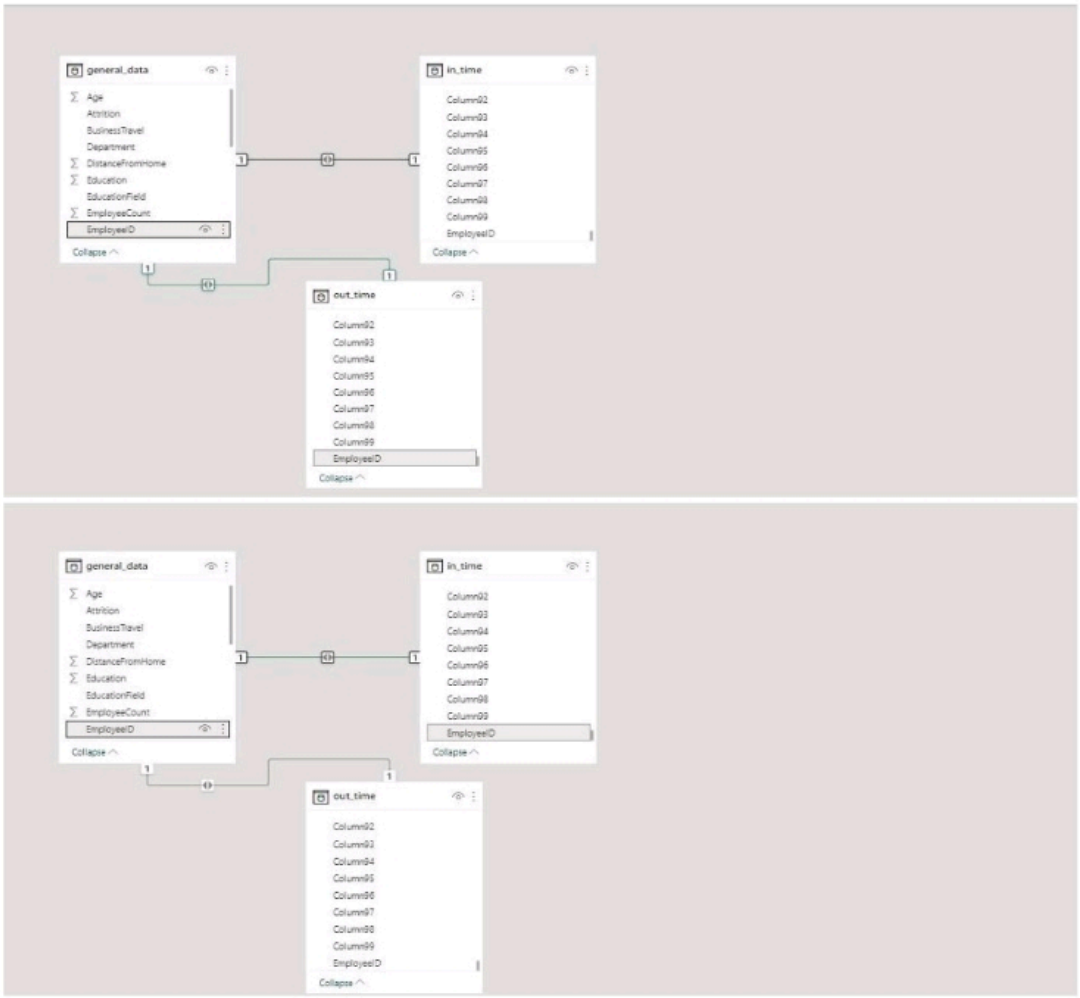


Age	Count of Age
18	24
19	27
20	33
21	39
22	48
23	42
24	78
25	78
26	117
27	144
28	144
29	204
30	180
31	207
32	183
33	174
34	231
35	234
36	207
37	150
38	174
39	126
40	171

