TASK 1 :- HR DATA ANALYSIS USING POWER BI AND MS EXCEL

NAME : SUPRIYA BHAT V
Data Analysis Project using
Power BI and MS Excel
Email ID: supriyabhatv@gmail.com

Human Resources (HR) Data Analysis is a crucial practice in modern organizations aimed at leveraging data-driven insights to enhance workforce management, employee satisfaction, and overall organizational performance. The analysis involves examining multiple datasets that collectively provide a comprehensive view of various aspects of the workforce. The datasets include general_data, employee_survey_data, manager_survey_data, in_time, and out_time.

1. **General_data:This dataset encompasses fundamental information about employees, such as their identification, department, position, joining date, salary, and other demographic details.**

2. **Employee_survey_data:This dataset captures the sentiment and feedback of employees through surveys. It includes valuable information on job satisfaction, workplace conditions, and suggestions for improvement.**

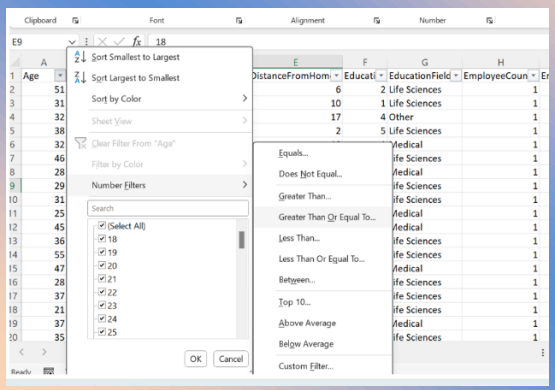
3. **Manager_survey_data:Similar to the employee survey data, this dataset focuses on managers or supervisors. It provides a perspective on leadership effectiveness, team satisfaction, and communication within the managerial hierarchy**.

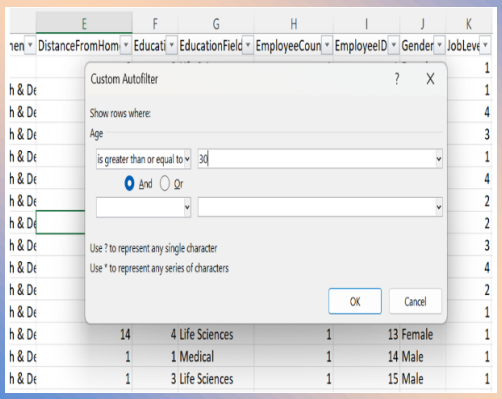
4. **In_time and out_time: The in_time and out_time datasets document the start and end times of employees' workdays. This temporal data is crucial for assessing punctuality, work hours, and potential trends in employee attendance.**

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

   - Select the column containing the ages (e.g., the "Age" column).
   - Go to the "Data" tab in the Excel ribbon.
   - Click on the "Filter" button. This will add filter dropdowns to the selected column headers.
   - Click on the filter dropdown arrow in the "Age" column.
   - Choose "Number Filters" from the menu.
   - From the sub-menu, select "Greater Than or Equal To."

   - In the dialog box that appears, enter "30" as the criteria.
   - Click "OK" to apply the filter.





Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeID	Gender
51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	1	Female
31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1	2	Female
32	No	Travel_Frequently	Research & Development	17	4	Other	1	3	Male
38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	4	Male
32	No	Travel_Rarely	Research & Development	10	1	Medical	1	5	Male
46	No	Travel_Rarely	Research & Development	8	3	Life Sciences	1	6	Female
31	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	9	Male
45	No	Travel_Rarely	Research & Development	17	2	Medical	1	11	Male
36	No	Travel_Rarely	Research & Development	28	1	Life Sciences	1	12	Male
55	No	Travel_Rarely	Research & Development	14	4	Life Sciences	1	13	Female
47	Yes	Non-Travel	Research & Development	1	1	Medical	1	14	Male
37	No	Travel_Rarely	Research & Development	1	3	Life Sciences	1	16	Male
37	No	Non-Travel	Research & Development	1	3	Medical	1	18	Male
35	No	Travel_Rarely	Sales	7	4	Life Sciences	1	19	Male
38	No	Travel_Rarely	Research & Development	8	3	Life Sciences	1	20	Female
50	No	Travel_Rarely	Sales	8	4	Life Sciences	1	22	Male
53	No	Travel_Rarely	Research & Development	11	4	Life Sciences	1	23	Female
42	No	Travel_Rarely	Research & Development	4	4	Life Sciences	1	24	Male

