SmartBridge Externship

Data Analytics

Ujjawal Srivastava - 19BIT0072 - VIT Vellore

Demo link :-

drive link :

https://drive.google.com/drive/folders/1cgxa1xvBgcTn_uIuEf6zp7hVr9M-DzYT?usp=sharing

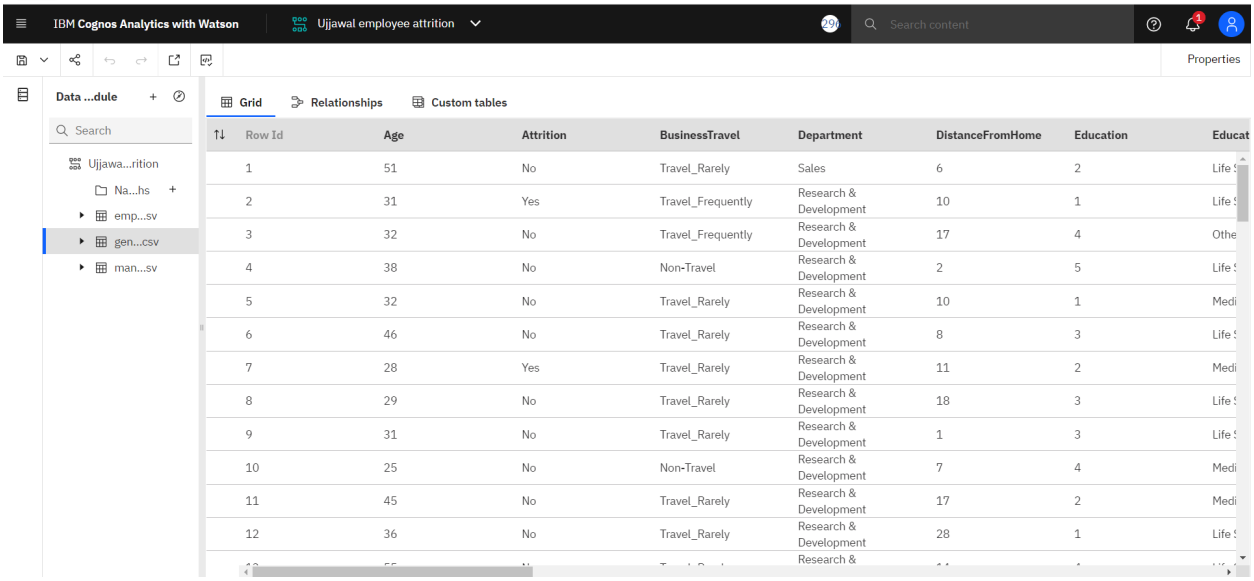
youtube link :

https://youtu.be/keaLm7h-DuI

About dataset :-

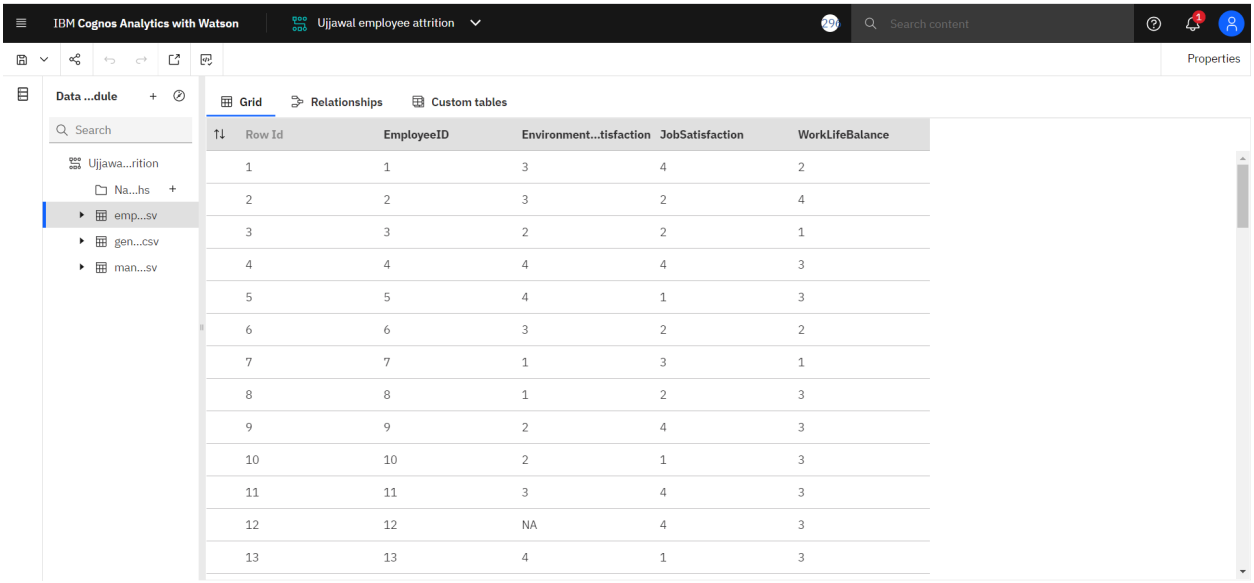
Attriton is a huge concern of every organization . By knowing the proper attrition of employee , an organization can save monay and effort both by not investing in them . As the work dependency and other factors of leaving employees can lead organization to a huge loss. Also, attrition causes distrust among existing employees which can itself become a major difficulty for the management of the organization.
Here using Cognos we prepare our table to perform data exploration to it . As there are 3 different csv files so we merge them by defining relations between them . Our aim to perform the identification of important factors that might be influential in determining which employee might leave the firm and who may not.

General_data.csv :-
There are total of 24 fields in this file that tells us various information about the employee.



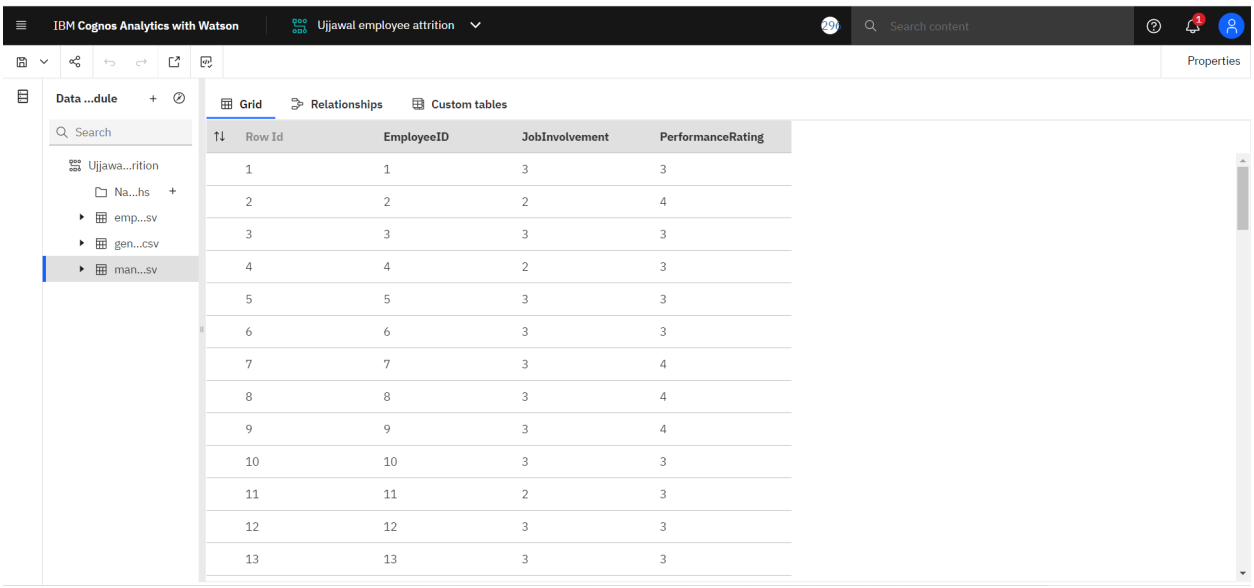
Employee_survey_data.csv :-

1. EmployeeID
2. EnvironmentSatisfaction(1 'Low' 2 'Medium' 3 'High' 4 'Very High')
3. JobSatisfaction(1 'Low' 2 'Medium' 3 'High' 4 'Very High')
4. WorkLifeBalance (1 'Bad', 2 'Good', 3 'Better', 4 'Best')