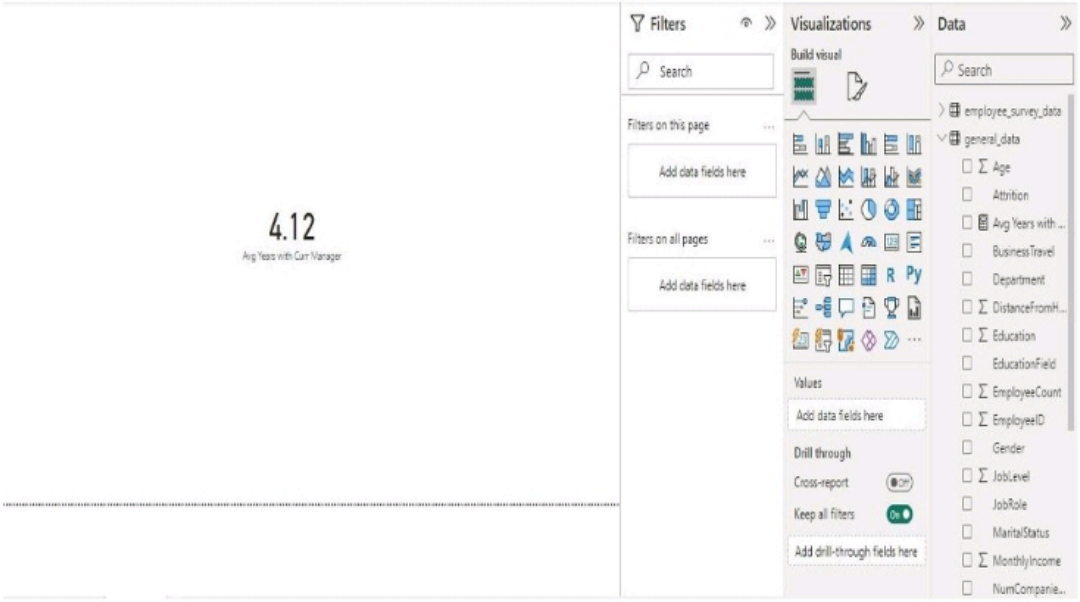
5. Identify and clean any missing or inconsistent data in the "Department" column.

No Missing or Inconsistent Data.

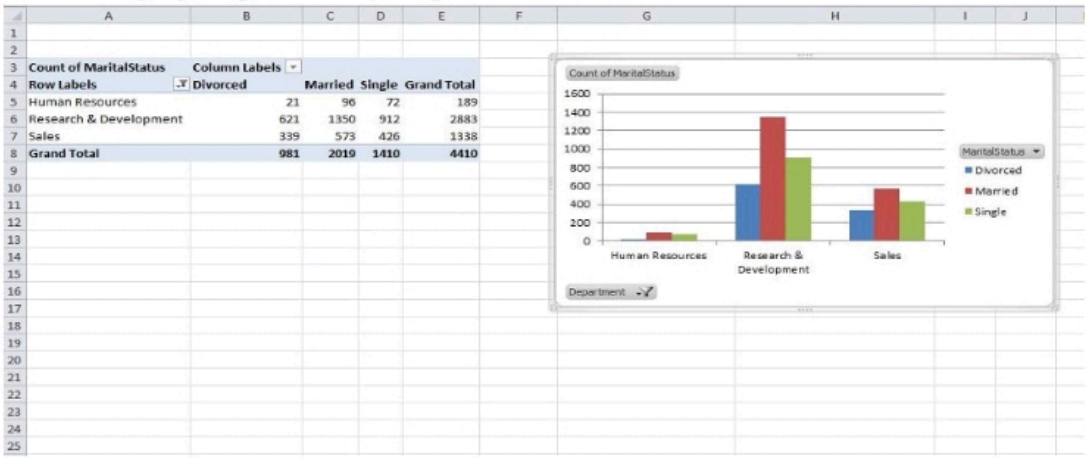
6. In Power BI, establish a relationship between the "EmployeeID" in the employee data and the "EmployeeID" in the time tracking data.



