NAME : B. RISHITHA

ROLL NO : 21X05A6710

COLLEGE : NARASIMHA REDDY ENGINEERING COLLEGE UGC AUTONOMOUS

BRANCH : DATA SCIENCE

YEAR : 4TH YEAR

GUIDER NAME : MR. B. LOKESH

PROJECT

TITLE :

Analysis of HR Datasets to help the organization take a meaningful right decision and find insights to excel their businesses.

Organization Name :

IBM India :

The International Business Machines Corporation (IBM), nicknamed Big Blue,is an American multinational technology corporation headquartered in Armonk, New York and is present in over 175 countries.It specializes in computer hardware, middleware, and software, and provides hosting and consulting services in areas ranging from mainframe computers to nanotechnology. IBM is the largest industrial research organization in the world, with 19 research facilities across a dozen countries, and has held the record for most annual U.S. patents generated by a business for 29 consecutive years from 1993 to 2021.

Tools and Technology :

Advanced Excel :

Advanced Excel skills include the ability to produce graphs and tables, use spreadsheets efficiently, and perform calculations and automations to process large volumes of data.

"Advanced Excel" refers to a higher level of proficiency and skill in using Microsoft Excel, a popular spreadsheet software application. While basic Excel knowledge includes tasks like creating simple spreadsheets, inputting data, and performing basic calculations, advanced Excel skills involve more complex and powerful features of the software. These skills are particularly valuable for data analysis, reporting, and decision-making in various industries and roles.

Power BI :

Power BI is a business intelligence (BI) tool developed by Microsoft. It is used for data visualization, interactive reporting, and business analytics. Power BI enables users to connect to various data sources, transform and shape the data, create visually appealing reports and dashboards, and share insights with others.

Power BI is widely used by businesses to gain insights from their data, make informed decisions, and communicate those insights effectively to stakeholders. It's a versatile tool for data professionals, analysts, and business users alike.

Problem Statement :

Analysis of HR Datasets to help the organization take a meaningful right decision and find insights to excel their businesses.

By observing the raw datasets there is so many columns present which is unwanted to output solution. Remove the columns which is unwanted in the data set.

Removing of columns from the dataset.

We can remove the columns from the dataset in excel by using commands that is Click on the column in the dataset and press ‘cntrl –‘then automatically the columns are removed from the dataset.

There are totally 1471 rows are present in dataset and 30 columns are present in the Dataset.

The unwanted columns which is to be removed from the dataset are by using some excel commands blank spaces are filled with the zeros.

The data which is represented with the dummy.

RAW DATASET :

