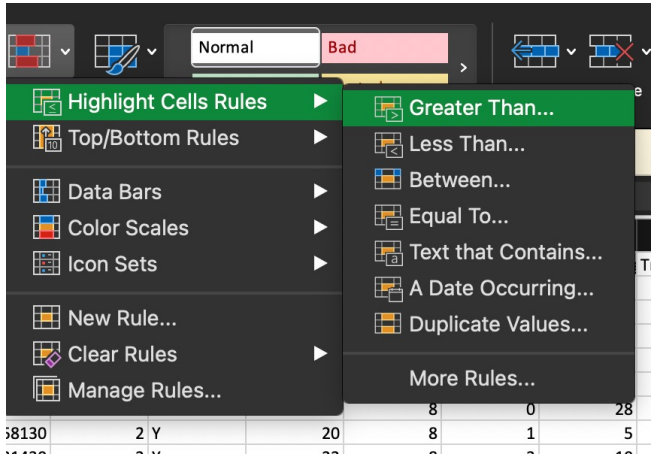
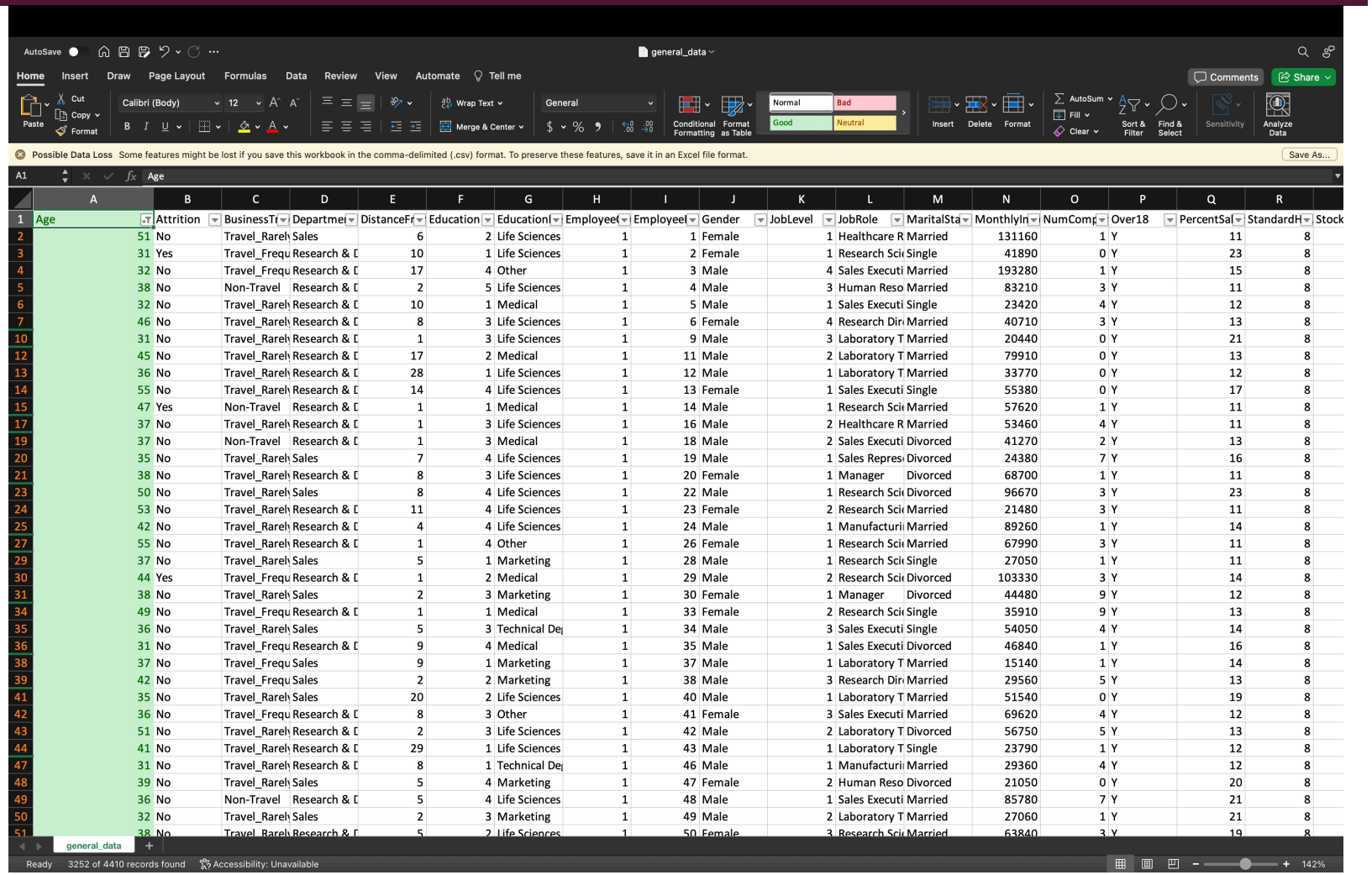
HR DATA ANALYSIS