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

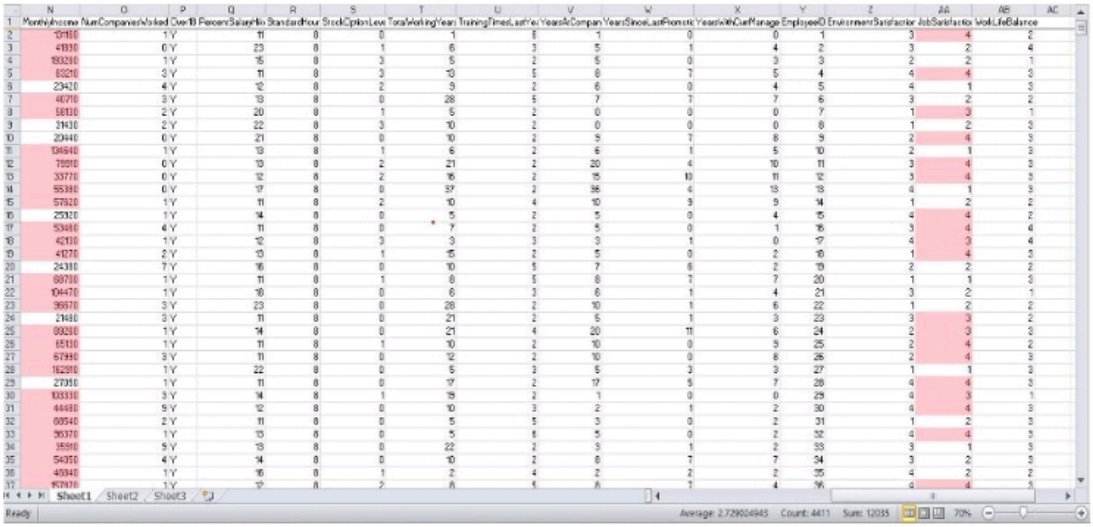


8. Using Excel, create a pivot table that displays the count of employees in each Marital Status category, segmented by Department.

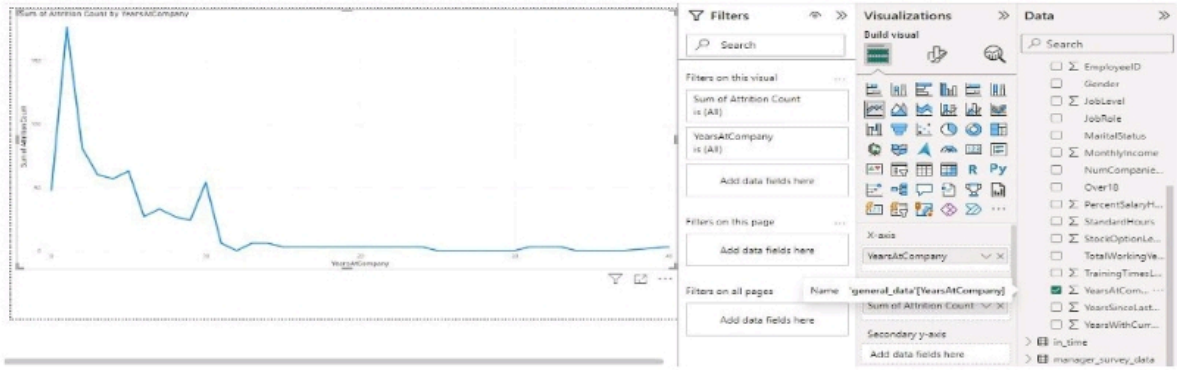


Count of MaritalStatus	Column Labels			
Row Labels	Divorced	Married	Single	Grand Total
Human Resources	21	96	72	189
Research & Development	621	1350	912	2883
Sales	339	573	426	1338
Grand Total	981	2019	1410	4410

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

- **Star Schema Components:**

  - **Fact Table**: This table holds the primary metrics or measures of interest.
    - **Primary Key : EmployeeID**

  - **Dimension Tables**: These tables contain descriptive information related to each dimension.
    - **Employee Details**
    - **Manager**
    - **Time**

Benefits of a Star Schema:

I. **Simplicity and Query Performance**: Star schemas simplify queries by separating measures from descriptive attributes. This structure streamlines querying and improves performance by minimizing the number of joins needed to retrieve information.

II. **Scalability**: It's easier to extend star schemas with new dimensions or measures. Adding new data doesn't require modifying existing tables extensively, making it more scalable as the dataset grows.

III. **Improved Data Analysis** : With a star schema, it's easier to perform complex analyses.

IV. **Easier to Understand and Navigate :** The star schema's structure is intuitive and user-friendly, making it easier for analysts and users to understand relationships between different data elements.

V. **Better for Data Warehousing**: Star schemas are commonly used in data warehousing environments, facilitating efficient data retrieval and analysis in decision-making processes.

