



DATA ANALYSIS INTERNSHIP

Task 1: HR DATA ANALYSIS ASSESSMENT

BY

Manav Malhotra

Email id :-

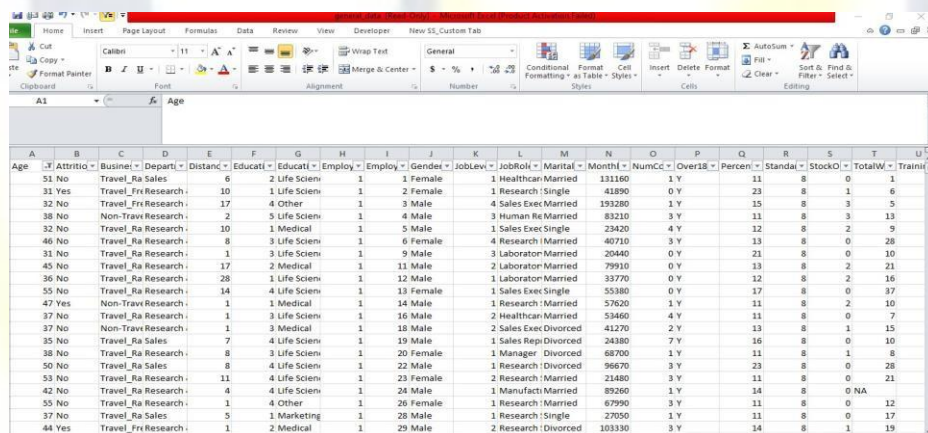
manavmalhotra173@gmail.com

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

Get the general_data.csv into Excel.

Go to the Home tab > Sort & Filter > Filter. Before applying filter make sure to choose

1. right column or CTRL+ SHIFT + L for applying filter
2. 2. Select Age column and click on drop down and select the Number filter > Greater than > Open Custom AutoFilter mention the age value click on OK.



Age	Attriti	Busine	Depart	Distanc	Educat	Educat	Employ	Employ	Gender	JobLev	JobRol	Marital	Monthl	NumC	Over18	Perce	Stand	Stock	TotalW	Train
51	No	Travel_Ra	Sales	6	2	Life Scien	1	1	Female	1	Healthcan	Married	131160	1	Y	11	8	0	1	
31	Yes	Travel_Fri	Research	10	1	Life Scien	1	2	Female	1	Research	Single	41890	0	Y	23	8	1	6	
32	No	Travel_Fri	Research	17	4	Other	1	3	Male	4	Sales Exec	Married	193280	1	Y	15	8	3	5	
38	No	Non-Travi	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	
31	No	Travel_Ra	Research	1	3	Life Scien	1	9	Male	3	Laborator	Married	20440	0	Y	21	8	0	10	
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-Travi	Research	1	1	Medical	1	14	Male	1	Research	Married	57620	1	Y	11	8	2	10	
37	No	Travel_Ra	Research	1	3	Life Scien	1	16	Male	2	Healthcan	Married	53460	4	Y	11	8	0	7	
37	No	Non-Travi	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	
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	Manufach	Married	89260	1	Y	14	8	0	NA	
55	No	Travel_Ra	Research	1	4	Other	1	26	Female	1	Research	Married	67990	3	Y	11	8	0	12	
37	No	Travel_Ra	Sales	5	1	Marketing	1	28	Male	1	Research	Single	27050	1	Y	11	8	0	17	
44	Yes	Travel_Fri	Research	1	2	Medical	1	29	Male	2	Research	Divorced	103330	3	Y	14	8	1	19	

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

2		
3	Job Role	Average of Monthly Income
4	Healthcare Representative	60983.74
5	Human Resources	58528.08
6	Laboratory Technician	66314.05
7	Manager	63395.88
8	Manufacturing Director	69183.72
9	Research Director	65473.13
10	Research Scientist	64975.68
11	Sales Executive	65186.69
12	Sales Representative	65370.96
14	Grand Total	65029.31