2. 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**

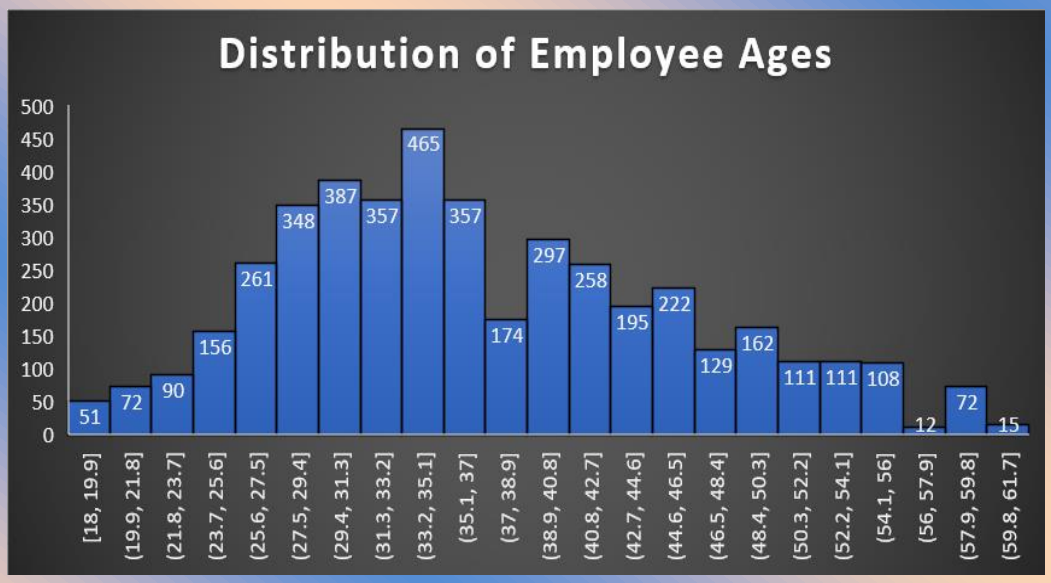
3. Apply conditional formatting to highlight employees with Monthly Income above the company's average income.

Average Monthly Income = 65029.31

The values which are formated to red colour represents the values that are greater than monthly average income of employees.

EmployeeID	JobRole	MonthlyIncome
1	Healthcare Representative	131160
2	Research Scientist	41890
3	Sales Executive	193280
4	Human Resources	83210
5	Sales Executive	23420
6	Research Director	40710
7	Sales Executive	58130
8	Sales Executive	31430
9	Laboratory Technician	20440
10	Laboratory Technician	134640
11	Laboratory Technician	79910
12	Laboratory Technician	33770
13	Sales Executive	55380
14	Research Scientist	57620
15	Manufacturing Director	25920
16	Healthcare Representative	53460
17	Laboratory Technician	42130
18	Sales Executive	41270
19	Sales Representative	24380

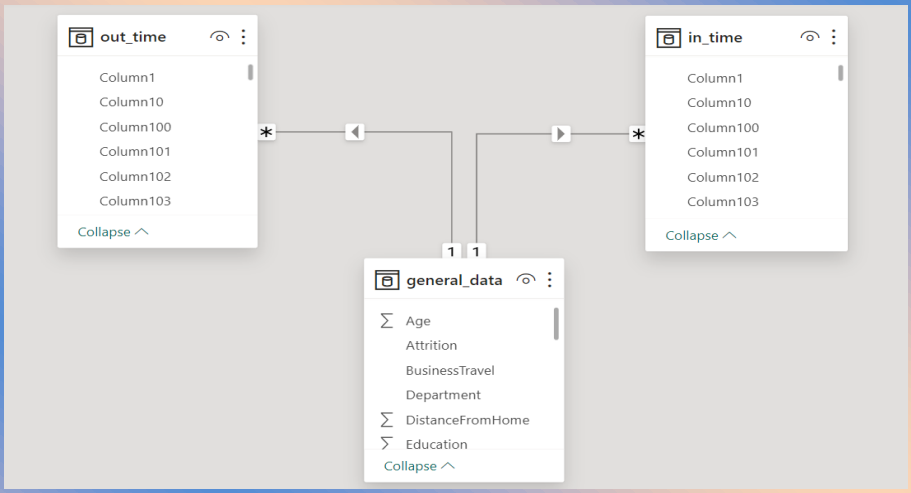
4. **Create a bar chart in Excel to visualize the distribution of employee ages.**



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

There are 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.

