Project Title: HR Employee Attrition Analysis Using MySQL

Objective

The goal of this project is to analyze employee attrition patterns using IBM HR analytics data. We aim to:

- Understand attrition rates by job role, age, department, etc.
- Identify high-risk attrition segments.
- Explore business insights to reduce employee turnover.

Tools Used

- MySQL 8+ (Workbench + CLI)
- CSV Dataset from IBM (HR Analytics)
- SQL Queries for cleaning and analysis

Dataset Overview

- **Source:** IBM HR Analytics Dataset
- **File Used:** hr_employee_attrition.csv
- **Original Rows:** 1,470
- **Final Cleaned Rows:** 1,470 (after checking missing values, constant columns, etc.)

Data Import Steps

CREATE DATABASE HRAnalytics;

USE HRAnalytics;

Data Cleaning Steps

USE HRAnalytics;

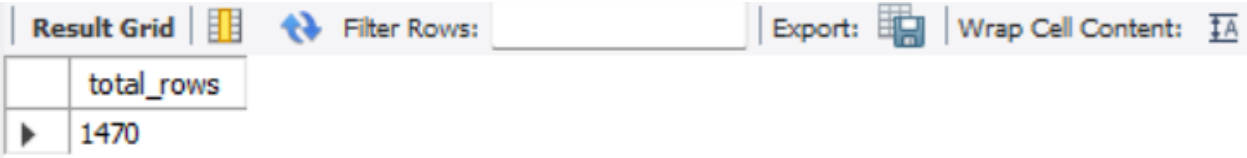
-- Fix BOM issue if first column name appears as ï»¿Age

ALTER TABLE hr_employee_attrition

CHANGE `ï»¿Age` Age INT;

-- 1. Total Rows

SELECT COUNT(*) AS total_rows FROM hranalytics.`hr_employee_attrition`;



total_rows
1470

-- 2. Check for NULL or 0 values

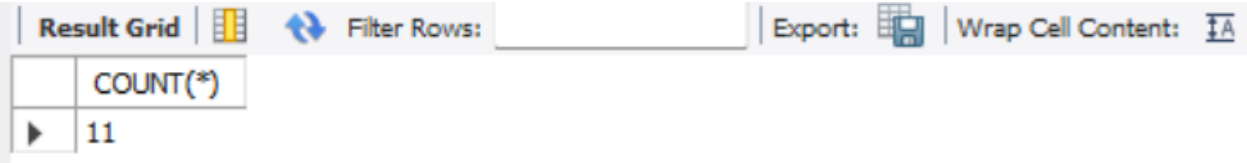
SELECT COUNT(*) FROM hranalytics.`hr_employee_attrition`

WHERE NumCompaniesWorked IS NULL OR TotalWorkingYears IS NULL;

COUNT(*)
0

SELECT COUNT(*) FROM hranalytics.`hr_employee_attrition`

WHERE MonthlyIncome = 0 OR TotalWorkingYears = 0;



COUNT(*)
11

-- 3. Replace missing values (if any found)

UPDATE hranalytics.`hr_employee_attrition` SET NumCompaniesWorked = 1

WHERE NumCompaniesWorked IS NULL;

UPDATE hranalytics.`hr_employee_attrition` SET TotalWorkingYears = 1

WHERE TotalWorkingYears IS NULL;

-- 4. Drop unnecessary columns (these are constant)

ALTER TABLE hranalytics.`hr_employee_attrition` DROP COLUMN Over18;

ALTER TABLE hranalytics.`hr_employee_attrition` DROP COLUMN StandardHours;

ALTER TABLE hranalytics.`hr_employee_attrition` DROP COLUMN EmployeeCount;

-- 5. Normalize boolean fields

```
ALTER TABLE hranalytics.`hr_employee_attrition` ADD is_attrition BOOLEAN;

UPDATE hranalytics.`hr_employee_attrition` SET is_attrition = (Attrition = 'Yes');


ALTER TABLE hranalytics.`hr_employee_attrition` ADD is_overtime BOOLEAN;

UPDATE hranalytics.`hr_employee_attrition` SET is_overtime = (OverTime = 'Yes');

-- Total Employees & Attrition

SELECT COUNT(*) AS total, SUM(is_attrition) AS attrited,
ROUND(SUM(is_attrition)/COUNT(*)*100, 2) AS attrition_rate

FROM hranalytics.`hr_employee_attrition`;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

total	attrited	attrition_rate
1470	237	16.12