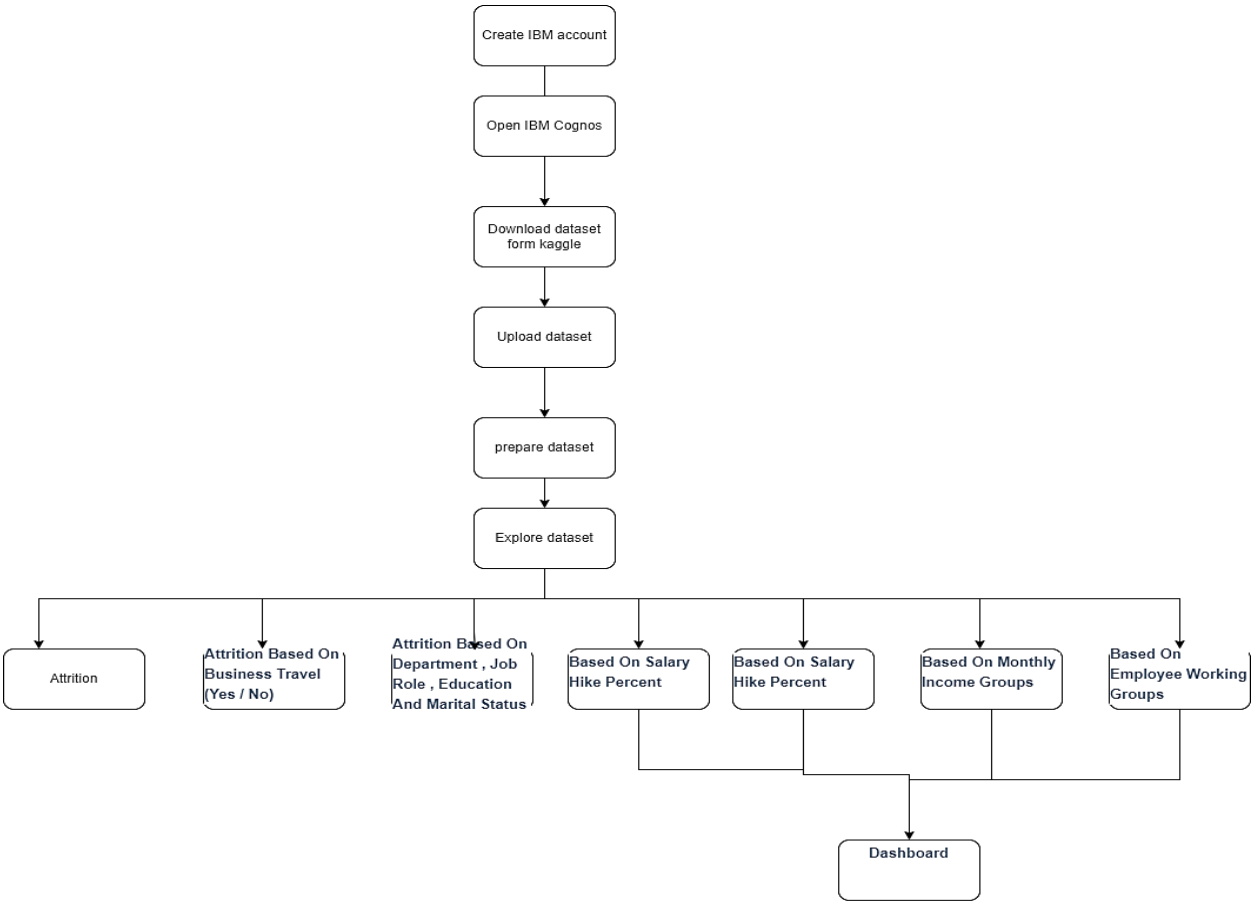
manager_survey_data.csv:-

1. EmployeeID
2. JobInvolvement (1 'Low' 2 'Medium' 3 'High' 4 'Very High')
3. PerformanceRating (1 'Low', 2 'Good', 3 'Excellent', 4 'Outstanding')



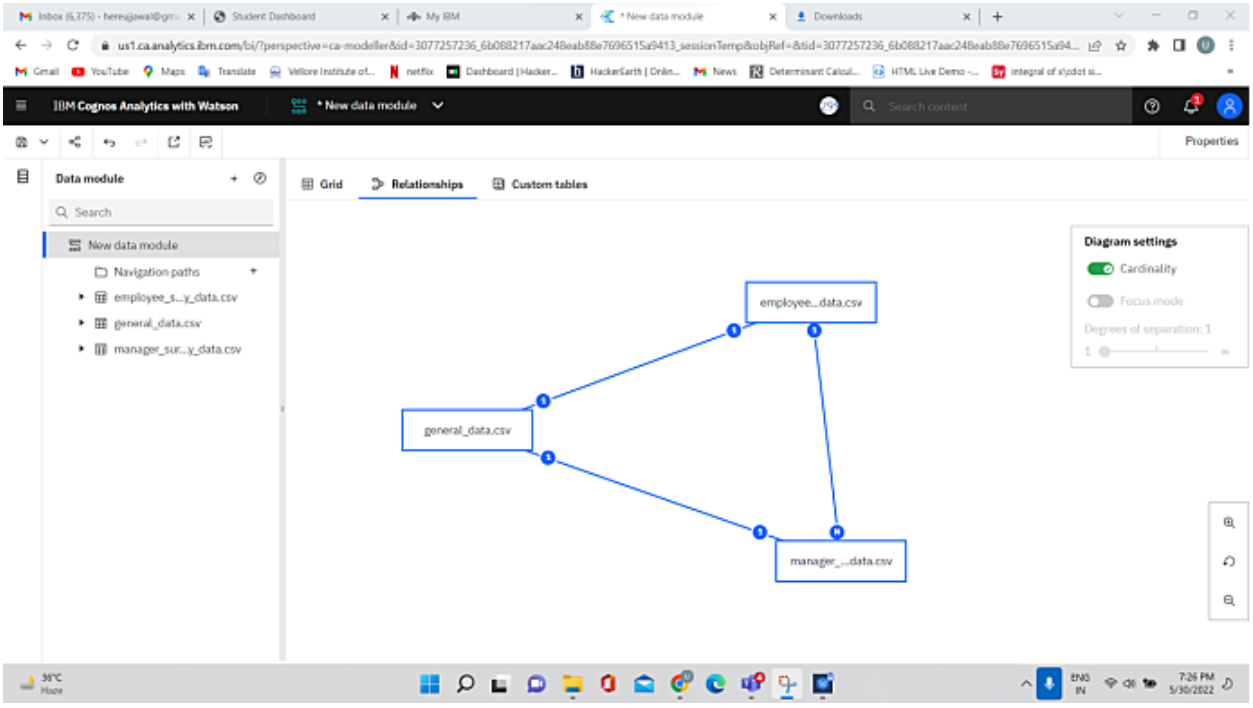
Row Id	EmployeeID	JobInvolvement	PerformanceRating
1	1	3	3
2	2	2	4
3	3	3	3
4	4	2	3
5	5	3	3
6	6	3	3
7	7	3	4
8	8	3	4
9	9	3	4
10	10	3	3
11	11	2	3
12	12	3	3
13	13	3	3