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

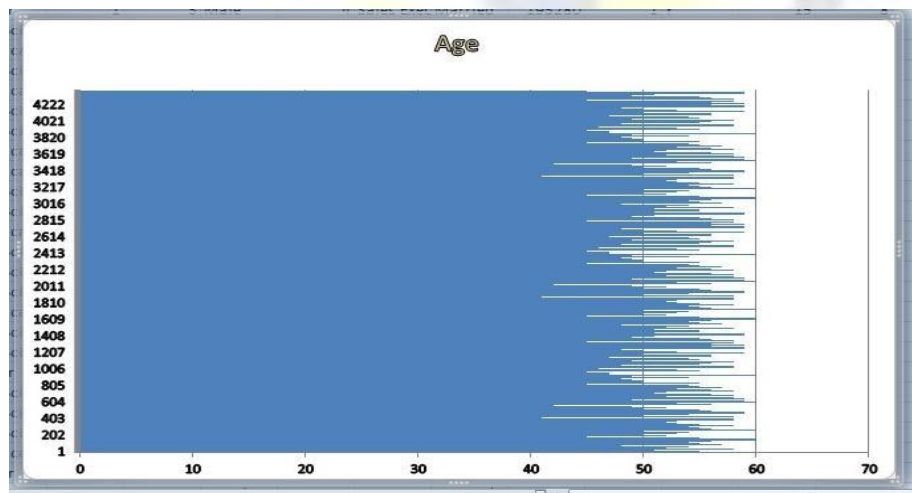
Employee	Gender	JobLevel	JobRole	MaritalSts	MonthlyIncome	NumCom	Over18	PercentSa	Standard	StockOpti	TotalWor	TrainingT1	YearsAtCc	YearsSino	YearsWithCurrManager	Average Monthly Income
1	Female	1	Healthcar	Married	131160	1	Y	11	8	0	1	6	1	0	0	
2	Female	1	Research	Single	41890	0	Y	23	8	1	6	3	5	1	4	
3	Male	4	Sales Exec	Married	193260	1	Y	15	8	3	5	2	5	0	3	
4	Male	4	Human Re	Married	83210	3	Y	11	8	3	13	5	8	7	5	
5	Male	5	Sales Exec	Single	23420	4	Y	12	8	2	9	2	6	0	4	
6	Female	4	Research	Married	40710	3	Y	13	8	0	28	5	7	7	7	
7	Male	7	Sales Exec	Single	58130	2	Y	20	8	1	5	2	0	0	0	
8	Male	2	Sales Exec	Married	31430	2	Y	22	8	3	10	2	0	0	0	
9	Male	3	Laborator	Married	20440	0	Y	21	8	0	10	2	9	7	8	
10	Female	4	Laborator	Divorced	134640	1	Y	13	8	1	6	2	6	1	5	
11	Male	2	Laborator	Married	79910	0	Y	13	8	2	21	2	20	4	10	
12	Male	1	Laborator	Married	33770	0	Y	12	8	2	16	2	15	10	11	
13	Female	1	Sales Exec	Single	55380	0	Y	17	8	0	37	2	36	4	13	
14	Male	1	Research	Married	57620	1	Y	11	8	2	10	4	10	9	9	
15	Male	1	Manufac	Married	25920	1	Y	14	8	0	5	2	5	0	4	
16	Male	2	Healthcar	Married	53460	4	Y	11	8	0	7	2	5	0	1	
17	Male	1	Laborator	Single	42130	1	Y	12	8	3	3	3	3	1	0	
18	Male	2	Sales Exec	Divorced	41270	2	Y	13	8	1	15	2	5	0	2	
19	Male	1	Sales Rep	Divorced	24380	7	Y	16	8	0	10	5	7	6	2	
20	Female	1	Manager	Divorced	68700	1	Y	11	8	1	8	5	8	7	7	
21	Male	2	Laborator	Divorced	104470	1	Y	18	8	0	6	3	6	1	4	

Select Monthly Income column and from Home tab > Conditional formatting > Top/Bottom Rules > Above Average.

