CHS
Unit-3 Task(Power Bi)

AMAN GUPTA(2021UCS1602) **GARIMA(2021UCS1605)** **TARUN MITTAL(2021UCS1608)**

INDEX

S.No	Questions
1.	**Explore Power View, Power Query** Create a table Employee(empid, gender, department, salary, country, year_of_joining) connect to the Employee data file. Remove missing gender and department values. Extract year_of_joining column and visualize number of employees w.r.t year of experience in the company. Perform self-join using Power Query. Aggregate salary with gender and Visualize using Pie chart.
2.	Visualize the result of any Machine Learning algorithm on any dataset of your choice in PowerBI.

Q1) Explore Power View, Power Query

Create a table Employee(empid, gender, department, salary, country, year_of_joining) connect to the Employee data file.

Remove missing gender and department values.

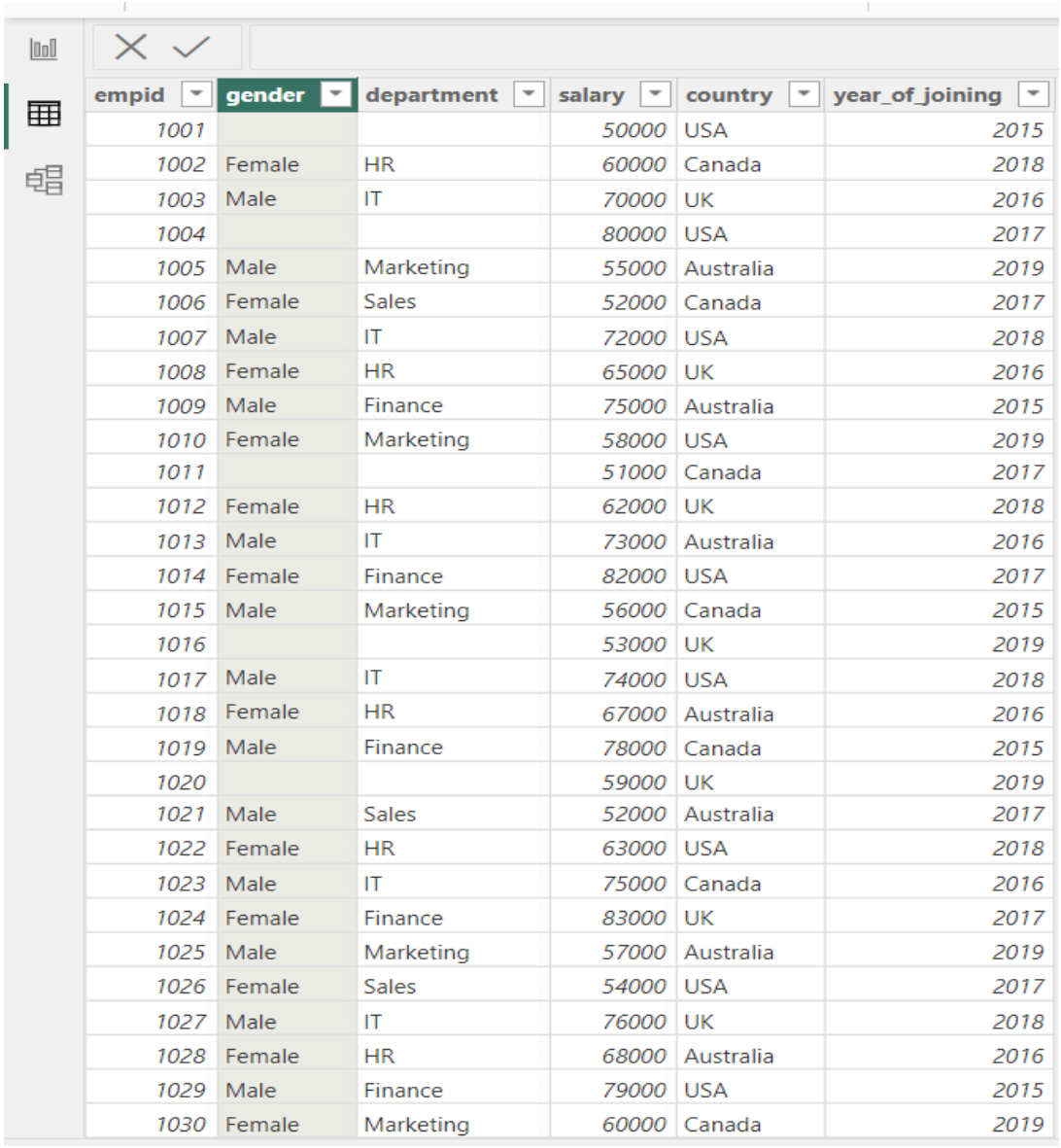
Extract year_of_joining column and visualize number of employees w.r.t year of experience in the company.

