

# SQL Project on HR Employee Attrition Dataset



Trillia Kumaresan  
Data Analyst

Date: 24/07/2024

## Table Index

| S.No | Content                           | Page Number |
|------|-----------------------------------|-------------|
| 1.   | Introduction                      | 2           |
| 2.   | Base Tables Present in a Database | 3           |
| 3.   | Exploratory Data Analysis (EDA)   | 3           |
| 4.   | Fundamental Data Analysis         | 6           |
| 5.   | Reasons for Attrition             | 15          |
| 6.   | Conclusion                        | 25          |
| 7.   | Recommendations to the Company    | 25          |
| 8.   | Reference                         | 26          |

## Introduction

In this SQL project, I have undertaken a comprehensive analysis of the HREmployeeAttrition dataset sourced from [www.kaggle.com](https://www.kaggle.com) to enhance my SQL proficiency and gain practical experience in data manipulation and analysis. This dataset provides detailed information about employees, including attributes such as job role, education, and attrition status. The project focuses on executing a series of SQL queries to explore various facets of the data, derive meaningful insights, and uncover trends related to employee attrition.

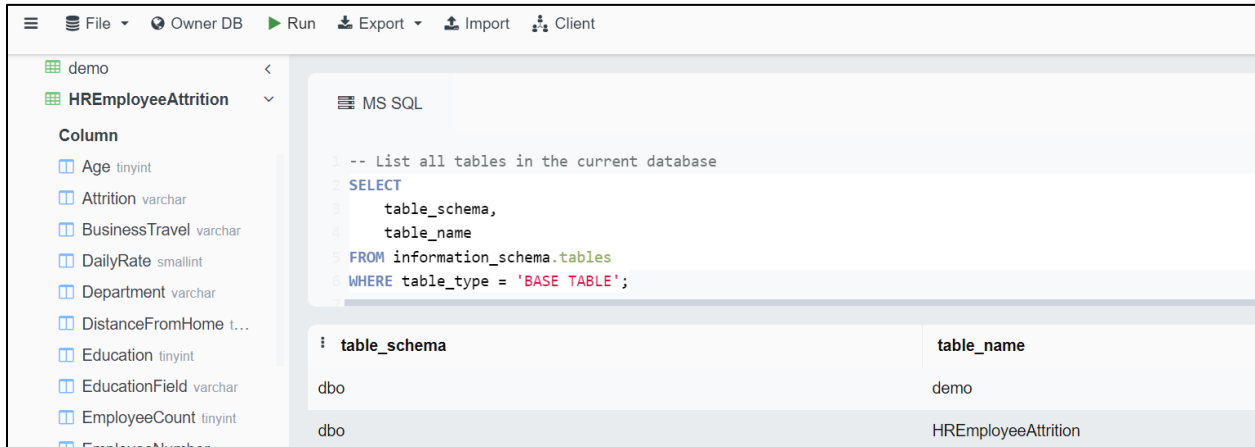
The primary objectives of this project include:

1. **Exploratory Data Analysis (EDA):** I have conducted preliminary investigations to understand the structure and content of the dataset. This involves identifying the number of records, and detecting any inconsistencies or anomalies.
2. **Descriptive Statistics:** I have used SQL queries to compute essential statistics, such as averages, minimums, and maximums, to summarize data distributions and understand key metrics like average monthly income and attrition rates across different departments and job roles.
3. **Trend Analysis:** The project involves examining trends and patterns in employee attrition based on factors such as job satisfaction, distance from home, and education level.
4. **Advanced Queries:** To deepen my SQL skills, I have implemented advanced queries using aggregation, grouping, conditional logic and window functions to analyze employee turnover by various attributes.

To summarize, this project serves as a valuable exercise in applying SQL techniques to real-world data scenarios, aiming to enhance my ability to draw actionable insights from complex datasets.

## Base Tables Present in a Database

Getting an overview of all tables present in the database, which is helpful for database administrators and developers. The SQL query you provided is used to retrieve information about all the base tables present in a database.