Flowchart :-



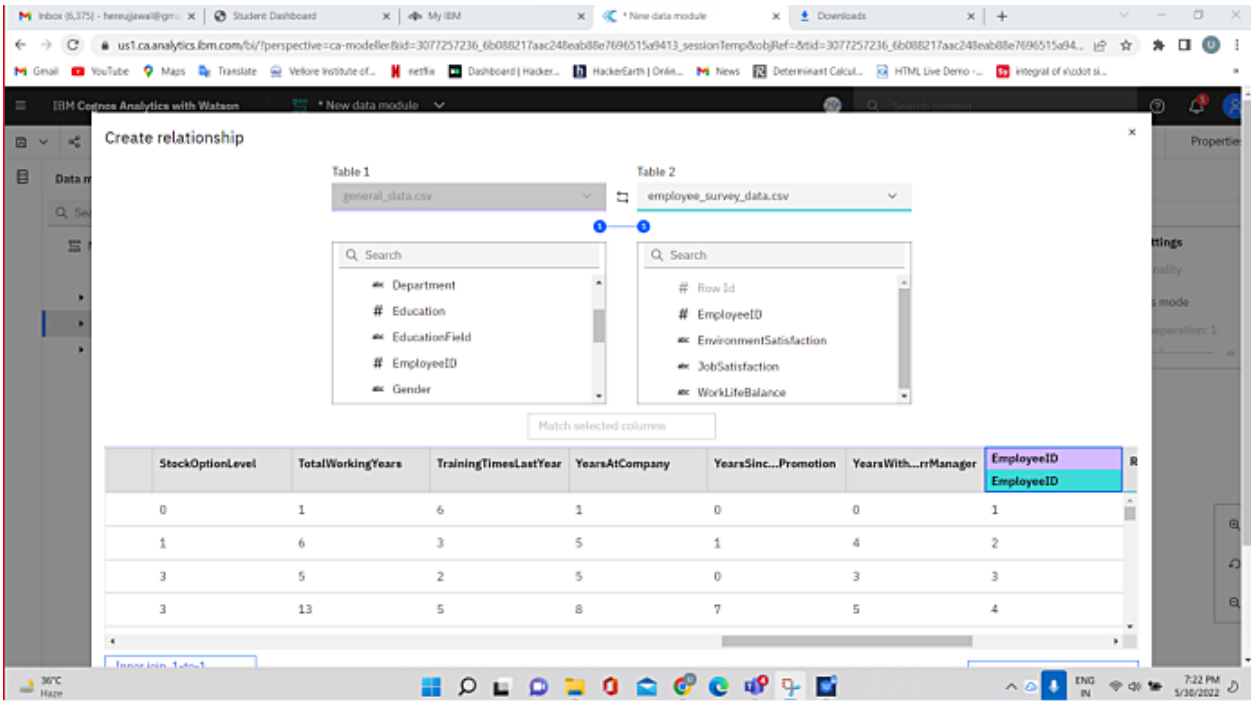
Working and outcome :-

After Loading and Understanding the dataset , I prepare the data for further process
In preparation of data I have establish the relationship between various table that is general CSV file employee CSV file and manager CSV file.

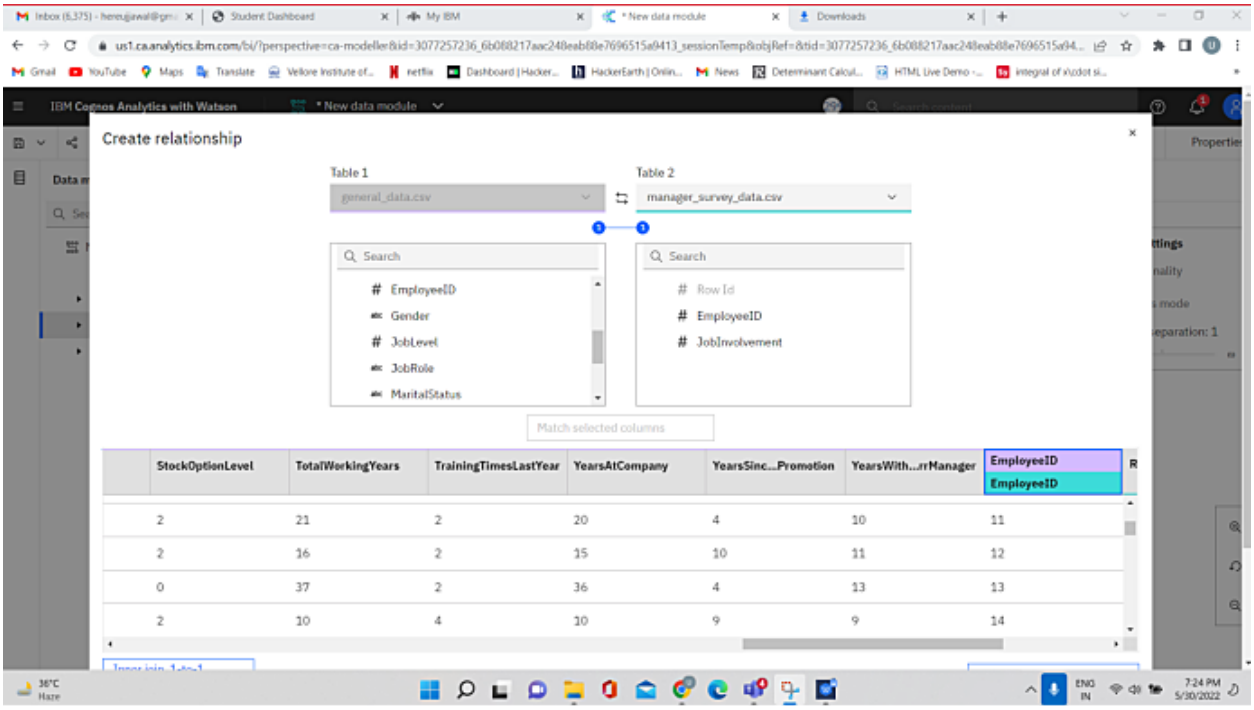


In the general_data.csv is having 1-to-1 relationship between employee_data.csv and manager_data.csv , whereas employee_data.csv and manager_data.csv is having 1-to-many relationship between them.
Defining relationship :-

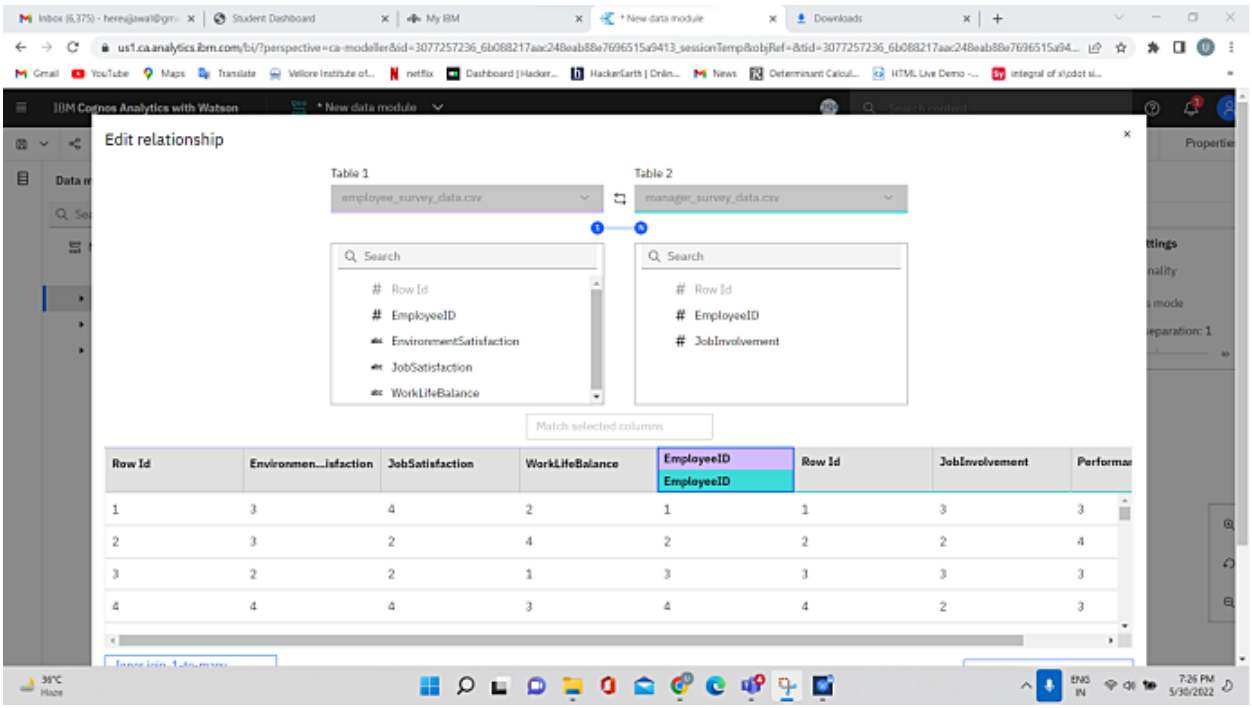
1) Between general_data.csv and employee_data.csv