Perform self-join using Power Query.

Aggregate salary with gender and Visualize using Pie chart.

Ans.

Create a table Employee(empid, gender, department, salary, country, year_of_joining) connect to Employee data file.

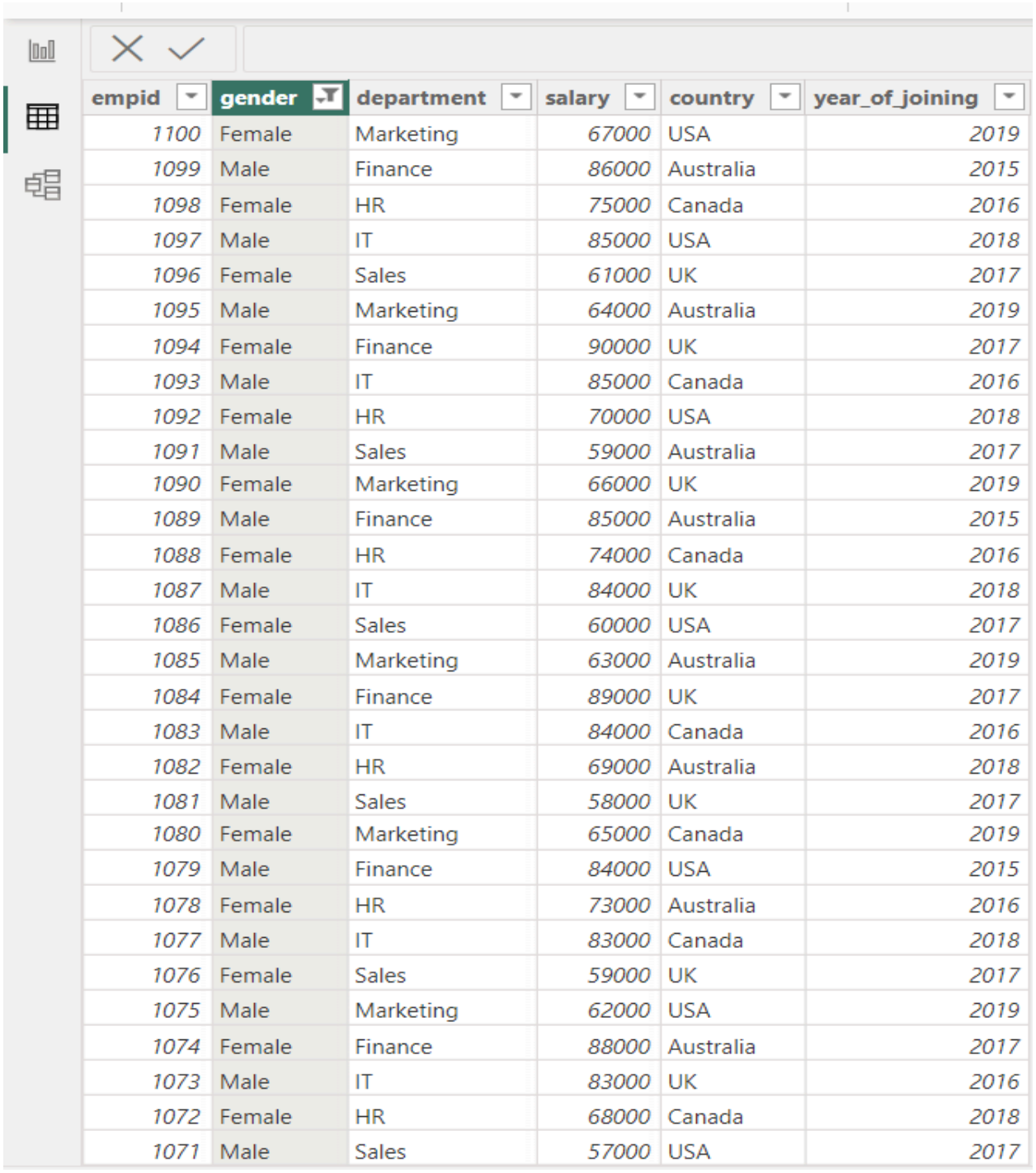


empid	gender	department	salary	country	year_of_joining
1001			50000	USA	2015
1002	Female	HR	60000	Canada	2018
1003	Male	IT	70000	UK	2016
1004			80000	USA	2017
1005	Male	Marketing	55000	Australia	2019
1006	Female	Sales	52000	Canada	2017
1007	Male	IT	72000	USA	2018
1008	Female	HR	65000	UK	2016
1009	Male	Finance	75000	Australia	2015
1010	Female	Marketing	58000	USA	2019
1011			51000	Canada	2017
1012	Female	HR	62000	UK	2018
1013	Male	IT	73000	Australia	2016
1014	Female	Finance	82000	USA	2017
1015	Male	Marketing	56000	Canada	2015
1016			53000	UK	2019
1017	Male	IT	74000	USA	2018
1018	Female	HR	67000	Australia	2016
1019	Male	Finance	78000	Canada	2015
1020			59000	UK	2019
1021	Male	Sales	52000	Australia	2017
1022	Female	HR	63000	USA	2018
1023	Male	IT	75000	Canada	2016
1024	Female	Finance	83000	UK	2017
1025	Male	Marketing	57000	Australia	2019
1026	Female	Sales	54000	USA	2017
1027	Male	IT	76000	UK	2018
1028	Female	HR	68000	Australia	2016
1029	Male	Finance	79000	USA	2015
1030	Female	Marketing	60000	Canada	2019

Remove missing gender and department values.

Steps:

- In the "Query Editor" window, select the "Employee" table.
- Click on the "Remove Rows" dropdown in the "Home" tab.
- Select "Remove Blank Rows" and choose the columns "gender" and "department".