ADITYA DUBE

1. USING EXCEL, HOW WOULD YOU FILTER THE DATASET TO ONLY SHOW EMPLOYEES AGED 30 AND ABOVE?

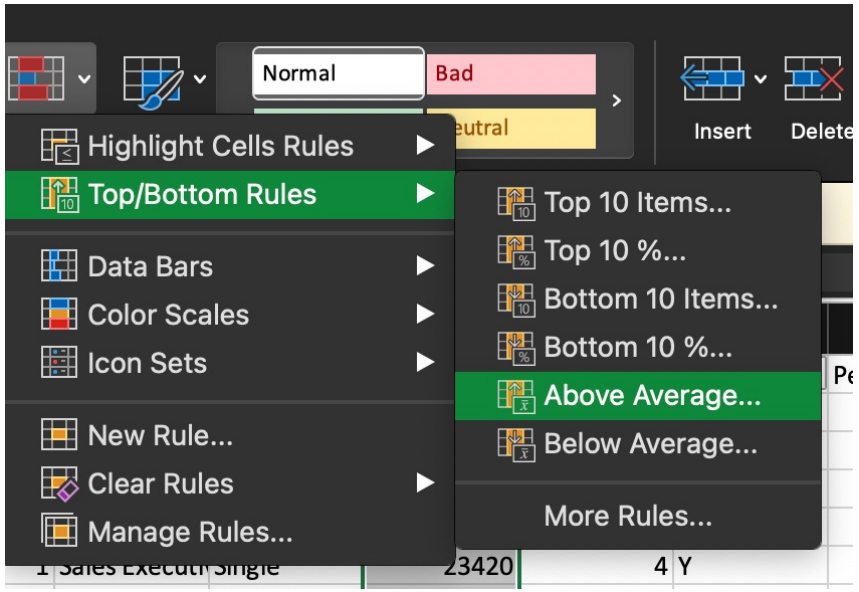


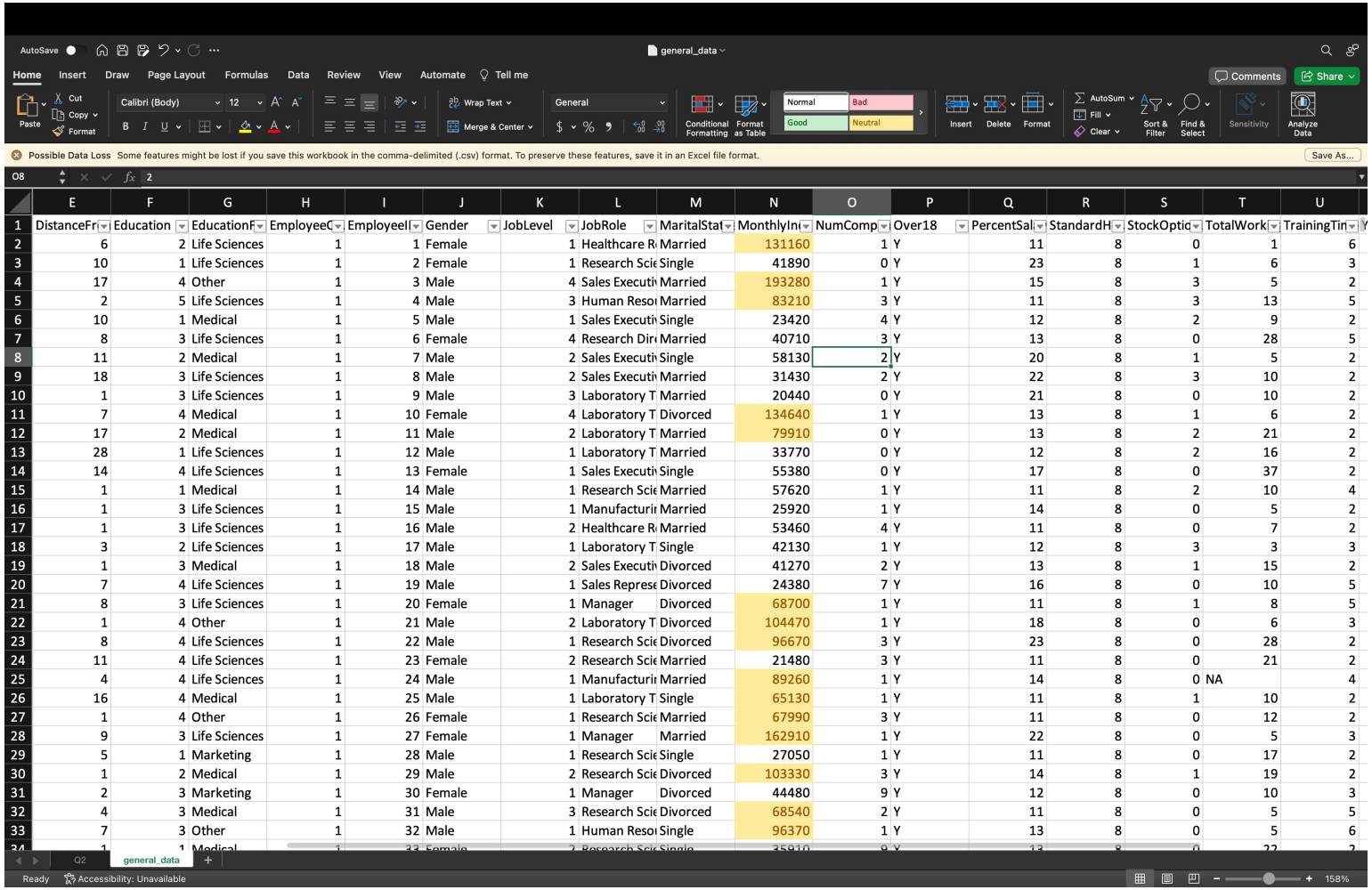


2. CREATE A PIVOT TABLE TO SUMMARIZE THE AVERAGE MONTHLY INCOME BY JOB ROLE.

Job Role	Average of MonthlyIncome
Healthcare Representative	$60,983.74
Human Resources	$58,528.08