2) Between general_data.csv and manager_data.csv

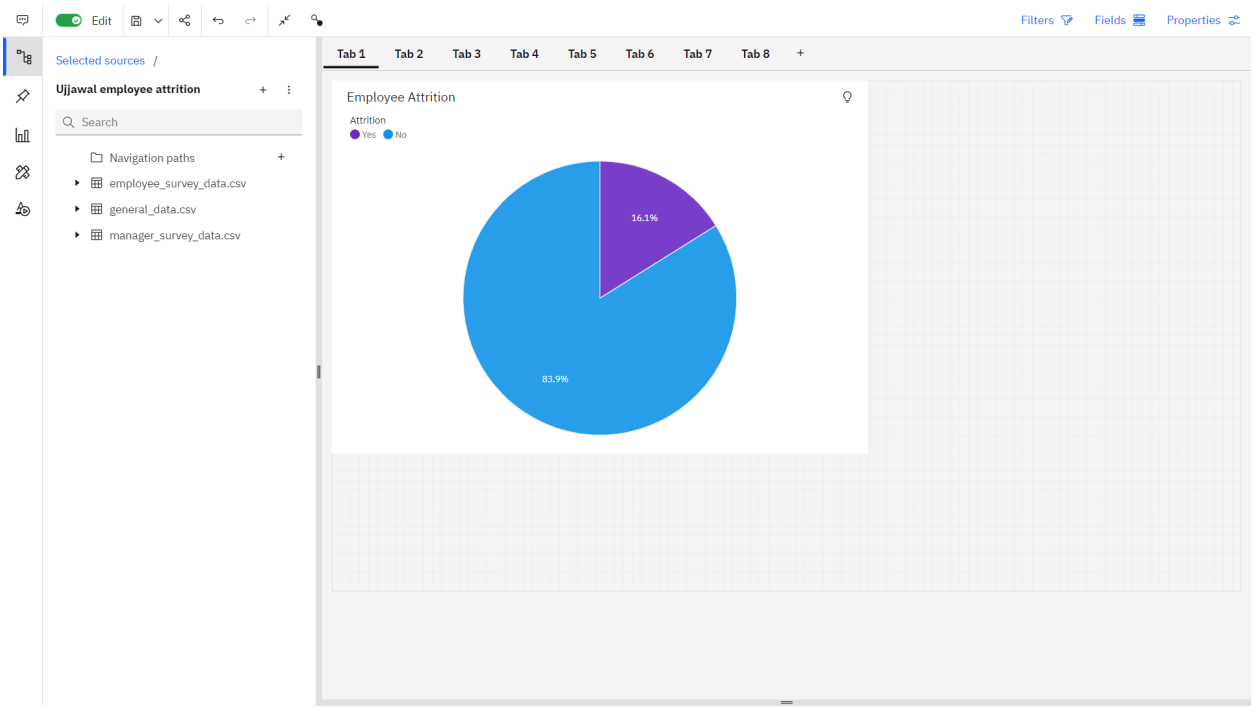


3) Between manager_data.csv and employee_data.csv



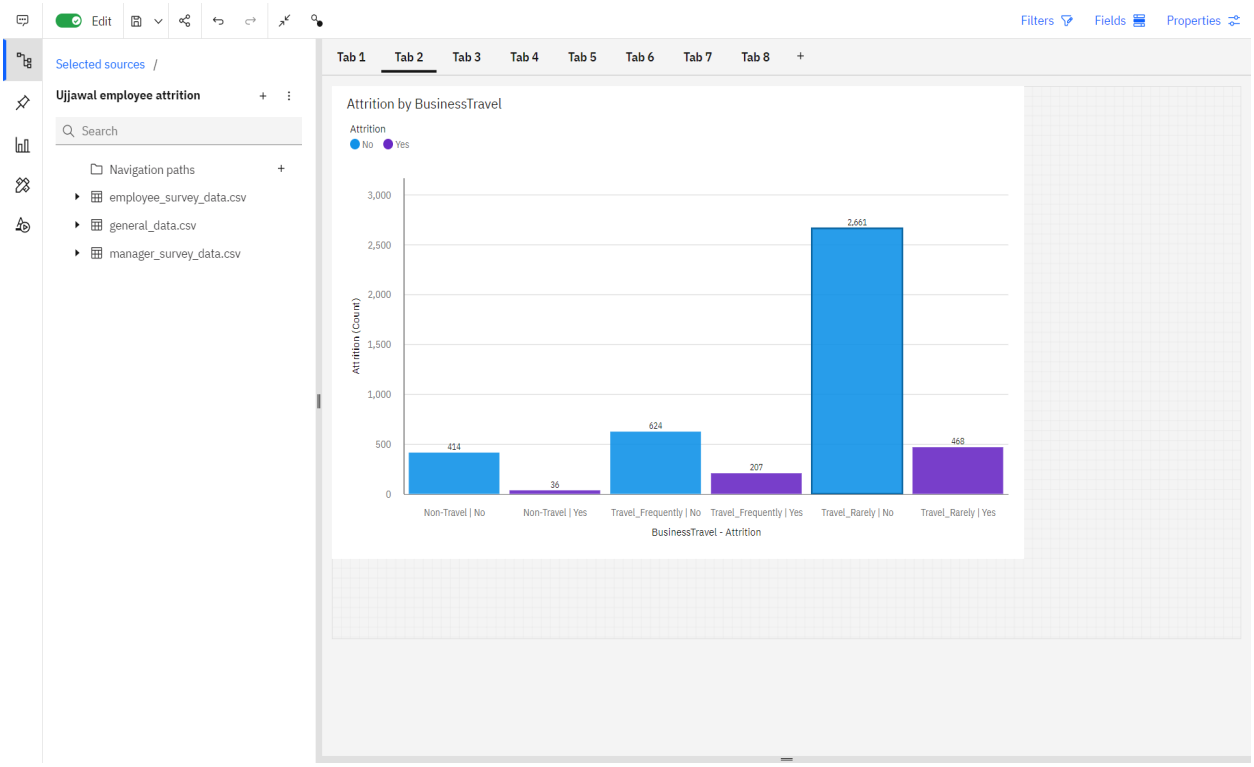
Hence after the preparation of data we start exploration (visualization) of data

1)**Attrition Status :-**



this tells us about the percentage of employees wanted to leave and stay in the company. hence here presented by pie chart