```
-- Attrition by Job Role

SELECT JobRole, COUNT(*) AS total, SUM(is_attrition) AS attrited,

ROUND(SUM(is_attrition)/COUNT(*)*100, 2) AS attrition_rate

FROM hranalytics.`hr_employee_attrition`

GROUP BY JobRole

ORDER BY attrition_rate DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

JobRole	total	attrited	attrition_rate
Sales Representative	83	33	39.76
Laboratory Technician	259	62	23.94
Human Resources	52	12	23.08
Sales Executive	326	57	17.48
Research Scientist	292	47	16.10

```
-- Attrition by Age Group
SELECT
  CASE
    WHEN Age < 30 THEN '<30'
    WHEN Age BETWEEN 30 AND 40 THEN '30-40'
    WHEN Age BETWEEN 41 AND 50 THEN '41-50'
    ELSE '50+' END AS age_group,
  COUNT(*) AS total, SUM(is_attrition) AS attrited,
  ROUND(SUM(is_attrition)/COUNT(*)*100, 2) AS attrition_rate
FROM hranalytics.`hr_employee_attrition`
GROUP BY age_group;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

age_group	total	attrited	attrition_rate
41-50	322	34	10.56
30-40	679	94	13.84
<30	326	91	27.91
50+	143	18	12.59

```
-- Attrition by Overtime
SELECT is_overtime, COUNT(*) AS total, SUM(is_attrition) AS attrited,
ROUND(SUM(is_attrition)/COUNT(*)*100, 2) AS attrition_rate
FROM hranalytics.`hr_employee_attrition`
GROUP BY is_overtime;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

is_overtime	total	attrited	attrition_rate
1	416	127	30.53
0	1054	110	10.44

```
-- Monthly Income Comparison
SELECT
  is_attrition,
```

```
  ROUND(AVG(MonthlyIncome), 2) AS avg_income,
  ROUND(AVG(TotalWorkingYears), 2) AS avg_experience
FROM hranalytics.`hr_employee_attrition`
GROUP BY is_attrition;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

is_attrition	avg_income	avg_experience
1	4787.09	8.24
0	6832.74	11.86

```
-- Attrition by Distance from Home
SELECT DistanceFromHome, COUNT(*) AS total, SUM(is_attrition) AS attrited,
ROUND(SUM(is_attrition)/COUNT(*)*100, 2) AS attrition_rate
FROM hranalytics.`hr_employee_attrition`
GROUP BY DistanceFromHome
ORDER BY attrition_rate DESC
LIMIT 10;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

DistanceFromHome	total	attrited	attrition_rate
24	28	12	42.86
22	19	6	31.58
13	19	6	31.58
12	20	6	30.00
17	20	5	25.00

```
-- Attrition by Education Field
SELECT EducationField, COUNT(*) AS total, SUM(is_attrition) AS attrited,
ROUND(SUM(is_attrition)/COUNT(*)*100, 2) AS attrition_rate
FROM hranalytics.`hr_employee_attrition`
GROUP BY EducationField
ORDER BY attrition_rate DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

	EducationField	total	attrited	attrition_rate
▶	Human Resources	27	7	25.93
	Technical Degree	132	32	24.24
	Marketing	159	35	22.01
	Life Sciences	606	89	14.69
	Medical	464	63	13.58

Summary

- **Overall Attrition Rate:** ~16.1% (237 out of 1,470 employees)
- **High Attrition Jobs:** Sales Representatives, Laboratory Technicians
- **Age Groups at Risk:** Under 30 had the highest attrition (~39%)
- **Overtime Employees:** Show significantly higher attrition (~30%) vs non-overtime (~10%)
- **Income Impact:** Employees who left had lower average salaries and less experience
- **Education Field:** Human Resources and Life Sciences showed more attrition

---

Conclusion

This project demonstrated how SQL can be used to:

- Transform HR data into actionable insights
- Highlight attrition risks across job functions and demographics
- Suggest business strategies (e.g., monitoring young or overworked employees more closely)