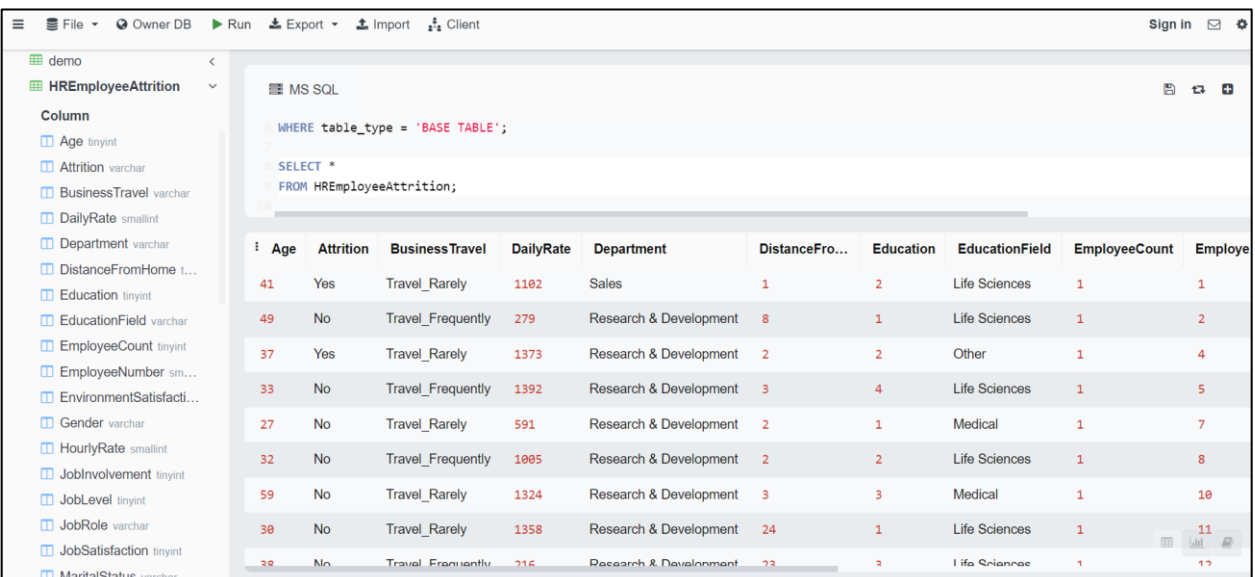


The screenshot shows a database client interface with a menu bar (File, Owner DB, Run, Export, Import, Client) and a sidebar with a tree view showing the database structure. The main pane displays the results of an SQL query. The query is: `-- List all tables in the current database`  
`SELECT`  
`table_schema,`  
`table_name`  
`FROM information_schema.tables`  
`WHERE table_type = 'BASE TABLE';`  
The results are displayed in a table with two columns: `table_schema` and `table_name`. The results show two tables: `dbo` and `demo`, both in the `dbo` schema.

| table_schema | table_name          |
|--------------|---------------------|
| dbo          | demo                |
| dbo          | HREmployeeAttrition |

## Exploratory Data Analysis (EDA):

- As a data analyst, Exploratory Data Analysis is the starting point of the data analysis. It allows to view all the data in the table, helping to understand the structure, column names, data types, and the general nature of the data.
- To check for data quality issues such as missing values, outliers, and inconsistencies visually.

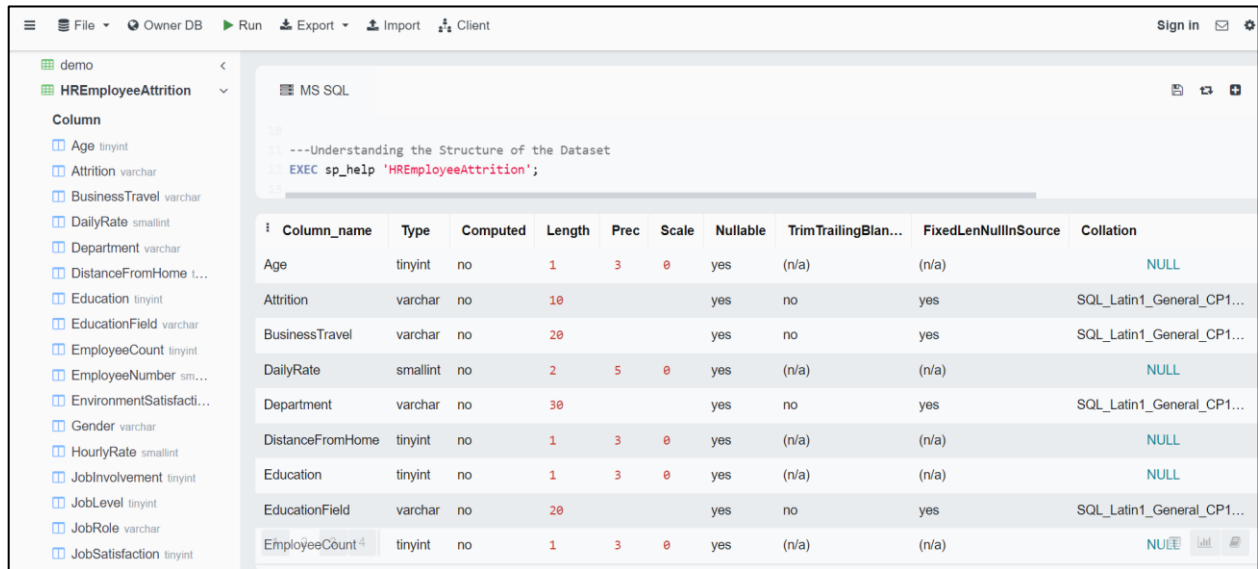


The screenshot shows a database client interface with a menu bar (File, Owner DB, Run, Export, Import, Client) and a sidebar with a tree view showing the database structure. The main pane displays the results of an SQL query. The query is: `WHERE table_type = 'BASE TABLE';`  
`SELECT *`  
`FROM HREmployeeAttrition;`  
The results are displayed in a table with 10 columns: `Age`, `Attrition`, `BusinessTravel`, `DailyRate`, `Department`, `DistanceFro...`, `Education`, `EducationField`, `EmployeeCount`, and `Employee`. The results show 10 rows of data.

| Age | Attrition | BusinessTravel    | DailyRate | Department             | DistanceFro... | Education | EducationField | EmployeeCount | Employee |
|-----|-----------|-------------------|-----------|------------------------|----------------|-----------|----------------|---------------|----------|
| 41  | Yes       | Travel_Rarely     | 1102      | Sales                  | 1              | 2         | Life Sciences  | 1             | 1        |
| 49  | No        | Travel_Frequently | 279       | Research & Development | 8              | 1         | Life Sciences  | 1             | 2        |
| 37  | Yes       | Travel_Rarely     | 1373      | Research & Development | 2              | 2         | Other          | 1             | 4        |
| 33  | No        | Travel_Frequently | 1392      | Research & Development | 3              | 4         | Life Sciences  | 1             | 5        |
| 27  | No        | Travel_Rarely     | 591       | Research & Development | 2              | 1         | Medical        | 1             | 7        |
| 32  | No        | Travel_Frequently | 1005      | Research & Development | 2              | 2         | Life Sciences  | 1             | 8        |
| 59  | No        | Travel_Rarely     | 1324      | Research & Development | 3              | 3         | Medical        | 1             | 10       |
| 30  | No        | Travel_Rarely     | 1358      | Research & Development | 24             | 1         | Life Sciences  | 1             | 11       |
| 38  | No        | Travel_Frequently | 116       | Research & Development | 23             | 2         | Life Sciences  | 1             | 12       |

## 1. Understanding the Structure of the Dataset

The EXEC sp\_help command is specific to Microsoft SQL Server. It is used to display detailed information about a database object, such as a table, view, or stored procedure.