For Calculating the Average Monthly income – =Average(range).Observation:

We came to know that there are 1479 employees having monthly income greater than company's average income is 65029.

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



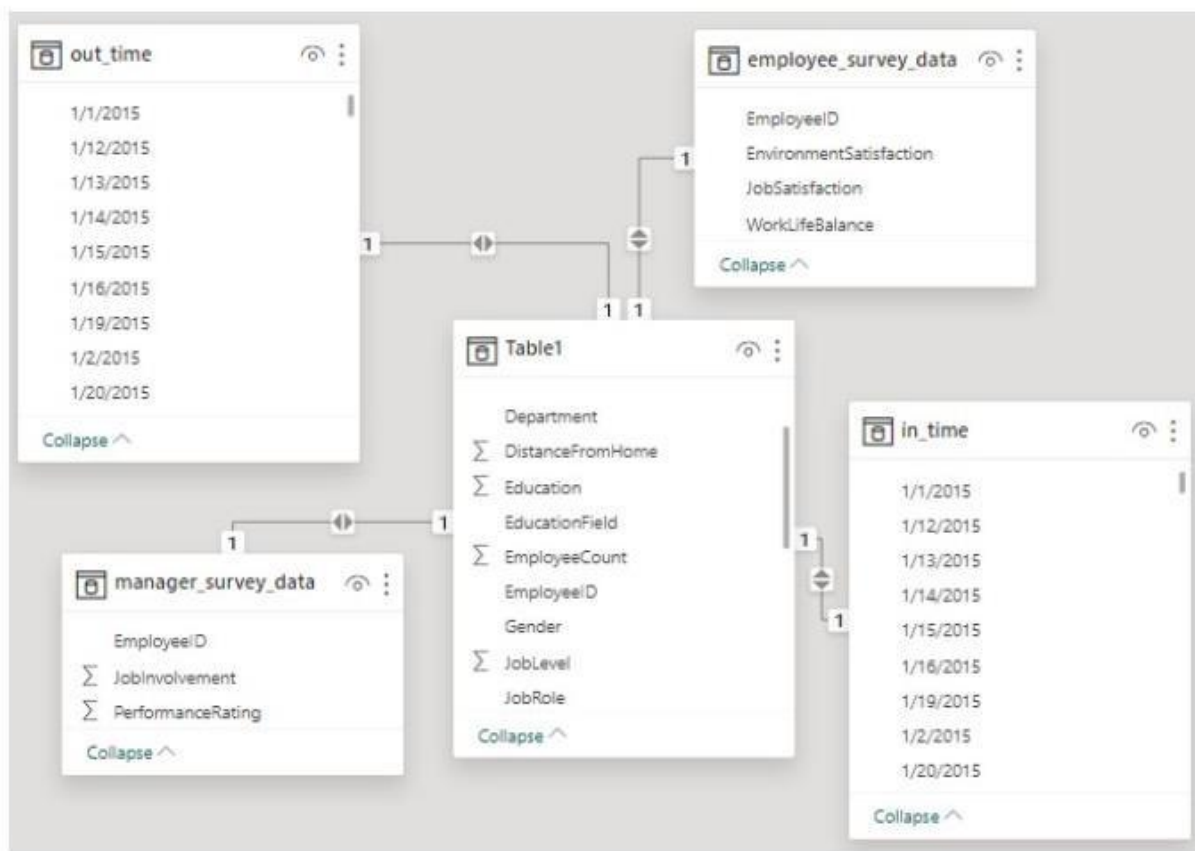
5. Identify and clean any missing or inconsistent data in the "Department" column.

While analyzing the data, there are neither missing values nor inconsistent data in the department column.

If we wanted to know about the missing values from a particular column. You can add filter and select the blank values OR you can use COUNTBLANK function-

=COUNTBLANK(given_range) OR from Home tab > conditional formatting > New rules > format only cells that contain > from format only cells with : select cell value dropdown select - blank dropdown.