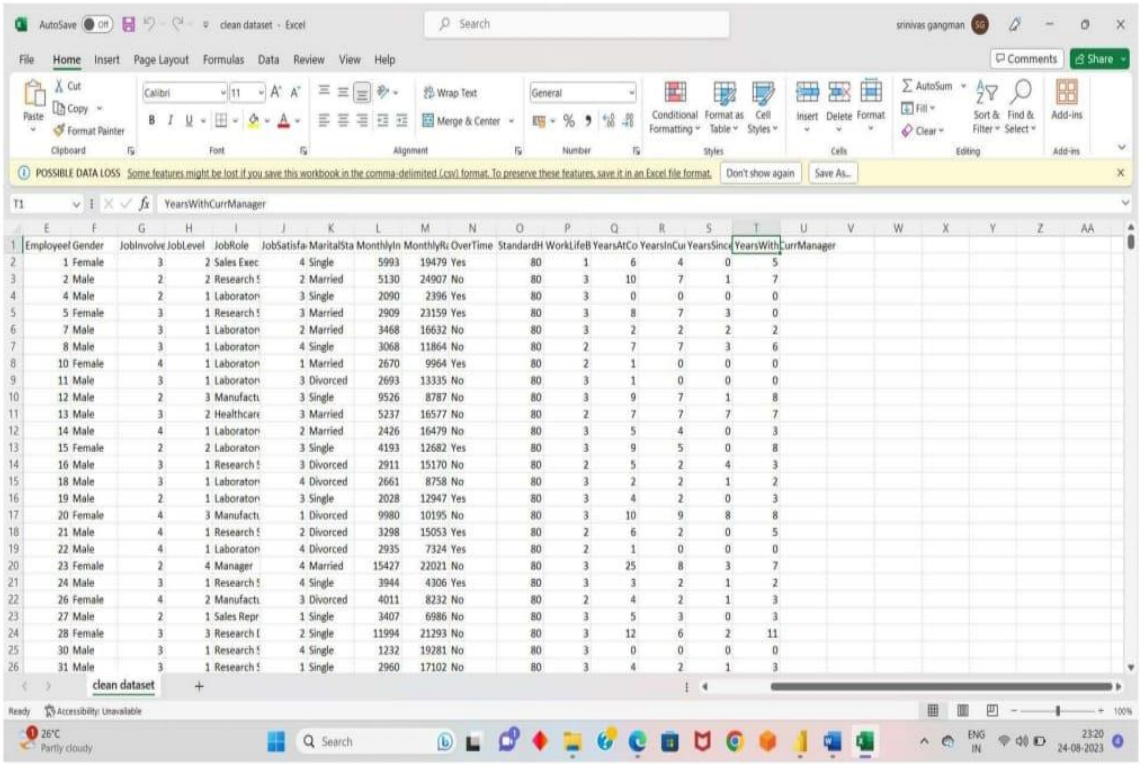
A raw dataset refers to a collection of data that has been collected or generated without undergoing any significant processing, transformation, or analysis. It's the initial, unprocessed state of data as it is collected or generated from its source.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfac	Gender	HourlyRate	JobInvolvement	JobLevel
41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	2	Female	94	3	
49	No	Travel_Frequently	279	Research & Develop	8	1	Life Sciences	1	2	3	Male	61	2	
37	Yes	Travel_Rarely	1373	Research & Develop	2	2	Other	1	4	4	Male	92	2	
33	No	Travel_Frequently	1392	Research & Develop	3	4	Life Sciences	1	5	4	Female	56	3	
27	No	Travel_Rarely	591	Research & Develop	2	1	Medical	1	7	1	Male	40	3	
32	No	Travel_Frequently	1005	Research & Develop	2	2	Life Sciences	1	8	4	Male	79	3	
59	No	Travel_Rarely	1324	Research & Develop	3	3	Medical	1	10	3	Female	81	4	
30	No	Travel_Rarely	1358	Research & Develop	24	1	Life Sciences	1	11	4	Male	67	3	
38	No	Travel_Frequently	216	Research & Develop	23	3	Life Sciences	1	12	4	Male	44	2	
36	No	Travel_Rarely	1299	Research & Develop	27	3	Medical	1	13	3	Male	94	3	
35	No	Travel_Rarely	809	Research & Develop	16	3	Medical	1	14	1	Male	84	4	
29	No	Travel_Rarely	153	Research & Develop	15	2	Life Sciences	1	15	4	Female	49	2	
31	No	Travel_Rarely	670	Research & Develop	26	1	Life Sciences	1	16	1	Male	31	3	
34	No	Travel_Rarely	1346	Research & Develop	19	2	Medical	1	18	2	Male	93	3	
28	Yes	Travel_Rarely	103	Research & Develop	24	3	Life Sciences	1	19	3	Male	50	2	
29	No	Travel_Rarely	1389	Research & Develop	21	4	Life Sciences	1	20	2	Female	51	4	
32	No	Travel_Rarely	334	Research & Develop	5	2	Life Sciences	1	21	1	Male	80	4	
22	No	Non-Travel	1123	Research & Develop	16	2	Medical	1	22	4	Male	96	4	
53	No	Travel_Rarely	1219	Sales	2	4	Life Sciences	1	23	1	Female	78	2	
38	No	Travel_Rarely	371	Research & Develop	2	3	Life Sciences	1	24	4	Male	45	3	
24	No	Non-Travel	673	Research & Develop	11	2	Other	1	26	1	Female	96	4	
36	Yes	Travel_Rarely	1218	Sales	9	4	Life Sciences	1	27	3	Male	82	2	
34	No	Travel_Rarely	419	Research & Develop	7	4	Life Sciences	1	28	1	Female	53	3	
21	No	Travel_Rarely	391	Research & Develop	15	2	Life Sciences	1	30	3	Male	96	3	
34	Yes	Travel_Rarely	699	Research & Develop	6	1	Medical	1	31	2	Male	83	3	
53	No	Travel_Rarely	1282	Research & Develop	5	3	Other	1	32	3	Female	58	3	
32	Yes	Travel_Frequently	1125	Research & Develop	16	1	Life Sciences	1	33	2	Female	72	1	
42	No	Travel_Rarely	691	Sales	8	4	Marketing	1	35	3	Male	48	3	
44	No	Travel_Rarely	477	Research & Develop	7	4	Medical	1	36	1	Female	42	2	
46	No	Travel_Rarely	705	Sales	2	4	Marketing	1	38	2	Female	83	3	