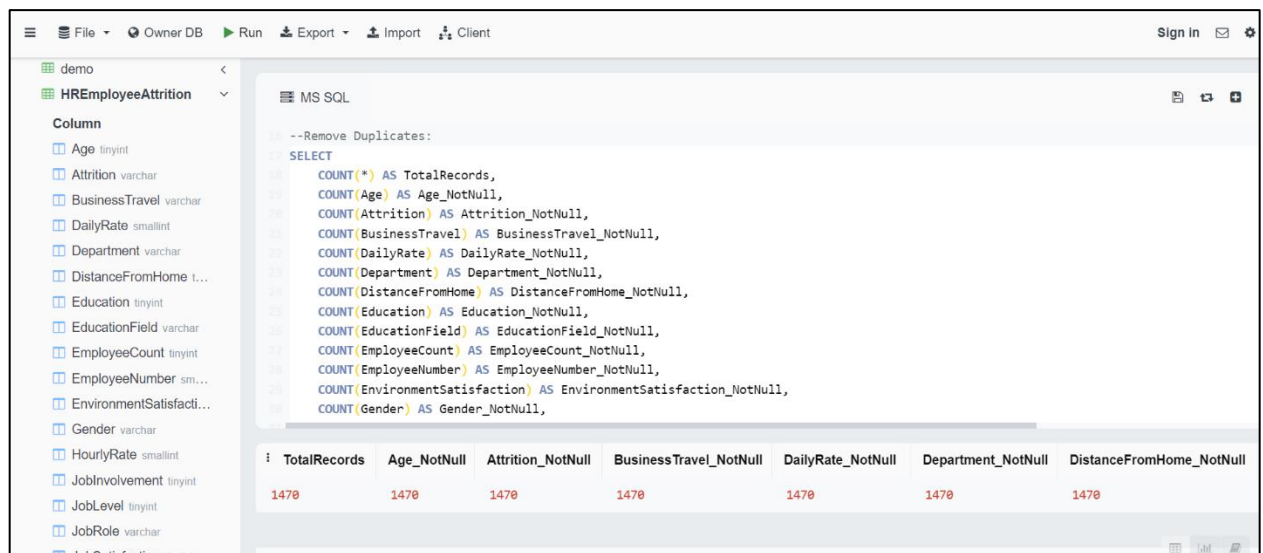


The screenshot shows the SQL Server Enterprise Manager interface. On the left, the 'demo' database is expanded, showing the 'HREmployeeAttrition' table. The 'Column' list on the left includes: Age (tinyint), Attrition (varchar), BusinessTravel (varchar), DailyRate (smallint), Department (varchar), DistanceFromHome (tinyint), Education (tinyint), EducationField (varchar), EmployeeCount (tinyint), EmployeeNumber (smallint), EnvironmentSatisfaction (tinyint), Gender (varchar), HourlyRate (smallint), JobInvolvement (tinyint), JobLevel (tinyint), JobRole (varchar), and JobSatisfaction (tinyint). The main pane shows the MS SQL query editor with the command 'EXEC sp\_help 'HREmployeeAttrition';'. Below the query, the table structure is displayed as a table with columns: Column\_name, Type, Computed, Length, Prec, Scale, Nullable, TrimTrailingBlanks, FixedLenNullInSource, and Collation.

| Column_name      | Type     | Computed | Length | Prec | Scale | Nullable | TrimTrailingBlanks | FixedLenNullInSource | Collation                 |
|------------------|----------|----------|--------|------|-------|----------|--------------------|----------------------|---------------------------|
| Age              | tinyint  | no       | 1      | 3    | 0     | yes      | (n/a)              | (n/a)                | NULL                      |
| Attrition        | varchar  | no       | 10     |      |       | yes      | no                 | yes                  | SQL_Latin1_General_CP1... |
| BusinessTravel   | varchar  | no       | 20     |      |       | yes      | no                 | yes                  | SQL_Latin1_General_CP1... |
| DailyRate        | smallint | no       | 2      | 5    | 0     | yes      | (n/a)              | (n/a)                | NULL                      |
| Department       | varchar  | no       | 30     |      |       | yes      | no                 | yes                  | SQL_Latin1_General_CP1... |
| DistanceFromHome | tinyint  | no       | 1      | 3    | 0     | yes      | (n/a)              | (n/a)                | NULL                      |
| Education        | tinyint  | no       | 1      | 3    | 0     | yes      | (n/a)              | (n/a)                | NULL                      |
| EducationField   | varchar  | no       | 20     |      |       | yes      | no                 | yes                  | SQL_Latin1_General_CP1... |
| EmployeeCount    | tinyint  | no       | 1      | 3    | 0     | yes      | (n/a)              | (n/a)                | NULL                      |

## 2. To find the missing values in a dataset

To find the missing values in a dataset with the specified columns using SQL, you can count the number of NULL values for each column.

