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.

Objective:

The objectives of this report are:

- To determine the primary factors contributing to employee attrition.
- To analyze the distribution and fairness of salaries across different roles, departments, and demographics.
- To examine the impact of age, education, and job involvement on employee satisfaction and turnover.
- To identify patterns in employee demographics, such as age, gender, education, and their effect on career progression
- To provide actionable insights for improving employee retention and satisfaction

Data Overview:

Column name	Data type
EmployeeID	int
Age	int
Age_category	text
Attrition	text
BusinessTravel	text
DailyRate	int
Department	text
DistanceFromHome	int
Education	int
EducationField	text
EnvironmentSatisfaction	int
Gender	text
HourlyRate	int
JobInvolvement	int
JobLevel	int
JobRole	text
JobSatisfaction	int
MaritalStatus	text
MonthlyIncome	int
Salary_category	text
MonthlyRate	int
NumCompaniesWorked	int
OverTime	text
PercentSalaryHike	int
PerformanceRating	int
RelationshipSatisfaction	int
StandardHours	int
Shift	int
TotalWorkingYears	int
TrainingTimesLastYear	int
WorkLifeBalance	int
YearsAtCompany	int
YearsInCurrentRole	int
YearsSinceLastPromotion	int
YearsWithCurrManager	int
EmployeeID	int

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 FEMALE AND MALE WORKERS
   -select Gender,count(*) as Count from employee_attirtion group by gender;

Gender	Count
Female	678
Male	998

2. AVERAGE AGE OF FEMALE MALE WORKERS
   -select Gender,avg(age) as Average_age from employee_attirtion group by gender;

Gender	Average_age
Female	37.3274
Male	36.5531

3. AVERAGE DAILYRATE OF THE PEOPLE WHO LEFT AND STAYED
   -select Attrition,avg(Dailyrate) as Average_Dailyrate from employee_attirtion group by Attrition;

Attrition	Average_Dailyrate
No	808.4997
Yes	741.6131

4. AGE CATEGORISED AND NUMBER OF PEOPLE IN EACH CATEGORY
   -alter table employee_attirtion add column Age_category text after Age;
   -update employee_attirtion set
   Age_category=case when Age between 16 and 24 then "Young Workers"
   when age between 25 and 34 then "Early Career"
   when age between 35 and 44 then "Mid-career"
   when age between 45 and 54 then "Established Career" else "Late Career" end;
   - select Age_category,count(*)as count from employee_attirtion where attrition='yes'
   group by Age_category order by count(*);

Age_category	Count
Late Career	76
Young Workers	109
Established Career	282
Mid-career	573
Early Career	636

5. NUMBER OF PEOPLE LEFT IN EACH AGE CATEGORY
   -select Age_category,count(*)as count from employee_attirtion where attrition='yes' group by Age_category order by count(*);

Age_category	Count
Late Career	7
Established Career	10
Mid-career	31
Young Workers	45
Early Career	106

6. TOTAL PEOPLE IN EACH BUSINESS TRAVEL CATEGORY
   - select businesstravel,count(*) as count from employee_attirtion group by businesstravel;

businesstrvel	count
Travel_Frequently	320
Travel_Rarely	1184
Non-Travel	172

7. COUNT OF MALE AND FEMALE LEFT FROM EACH BUSINESSTRAVEL CATEGORY
   -select businesstravel,count(*)as male_count from employee_attirtion where gender="male" and attrition ='yes' group by businesstravel;
   -select businesstravel,count(*)as female_count from employee_attirtion where gender="female" and attrition ='yes' group by businesstravel;

businesstravel	male_count	female_count
Travel_Rarely	77	49
Non-Travel	10	6
Travel_Frequently	26	31

8. NUMBER OF PEOLE WITH HIGH,LOW,AVERAGE SALARY
   -alter table employee_attirtion add column salary_category text after monthlyincome;
   -update employee_attirtion 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(*)as count from employee_attirtion group by salary_category;

salary_category	count
Low salary	859
High salary	325
Average salary	492

9. GENTER-BASED SALARY DISPARITY
   -select Gender,avg(MonthlyIncome) as AverageMonthlyIncome,
   min(MonthlyIncome) as MinMonthlyIncome,max(MonthlyIncome) as
   MaxMonthlyIncome from Employee_attirtion group by Gender
   order by AverageMonthlyIncome desc;

Gender	AverageMonthlyIncome	MinMonthlyIncome,	MaxMonthlyIncome
Female	6702.8746	1129	19973
Male	6389.9058	1009	19999