Laboratory Technician	$66,314.05
Manager	$63,395.88
Manufacturing Director	$69,183.72
Research Director	$65,473.13
Research Scientist	$64,975.68
Sales Executive	$65,186.69
Sales Representative	$65,370.96
Grand Total	**$65,029.31**

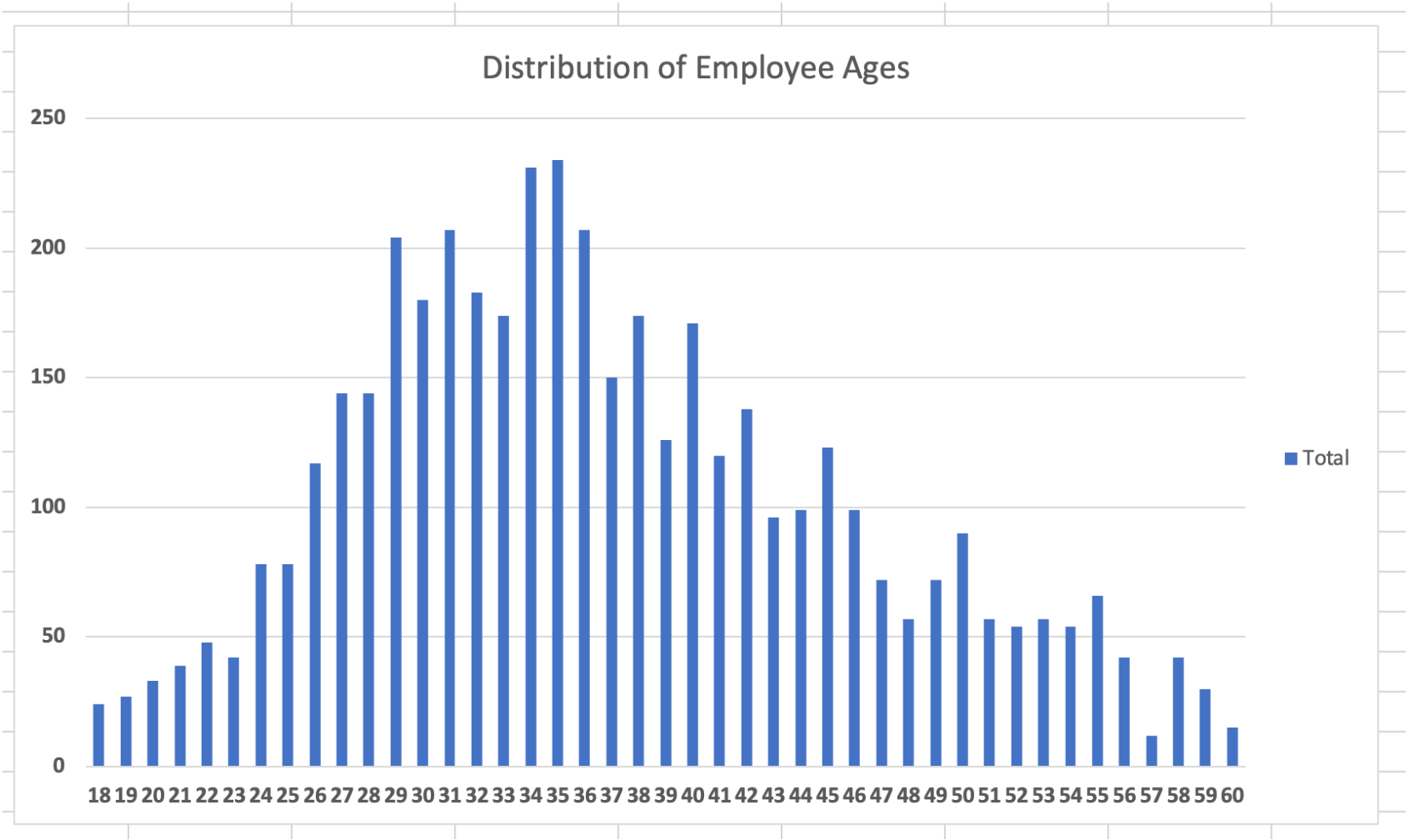
3. APPLY CONDITIONAL FORMATTING TO HIGHLIGHT EMPLOYEES WITH MONTHLY INCOME ABOVE THE COMPANY'S AVERAGE INCOME.