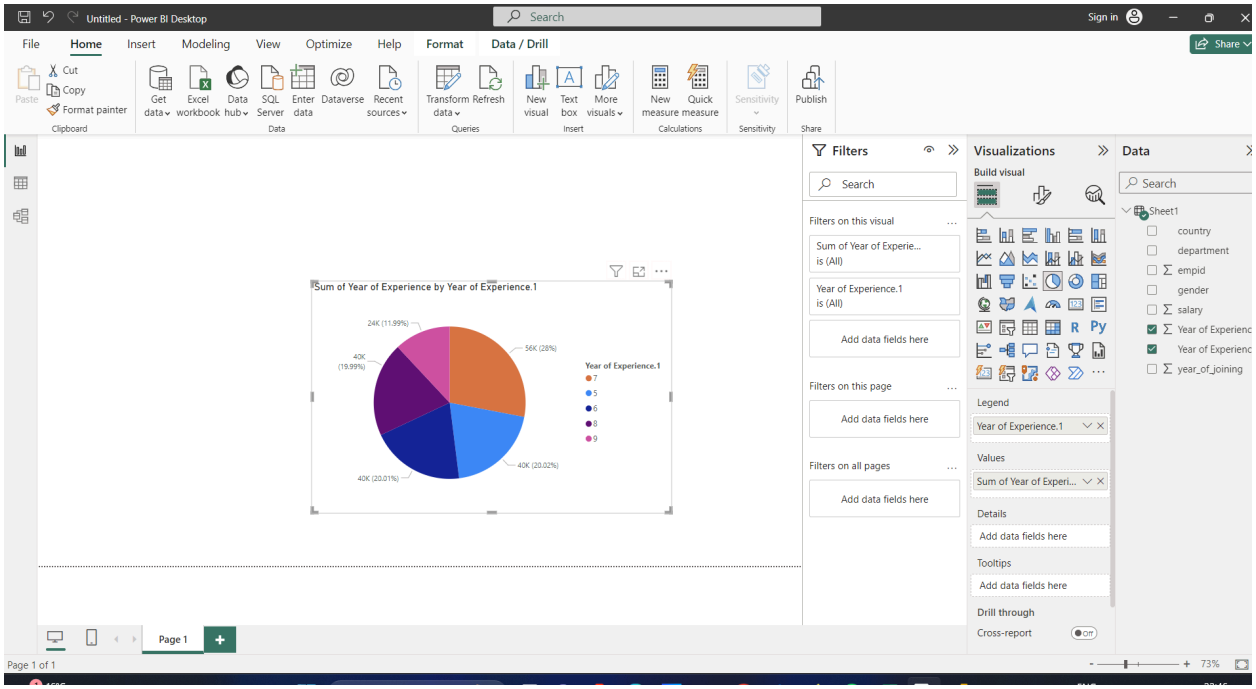
empid	gender	department	salary	country	year_of_joining
1100	Female	Marketing	67000	USA	2019
1099	Male	Finance	86000	Australia	2015
1098	Female	HR	75000	Canada	2016
1097	Male	IT	85000	USA	2018
1096	Female	Sales	61000	UK	2017
1095	Male	Marketing	64000	Australia	2019
1094	Female	Finance	90000	UK	2017
1093	Male	IT	85000	Canada	2016
1092	Female	HR	70000	USA	2018
1091	Male	Sales	59000	Australia	2017
1090	Female	Marketing	66000	UK	2019
1089	Male	Finance	85000	Australia	2015
1088	Female	HR	74000	Canada	2016
1087	Male	IT	84000	UK	2018
1086	Female	Sales	60000	USA	2017
1085	Male	Marketing	63000	Australia	2019
1084	Female	Finance	89000	UK	2017
1083	Male	IT	84000	Canada	2016
1082	Female	HR	69000	Australia	2018
1081	Male	Sales	58000	UK	2017
1080	Female	Marketing	65000	Canada	2019
1079	Male	Finance	84000	USA	2015
1078	Female	HR	73000	Australia	2016
1077	Male	IT	83000	Canada	2018
1076	Female	Sales	59000	UK	2017
1075	Male	Marketing	62000	USA	2019
1074	Female	Finance	88000	Australia	2017
1073	Male	IT	83000	UK	2016
1072	Female	HR	68000	Canada	2018
1071	Male	Sales	57000	USA	2017

Extract year_of_joining column and visualize number of employees w.r.t year of experience in the company.

Steps:

Extract Year of Joining Column:

- Go to the "Transform Data" view by clicking on the "Transform Data" button in the Home tab.
- Select the "year_of_joining" column.
- Right-click on the column header and choose "Duplicate Column".
- Rename the duplicated column to "Year of Experience".
- Calculate the years of experience by subtracting the "year_of_joining" from the current year.
- You can do this by adding a new column with the formula: 'Date.Year(DateTime.LocalNow()) [year_of_joining]'.
- Close and apply the changes.



Perform self-join using Power Query.

Steps:

Connect to Data Source: Connect Power BI to your data source containing the table you want to perform the self-join on.

Open Power Query Editor:

- In Power BI Desktop, go to the "Home" tab.
- Click on "Transform Data" to open the Power Query Editor.

Duplicate the Table:

- In the Power Query Editor, right-click on the table you want to perform the self-join on in the Queries pane (usually on the left side).

- Select "Duplicate" to create a copy of the table.

Rename Duplicate Table:

- Right-click on the duplicated table.
- Select "Rename" and give it a meaningful name to distinguish it from the original table.

