EMPLOYEE ATTRITION DATASET ANALYSIS

INTRODUCTION

Employee attrition, or turnover, is a critical concern for organizations as it directly impacts productivity, morale, and the financial health of a company. Understanding the factors contributing to employee attrition is essential for developing strategies to retain talent and reduce turnover rates.

This analysis aims to explore an employee attrition dataset to identify key factors influencing employees' decisions to leave the organization. By leveraging statistical methods and machine learning techniques, we will examine variables such as job satisfaction, work-life balance, compensation, career progression, and organizational culture, among others.

AIM

The aim of analysing an employee attrition dataset is to identify the key factors that contribute to why employees leave an organization, such as job satisfaction, work environment, compensation, and career growth opportunities.

By developing predictive models, the analysis seeks to pinpoint employees who are at a higher risk of leaving, enabling organizations to take proactive measures to retain valuable talent. Additionally, the analysis provides data-driven insights to inform and optimize human resource strategies aimed at reducing turnover and enhancing employee retention. Ultimately, this analysis supports decision-making processes that foster a positive work environment and improve employee satisfaction.

OBJECTIVES

1. Identify Key Drivers: Determine the primary factors that influence employee attrition within the organization.

2. Predict Attrition: Designed and executed SQL queries to extract relevant information from the database to forecast which employees are most likely to leave.

3. Profile High-Risk Groups: Identify specific employee segments (e.g., by department, role) that are more prone to attrition.

4. Inform HR Strategies: Provide actionable insights to help HR develop and refine strategies to improve employee retention.

5. Enhance Employee Satisfaction: Use findings to recommend changes in policies or practices that can increase job satisfaction and reduce turnover.

DATA OVERVIEW

Columns	Datatype
EmployeeID	int
Age	int
Attrition	text
BusinessTravel	text
DailyRate	int
Department	text
DistanceFromHome	int
Education	int
EducationField	text
EmployeeCount	int
EnvironmentSatisfaction	int
Gender	text
HourlyRate	int
JobInvolvement	int
JobLevel	int
JobRole	text
JobSatisfaction	int
MaritalStatus	text
MonthlyIncome	int
MonthlyRate	int
?	text
salary_category	varchar(50)
age_category	varchar

DATA ANALYSIS

1. Descriptive Analysis: Summarize data to understand overall attrition trends, average tenure, and most affected departments.

2. Trend Analysis: Identify patterns over time, such as monthly or yearly attrition rates and changes in employee satisfaction.

3. Employee Segmentation: Group employees based on demographics or job roles to tailor retention strategies.

4. Performance Analysis: Analyze performance metrics, including average productivity and its correlation with attrition.

5. Attrition Reasons Analysis: Understand the reasons and frequency of employee attrition across different segments.

QUESTIONS

1.NUMBER OF MALE AND FEMALE WORKERS

select gender,count(*) from employee_attirtion_new group by gender;

gender	count(*)
Female	678
Male	998

2.AVERAGE AGE OF FEMALE AND MALE WORKERS

select gender,avg(age) from employee_attirtion_new group by gender;

gender	avg(age)
Female	37.3274
Male	36.5531

3.AVERAGE DAILYRATE FOR THE PEOPLE WHO LEFT AND STAYED

```
select attrition,avg(dailyrate) from
employee_attirtion_new group by attrition;
```

attrition	avg(dailyrate)
No	808.4997
Yes	741.6131

4.NUMBER OF PEOLE WITH HIGH,LOW,AVERAGE SALARY

```
update employee_attirtion_new set salary_category=case
when monthlyincome<5000 then 'Low salary'

when monthlyincome>=5000 and
monthlyincome<=10000 then 'Average salary'

else 'High salary' end ;

select salary_category,count(*) from
employee_attirtion_new group by salary_category;
```

salary_category	count(*)
Average salary	492
Low salary	859
High salary	325

5.COUNT OF PEOPLE LEFT FROM EACH SALARY_CATEGORY

```
select salary_category ,count(attrition) as p_left from employee_attirtion_new where attrition='Yes' group by salary_category;
```

salary_category	p_left
Low salary	153
Average salary	34
High salary	12

6.COUNT OF MALE AND FEMALE LEFT FROM EACH SALARY_CATEGORY

```
select salary_category, count(*) as m_left from employee_attirtion_new where attrition='Yes'and gender='Male' group by salary_category;

select salary_category ,count(*) as f_left from employee_attirtion_new where attrition='Yes'and gender='Female' group by salary_category;
```

salary_category	m_left	f_left
Low salary	81	72
Average salary	22	12
High salary	10	2