Formula Used :

Average years with current manager =
CALCULATE(AVERAGEX(FILTER(general_data,general_data[YearsWithCurrManager]>0), general_data[YearsWithCurrManager]), ALLEXCEPT(general_data,general_data[EmployeeID]))

= 5.02

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

Count of EmployeeID	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**

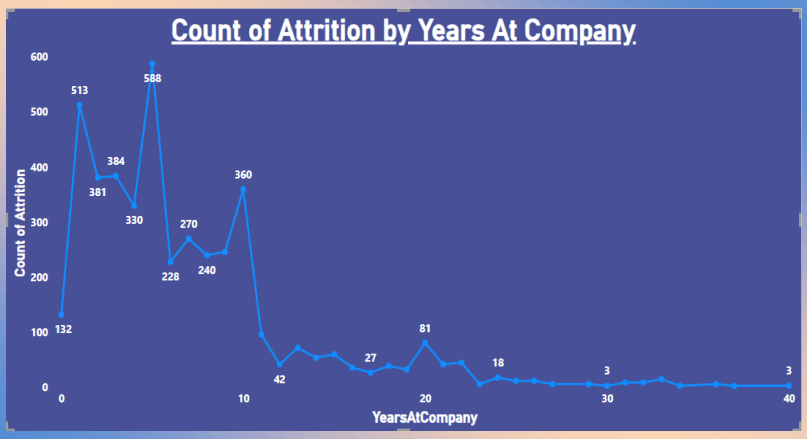
9. Apply conditional formatting to highlight employees with both above-average Monthly Income and above-average Job Satisfaction.

Formula Used:

=AND($B2>AVERAGE($B$2:$B$4411), $C2>AVERAGE($C$2:$C$4411))

EmployeeID	MonthlyIncome	JobSatisfaction
1	131160	4
2	41890	2
3	193280	2
4	83210	4
5	23420	1
6	40710	2
7	58130	3
8	31430	2
9	20440	4
10	134640	1
11	79910	4
12	33770	4
13	55380	1
14	57620	2
15	25920	4
16	53460	4
17	42130	3
18	41270	4
19	24380	2
20	68700	1
21	104470	2
22	96670	2
23	21480	3
24	89260	3
25	65130	4
26	67990	4
27	162910	1
28	27050	4
29	103330	3
30	44480	4
31	68540	2

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.**

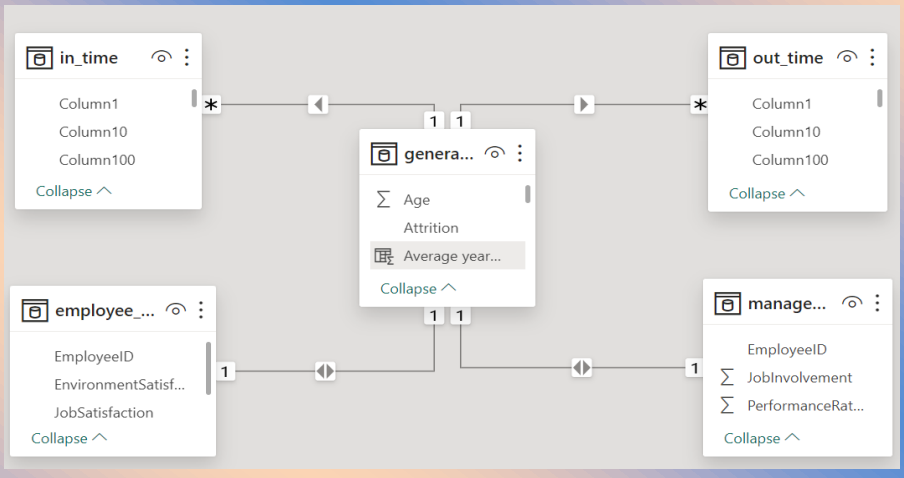
Consider general_data table as a fact table and rest other tables as a Dimension tables, by connecting employee id column of fact table with all the dimension table of primary key(i.e,employee id) .

Benefits

1. **Data Consistency:**

By connecting on the common attribute (EmployeeID), you maintain data consistency across tables, preventing inconsistencies in employee-related information.

2. **Enhanced Analysis:**
Users can now easily analyze the general data along with employee details, manager information, and department specifics without complex joins.
3. **Simplified Queries:**
Queries involving metrics and related employee, manager, or department details become simpler and more intuitive.
4. **Performance Improvement:**
Joining on primary keys typically results in better query performance, making data retrieval more efficient.



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

From Given Data Source ,
Employees having only one Monthly income value . So it's not possible to calculate the Rolling 3 months average for employees.