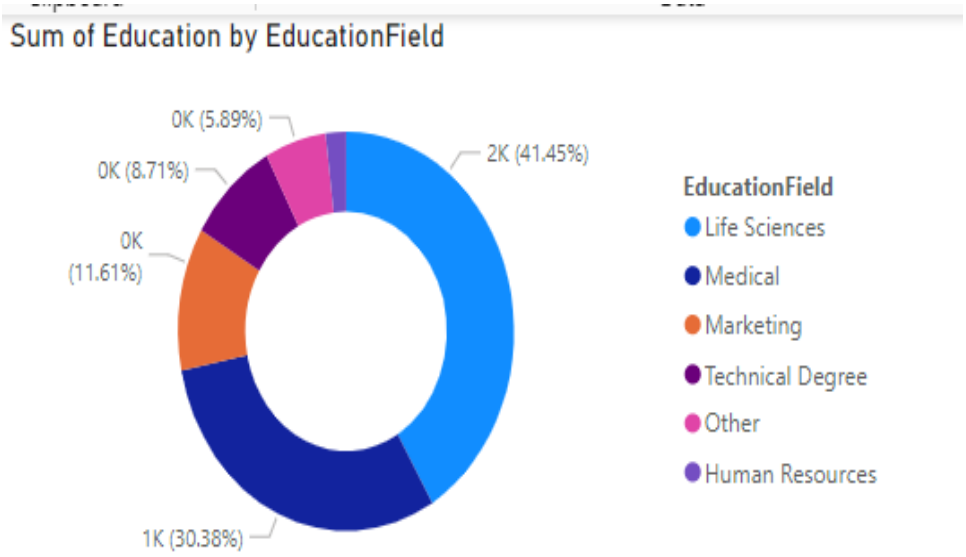
CLEANED OPTIMIZED DATASET :