10. COUNT OF PEOPLE LEFT FROM EACH SALARY_CATEGORY
    - select salary_category ,count(attrition) as count from employee_attirtion where
      attrition='Yes' group by salary_category;

salary_category	count
Low salary	153
Average salary	34
High salary	12

11. COUNT OF MALE AND FEMALE LEFT FROM EACH SALARY_CATEGORY
    - select salary_category, count(*) as male_count from employee_attirtion where
      attrition='Yes'and gender='Male' group by salary_category;
    - select salary_category ,count(*) as female_count from employee_attirtion where
      attrition='Yes'and gender='Female' group by salary_category;

salary_category	male_count	Female_count
Low salary	81	72
Average salary	22	12
High salary	10	2

12. HIGHEST COUNT OF PEOPLE WHO LEFT BASED OF PERCENTSALARYHIKE
    - select PercentSalaryHike,count(*) as count from employee_attirtion where
      attrition="yes" group by PercentSalaryHike order by count(*) desc limit 4;

PercentSalaryHike	count
13	34
11	33
12	24
14	20

13. LOWEST COUNT OF PEOPLE WHO LEFT BASED OF PERCENTSALARYHIKE
    - select PercentSalaryHike,count(*) as count from employee_attirtion where attrition="yes" group by PercentSalaryHike order by count(*) limit 4;

PercentSalaryHike	count
25	1
24	4
21	4
23	6

14. COMPARE JOB SATISFACTION
    - select Department,avg(JobSatisfaction) as AverageJobSatisfaction,count(EmployeeID) as TotalEmployees from Employee_attirtion group by Department order by AverageJobSatisfaction desc;

Department	AverageJobSatisfaction	TotalEmployees
Neurology	2.7622	349
Cardiology	2.7382	531
Maternity	2.7286	796

15. NUMBER OF PEOPLE LEFT FROM EACH DEPARTMENT
    - select department,count(attrition) as num_left from employee_attirtion where attrition='Yes' group by department;

department	num_left
Maternity	98
Cardiology	74
Neurology	27

16. NUMBER OF MALE AND FEMALE LEFT
    - select gender,count(*) from employee_attirtion where attrition='Yes' group by gender;

gender	count(*)
Male	113
Female	86

17. EVALUVATE THE IMPACT OF ATTRITION OF FINANCIALS BY ANALYSING THE COST
    - select JobRole,sum(MonthlyIncome) as TotalAttritionCost from Employee_attirtion where Attrition = 'Yes' group by JobRole order by TotalAttritionCost desc;

JobRole	TotalAttritionCost
Nurse	487514
Other	261312

Therapist	32753
Administrative	19246

18. ANALYZE THE MONTHLYINCOME, HOURLYRATE, AND DAILYRATE IN CONJUNCTION WITH JOBROLE, DEPARTMENT, EDUCATIONFIELD, JOBLEVEL, AND GENDER TO IDENTIFY PAY EQUITY AND ANY DISPARITIES
    - select JobRole,Department,gender,AVG(MonthlyIncome) as AverageMonthlyIncome, MIN(MonthlyIncome) as MinMonthlyIncome,MAX(MonthlyIncome) as MaxMonthlyIncome from Employee_attirtion
    group by JobRole, Department,gender order by AverageMonthlyIncome desc;

JobRole	Department	gender	AverageMonthly Income	MinMonthly Income	MaxMonthly Income
Admin	Maternity	Female	19187.0000	19187	19187
Admin	Cardiology	Male	18213.0000	18213	18213
Administrative	Neurology	Male	17817.2143	13348	19717
Admin	Neurology	Male	17813.0000	16437	19189
Administrative	Cardiology	Female	17016.2632	12504	19845
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

19. CORRELATION BETWEEN ATTRITIONAND DISTANCE
    - select attrition,avg(jobsatisfaction) from employee_attirtion group by attrition;

attrition	avg(jobsatisfaction)
No	2.7718
Yes	2.4925

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. From the workers who left from the high salary category most of them are men and only 2 are women.
6. There are 3 departments Maternity , Cardiology and Neurology with number of workers who left the company to be 98,74 and 27.
7. From each category of age most of the workers are found to be middle aged.

8. From the middle aged people almost 80 left more people are left from an age group of young adults.
9. Most of the people travel rarely and 172 workers do not travel.
10. 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.
11. Most of the workers are not quite satisfied with their overall assistance provided by the company.
12. Admin from neurology department have the highest average monthly income
13. Cardiology department workers have the lowest average monthly income