2) **Attrition Based On Business Travel (Yes / No)**

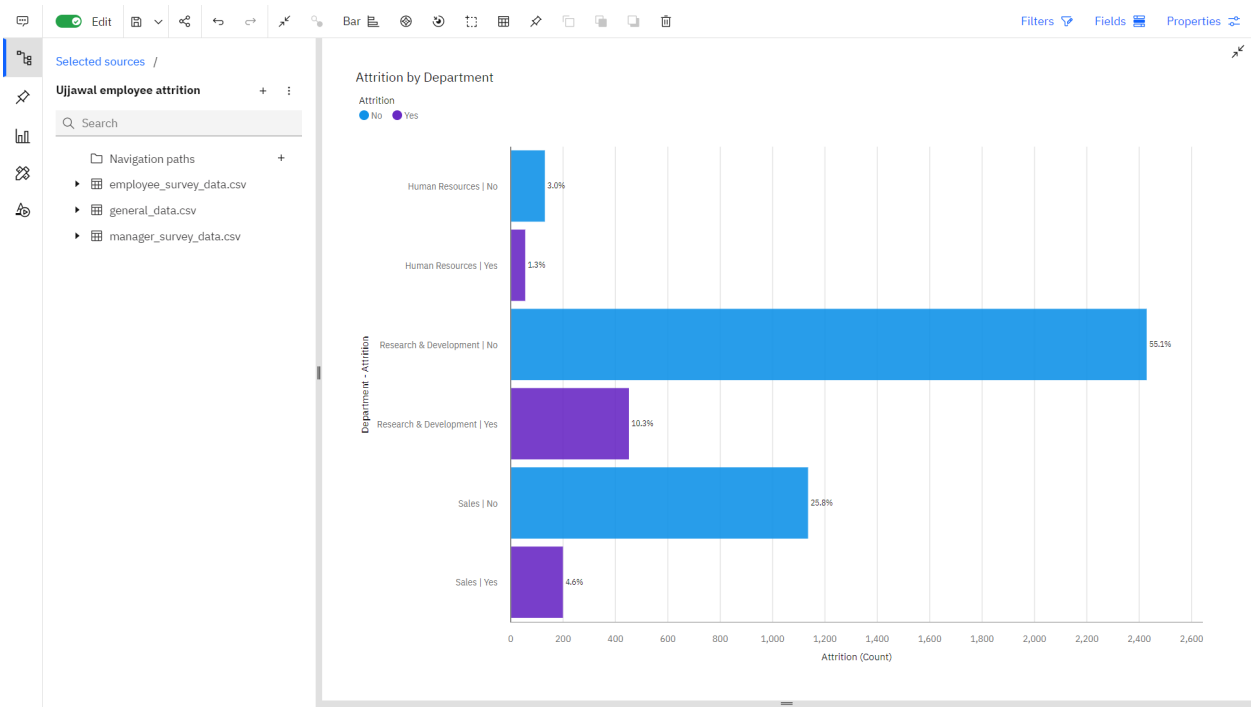


Business travel vs attrition is shown using column graph. hence rearly travel employees trends to leave company more than others.

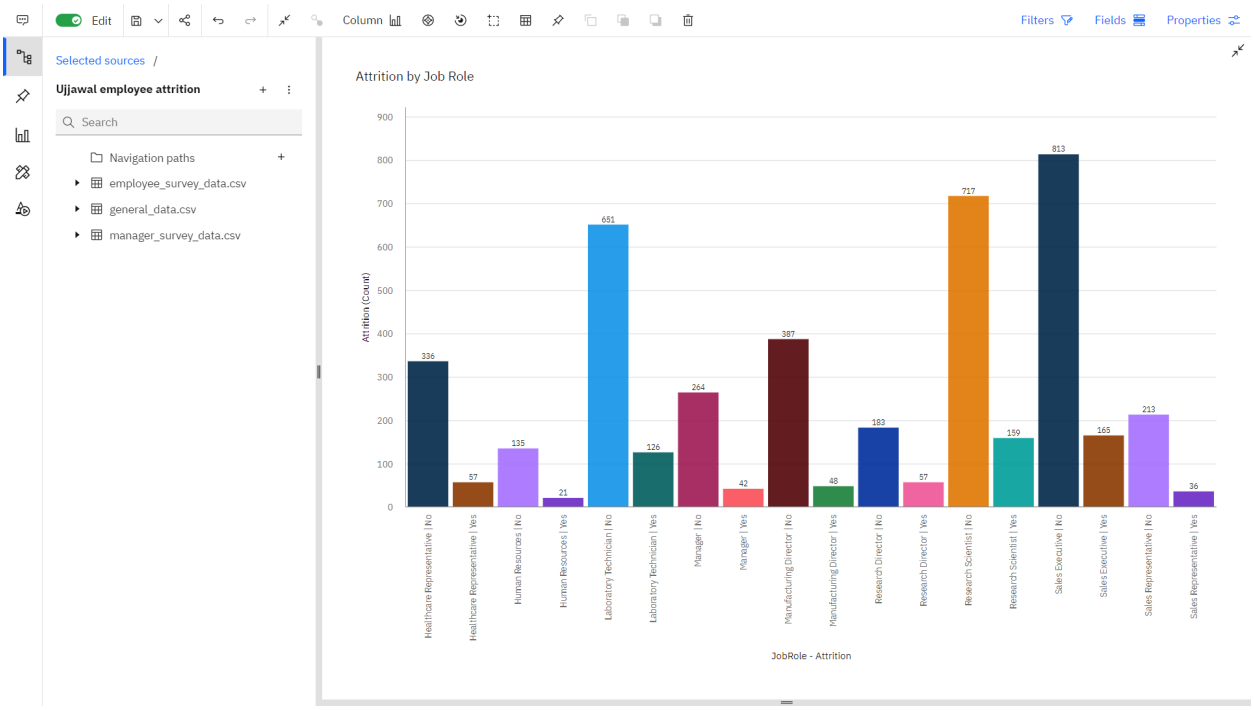
3) Attrition Based On Department , Job Role , Education And Marital Status

here 4 visualizations are presented in a single template

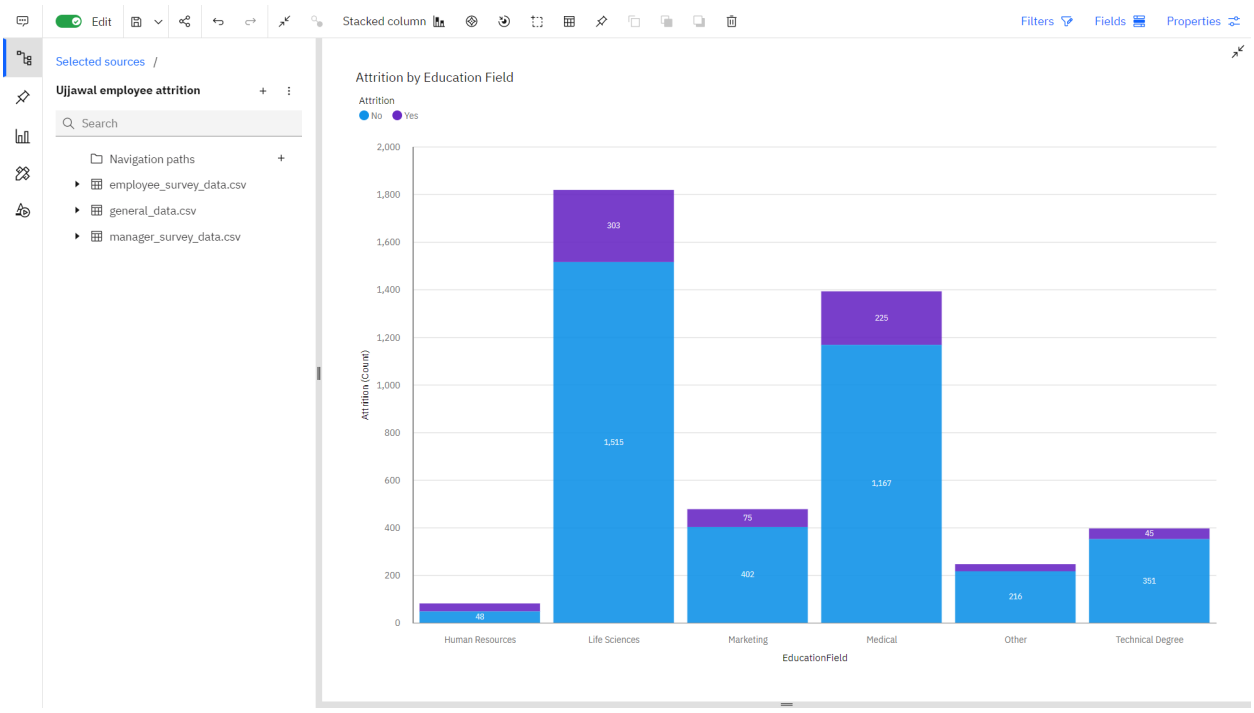
> Attrition by department : Used bar chart for this visualization



