HR Analytics

MySQL

www.linkedin.com/in/reshma-reghunath

ABOUT DATASET

The HR Analytics dataset contains detailed employee-level information collected from an organization’s Human Resources department.

It is designed to analyze employee attrition, satisfaction, compensation, and career progression — helping identify the key drivers that influence employee turnover and retention.

.

Dataset Overview:

- The dataset contains detailed HR records such as:
- Employee demographics (Age, Gender, MaritalStatus)
- Job information (Department, JobRole, YearsAtCompany)
- Compensation details (MonthlyIncome, Overtime)
- Satisfaction metrics (JobSatisfaction, EnvironmentSatisfaction, WorkLifeBalance)
- Attrition status (Yes/No)

OBJECTIVE

The primary objective of this project is to analyze employee data using advanced SQL techniques to uncover key insights about workforce behavior, attrition patterns, and performance trends.

By leveraging SQL for data transformation, aggregation, and analytics, this project aims to help the HR department make data-driven decisions that reduce employee turnover and improve overall satisfaction and productivity.

1. Identify Top 5 Departments by Attrition Rate

```
SELECT
    Department,
    COUNT(CASE WHEN Attrition = 'Yes' THEN 1 END) * 100.0 / COUNT(*) AS AttritionRate
FROM hr_analytics_data
GROUP BY Department
ORDER BY AttritionRate DESC
LIMIT 5;
```

Output

Department	AttritionRate
Sales	20.62780
Human Resources	19.04762
Research & Development	13.83975

2. Average Monthly Income by JobRole (Aggregation + HAVING)

```
SELECT
    JobRole,
    ROUND(AVG(MonthlyIncome), 2) AS AvgSalary
FROM hr_analytics_data
GROUP BY JobRole
HAVING AVG(MonthlyIncome) > 5000
ORDER BY AvgSalary DESC;
```

Output

JobRole	AvgSalary
Manager	17181.68
Research Director	16033.55
Healthcare Representative	7528.76
Manufacturing Director	7295.14
Sales Executive	6924.28

3. Attrition Trend by Work Experience (CASE + Aggregate)

```
SELECT
    CASE
        WHEN YearsAtCompany < 3 THEN 'New (<3 years)'
        WHEN YearsAtCompany BETWEEN 3 AND 7 THEN
'Mid (3-7 years)'
        ELSE 'Senior (>7 years)'
    END AS ExperienceLevel,
    COUNT(CASE WHEN Attrition = 'Yes' THEN 1 END) AS
EmployeesLeft,
    COUNT(*) AS TotalEmployees,
    ROUND(COUNT(CASE WHEN Attrition = 'Yes' THEN 1
END) * 100.0 / COUNT(*), 2) AS AttritionRate
FROM hr_analytics_data
GROUP BY ExperienceLevel
ORDER BY
    CASE
        WHEN ExperienceLevel = 'New (<3 years)' THEN 1
        WHEN ExperienceLevel = 'Mid (3-7 years)' THEN 2
        WHEN ExperienceLevel = 'Senior (>7 years)' THEN 3
    END ASC;
```

Output

	ExperienceLevel	EmployeesLeft	TotalEmployees	AttritionRate
▶	New (<3 years)	102	342	29.82
	Mid (3-7 years)	80	600	13.33
	Senior (>7 years)	55	528	10.42

4. Create a View for Attrition Risk (JOIN + Derived Columns)

```
CREATE VIEW Attrition_Risk_View AS
SELECT

    Department,
    JobRole,
    MonthlyIncome,
    OverTime,
    JobSatisfaction,
    YearsSinceLastPromotion,
    CASE
        WHEN OverTime = 'Yes' AND JobSatisfaction <= 2
THEN 'High Risk'
        WHEN OverTime = 'No' AND JobSatisfaction <= 2
THEN 'Medium Risk'
        ELSE 'Low Risk'
    END AS RiskLevel
FROM hr_analytics_data;
SELECT * FROM Attrition_Risk_View WHERE
RiskLevel='High Risk';
```

Department	JobRole	MonthlyIncome	OverTime	JobSatisfaction	YearsSinceLastPromotion	RiskLevel
Research & Development	Laboratory Technician	2670	Yes	1	0	High Risk
Research & Development	Research Scientist	3298	Yes	2	0	High Risk
Research & Development	Research Scientist	3919	Yes	1	6	High Risk
Research & Development	Research Scientist	3419	Yes	1	0	High Risk
Sales	Sales Executive	5454	Yes	1	1	High Risk
Research & Development	Healthcare Representative	9884	Yes	1	2	High Risk
Research & Development	Laboratory Technician	4014	Yes	1	2	High Risk
Research & Development	Research Director	13245	Yes	1	0	High Risk

5. Combine Two Categories using UNION (Attrition vs Active Employees)

```
SELECT
    'Attrited Employees' AS Category,
    AVG(MonthlyIncome) AS
AvgIncome,
    AVG(YearsAtCompany) AS
AvgTenure
FROM hr_analytics_data
WHERE Attrition = 'Yes'

UNION ALL

SELECT
    'Active Employees' AS Category,
    AVG(MonthlyIncome),
    AVG(YearsAtCompany)
FROM hr_analytics_data
WHERE Attrition = 'No';
```

Output

	Category	AvgIncome	AvgTenure
▶	Attrited Employees	4787.0928	5.1308
	Active Employees	6832.7397	7.3690

6.Calculate the number of employees who left, the total number of employees, and the attrition rate (%) for each department and gender combination.

```
SELECT
    Department,
    Gender,
    COUNT(CASE WHEN Attrition = 'Yes'
THEN 1 END) AS EmployeesLeft,
    COUNT(*) AS TotalEmployees,
    ROUND(
        COUNT(CASE WHEN Attrition =
'Yes' THEN 1 END) * 100.0 / COUNT(*),
        2
    ) AS AttritionRate
FROM hr_analytics_data
GROUP BY Department, Gender
ORDER BY Department;
```

Output

	Department	Gender	EmployeesLeft	TotalEmployees	AttritionRate
	Human Resources	Male	6	43	13.95
	Research & Development	Female	43	379	11.35
	Research & Development	Male	90	582	15.46
	Sales	Female	38	189	20.11
	Sales	Male	54	257	21.01