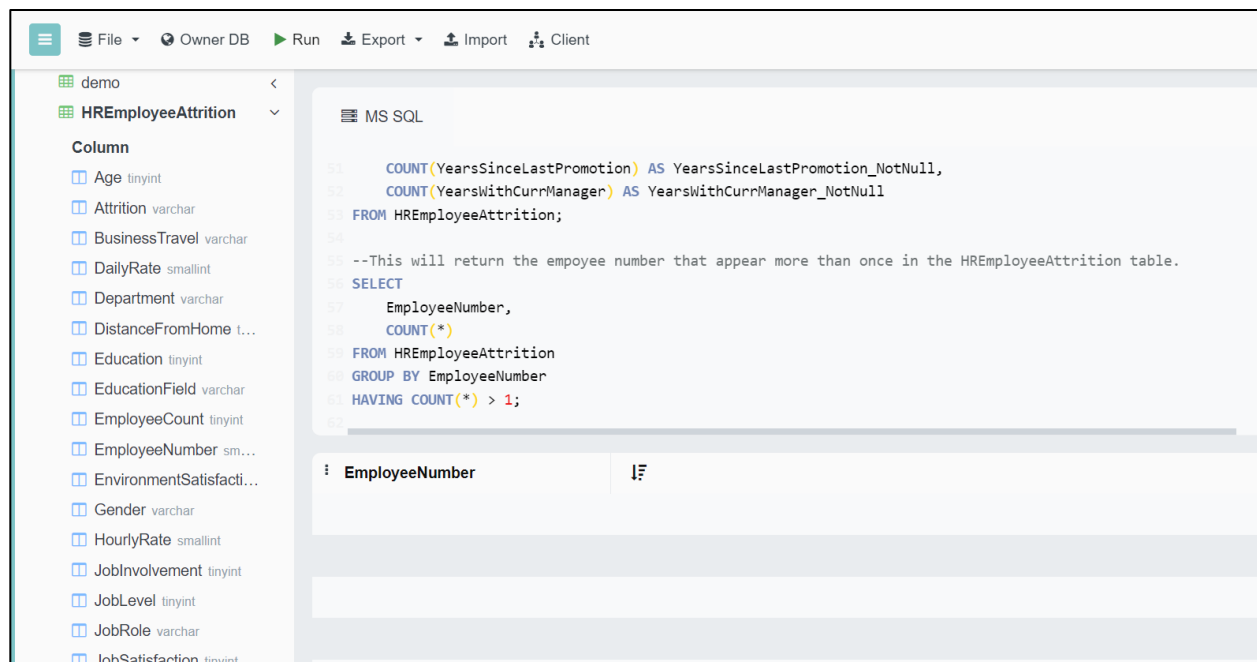


The screenshot shows the SQL Server Enterprise Manager interface. The main pane shows the MS SQL query editor with the following query: 'SELECT COUNT(\*) AS TotalRecords, COUNT(Age) AS Age\_NotNull, COUNT(Attrition) AS Attrition\_NotNull, COUNT(BusinessTravel) AS BusinessTravel\_NotNull, COUNT(DailyRate) AS DailyRate\_NotNull, COUNT(Department) AS Department\_NotNull, COUNT(DistanceFromHome) AS DistanceFromHome\_NotNull, COUNT(Education) AS Education\_NotNull, COUNT(EducationField) AS EducationField\_NotNull, COUNT(EmployeeCount) AS EmployeeCount\_NotNull, COUNT(EmployeeNumber) AS EmployeeNumber\_NotNull, COUNT(EnvironmentSatisfaction) AS EnvironmentSatisfaction\_NotNull, COUNT(Gender) AS Gender\_NotNull;'. Below the query, the results are displayed as a table with columns: TotalRecords, Age\_NotNull, Attrition\_NotNull, BusinessTravel\_NotNull, DailyRate\_NotNull, Department\_NotNull, and DistanceFromHome\_NotNull.

| TotalRecords | Age_NotNull | Attrition_NotNull | BusinessTravel_NotNull | DailyRate_NotNull | Department_NotNull | DistanceFromHome_NotNull |
|--------------|-------------|-------------------|------------------------|-------------------|--------------------|--------------------------|
| 1470         | 1470        | 1470              | 1470                   | 1470              | 1470               | 1470                     |

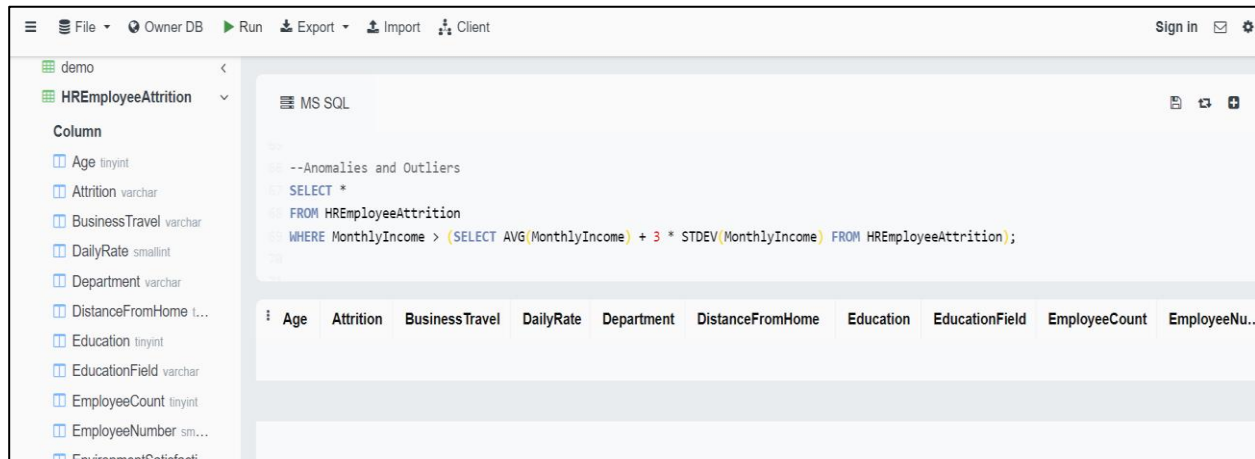
### 3. Finding missing values in EmployeeNumber attribute(Primary Key)

