# Employee Attrition Analysis using SQL

Prepared by: **Rawan Sotohy**



## 🔍 Project Overview

Employee attrition is a critical HR metric that can reveal deep insights about organizational health. In this project, we analyzed employee data using SQL and Excel to uncover factors affecting attrition rates across departments, genders, job roles, overtime, age groups, and more.

---

## 🎯 Project Goal

To identify key patterns and risk factors contributing to employee attrition using structured SQL queries and data visualization in Excel. The aim is to support decision-making in HR strategies and employee retention planning.

## 🧹 Dataset Review & Cleaning

Before starting the analysis, the structure of the dataset was thoroughly examined. It was found that the data is clean and free from common issues such as missing values, duplicates, or inconsistent entries.

✅ Data Cleanliness Notes:

- No missing values (Null/NA) in any of the key columns

- No duplicate rows

- All columns contain logical and consistent data

- Data types are appropriate for analysis (e.g., numeric values, text)

As a result, no cleaning or preprocessing was required prior to analysis.

## Columns Description📊

| Column Name | Description | Type |
|---|---|---|
| EmployeeID | Unique identifier for each employee | Integer |
| Age | Employee's age | Integer |
| Gender | Male / Female | String |
| Department | Department where the employee works | String |

| | | |
|---|---|---|
| JobRole | Specific job title | String |
| MonthlyIncome | Monthly salary in USD | Integer |
| JobSatisfaction | Satisfaction level from 1 (Low) to 4 (High) | Integer |
| OverTime | Whether the employee works overtime (Yes/No) | String |
| YearsAtCompany | Tenure in years | Integer |
| Attrition | Whether the employee left the company (Yes/No) | String |
| EducationField | Field of highest education | String |
| TrainingTimesLastYear | Number of training sessions attended last year | Integer |
| Education | Education level from 1 to 5 | Integer |
| Age | Used to derive age group | Integer |

# 🧩 Key Analytical Questions & Insights

---

## 1. What is the overall attrition rate?

```sql
SELECT
    ROUND(SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) * 100.0 /
    COUNT(*), 2) AS attrition_rate_percentage
FROM "HR-Analytics";
```

✅ **Result**: **16.12%**

💡 **Insight**: A relatively high attrition rate, which may indicate potential issues in retention strategies.

---

## 2. What is the attrition rate by gender?

```sql
SELECT
    Gender,
    ROUND(SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) * 100.0 /
    COUNT(*),   2) AS attrition_rate_percentage
FROM "HR-Analytics"
GROUP BY Gender;
```

✅ **Result**:

| Gender | Attrition Rate (%) |
| --- | --- |
| Female | 14.80 |
| Male | 17.01 |

💡 **Insight**: Males are more likely to leave the company than females.

---

## 3. Average Income and Job Satisfaction for Those Who Left

```sql
SELECT
    ROUND(AVG(MonthlyIncome), 2) AS avg_income,
    ROUND(AVG(JobSatisfaction), 2) AS avg_job_satisfaction
FROM "HR-Analytics"
WHERE Attrition = 'Yes';
```

✅ **Result**:

| Metric | Value |
|---|---|
| Avg. Income | $4787.09 |
| Job Satisfaction | 2.47 (out of 4) |

💡 **Insight**: Employees who left had moderate satisfaction and lower average salaries than long-term employees in many organizations.

---

## 4. Attrition Rate by Department

```sql
SELECT
    Department,
    ROUND(SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) * 100.0 /
    COUNT(*), 2) AS attrition_rate_percentage
FROM "HR-Analytics"
GROUP BY Department
ORDER BY attrition_rate_percentage DESC;
```

✅ **Result**:

| Department | Attrition Rate (%) |
| --- | --- |
| Sales | 20.63 |
| Human Resources | 19.05 |
| R&D | 13.84 |

💡 **Insight**: Sales and HR departments are more prone to attrition, possibly due to high stress or turnover-prone roles.