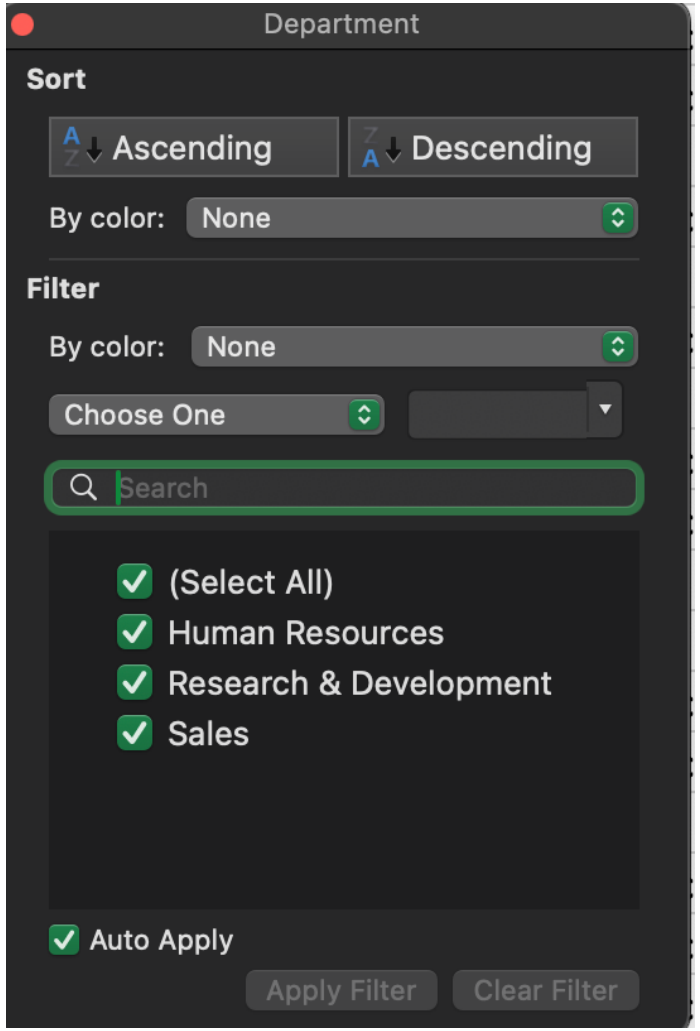


	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	DistanceFr	Education	EducationF	EmployeeC	EmployeeI	Gender	JobLevel	JobRole	MaritalStat	MonthlyIn	NumComp	Over18	PercentSal	StandardH	StockOptio	TotalWork	TrainingTi
2	6	2	Life Sciences	1	1	Female	1	Healthcare R	Married	131160	1	Y	11	8	0	1	6
3	10	1	Life Sciences	1	2	Female	1	Research Scie	Single	41890	0	Y	23	8	1	6	3
4	17	4	Other	1	3	Male	4	Sales Executi	Married	193280	1	Y	15	8	3	5	2
5	2	5	Life Sciences	1	4	Male	3	Human Reso	Married	83210	3	Y	11	8	3	13	5
6	10	1	Medical	1	5	Male	1	Sales Executi	Single	23420	4	Y	12	8	2	9	2
7	8	3	Life Sciences	1	6	Female	4	Research Dir	Married	40710	3	Y	13	8	0	28	5
8	11	2	Medical	1	7	Male	2	Sales Executi	Single	58130	2	Y	20	8	1	5	2
9	18	3	Life Sciences	1	8	Male	2	Sales Executi	Married	31430	2	Y	22	8	3	10	2
10	1	3	Life Sciences	1	9	Male	3	Laboratory T	Married	20440	0	Y	21	8	0	10	2
11	7	4	Medical	1	10	Female	4	Laboratory T	Divorced	134640	1	Y	13	8	1	6	2
12	17	2	Medical	1	11	Male	2	Laboratory T	Married	79910	0	Y	13	8	2	21	2
13	28	1	Life Sciences	1	12	Male	1	Laboratory T	Married	33770	0	Y	12	8	2	16	2
14	14	4	Life Sciences	1	13	Female	1	Sales Executi	Single	55380	0	Y	17	8	0	37	2
15	1	1	Medical	1	14	Male	1	Research Scie	Married	57620	1	Y	11	8	2	10	4
