HR Employee Attrition Analysis

1. Project Overview

This project analyzes employee attrition using the IBM HR Analytics dataset. The objective is to identify patterns and factors contributing to employee turnover and provide insights to help HR teams reduce attrition and improve retention strategies.

2. Dataset Summary

- **Source:** IBM HR Analytics Employee Attrition & Performance Dataset
- **Rows:** 1,470
- **Columns:** 35
- **Target Variable:** Attrition (Yes / No)
- **Key Features:**
  - Employee demographics (Age, Gender, Marital Status)
  - Job information (Department, Job Role, Job Level)
  - Compensation (Monthly Income, Salary Hike)
  - Work conditions (Overtime, Years at Company, Work-Life Balance)

3. Exploratory Data Analysis using Python

We began with data preparation and cleaning in Python:

- **Data Loading:** Imported the dataset using `pandas`.

- **Initial Exploration:** Used `df.info()` to check structure and `.describe()` for summary statistics.

	age	attrition	business_travel	department	distance_from_home	education
count	1470.000000	1470.000000	1470	1470	1470.000000	1470.000000
unique	NaN	NaN	3	3	NaN	NaN
top	NaN	NaN	Travel_Rarely	Research & Development	NaN	NaN
freq	NaN	NaN	1043	961	NaN	NaN
mean	36.923810	0.161224	NaN	NaN	9.192517	2.912925
std	9.135373	0.367863	NaN	NaN	8.106864	1.024165
min	18.000000	0.000000	NaN	NaN	1.000000	1.000000
25%	30.000000	0.000000	NaN	NaN	2.000000	2.000000
50%	36.000000	0.000000	NaN	NaN	7.000000	3.000000
75%	43.000000	0.000000	NaN	NaN	14.000000	4.000000
max	60.000000	1.000000	NaN	NaN	29.000000	5.000000

```
Data columns (total 35 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   age                       1470 non-null   int64
 1   attrition                 1470 non-null   object
 2   business_travel           1470 non-null   object
 3   daily_rate                1470 non-null   int64
 4   department                1470 non-null   object
 5   distance_from_home        1470 non-null   int64
 6   education                 1470 non-null   int64
 7   education_field           1470 non-null   object
 8   employee_count            1470 non-null   int64
 9   employee_number           1470 non-null   int64
 10  environment_satisfaction  1470 non-null   int64
 11  gender                    1470 non-null   object
```

- **Missing Data Handling:** Checked for null values across employee attributes such as age, income, job role, and work-life balance. No critical missing values affecting attrition analysis were found.

- **Column Standardization:** Renamed columns to **snake case** for better readability and documentation.

- **Feature Engineering:**

    - Created **age_group** column by binning employee ages into meaningful ranges for attrition comparison.
    - Created **salary_range** column based on monthly income to analyze attrition across income levels.
    - Derived **experience_group** using years at company to study attrition by tenure.

- **Data Consistency Check:** Verified binary and categorical columns such as **attrition** and **over_time** for valid and consistent values.

- **Database Integration:** Connected the cleaned Python Data Frame to **MySQL** and loaded the data into the database for further SQL-based analysis.

4. Data Analysis using SQL (Business Transactions)

We performed structured analysis in MYSQL to answer key business questions:

1. **Overall Attrition rate** – Calculated the overall percentage of employees who left the organization.

```
[12]: attrition
      0    83.877551
      1    16.122449
      Name: proportion, dtype: float64
```

2. **Attrition by Department** – Identified departments with higher employee attrition.

department	attrition_percentage
Sales	20.63
Research & Development	13.84
Human Resources	19.05

3. **Gender-wise Total Employees** – Shows workforce distribution by gender.

gender	total_employees
0	588
1	882

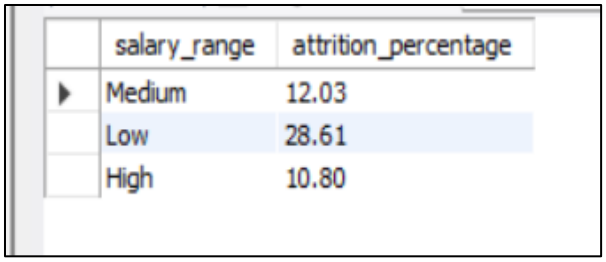
4. **High-Attrition Job Roles** – Identified job roles where attrition exceeds 20%

job_role	attrition_percentage
Laboratory Technician	23.94
Sales Representative	39.76
Human Resources	23.08

5. **Impact of Overtime on Attrition** – Analyzed whether working overtime is associated with higher attrition.

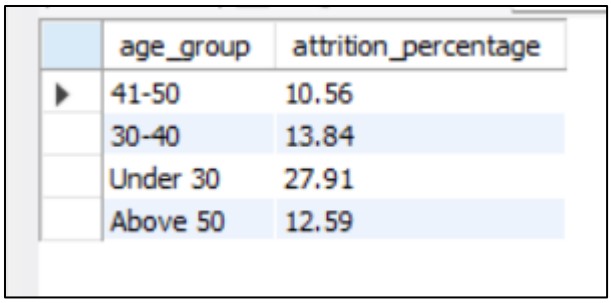
over_time	attrition_percentage
1	30.53
0	10.44

6. **Attrition by Salary Range** – Analyzed attrition trends across different income levels.