It is used to find duplicate entries in the HREmployeeAttrition table based on the EmployeeNumber field. If EmployeeNumber having multiple records with the same EmployeeNumber might indicate a data quality issue, such as duplicate records that need to be investigated and resolved.



### 4. Finding Anomalies and Outliers

Finding outlier can be useful in various business scenarios, such as detecting unusually high salaries that may need further investigation or validation. Here, below query is used to identify employees with exceptionally high monthly incomes. The choice of "three standard deviations" as a threshold is based on statistical conventions, as data points beyond this range are often considered outliers in a normal distribution.



## Fundamental Data Analysis

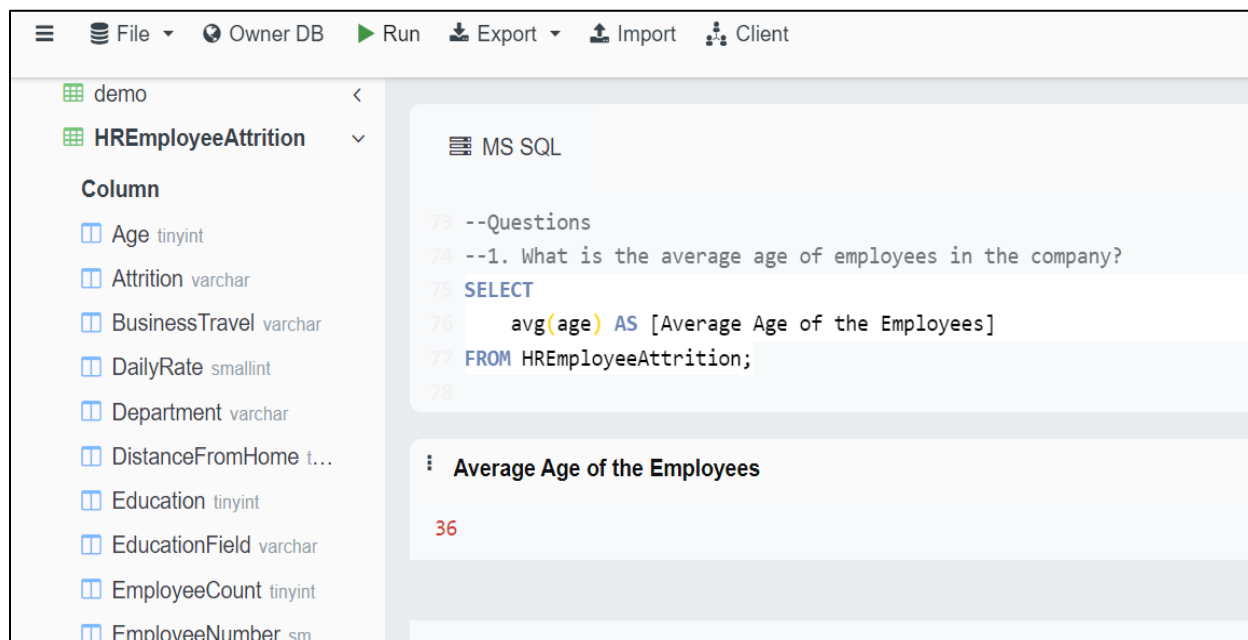
Fundamental key checks such as data retrieval, analysis, and summarization within a database are the part of data exploration and analyzation of data stored in relational databases.

- **Selection Queries:** Used to select specific data from one or more tables in a database.
- **Aggregation Queries:** Used to summarize data in the table.
- **Grouping Queries:** These queries group data based on one or more columns and often include aggregate functions to summarize data within each group.
- **Filtering Queries:** It retrieves data based on specific conditions using the WHERE clause.
- **Sorting Queries:** Used to order the results based on one or more columns.
- **Limiting Queries:** To limit the number of rows returned by a query.

### Questions:

#### 1. What is the average age of employees in the company?

Computing the **average age** of employees in the HREmployeeAttrition table is a basic but essential analysis that helps in understanding the overall demographic profile of the workforce. Hence, the average age distribution of the employees in the company is 36(age).



This information is useful for understanding the composition of the workforce, planning for succession, and making decisions related to hiring or employee development.

## 2. How many employees are in each department?

The aim is to count the number of employees in each department. It helps in understanding the distribution of staff across various departments, which can provide insights into resource allocation, department size, and workforce management.