---

## 5. Does Overtime Affect Attrition?

```sql
SELECT
    OverTime,
    COUNT(*) AS total_employees,
    SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) AS attritions,
    ROUND(SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) * 100.0 /
    COUNT(*), 2) AS attrition_rate_percentage
FROM "HR-Analytics"
GROUP BY OverTime;
```

✅ **Result**:

| OverTime | Employees | Attritions | Attrition Rate (%) |
| --- | --- | --- | --- |
| No | 1054 | 110 | 10.44 |
| Yes | 416 | 127 | 30.53 |

💡 **Insight**: Employees who work overtime are **3x more likely** to leave. Major red flag for burnout.

---

## 6. Average Tenure (Years at Company)

```sql
SELECT
    Attrition,
    ROUND(AVG(YearsAtCompany), 2) AS avg_years_at_company
FROM "HR-Analytics"
GROUP BY Attrition;
```

✅ **Result**:

| Attrition | Avg. Years |
|-----------|-----------|
| No | 7.37 |
| Yes | 5.13 |

💡 **Insight**: Employees who leave tend to do so earlier in their career — before reaching long-term loyalty.

---

## 7. Attrition Rate by Job Role

```sql
SELECT
    JobRole,
    ROUND(SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) * 100.0 /
 COUNT(*), 2) AS attrition_rate_percentage
FROM "HR-Analytics"
GROUP BY JobRole
ORDER BY attrition_rate_percentage DESC;
```

✅ **Result**:

| Job Role | Attrition Rate (%) |
|---|---|
| Sales Representative | 39.76 |
| Lab Technician | 23.94 |
| HR | 23.08 |
| Research Director | 2.5 |

💡 **Insight**: Sales reps and technicians have the **highest risk**, while managers and directors are highly stable.

---

## 8. Attrition by Age Group

```sql
SELECT
    CASE
        WHEN Age < 30 THEN 'Under 30'
        WHEN Age BETWEEN 30 AND 40 THEN '30-40'
        WHEN Age BETWEEN 41 AND 50 THEN '41-50'
        ELSE '51+'
    END AS age_group,
    ROUND(SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) * 100.0 /
    COUNT(*), 2) AS attrition_rate_percentage
FROM "HR-Analytics"
GROUP BY age_group;
```

✅ **Result**:

| Age Group | Attrition Rate (%) |
|---|---|
| Under 30 | 27.91 |
| 30–40 | 13.84 |
| 41–50 | 10.56 |
| 51+ | 12.59 |

💡 **Insight**: Younger employees (<30) are nearly **twice as likely** to leave. Gen Z retention strategies may be needed.

---

## 9. Average Training Time for Leavers

```sql
SELECT
    ROUND(AVG(TrainingTimesLastYear), 2) AS avg_training_time
FROM "HR-Analytics"
WHERE Attrition = 'Yes';
```

✅ **Result**: **2.62 sessions/year**

💡 **Insight**: Minimal training could correlate with dissatisfaction or lack of career development.

---

## 10. Attrition by Education Field

```sql
SELECT
    EducationField,
    ROUND(SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) * 100.0 /
    COUNT(*), 2) AS attrition_rate_percentage
FROM "HR-Analytics"
GROUP BY EducationField
ORDER BY attrition_rate_percentage DESC;
```

✅ **Result:**

| Education Field | Attrition Rate (%) |
|---|---|
| HR | 25.93 |
| Technical Degree | 24.24 |
| Marketing | 22.01 |

| Medical | 13.58 |
| --- | --- |

💡 **Insight**: HR and Marketing background employees tend to leave more often — possibly due to job stress or misalignment.

---

## 📌 Final Thoughts

This analysis offers a data-driven lens into attrition drivers. Some key takeaways:

- **Overtime, job role, and age** are strong attrition indicators.

- **Younger and entry-level employees** need better retention programs.

- **Departments like Sales and HR** may benefit from wellness and engagement interventions.

---

👩🏽‍💼 Guided and mentored by **Dina Mohsen** during my summer internship at **NTI**