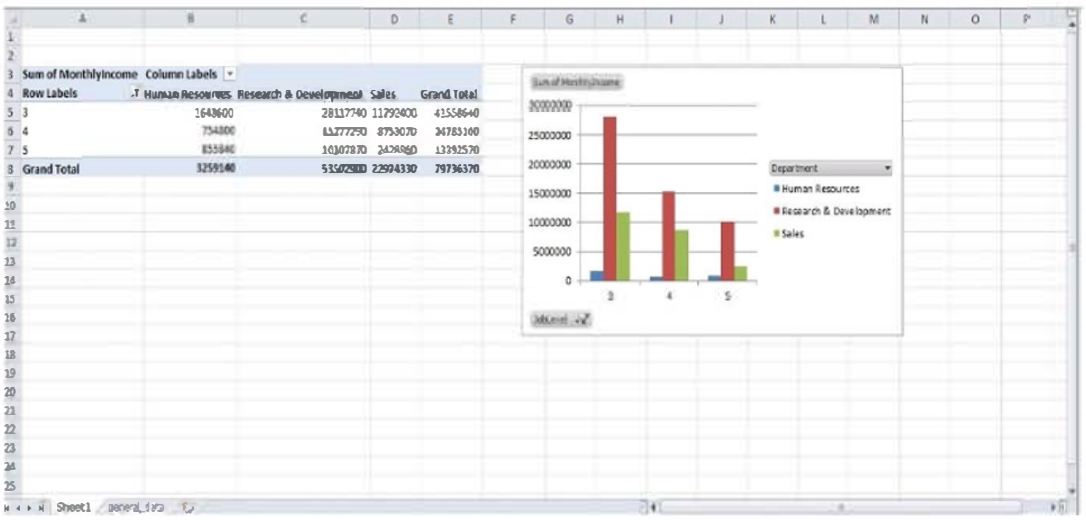
12. Create a hierarchy in Power BI that allows users to drill down from Department to Job Role to further narrow their analysis?

13. How can you set up parameterized queries in Power BI to allow users to filter data based on the Distance from Home column?

- Create a Parameter
- Use the parameters in Queries
- Apply filtering based on Parameters
- Interacting with the parameter

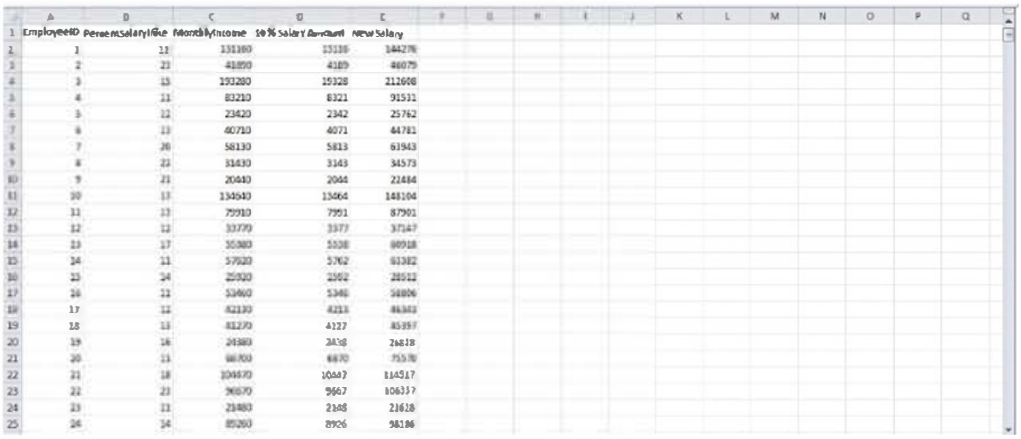
➢ This setup enables users to dynamically filter data based on the distance from the Home column, providing a user-friendly way to interact with the report and customize data analysis based on varying distance criteria.

14. 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	Human Resources	Research & Development	Sales	Grand Total
3	1648600	28117740	11792400	41558640
4	754800	43277250	8753070	34783100
5	855640	10307870	2432860	13592570
Grand Total	3259140	51502860	22924330	79736370

15. 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.



EmployeeID	PercentSalaryHike	MonthlyIncome	10% Salary Amount	New Salary
1	11	131160	13116	144276
2	23	41890	4189	46079
4	15	193280	19328	212608
5	11	83210	8321	91531
6	12	23420	2342	25762
7	13	40710	4071	44781
8	20	58130	5813	63943
9	22	31430	3143	34573
10	21	20440	2044	22484
11	13	134640	13464	148104
12	13	79910	7991	87901
13	12	33770	3377	37147
14	17	55380	5538	60918
15	11	57620	5762	63382
16	14	25920	2592	28512
17	22	53460	5346	58806
18	12	42130	4213	46343
19	13	41270	4127	45397
20	16	24380	2438	26818
21	15	68700	6870	75570
22	18	104670	10467	115137
23	22	96670	9667	106337
24	13	23480	2348	25828
25	14	89260	8926	98186

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

- **Understand the Predefined Schema:** Review the predefined schema or data model to understand the expected structure, data types, relationships, constraints, and other specifications

- **Examine the Data:**
  - Analyze the actual data against the predefined schema.
  - Check column names, data types, null values, unique constraints, relationships between tables (if applicable), and overall data consistency

- **Use Automated Checks:**

- Employ automated tools or scripts to compare the actual data against the schema.
- Utilize validation scripts or functions to check for data type mismatches, missing values, unexpected values, or any inconsistencies.