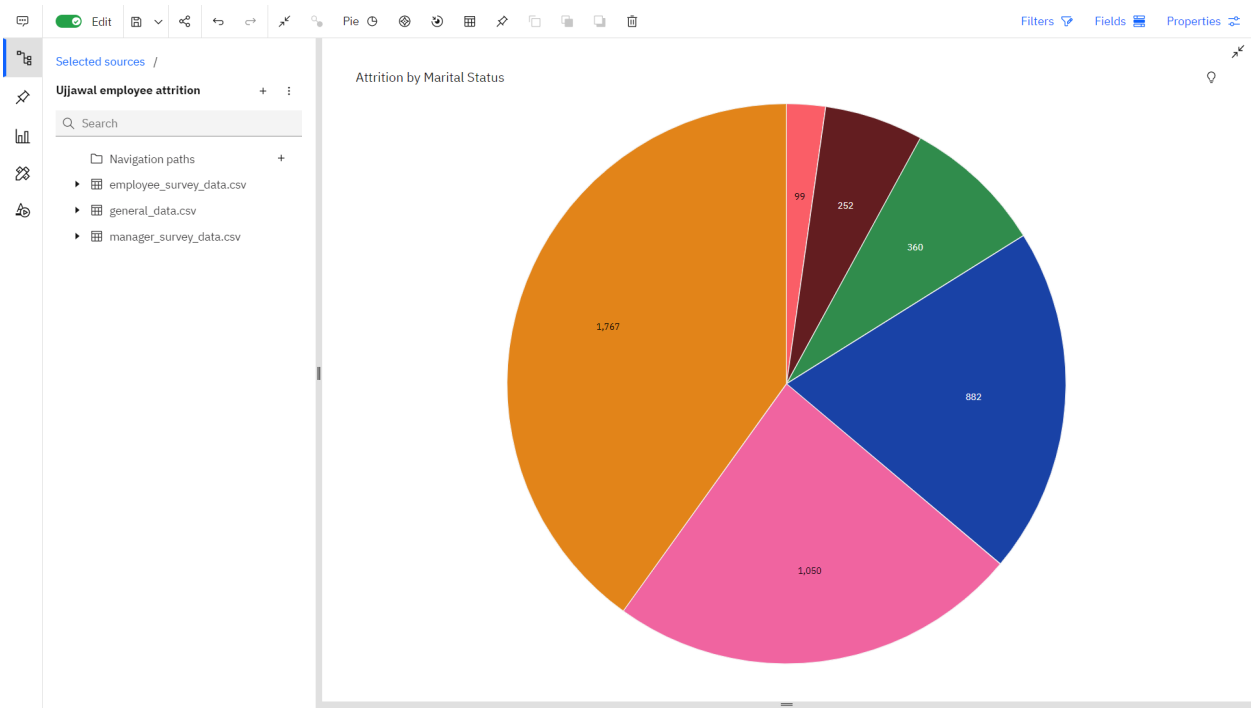
> Attrition by Job Role: used Columns graph for this



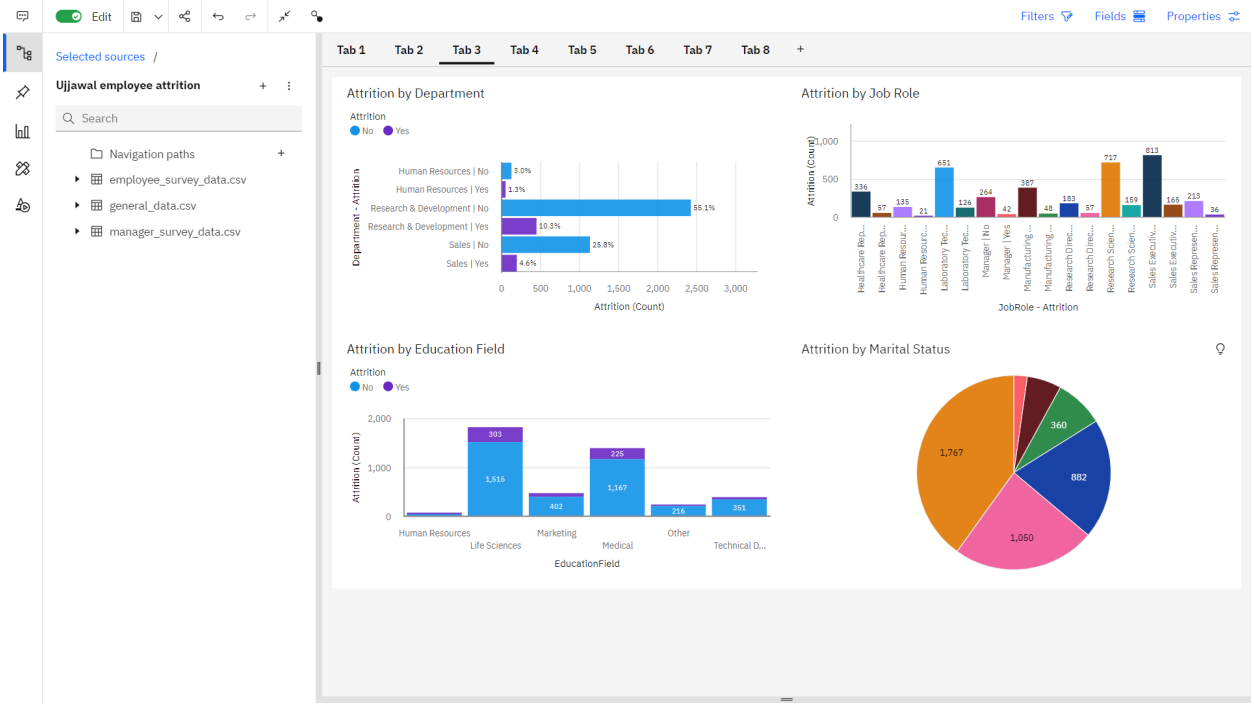
> Attrition by Education : Used stacked column chart for this



> Attrition by marital status : used pie chart for this

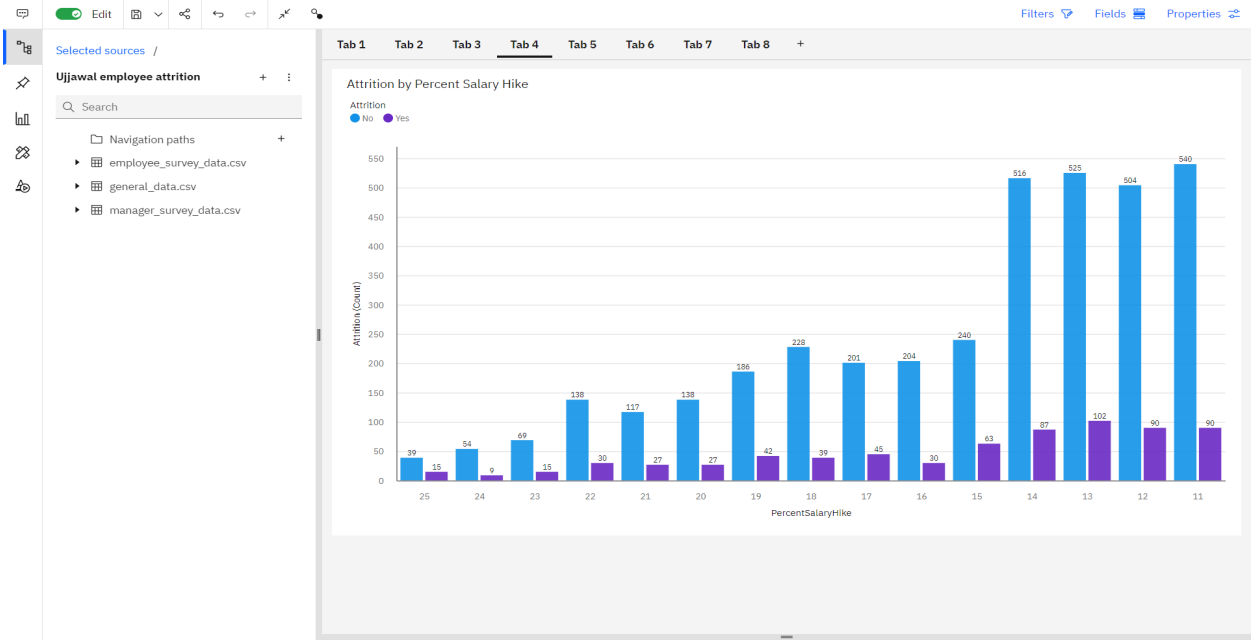


After Combining all four visualization in a single template : to merge all 4 we have to select 4 charts template in the starting