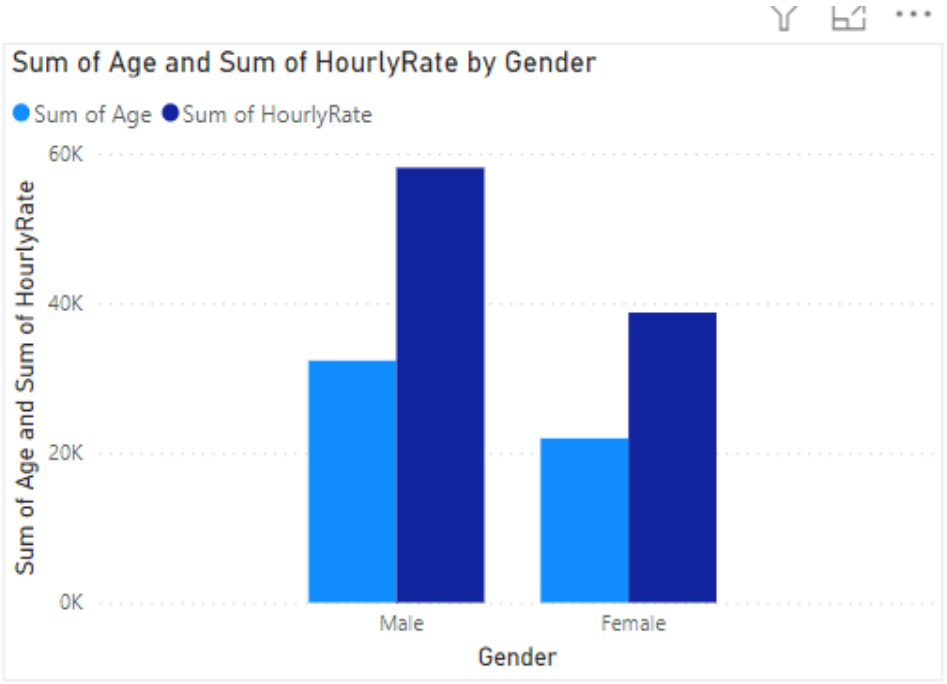
A cleaned and optimized dataset refers to a collection of data that has undergone various preprocessing steps to ensure its quality, consistency, and suitability for analysis or other purposes. This dataset has been refined, corrected, and structured in a way that makes it easier to work with and yields more accurate and meaningful results.

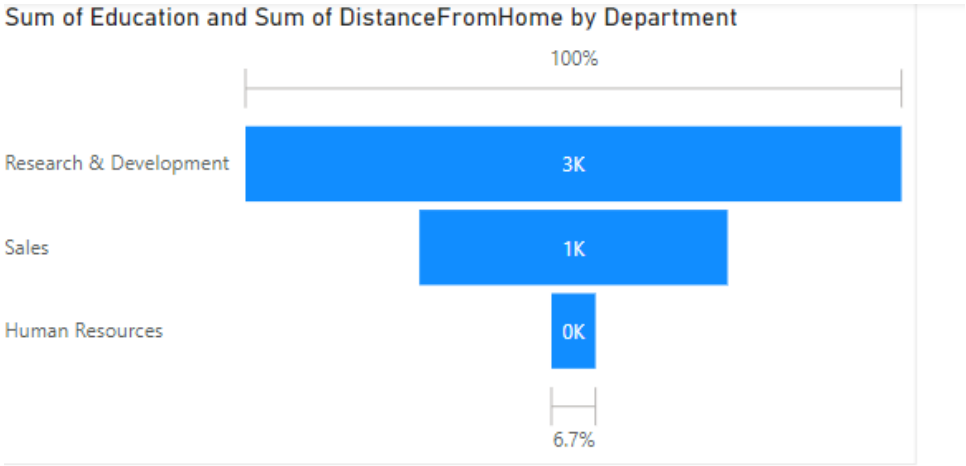
After removing the unwanted columns, the optimized data is enabled with the 19 columns.