Merge Tables:

- In the duplicated table, click on the column header that you want to use as the key for the self-join.
- Go to the "Home" tab.
- Click on "Merge Queries" in the "Combine" group.
- Select the original table from the dropdown menu.
- Choose the matching column from both tables.
- Specify the type of join you want to perform (e.g., inner join, left outer join, etc.).
- Click "OK".

Expand the Merged Table:

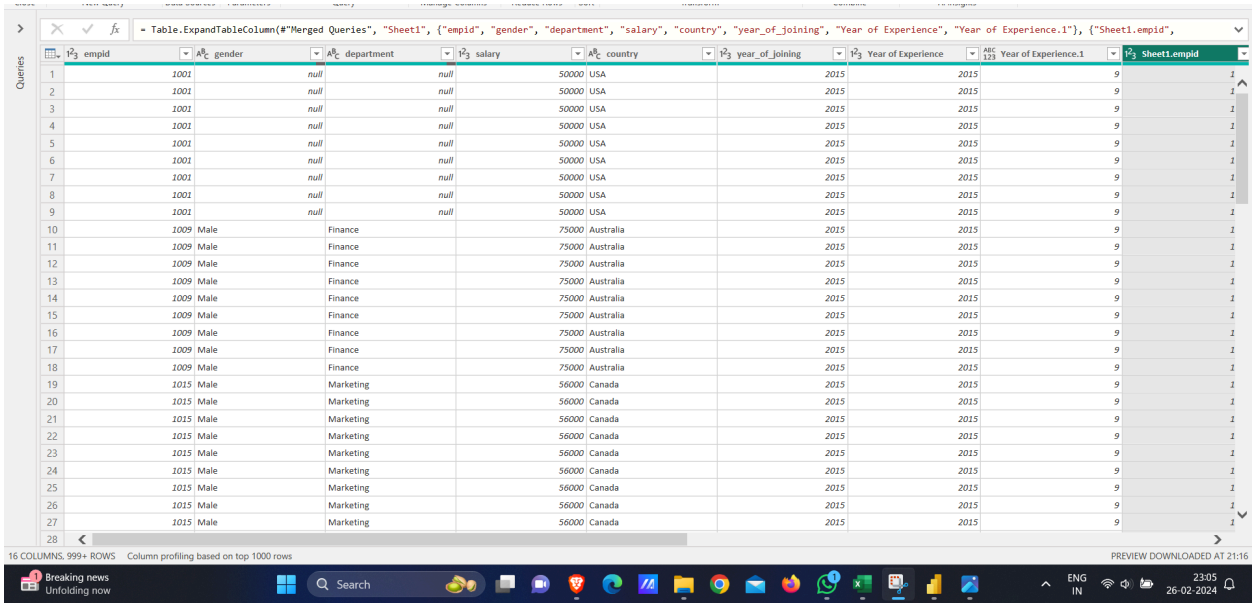
- After merging, a new column will be added containing the original table.
- Click on the expand icon in the column header of the merged table.
- Choose which columns you want to expand from the original table.
- Click "OK".