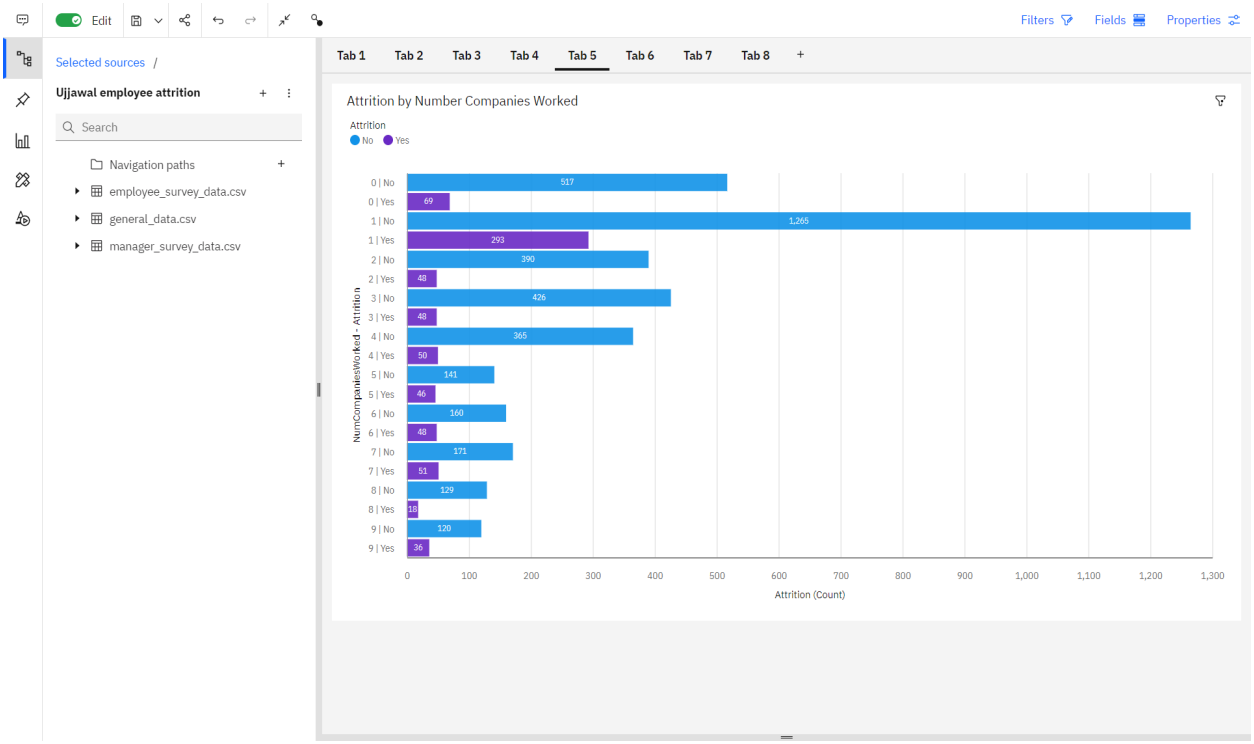
4) Based On Salary Hike Percent :-

here column graph is used to demonstrate the Attrition by Hike percent in salary :



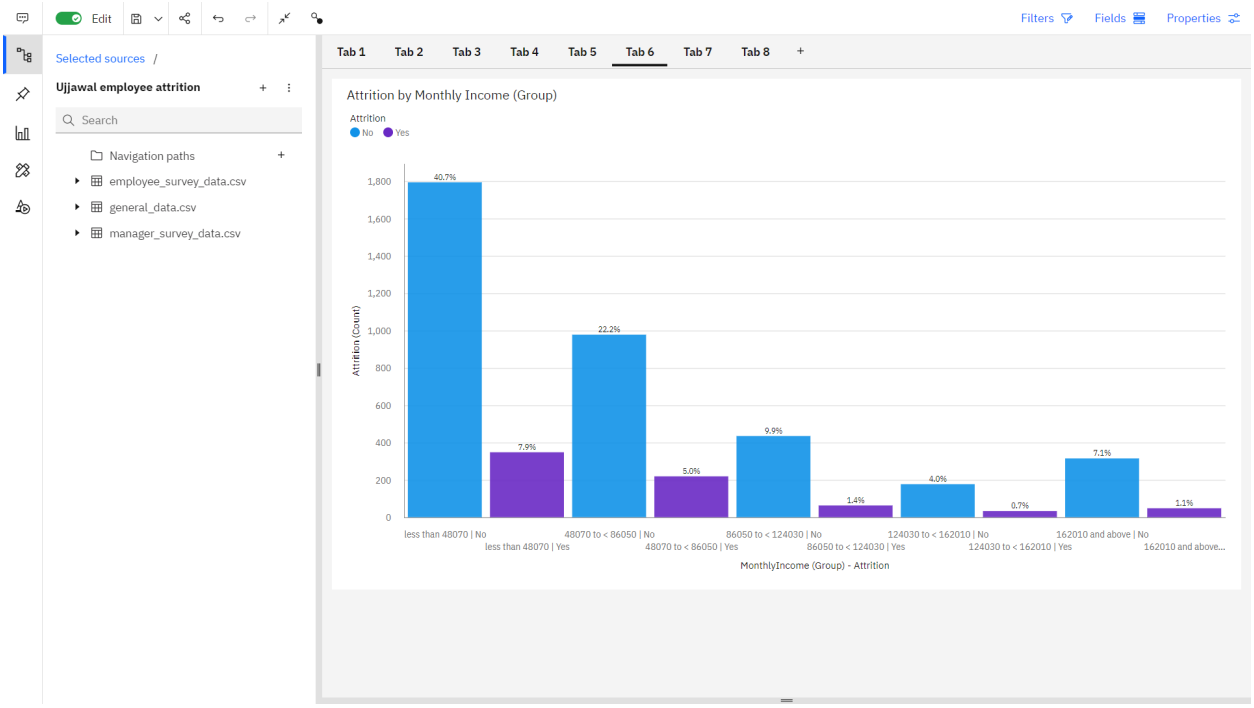
5) Based On No Of Companies Worked

here the visualization of attrition by Number of companies that employee worked is done by bar chart .