7.NUMBER OF PEOPLE LEFT FROM EACH DEPARTMENT

```
select department,count(attrition) as num_left from
employee_attirtion_new where attrition='Yes' group by
department;
```

department	num_left
Maternity	98
Cardiology	74
Neurology	27

8.NUMBER OF MALE AND FEMALE LEFT

```
select gender,count(attrition) from
employee_attirtion_new where attrition='Yes' group by
gender;
```

Gender	p_left
Female	86
Male	113

9.AGE CATEGORISED AND NUMBER OF PEOPLE IN EACH CATEGORY

```
alter table employee_attirtion_new add column
age_category varchar(50);
```

```
update employee_attirtion_new set age_category=case
when age between 18 and 30 then 'Young adults'
when age between 30 and 50 then 'Middle aged adults'
else 'Older adults' end ;
select age_category,count(*) from employee_attirtion_new
group by age_category;
```

age_category	count(*)
Middle aged adults	1066
Young adults	448
Older adults	162

10.NUMBER OF PEOPLE LEFT IN EACH AGE CATEGORY

```
select age_category,count(attrition) as num_left from
employee_attirtion_new where attrition='Yes' group by
age_category;
```

age_category	num_left
Middle aged adults	80
Young adults	109
Older adults	10

11.TOTAL PEOPLE IN EACH BUSINESS TRAVEL CATEGORY

select businesstravel,count(*) as count from employee_attirtion_new group by businesstravel;

businesstravel	count(*)
Travel_Rarely	1184
Travel_Frequently	320
Non-Travel	172

12.COUNT OF MALE AND FEMALE LEFT FROM EACH BUSINESSTRAVELCATEGORY

select businesstravel,count(*) as F_count from employee_attirtion_new where gender='Female' and attrition='Yes' group by businesstravel;

select businesstravel,count(*) as M_count from employee_attirtion_new where gender='Male' and attrition='Yes' group by businesstravel;

businesstravel	f_count	m_count
Travel_Frequently	31	22
Travel_Rarely	49	77
Non-Travel	6	10

13.CORRELATION BETWEEN ATTRITIONAND DISTANCE

SELECT attrition,AVG(distancefromhome) AS average_distance FROM employee_attirtion_new GROUP BY attrition;

attrition	average_distance
No	8.9059
Yes	11.5678

14.COUNT OF WORKERS WHO LEFT IN EACH CATEGORY JOB SATISFACTION

select jobsatisfaction,count(*)as count_yes from employee_attirtion_new where attrition='Yes' group by jobsatisfaction;

jobsatisfaction	count_yes
2	42
4	45
1	52
3	60

15.CORRELATION OF JOB SATISFACTION TO ATTRITION

select attrition,avg(jobsatisfaction) from employee_attirtion_new group by attrition;

attrition	avg(jobsatisfaction)
No	2.4925
Yes	2.7718

CONCLUSION

1. The total number of female workers are 678 and male workers are 998.

2. The average age of the females and male workers are 37 and 36.

3. The average of daily rate given to the workers who left is lower than the people who stayed. By increasing the wage of the workers can affect the decision the workers whether to stay or not.

4. There is only 325 workers with salary higher than 10k and almost 12 workers have already left.

5. Most of the workers with a low salary is found to have left the company because of their low income .

6. From the workers who left from the high salary category most of them are men and only 2 are women.

7. There are 3 departments Maternity , Cardiology and Neurology with number of workers who left the company to be 98,74 and 27.

8. From each category of age most of the workers are found to be middle aged.

9. From the middle aged people almost 80 left more people are left from an age group of young adults.

10. Most of the people travel rarely and 172 workers do not travel .

11. The people who left from the company has the longest distance from home to their workplace. If necessary stay is provided near the workplace these number can be reduced.

12. Most of the workers are not quite satisfied with their overall assistance provided by the company