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.

CurrManager	EmployeeID	EnvironmentSatisfaction	JobSatisfaction	WorkLifeBalance	JobInvolvement	PerformanceRating	AttritionCount	AgeGroup	AvgYearsWithCurrManager
0	106	4	3	3	1	4	1	26-35	4.12312925170068
0	113	3	2	3	3	3	1	18-25	4.12312925170068
0	124	3	2	3	3	3	0	26-35	4.12312925170068
0	159	4	4	3	3	3	0	36-45	4.12312925170068
0	219	4	2	3	1	3	0	18-25	4.12312925170068
0	244	3	3	3	3	3	0	26-35	4.12312925170068
0	258	3	3	3	4	3	0	26-35	4.12312925170068
0	289	3	3	3	3	3	0	26-35	4.12312925170068
0	329	3	2	3	3	3	0	18-25	4.12312925170068
0	332	2	1	3	4	3	1	26-35	4.12312925170068
0	346	2	3	3	3	3	0	36-45	4.12312925170068
0	413	2	3	3	3	3	0	26-35	4.12312925170068
0	426	3	2	3	2	3	1	18-25	4.12312925170068
0	432	1	1	3	3	3	1	18-25	4.12312925170068
0	474	4	1	3	3	3	1	18-25	4.12312925170068
0	506	3	3	3	2	3	0	26-35	4.12312925170068
0	507	4	2	3	4	3	0	26-35	4.12312925170068
0	543	1	2	3	3	3	1	26-35	4.12312925170068
0	593	3	1	3	2	3	0	26-35	4.12312925170068
0	596	3	1	3	3	3	0	18-25	4.12312925170068
0	755	2	3	3	3	3	1	26-35	4.12312925170068

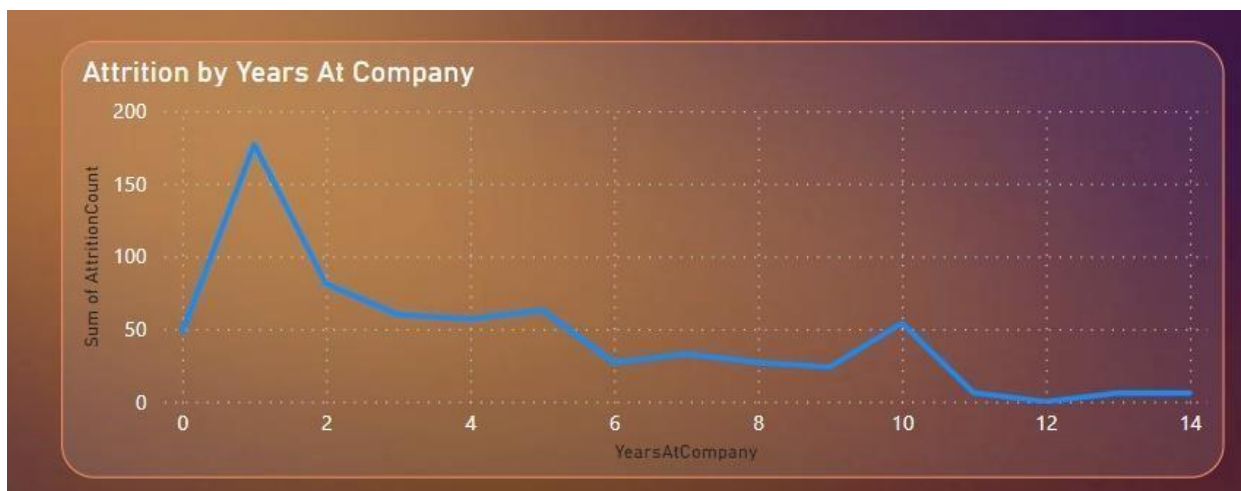
8. Using Excel, create a pivot table that displays the count of employees in each Marital Status category, segmented by Department.