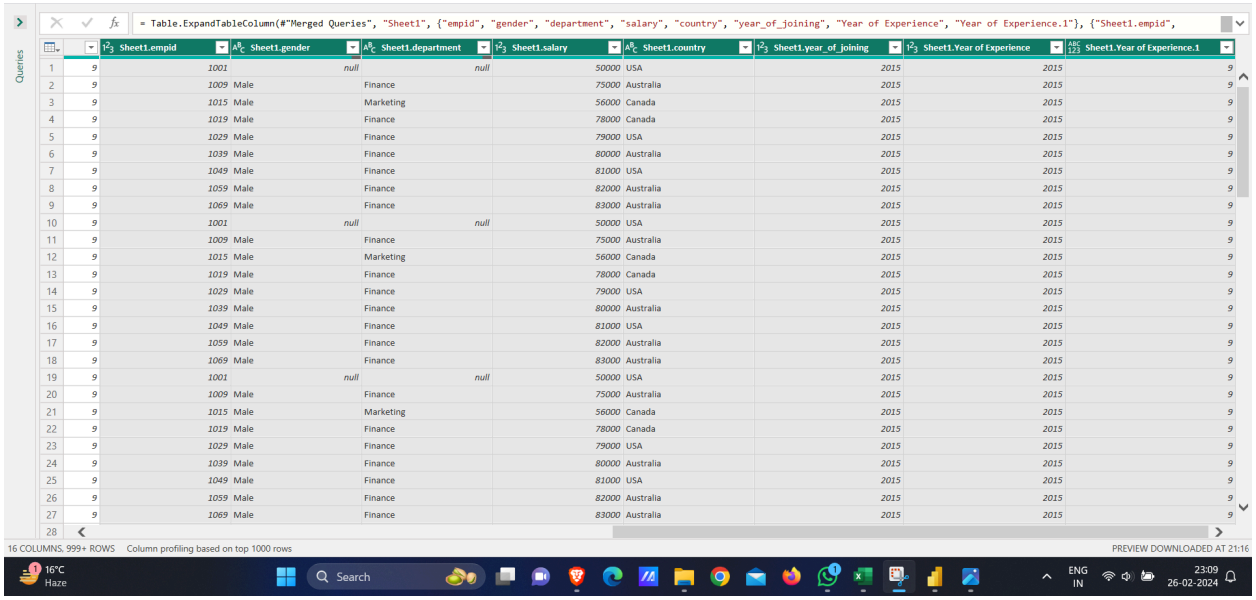
Clean Up:

- Rename columns if necessary to distinguish between columns from the original table and the merged table.
- Remove any unnecessary columns.

Close and Apply Changes:

- Once you're done, click "Close & Apply" to apply the changes and close the Power Query Editor.





Aggregate salary with gender and Visualize using Pie chart.

Steps:

Connect to Data Source: Connect Power BI to your data source containing the Employee table.

- Open Power Query Editor:
- In Power BI Desktop, go to the "Home" tab.
- Click on "Transform Data" to open the Power Query Editor.
- Group and Aggregate Data:
- In the Power Query Editor, select the Employee table.
- Go to the "Home" tab.
- Click on "Group By" in the "Transform" group.

In the Group By window:

- Group By: Select the 'gender' column.
- New column name: Provide a name for the new column, e.g., TotalSalary.
- Operation: Choose "Sum" for the Operation and select the 'salary' column.
- Click "OK" to apply the grouping and aggregation.

Close and Apply Changes:

- Once you're done, click "Close & Apply" to apply the changes and close the Power Query Editor.

Create Pie Chart:

- In the Power BI report view, click on the "Insert" tab.
- Choose "Pie Chart" from the Visualization section.

Configure Pie Chart:

- Drag the 'gender' column from the Fields pane into the Legend field of the Pie chart visualization.
- Drag the 'TotalSalary' column (or whatever you named it) into the Values field of the Pie chart.