For each unique department value, the query aggregates the data and calculates the count of employees within that department.



demo < HREmployeeAttrition

Column

- Age tinyint
- Attrition varchar
- BusinessTravel varchar
- DailyRate smallint
- Department varchar
- DistanceFromHome t...
- Education tinyint
- EducationField varchar
- EmployeeCount tinyint
- EmployeeNumber sm...
- EnvironmentSatisfacti...
- Gender varchar
- HourlyRate smallint
- JobInvolvement tinyint
- JobLevel tinyint
- JobRole varchar

MS SQL

```
--2. How many employees are in each department?
SELECT
    department,
    COUNT(*) AS [Number of Employees]
FROM HREmployeeAttrition
GROUP BY department;
```

| department             | Number of Employees |
|------------------------|---------------------|
| Human Resources        | 63                  |
| Sales                  | 446                 |
| Research & Development | 961                 |

### 3. What is the average monthly income by job role?

demo < HREmployeeAttrition

Column

- Age tinyint
- Attrition varchar
- BusinessTravel varchar
- DailyRate smallint
- Department varchar
- DistanceFromHome t...
- Education tinyint
- EducationField varchar
- EmployeeCount tinyint
- EmployeeNumber sm...
- EnvironmentSatisfacti...
- Gender varchar
- HourlyRate smallint
- JobInvolvement tinyint
- JobLevel tinyint
- JobRole varchar
- JobSatisfaction tinyint
- MaritalStatus varchar

MS SQL

```
--3. What is the average monthly income by job role?
SELECT
    jobrole,
    avg(monthlyincome) AS [Average Monthly Income]
FROM HREmployeeAttrition
GROUP BY jobrole;
```

| jobrole                   | Average Monthly Income |
|---------------------------|------------------------|
| Sales Executive           | 6924                   |
| Sales Representative      | 2626                   |
| Research Scientist        | 3239                   |
| Human Resources           | 4235                   |
| Healthcare Representative | 7528                   |
| Research Director         | 16033                  |
| Manager                   | 17181                  |

The screenshot shows a database client interface with a menu bar (File, Owner DB, Run, Export, Import, Client) and a sidebar listing columns for 'demo' and 'HREmployeeAttrition'. The main area displays an SQL query and its results.

**MS SQL**

```

87 --3. What is the average monthly income by job role?
88 SELECT
89     jobrole,
90     avg(monthlyincome) AS [Average Monthly Income]
91 FROM HREmployeeAttrition
92 GROUP BY jobrole;
93

```

| jobrole                   | Average Monthly Income |
|---------------------------|------------------------|
| Healthcare Representative | 7528                   |
| Research Director         | 16033                  |
| Manager                   | 17181                  |
| Manufacturing Director    | 7295                   |
| Laboratory Technician     | 3237                   |

It calculates the average monthly income for employees in each job role. This analysis helps in understanding the compensation structure and disparities among different job roles within the organization. Comparing average incomes across job roles can provide insights into whether compensation levels are aligned with industry standards or company expectations.

#### 4. How many employees have left the company (Attrition)?

The query counts the number of employees who have exited the company, which is crucial for understanding employee turnover and attrition rates. It ensures that only employees who have left the company (indicated by 'Yes' in the attrition column) are included in the count.

Knowing the number of employees who have left can inform workforce planning and management decisions in terms of understanding staffing needs, planning for recruitment, and developing retention strategies.

The screenshot shows a database client interface with a menu bar (File, Owner DB, Run, Export, Import, Client) and a sidebar listing columns for a table named 'HREmployeeAttrition'. The main pane displays an MS SQL query and its result.

**Columns:**

- Age tinyint
- Attrition varchar
- BusinessTravel varchar
- DailyRate smallint
- Department varchar
- DistanceFromHome t...
- Education tinyint
- EducationField varchar
- EmployeeCount tinyint
- EmployeeNumber sm...
- EnvironmentSatisfacti...
- Gender varchar
- HourlyRate smallint
- JobInvolvement tinyint
- JobLevel tinyint
- JobRole varchar

**MS SQL Query:**

```
--4. How many employees have left the company (Attrition)?
SELECT
  COUNT(*) AS 'Number of Employees Left the Company'
FROM HREmployeeAttrition
WHERE attrition='Yes';
```

**Result:**

| Number of Employees Left the Company |
|--------------------------------------|
| 237                                  |

## 5. What is the distribution of EducationField among employees?

The number of employees in each educational field, providing insights into the educational background of the workforce. Insights into the educational background of employees can inform recruitment strategies and training programs. For example, if a certain educational field is underrepresented, the company might target recruitment efforts to attract more candidates from that field.

The screenshot shows a database client interface with a menu bar (File, Owner DB, Run, Export, Import, Client) and a sidebar listing columns for 'demo' and 'HREmployeeAttrition'. The main area displays an MS SQL query and its results.

**MS SQL Query:**

```
--5. What is the distribution of EducationField among employees?
1001 SELECT
1002     EducationField,
1003     COUNT(*) AS [Number of Employees]
1004 FROM HREmployeeAttrition
1005 GROUP BY educationfield;
```

**Results:**

| EducationField   | Number of Employees |
|------------------|---------------------|
| Technical Degree | 132                 |
| Other            | 82                  |
| Human Resources  | 27                  |
| Medical          | 464                 |
| Marketing        | 159                 |
| Life Sciences    | 606                 |

## 6. What is the average number of years employees have been with the company?

The mean number of years that employees have been with the company, providing insights into employee tenure and stability. Comparing the average no. of years at the company with industry standards can provide how the company's employee tenure compares with competitors.

The screenshot shows the same database client interface as above, but with a different query and result.