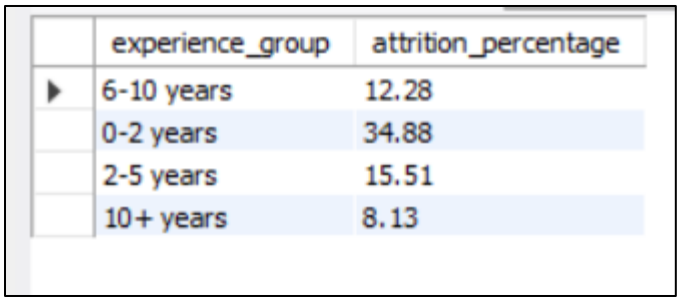
salary_range	attrition_percentage
Medium	12.03
Low	28.61
High	10.80

7. **Attrition by Age Group** – Examined attrition patterns across different age groups.



age_group	attrition_percentage
41-50	10.56
30-40	13.84
Under 30	27.91
Above 50	12.59

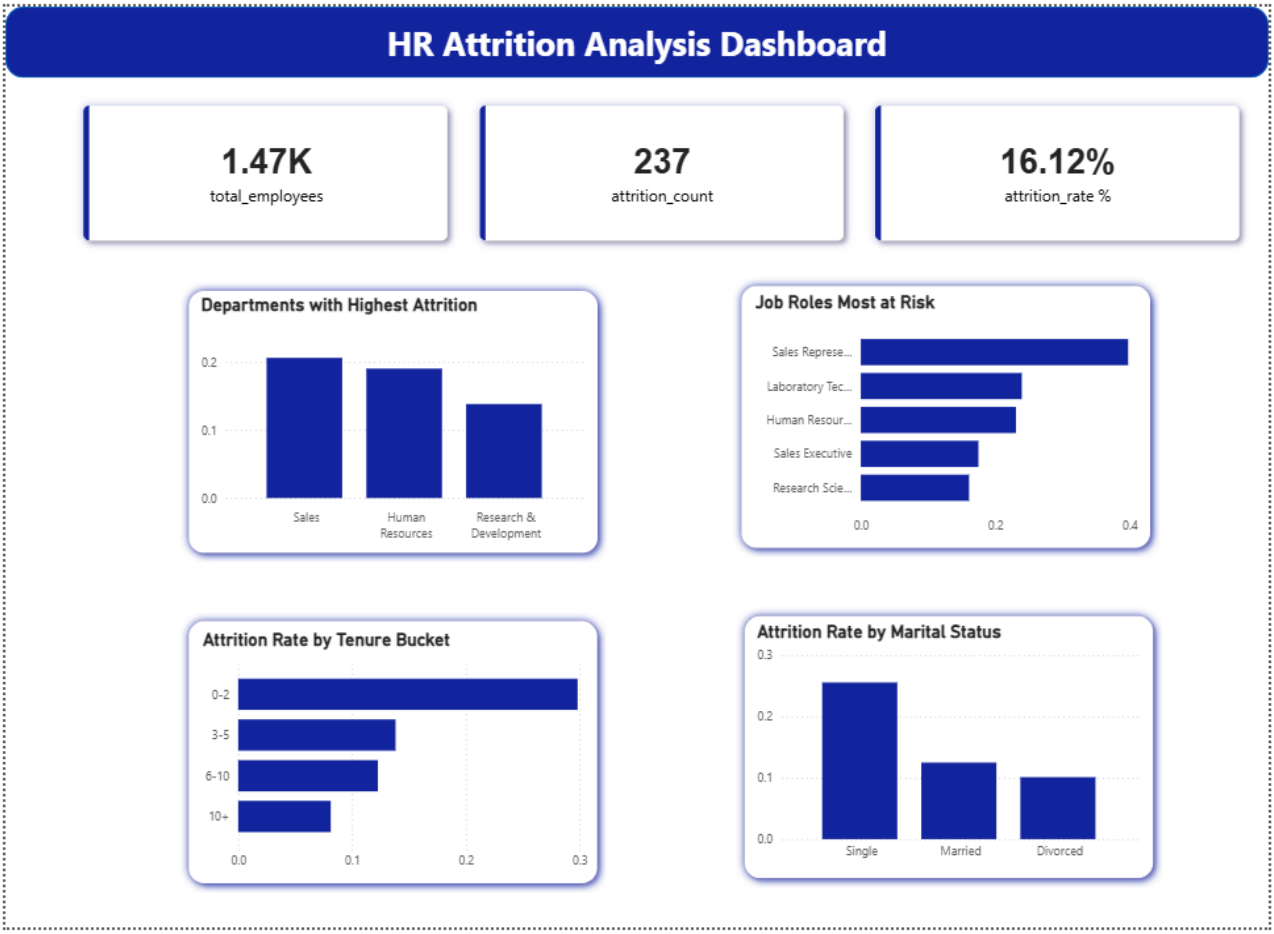
8. **Attrition by Years at Company** – Analyzed how employee tenure impacts attrition.



experience_group	attrition_percentage
6-10 years	12.28
0-2 years	34.88
2-5 years	15.51
10+ years	8.13

5. Dashboard in Power BI

Finally, we built an interactive dashboard in **Power BI** to present insights visually.



6. Business Recommendations

- **Reduce Excessive Overtime** – Employees working overtime show higher attrition; optimize workload distribution and enforce reasonable work hours.
- **Target High-Attrition Job Roles** – Focus retention strategies such as role redesign, career progression plans, and mentoring for job roles with consistently high attrition rates.
- **Review Compensation Structure** – Address salary dissatisfaction by reviewing pay bands, especially for low and mid income employee groups with higher attrition.
- **Improve Work–Life Balance Initiatives** – Strengthen flexible work policies and wellness programs to improve work–life balance scores among employees.
- **Early Tenure Retention Programs** – Implement onboarding support and engagement programs for employees in their first 0–2 years, where attrition risk is highest.
- **Department-Specific HR Interventions** – Design customized retention strategies for departments with above-average attrition instead of applying a one-size-fits-all approach.