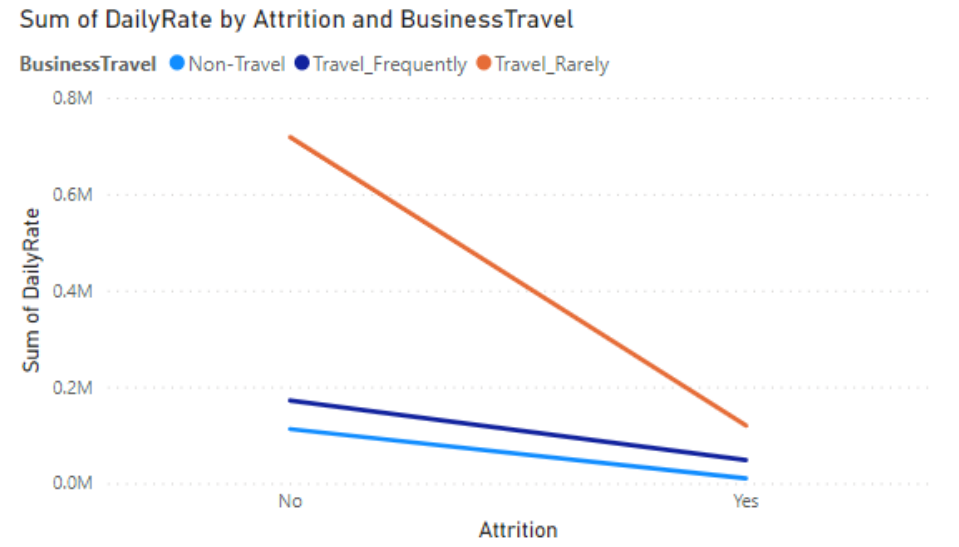


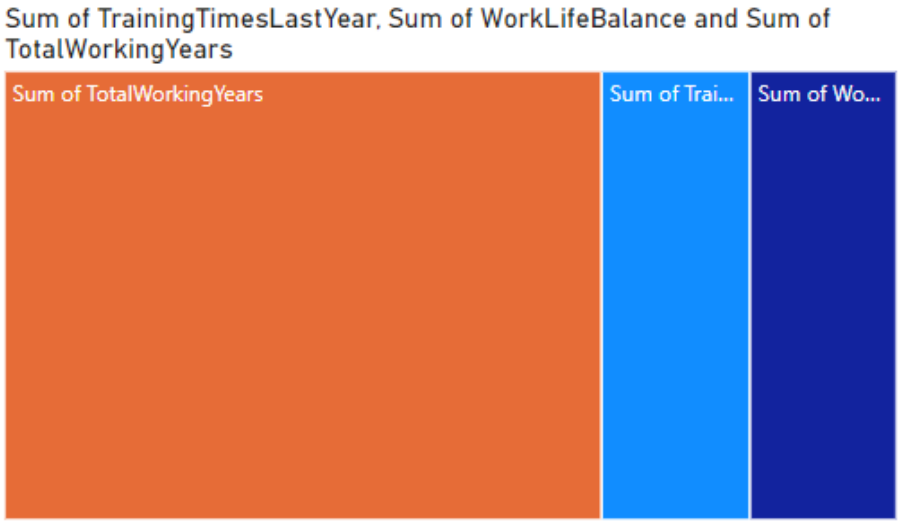
DASH BOARDS :

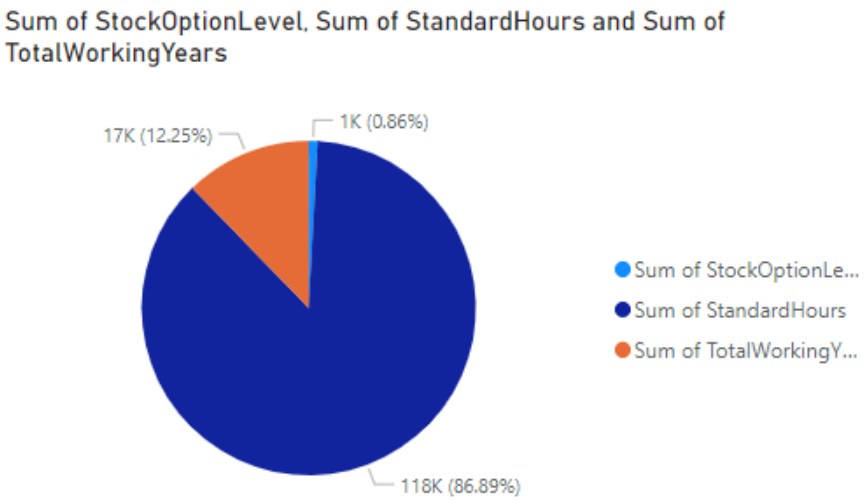

Sum of Education by EducationField
0K (5.89%)
2K (41.45%)
0K (8.71%)
0K
(11.61%)
1K (30.38%)
EducationField
Life Sciences
Medical
Marketing
Technical Degree
Other
Human Resources


Sum of Age and Sum of HourlyRate by Gender
Sum of Age
Sum of HourlyRate
60K
40K
20K
0K
Sum of Age and Sum of HourlyRate
Male
Female
Gender