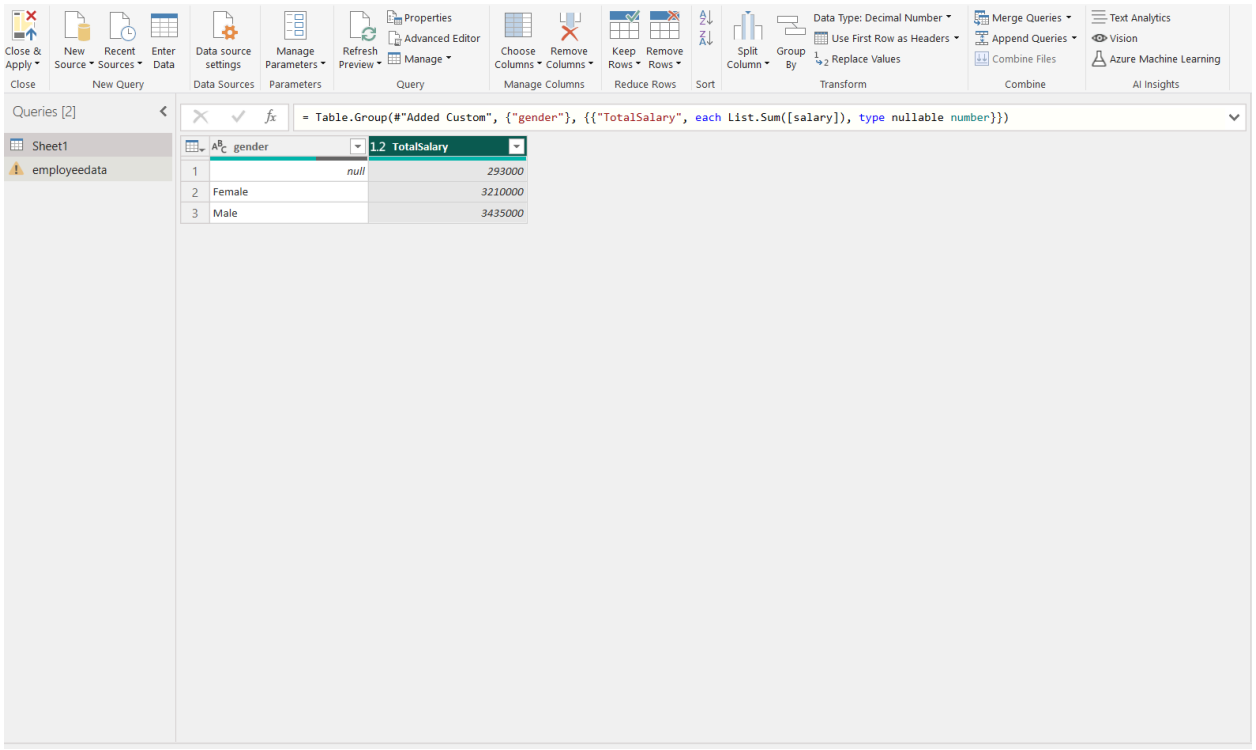
- Power BI will automatically sum the salary based on gender and display it in the Pie chart.

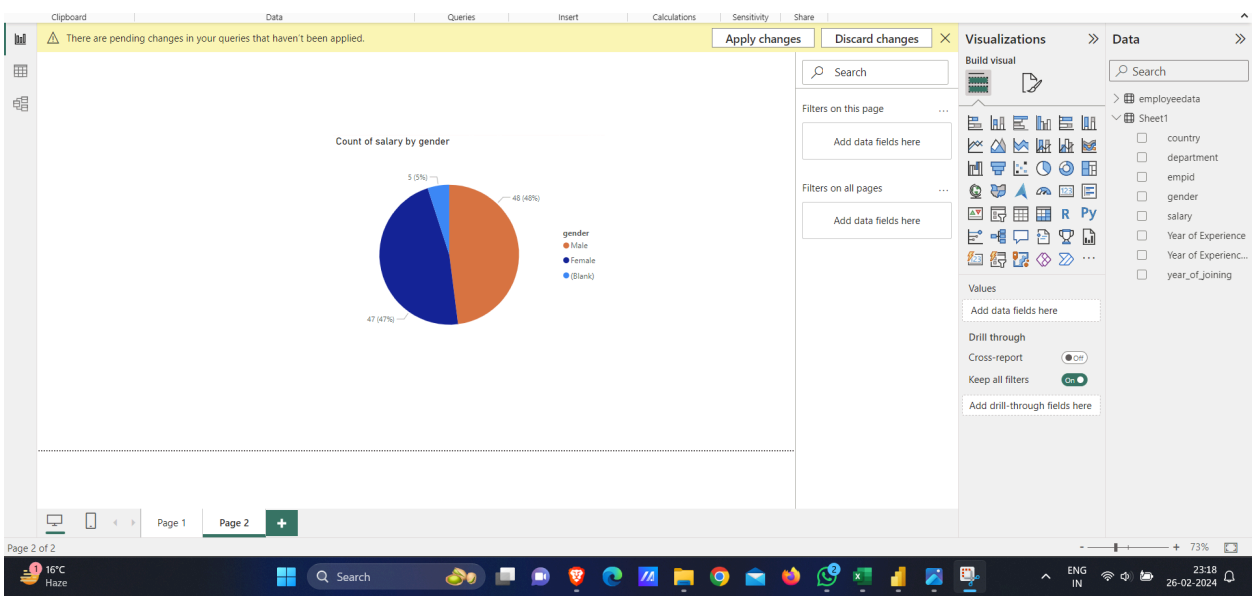
Customize Pie Chart (Optional):