Incase, IF we are given with the past 3 months monthly income then average of 3 months can be calculated as follows :

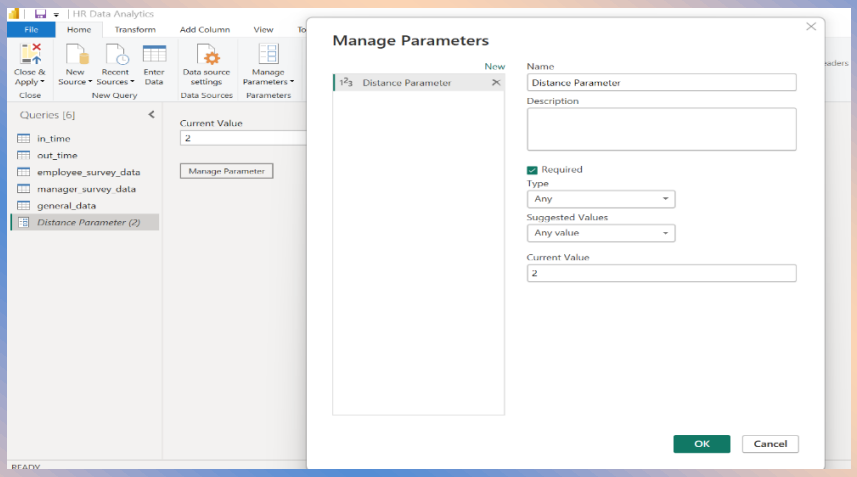
```
RollingAverage =
VAR EmployeeID = 'general_data'[EmployeeID]
VAR CurrentIndex = 'general_data'[Index]

RETURN
        CALCULATE(
            AVERAGE('general_data'[MonthlyIncome]),
            FILTER(
                ALL('general_data'),
                'general_data'[EmployeeID] = EmployeeID &&
                'general_data'[Index] >= CurrentIndex - 2 &&
                'general_data'[Index] <= CurrentIndex
        )
    )
```

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

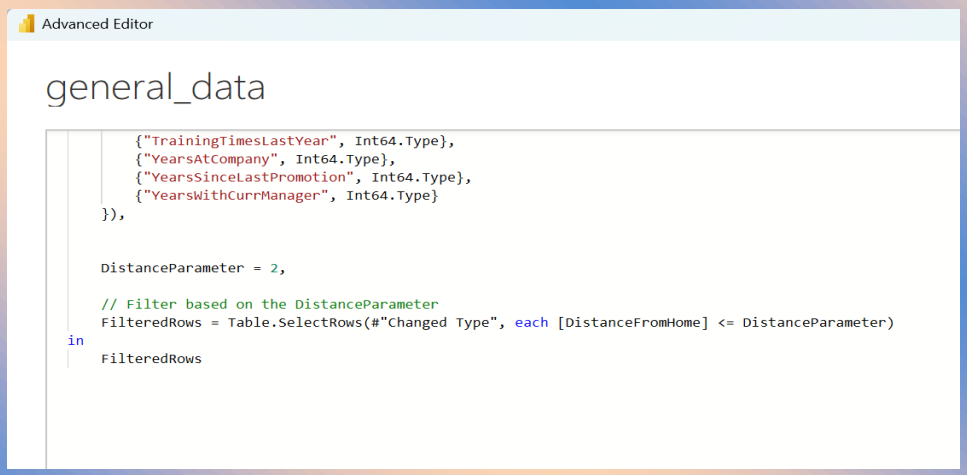
Department	Count of EmployeeCount	Average of MonthlyIncome
Human Resources	**189**	**57904.44**
Healthcare Representative	9	92360.00
Human Resources	3	20140.00
Laboratory Technician	39	77731.54
Manager	9	87603.33
Manufacturing Director	24	55690.00
Research Director	3	74460.00
Research Scientist	36	52776.67
Sales Executive	54	45062.22
Sales Representative	12	28255.00
Research & Development	**2883**	**67187.96**
Healthcare Representative	261	59380.57
Human Resources	108	64283.89
Laboratory Technician	492	67488.23
Manager	207	67902.46
Manufacturing Director	285	71635.37
Research Director	156	67864.04
Research Scientist	564	68052.45
Sales Executive	630	67639.90
Sales Representative	180	66690.50
Sales	**1338**	**61384.48**
Healthcare Representative	123	62089.76
Human Resources	45	47273.33
Laboratory Technician	246	62155.61
Manager	90	50610.00
Manufacturing Director	126	66208.57
Research Director	81	60535.56
Research Scientist	276	60279.57
Sales Executive	294	63626.12
Sales Representative	57	69017.89
Total	**4410**	**65029.31**