Sum of Education and Sum of DistanceFromHome by Department
100%
Research & Development
3K
Sales
1K
Human Resources
0K
6.7%


Sum of DailyRate by Attrition and BusinessTravel
BusinessTravel
Non-Travel
Travel_Frequently
Travel_Rarely
0.8M
0.6M
0.4M
0.2M
0.0M
Sum of DailyRate
No
Yes
Attrition


Sum of TrainingTimesLastYear, Sum of WorkLifeBalance and Sum of TotalWorkingYears
Sum of TotalWorkingYears
Sum of Trai...
Sum of Wo...



.....

INSIGHTS :

1. Sum of Education by EducationField :

In this visualization they have sum of Education by EducationField. In EducationField consists of Life Sciences, Medical, Marketing, Technical Degree, Other and Human resources.

Life Sciences has the 1775 (41.45%), Medical has the 1301 (30.38%), Marketing has the 497 (11.61%), Technical Degree has the 373 (8.71%), Other has the 252 (5.89%), Human Resources has 84(1.96%).

Life sciences has the highest percentage of education than the Human Resources. So my recommendation is to give the best services to the life sciences.

2. Sum of Age and Sum of HourlyRate by Gender :

In this visualization they have Sum of Age and Sum of HourlyRate by Gender. In gender consists of Male and Female.

Male has the 58109 of Sum of Hourly Rate and 32328 of Sum of Age. Whereas Female has the 38751 of Sum of Hourly Rate and 21950 of Sum of Age.

Male has the highest Sum of Hourly Rate and Sum of Age than the Female.

3. Sum of Education and Sum of DistanceFromHome by Department :

In this visualization they have Sum of Education and Sum of DistanceFromHome by Department, Department consists of Research Development, Sales, and Human Resources.

Research Development has 3Kms, Sales has 1K, and Human Resources has 0Km from home. Total Distance is 100%, Research Development has 100%, Human Resources has 6.7%.

Research Development has the highest distance than the Human Resources.

4. **Sum of Daily Rate by attrition and Business Travel :**

In this visualization Sum of Daily Rate by attrition and Business Travel,

Business Travel consists of Non_Travel, Travel_Frequently, Travel_Rarely, whereas attrition consists of Yes and No.

Non_Travel people have 10163, Travel_Frequently people have 48193, whereas Travel_Rarely people have 119479. So the Travel_Rarely people has the highest than the Non_Travel people.

5. **Sum of Training TimesLastYear, Sum of WorkLifeBalance and Sum of TotalWorkingYears :**

In this visualization they have Sum of Training TimesLastYear, Sum of WorkLifeBalance and Sum of TotalWorkingYears.

Sum of Training TimesLastYear have 4115 people, Sum of WorkLifeBalance have 4051 people and Whereas Sum of TotalWorkingYears have 16181 people.

Sum of TotalWorkingYears have the highest people than the Sum of WorkLifeBalance.

6. **Sum of StockOptionalLevel, Sum of standardHours and Sum of TotalWorkingYears :**

In this visualization they have Sum of StockOptionalLevel, Sum of standardHours and Sum of TotalWorkingYears.

Sum of TotalWorkingYears has 16581 (12.25%), Sum of standardHours has 117600 (86.89%), whereas Sum of StockOptionalLevel has 1167 (0.86%).

Sum of standardHours has the highest percentage than the StockOptionalLevel. So my recommendation is to give the standardHours to the employees.

RECOMMENDATIONS TO THE IBM BASED UPON THE VISUALS WHICH WE OBSERVED EFFICIENTLY :

1. **Clustered Column Chart :**

Clustered column chart talks about the age and gender. By observing the clustered column chart efficiently and from the chart the 30 aged people are more.

2. **Line and Clustered Column Chart :**

My recommendation to the IBM is sales for the technology is very high. The technology must be increased so that organization sales will also increases effieciently then company stands first position in the market.

Product manufacturing must be decreased to avoid the organization from the loss.

THANK YOU !