16	1	3	Life Sciences	1	15	Male	1	Manufacturi	Married	25920	1	Y	14	8	0	5	2
17	1	3	Life Sciences	1	16	Male	2	Healthcare R	Married	53460	4	Y	11	8	0	7	2
18	3	2	Life Sciences	1	17	Male	1	Laboratory T	Single	42130	1	Y	12	8	3	3	3
19	1	3	Medical	1	18	Male	2	Sales Executi	Divorced	41270	2	Y	13	8	1	15	2
20	7	4	Life Sciences	1	19	Male	1	Sales Represe	Divorced	24380	7	Y	16	8	0	10	5
21	8	3	Life Sciences	1	20	Female	1	Manager	Divorced	68700	1	Y	11	8	1	8	5
22	1	4	Other	1	21	Male	2	Laboratory T	Divorced	104470	1	Y	18	8	0	6	3
23	8	4	Life Sciences	1	22	Male	1	Research Scie	Divorced	96670	3	Y	23	8	0	28	2
24	11	4	Life Sciences	1	23	Female	2	Research Scie	Married	21480	3	Y	11	8	0	21	2
25	4	4	Life Sciences	1	24	Male	1	Manufacturi	Married	89260	1	Y	14	8	0	NA	4
26	16	4	Medical	1	25	Male	1	Laboratory T	Single	65130	1	Y	11	8	1	10	2
27	1	4	Other	1	26	Female	1	Research Scie	Married	67990	3	Y	11	8	0	12	2
28	9	3	Life Sciences	1	27	Female	1	Manager	Married	162910	1	Y	22	8	0	5	3
29	5	1	Marketing	1	28	Male	1	Research Scie	Single	27050	1	Y	11	8	0	17	2
30	1	2	Medical	1	29	Male	2	Research Scie	Divorced	103330	3	Y	14	8	1	19	2
31	2	3	Marketing	1	30	Female	1	Manager	Divorced	44480	9	Y	12	8	0	10	3
32	4	3	Medical	1	31	Male	3	Research Scie	Divorced	68540	2	Y	11	8	0	5	5
33	7	3	Other	1	32	Male	1	Human Reso	Single	96370	1	Y	13	8	0	5	6