**MS SQL Query:**

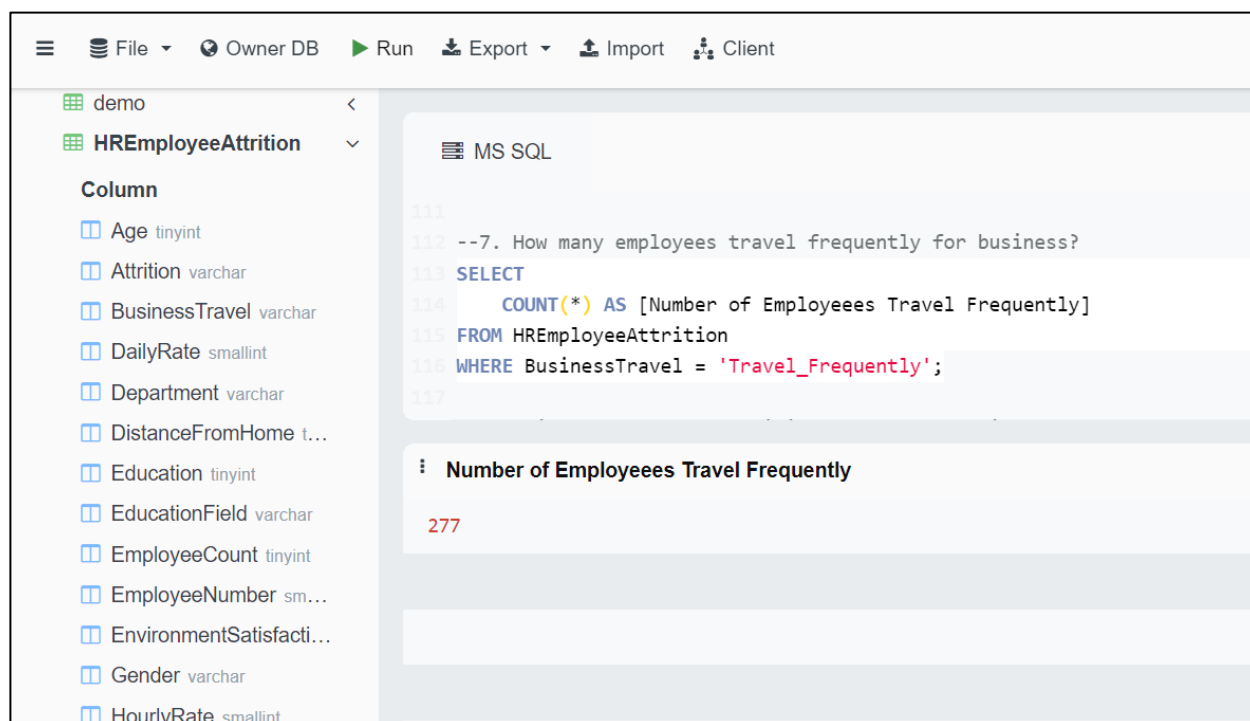
```
1006
1007 --6. What is the average number of years employees have been with the company?
1008 SELECT
1009     avg(yearsatcompany) AS 'Average Number of Years at Company'
1010 FROM HREmployeeAttrition;
1011
1012 --7. How many employees travel frequently for business?
```

**Results:**

| Average Number of Years at Company |
|------------------------------------|
| 7                                  |

## 7. How many employees travel frequently for business?

The number of employees whose business travel status is categorized as "Travel\_Frequently". This helps in understanding the proportion of employees with high travel requirements within the organization.



Understanding how many employees travel frequently helps in analyzing travel-related policies and expenses. It can be used to assess the need for travel management solutions or support for frequent travelers.

## 8. Which employees are eligible for stock options and at what levels?

The query provides a list of employees who are eligible for stock options and specifies their respective stock option levels. This information is crucial for understanding compensation packages and benefits.

Knowing which employees have stock options and at what levels can aid in strategic planning, especially when considering employee retention and reward strategies.

File Owner DB Run Export Import Client

demo

HREmployeeAttrition

Column

- Age tinyint
- Attrition varchar
- BusinessTravel varchar
- DailyRate smallint
- Department varchar
- DistanceFromHome t...
- Education tinyint
- EducationField varchar
- EmployeeCount tinyint
- EmployeeNumber sm...
- EnvironmentSatisfacti...
- Gender varchar
- HourlyRate smallint
- JobInvolvement tinyint
- JobLevel tinyint
- JobRole varchar
- JobSatisfaction tinyint
- MaritalStatus varchar

MS SQL

```
--8. Which employees are eligible for stock options and at what levels?
SELECT jobrole, stockoptionlevel, COUNT(*) AS [Number of Employees]
FROM HREmployeeAttrition
GROUP BY jobrole, stockoptionlevel HAVING stockoptionlevel > '0'
ORDER BY stockoptionlevel;
```

| jobrole                   | stockoptionlevel | Number of Employees |
|---------------------------|------------------|---------------------|
| Healthcare Representative | 1                | 59                  |
| Human Resources           | 1                | 21                  |
| Laboratory Technician     | 1                | 110                 |
| Manager                   | 1                | 57                  |
| Manufacturing Director    | 1                | 60                  |
| Research Director         | 1                | 31                  |
| Research Scientist        | 1                | 104                 |
| Sales Executive           | 1                | 125                 |
| Sales Representative      | 1                | 20                  |

File Owner DB Run Export Import Client

demo

HREmployeeAttrition

Column

- Age tinyint
- Attrition varchar
- BusinessTravel varchar
- DailyRate smallint
- Department varchar
- DistanceFromHome t...
- Education tinyint
- EducationField varchar
- EmployeeCount tinyint
- EmployeeNumber sm...
- EnvironmentSatisfacti...
- Gender varchar
- HourlyRate smallint
- JobInvolvement tinyint
- JobLevel tinyint
- JobRole varchar
- JobSatisfaction tinyint
- MaritalStatus varchar