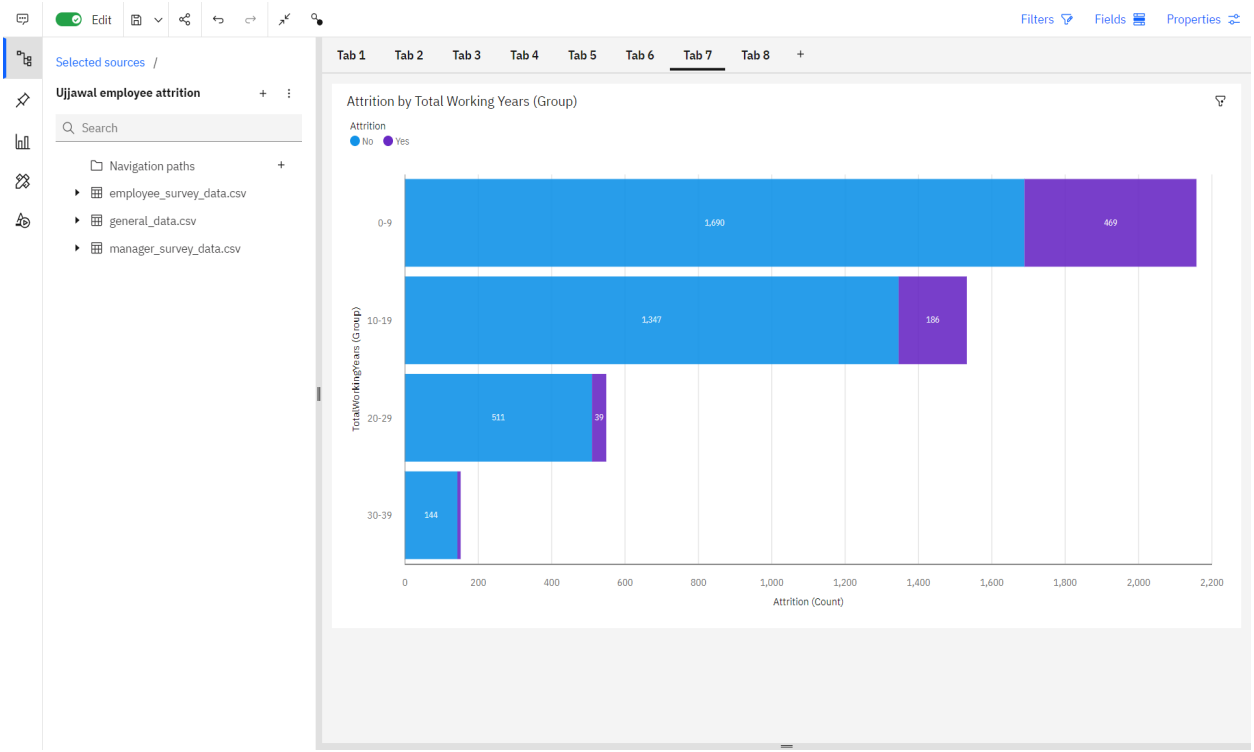
6)Based On Monthly Income Groups

here we use grouping of data . in this visualization group of 5 has been done .We have used auto grouping here.



7) Based On Employee Working Groups

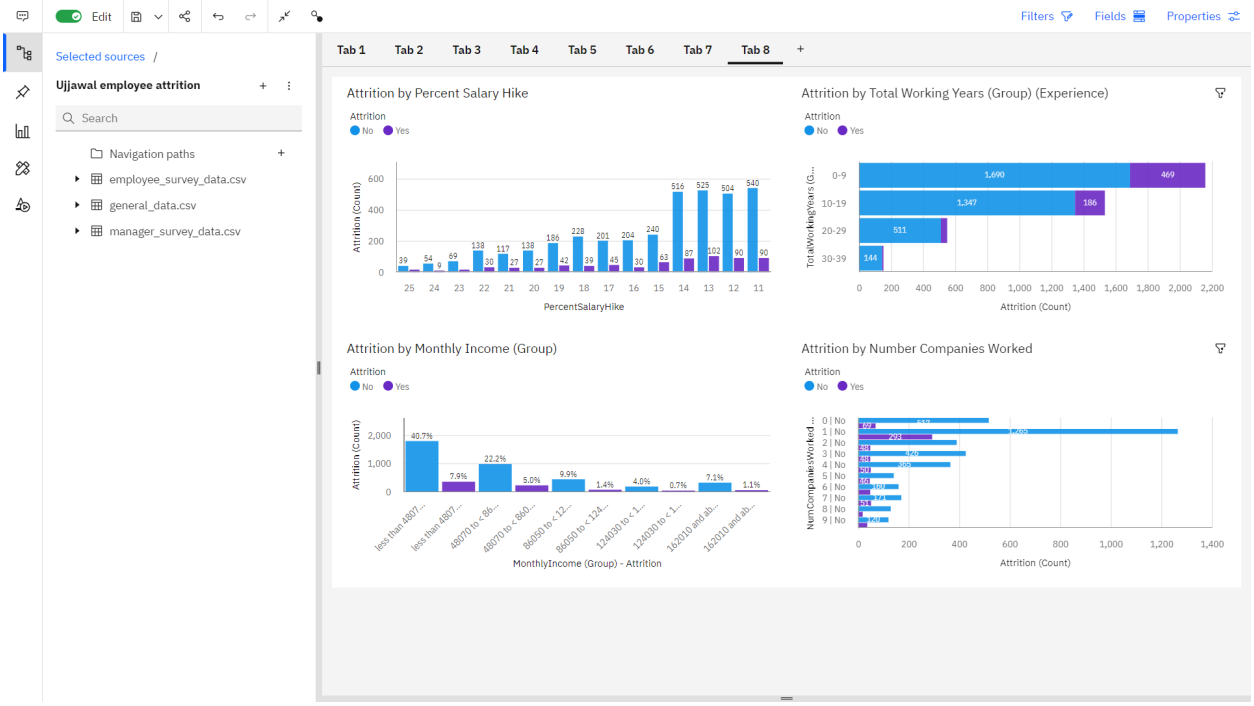
here I have done manual grouping by creating groups and manually putting data in it.



8) making dashboard :

in this we have copied last four visualization and pasted in a single template .

here while making this dashboard we can the percentage of area of each visualzation.



link of whole dashboard :-
https://us1.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FEmployee%2BAttrition%2BDashboard&action=view&mode=dashboard&subView=model0000018115a83d5b_00000000

Advantages:-

1. IBM cognos allows wide range of file types of upload as data assests.
2. Optimization dashboards allow sharing of the models' results with the team to enhance collaboration.
3. It has AI assistant and forecasting using AI
4. Connects to hundreds of data sources.
5. Easy to create charts, graphs and reports.
6. Users are able to schedule there own reports.
7. It offers data connectors including SQL Server, CSV, Excel, and more.
8. It is a reliable and scalable platform
9. Data security. It doesn't matter who the user is; when they log in and run a report, they will only see the data they are permitted to see.
10. Performant reports on large data. Many tools have issues running on tables containing only a few million rows. Cognos can consistently run reports on multi-billion row tables without issue.

Advantages of perforing Empployee Attrition analysis using IBM cognos are :
while working with IBM cognos , it is very easy to prepare data and there is a feature to establish the relationship between tables hence we can use many files to analyze the behaviour of employees and their needs .
Using cognos the visualization of data is also very easy and it provides a lot of charts and graph to make our customize dashboard very explainatory.

Disadvantages

1. Search engine displaying data elements is still not available on version 11.1.7.
2. Not a good IDE for Python/R
3. Data modules still can't match the flexibility of the framework manager. When dealing with large data models, or multi-grain, FM still wins.
4. Some functionality, like version control, is only available through third parties. Motio or BSP.
5. Better compatibility with available browsers.
6. Does not support multidimensional analysis

While working with cognos , I have found as disadvantage is the flexibity to share my working with anyone , if I want to share my work with non cognos user , I won't be able to do that when I share the link it ask for cognos login before displaying the dashboard , hence it might not be efficient we an employee wants to display the analysis report to stakeholders or to the clients.

Conclusion:-

After successfully performing the experiment and creating all the visualizations using IBM cognos , Here I can conclude that visualizations areready to be analyzed very easily and data is well prepared for further analyzation.
In first visualization that is attrition yes or no we have generated a pie chart in which the employees leaving the company and is staying the company well mark by percentage We can see that 83.9% of employees are staying in the company while 16.1% of employees are leaving the company.On further analyzing the visualizations .In *attrition by departmen*t , I have seen that *research and development and sale department* tends to have most employees leaving the company where *human resource* have the least employee leaving the company. Similarly *attrition by job role* , I've seen that the *research scientist and sales executive* are the most who are leaving the company.
Hence here the management of the company should focus on research department and sales department more than other. Similarly from attrition by salary hike I have seen that the more salary hicked Employees hare the least to leave the company.Similarly in *attrition by monthly income* ,I have seen that the employee who is getting less monthly income are more those are

leaving the company.In *attrition by marital status,* I have seen that *single employees* (non married employees)are trend to leave the company more than divorced or married employee.

Hence as final conclusion I can conclude that the management of company should not give critical tasks to the employees of research department or sales department who are non married and getting less salary in monthly income and from medical or life science educational background and also should not give critical tasks to employees those who have less hiked in their salary.

Thus from raw data of general_data.csv , employee_data.csv and manager_data.csv , we are able to extract these information using IBM cognos data analysis .