4. CREATE A BAR CHART IN EXCEL TO VISUALIZE THE DISTRIBUTION OF EMPLOYEE AGES.

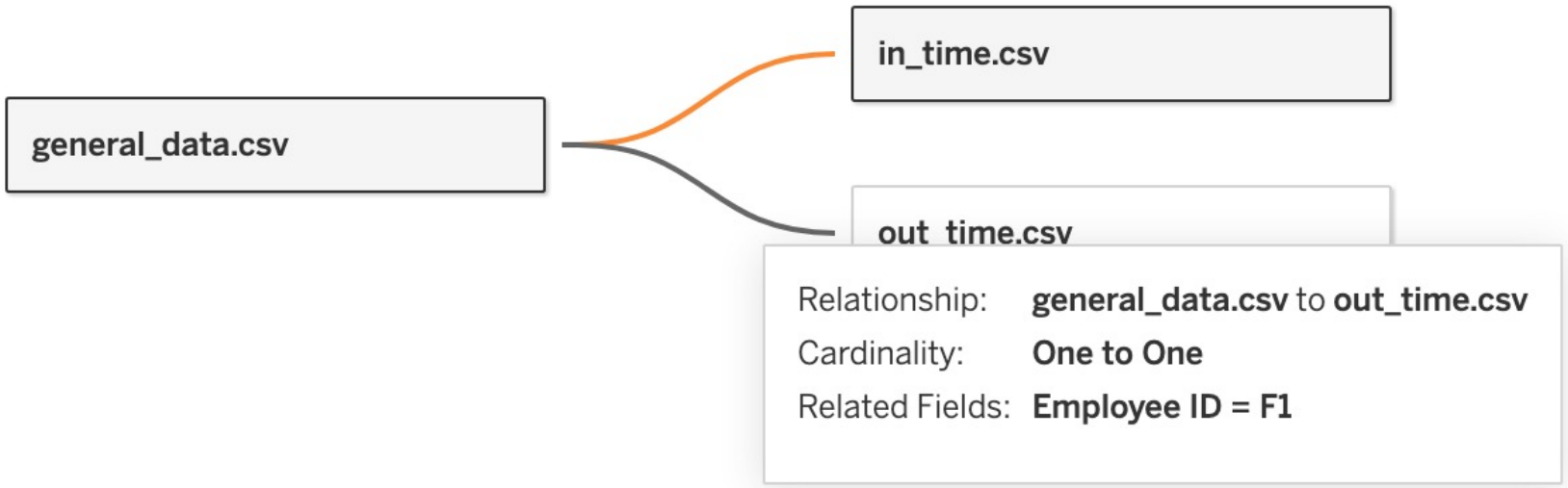


5. IDENTIFY AND CLEAN ANY MISSING OR INCONSISTENT DATA IN THE "DEPARTMENT" COLUMN.



All the data in this column is clean.

6. IN TABLEAU, ESTABLISH A RELATIONSHIP BETWEEN THE "EMPLOYEEID" IN THE EMPLOYEE DATA AND THE "EMPLOYEEID" IN THE TIME TRACKING DATA.



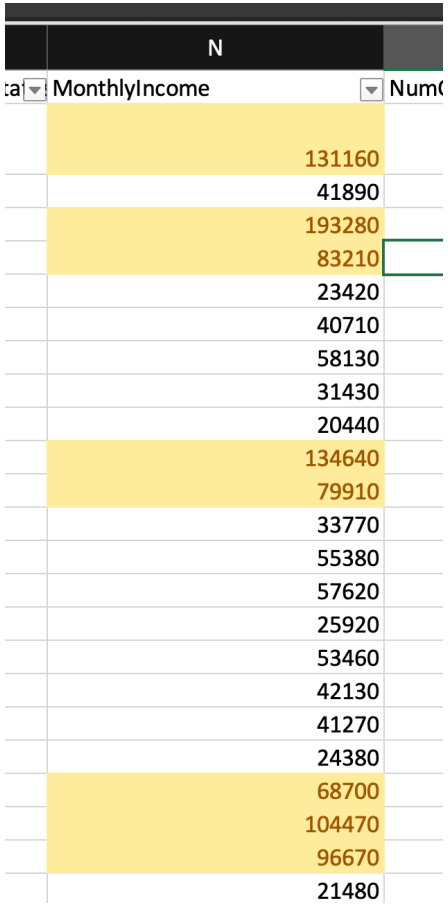
7. CREATE A CALCULATED COLUMN THAT CALCULATES THE AVERAGE YEARS AN EMPLOYEE HAS SPENT WITH THEIR CURRENT MANAGER.

Average Years with Current Manager	4.123129252

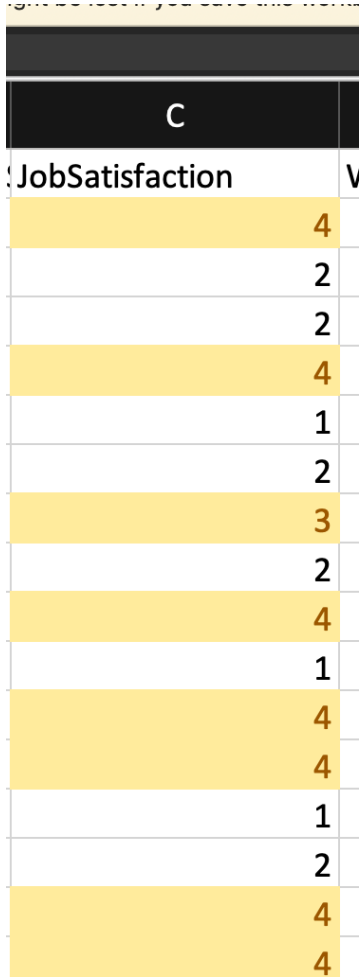
8. USING EXCEL, CREATE A PIVOT TABLE THAT DISPLAYS THE COUNT OF EMPLOYEES IN EACH MARITAL STATUS CATEGORY, SEGMENTED BY DEPARTMENT.