- You can customize the colors, labels, and other properties of the Pie chart as per your preference using the formatting options available in the Visualizations pane.

Save and Publish:

- Save your Power BI file and publish it to Power BI service if needed.





Q1) Visualize the result of any Machine Learning algorithm on any dataset of your choice in PowerBI.

Machine Learning Algorithm used : Linear Regression

Dataset :

	A	B	C	D	E	F
1	R&D Sp	Administ	Marketin	State	Profit	
2	165349	136898	471784	New Yor	192262	
3	162598	151378	443899	Californi	191792	
4	153442	101146	407935	Florida	191050	
5	144372	118672	383200	New Yor	182902	
6	142107	91392	366168	Florida	166188	
7	131877	99815	362861	New Yor	156991	
8	134615	147199	127717	Californi	156123	
9	130298	145530	323877	Florida	155753	
10	120543	148719	311613	New Yor	152212	
11	123335	108679	304982	Californi	149760	
12	101913	110594	229161	Florida	146122	
13	100672	91791	249745	Californi	144259	
14	93864	127320	249839	Florida	141586	
15	91992	135495	252665	Californi	134307	
16	119943	156547	256513	Florida	132603	
17	114524	122617	261776	New Yor	129917	
18	78013	121598	264346	Californi	126993	
19	94657	145078	282574	New Yor	125370	
20	91749	114176	294920	Florida	124267	
21	86420	153514	0	New Yor	122777	
22	76254	113867	298664	Californi	118474	
23	78389	153773	299737	New Yor	111313	
24	73995	122783	303319	Florida	110352	
25	67533	105751	304769	Florida	108734	
26	77044	99281	140575	New Yor	108552	
27	64665	139553	137963	Californi	107404	
28	75329	144136	134050	Florida	105734	
29	72108	127865	353184	New Yor	105008	
30	66052	182646	118148	Florida	103282	
31	65605	153032	107138	New Yor	101005	
32	61994	115641	91131	Florida	99938	
33	61136	152702	88218	New Yor	97484	
34	63409	129220	46085	Californi	97428	
35	55494	103057	214635	Florida	96779	
36	46426	157694	210798	Californi	96713	
37	46014	85047	205518	New Yor	96480	
38	28664	127056	201127	Florida	90708	
39	44070	51283	197029	Californi	89949	
40	20230	65948	185265	New Yor	81229	
41	38559	82982	174999	Californi	81006	
42	28754	118546	172796	Californi	78240	
43	27893	84711	164471	Florida	77799	
44	23641	96190	148001	Californi	71498	
45	15506	127382	35534	New Yor	69759	
46	22178	154806	28335	Californi	65200	
47	1000.2	124153	1903.9	New Yor	64926	
48	1315.5	115816	297114	Florida	49491	
49	0	135427	0	Californi	42560	
50	542.05	51743	0	New Yor	35673	
51	0	116984	45173	Californi	14681	

Python Script :

Python script editor

```
# The following code to create a dataframe and remove duplicated rows is always executed and acts as a preamble for your script:

# dataset = pandas.DataFrame(Administration, Marketing Spend, Profit, State, R&D Spend)
# dataset = dataset.drop_duplicates()

# Paste or type your script code here:
import seaborn as sns
import matplotlib.pyplot as plt
sns.pairplot(data=dataset)
plt.show()
```

Visualization :

Administration, Marketing Spend, Profit, State and R&D Spend