	Marital Status	Count of Department
4	Divorced	981
5	Human Resources	21
6	Research & Development	621
7	Sales	339
8	Married	2019
9	Human Resources	96
10	Research & Development	1350
11	Sales	573
12	Single	1410
13	Human Resources	72
14	Research & Development	912
15	Sales	426
16	Grand Total	4410

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

I	N	Y	Z	AA
Employee	MonthlyIr	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		

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

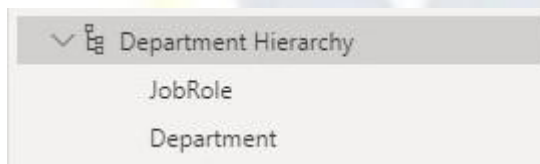


11. Describe how you would create a star schema for this dataset, explaining the benefits of doing so.

The Benefits of Star Schema

- ❖ Star schema are efficient for storing data and updating data by reducing the duplication.
- ❖ It is extremely simple to understand and build.
- ❖ No need for complex joins when querying data.

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 2 of 2 on the Distance from Home column?

- ❖ Select Home > Transform data > Transform data to open the Power Query Editor.
- ❖ In the Power Query Editor, select New Parameters under Manage Parameters in the ribbon.
- ❖ In the Manage Parameters window, fill out the information about the parameter

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.

The screenshot shows an Excel spreadsheet with a PivotTable and the PivotTable Field List task pane. The PivotTable is located in the range A3:M16 and has the following data:

Row Labels	Sum of MonthlyIncome
Human Resources	3259140
3	1648500
4	754800
5	855840
Research & Development	53502900
3	28117740
4	15277290
5	10107870
Sales	22974330
3	11792400
4	8753070
5	2428860
Grand Total	79736370

The PivotTable Field List task pane on the right shows the following settings:

- Choose fields to add to report:
 - ☐ EducationField
 - ☐ EmployeeCount
 - ☐ EmployeeID
 - ☐ Gender
 - ☒ JobLevel
 - ☐ JobRole
 - ☐ MaritalStatus
 - ☒ MonthlyIncome
 - ☐ NumCompaniesWorked
 - ☐ Over 18
- Drag fields between areas below:
 - Report Filter: (empty)
 - Column Labels: (empty)
 - Row Labels: Department
 - Σ Values: Sum of MonthlyIncome

15. Explain how to perform a What-If analysis in Excel to understand the

- Go to the Data Tab
- In the "Data Tools" group, click "What-If Analysis" and choose "Data Table."
- In "Row input cell," enter the reference to the cell with Percent Salary Hike.
- Click OK

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

Steps to Verify Data Adherence to a Predefined Schema:

- Understand Schema: Define expected data types and constraints.

- Use Profiling Tools: Analyse data characteristics with profiling tools.
- Check Data Types: Verify data types match the schema.
- Validate Constraints: Ensure constraints are not violated.
- Identify Missing Values: Check for missing values in mandatory fields.
- Look for Outliers: Investigate unexpected values or outliers.
- Validate Relationships: Confirm valid relationships between tables.

Actions if Inconsistencies are Found:

- Data Cleansing: Correct inaccurate data.
- Communicate: Discuss inconsistencies with stakeholders.
- Document: Document issues and resolutions.
- Implement Validation Rules: Enforce data validation rules.
- Review ETL Processes: Ensure ETL processes align with schema.
- Implement Quality Checks: Integrate automated data quality checks.
- Consider Governance: Evaluate and improve data governance practices.
- Training and Documentation: Train personnel and update documentation

DATA INSIGHTS ➤

Among 4410 employees, there are 2650 males and 1760 females. Also, 3699 active employees with average age comes with 37.

- The average monthly income of the employees is 65030 and there are 711 people left the company.
- Around 453 people from the dept. Research and development left the company, 201 employees from the dept. sales and less amount of 57 from the dept. Human Resources.
- Attrition rate is max for Sales Executives dept
- Sales department employees are mostly satisfied in there work.
- Maximum peoples are working for 1-2 years



THANK YOU

Name:- Manav Malhotra