Marital Status	Sum of EmployeeCount
Divorced	**981**
Human Resources	21
Research & Development	621
Sales	339
Married	**2019**
Human Resources	96
Research & Development	1350
Sales	573
Single	**1410**
Human Resources	72
Research & Development	912
Sales	426
Grand Total	**4410**

9. APPLY CONDITIONAL FORMATTING TO HIGHLIGHT EMPLOYEES WITH BOTH ABOVE-AVERAGE MONTHLY INCOME AND ABOVE-AVERAGE JOB SATISFACTION.

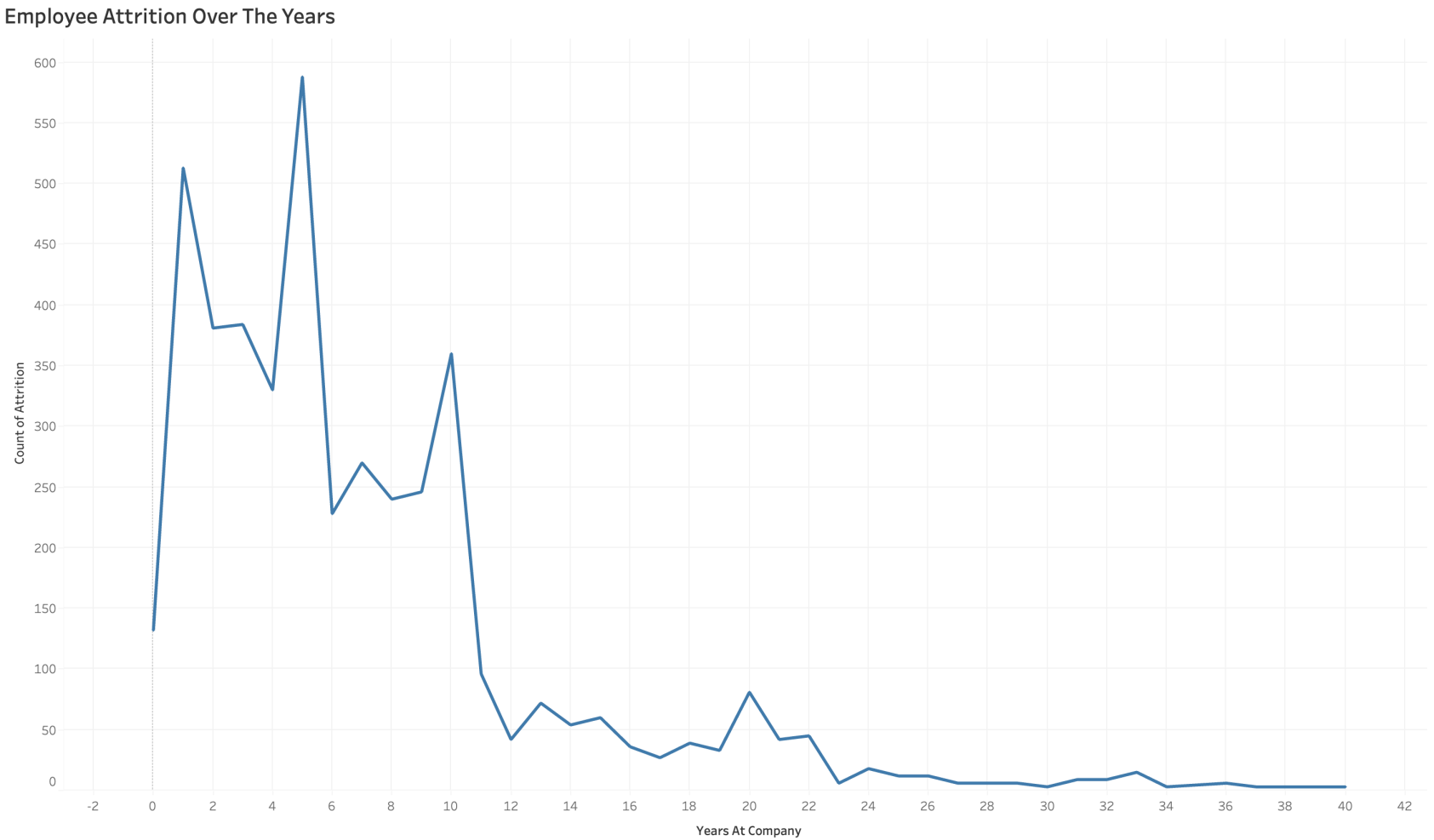


N
MonthlyIncome
131160
41890
193280
83210
23420
40710
58130
31430
20440
134640
79910
33770
55380
57620
25920
53460
42130
41270
24380
68700
104470
96670
21480