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

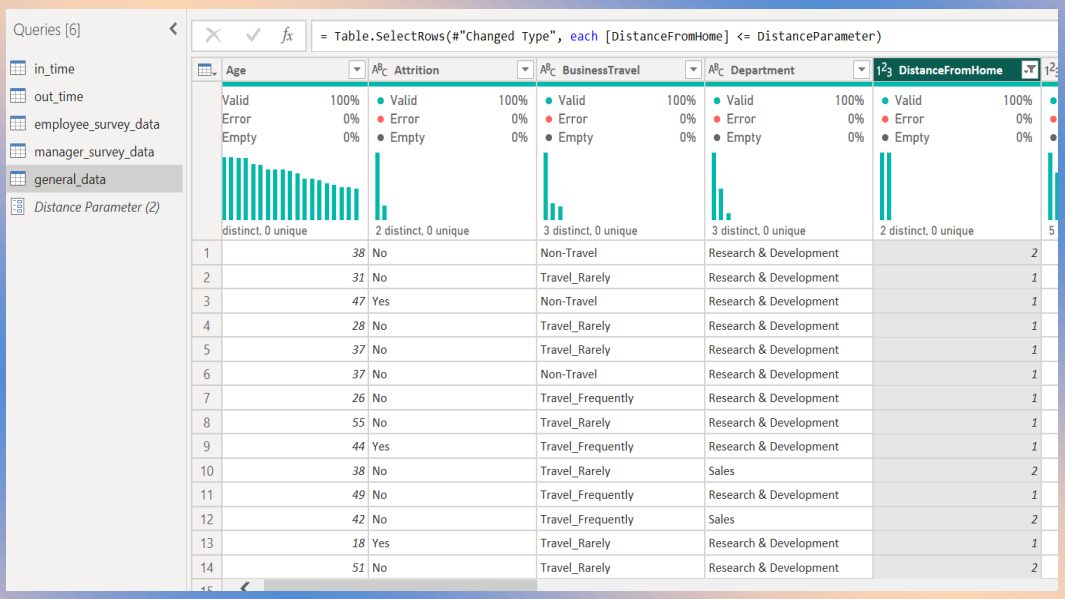
Firstly ,we have to set the parameters under manage parameters section



Then, in advanced editor we need to create a filter using the parameter.



This will dynamically filter the data based on that 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.

JobLevel	(Multiple Items)
Row Labels	**Sum of MonthlyIncome**
Human Resources	3259140
Research & Development	53502900
Sales	22974330
Grand Total	**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.

To understand the impact of a 10% increase in the Percent Salary Hike on monthly income, follow these steps in Excel:

1. **Prepare your data:**
   - **Ensure your table includes columns for:**
   - **Employee Name: To distnguish individuals.**
   - **Current Salary: Monthly salary before the hike.**
   - **Percent Salary Hike: Percentage increase applied to the base salary**
   - **(currently set at 0%).**

2. **Calculate New Salary:**
   - **Insert a new column named "New Salary."**
   - **enter the formula: =Current Salary * 1+Percent Salary Hike/100.**

3. **Calculate Increased Income:**
   - **Insert another column named "Increased Income."**
   - **formula: =New Salary - Current Salary.**

4. **Explore Scenarios:**
   - **Change the value in the "Percent Salary Hike" cell to analyze different scenarios (e.g., 5%, 10%,15%).**
   - **Use Goal Seek to find the salary hike needed to achieve a desired "Increased Income" for a specific employee.**

By following these steps, we can effectively perform a What-If analysis in Excel to understand the impact of a 10% increase (or any other percentage) in Percent Salary Hike on monthly income for all the employees. This versatile tool allows to explore different scenarios and gain valuable insights for informed decision-making.

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

Yes, the data is completely adhered to a predefined schema.

Actions to Take if Inconsistencies are Found:

1. **Data Cleaning:**
   Correct inconsistent or incorrect data. This may involve cleaning and standardizing values, removing duplicates, or filling missing values.

2. **Alert Stakeholders:**
   Notify relevant stakeholders, such as data owners, analysts, or decision-makers, about the inconsistencies. Provide details on the issues discovered and the actions taken to address them.

3. **Update Documentation:**
   Revise and update documentation to reflect any changes made to the schema or data cleaning procedures. This ensures that future users are aware of the data modifications.

4. **Implement Data Validation Rules:**
   Establish and enforce data validation rules within your database or data processing pipeline to prevent similar inconsistencies in the future. This could involve setting constraints, triggers, or using validation scripts.

5. **Audit and Monitor:**
   Implement regular data audits to continuously monitor adherence to the schema. Set up alerts or notifications for any deviations from the expected data structure.