C
JobSatisfaction
4
2
2
4
1
2
3
2
4
1
4
4
1
2
4
4

10. IN TABLEAU, 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.

1. Fact Table: Employee Attrition Fact (Employee ID, Year ID, Attrition Count).
2. Dimensions: a. **Employee Dimension**: Employee (Employee ID, Employee Name, Department, Marita Status, etc.). b. **Time Dimension**: Time (Year ID, Year, Quarter, Month, Date-related attributes).
3. Benefits of Star Schema:

- Simplifies structure with clear fact and dimension separation.
- Improves query performance through optimized design.
- Scales efficiently for large datasets.
- Facilitates easy maintenance of dimensions.
- Better support for BI tools like Power BI.
- Ideal for data warehousing environments.

12. CALCULATE THE ROLLING 3-MONTH AVERAGE OF MONTHLY INCOME FOR EACH EMPLOYEE.

Rolling 3-Month Average
13963.33333
64426.66667
27736.66667
7806.666667
13570
19376.66667
10476.66667
6813.333333
44880
26636.66667
11256.66667
18460
19206.66667

13. CREATE A HIERARCHY IN TABLEAU THAT ALLOWS USERS TO DRILL DOWN FROM DEPARTMENT TO JOB ROLE TO FURTHER NARROW THEIR ANALYSIS.

Department	Job Role	
Human Resources	Healthcare Representative	Abc
	Human Resources	Abc
	Laboratory Technician	Abc
	Manager	Abc
	Manufacturing Director	Abc
	Research Director	Abc
	Research Scientist	Abc
	Sales Executive	Abc
	Sales Representative	Abc
Research & Development	Healthcare Representative	Abc
	Human Resources	Abc
	Laboratory Technician	Abc
	Manager	Abc
	Manufacturing Director	Abc
	Research Director	Abc
	Research Scientist	Abc
	Sales Executive	Abc
	Sales Representative	Abc
Sales	Healthcare Representative	Abc
	Human Resources	Abc
	Laboratory Technician	Abc
	Manager	Abc
	Manufacturing Director	Abc
	Research Director	Abc
	Research Scientist	Abc
	Sales Executive	Abc
	Sales Representative	Abc

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 allow users to dynamically filter data based on predefined parameters. To set up a parameter, create one in the Manage Parameters dialog, assign it a data type, and incorporate it into your queries' filtering logic. By creating a slicer visualization linked to the parameter, users can interactively adjust the filter, enabling dynamic analysis of the dataset. This approach enhances flexibility and empowers users to explore and tailor their insights based on specific criteria, such as filtering data on the "Distance from Home" column.

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.

Department and Job Level	Sum of MonthlyIncome
Human Resources	**$3,259,140.00**
3	$1,648,500.00
4	$754,800.00
5	$855,840.00
Research & Development	**$53,502,900.00**
3	$28,117,740.00
4	$15,277,290.00
5	$10,107,870.00
Sales	**$22,974,330.00**
3	$11,792,400.00
4	$8,753,070.00
5	$2,428,860.00
Grand Total	**$79,736,370.00**

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 perform What If analysis in Excel we can do that from the Data tab and choose what if analysis and choosing scenario manager and adding a scenario with the first monthly income cell as an Example and this is the summary of this.

MonthlyIncome	10% Monthly Income
131160	144276
41890	46079
193280	212608
83210	91531
23420	25762
40710	44781
58130	63943
31430	34573
20440	22484
134640	148104
79910	87901
33770	37147

17. VERIFY IF THE DATA ADHERES TO A PREDEFINED SCHEMA. WHAT ACTIONS WOULD YOU TAKE IF YOU FIND INCONSISTENCIES?

- Yes, the data appears as a star schema
- We should identify inconsistencies which they are human errors, system errors, data entry errors, data transfer errors, or data corruption. Some common types are missing values, outliers, duplicates, typos, formatting errors, or inconsistent units or scales.
- Choose an appropriate strategy to handle data problems, depending on the nature and extent of the problem and here we find the EmployeeID cell empty in the in_time and out_time tables so we should complete them to make data consistent and ready for analysis, also we found that there isn't any date columns in the dataset so we should create new one to help us analyze data easily.