MS SQL

```
--8. Which employees are eligible for stock options and at what levels?
SELECT jobrole, stockoptionlevel, COUNT(*) AS [Number of Employees]
```

| jobrole                   | stockoptionlevel | Number of Employees |
|---------------------------|------------------|---------------------|
| Healthcare Representative | 2                | 13                  |
| Human Resources           | 2                | 3                   |
| Laboratory Technician     | 2                | 21                  |
| Manager                   | 2                | 5                   |
| Manufacturing Director    | 2                | 20                  |
| Research Director         | 2                | 11                  |
| Research Scientist        | 2                | 34                  |
| Sales Executive           | 2                | 44                  |
| Sales Representative      | 2                | 7                   |
| Healthcare Representative | 3                | 8                   |

The screenshot shows a database client interface with a menu bar (File, Owner DB, Run, Export, Import, Client) and a sidebar with a tree view containing 'demo' and 'HREmployeeAttrition'. The main pane displays an MS SQL query and its results.

**MS SQL**

```
--8. Which employees are eligible for stock options and at what levels?
SELECT jobrole, stockoptionlevel, COUNT(*) AS [Number of Employees]
```

| jobrole                | stockoptionlevel | Number of Employees |
|------------------------|------------------|---------------------|
| Human Resources        | 3                | 4                   |
| Laboratory Technician  | 3                | 20                  |
| Manager                | 3                | 3                   |
| Manufacturing Director | 3                | 6                   |
| Research Director      | 3                | 5                   |
| Research Scientist     | 3                | 18                  |
| Sales Executive        | 3                | 18                  |
| Sales Representative   | 3                | 3                   |

## 9. What is the correlation between JobRole and JobLevel?

The COUNT(\*) function will be applied to each unique combination of jobrole and joblevel, giving the number of employees for each specific role and level combination.

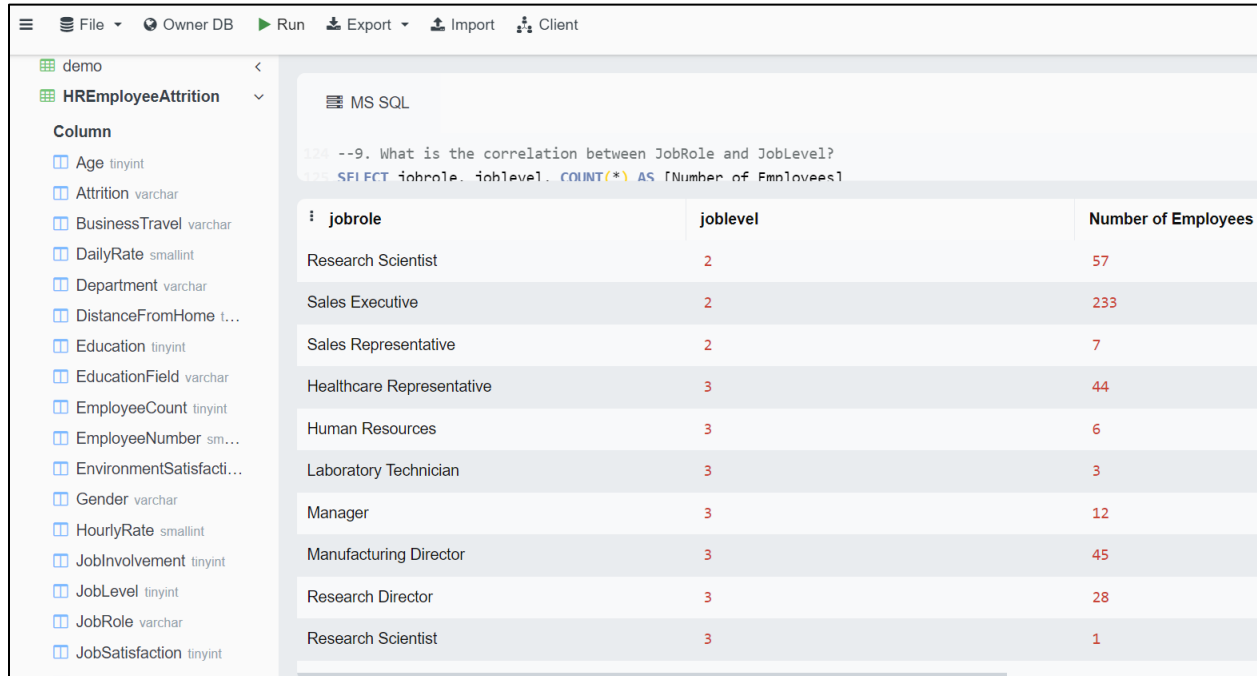
The screenshot shows the same database client interface as above, but with a different query and results.

**MS SQL**

```
--9. What is the correlation between JobRole and JobLevel?
SELECT jobrole, joblevel, COUNT(*) AS [Number of Employees]
FROM HREmployeeAttrition
GROUP BY jobrole, joblevel
ORDER BY joblevel;
```

| jobrole                   | joblevel | Number of Employees |
|---------------------------|----------|---------------------|
| Human Resources           | 1        | 33                  |
| Laboratory Technician     | 1        | 200                 |
| Research Scientist        | 1        | 234                 |
| Sales Representative      | 1        | 76                  |
| Healthcare Representative | 2        | 78                  |
| Human Resources           | 2        | 13                  |
| Laboratory Technician     | 2        | 56                  |
| Manufacturing Director    | 2        | 90                  |

This query provides insights into how employees are distributed across various job roles and levels. It helps in understanding which job roles and levels have the most or least number of employees.

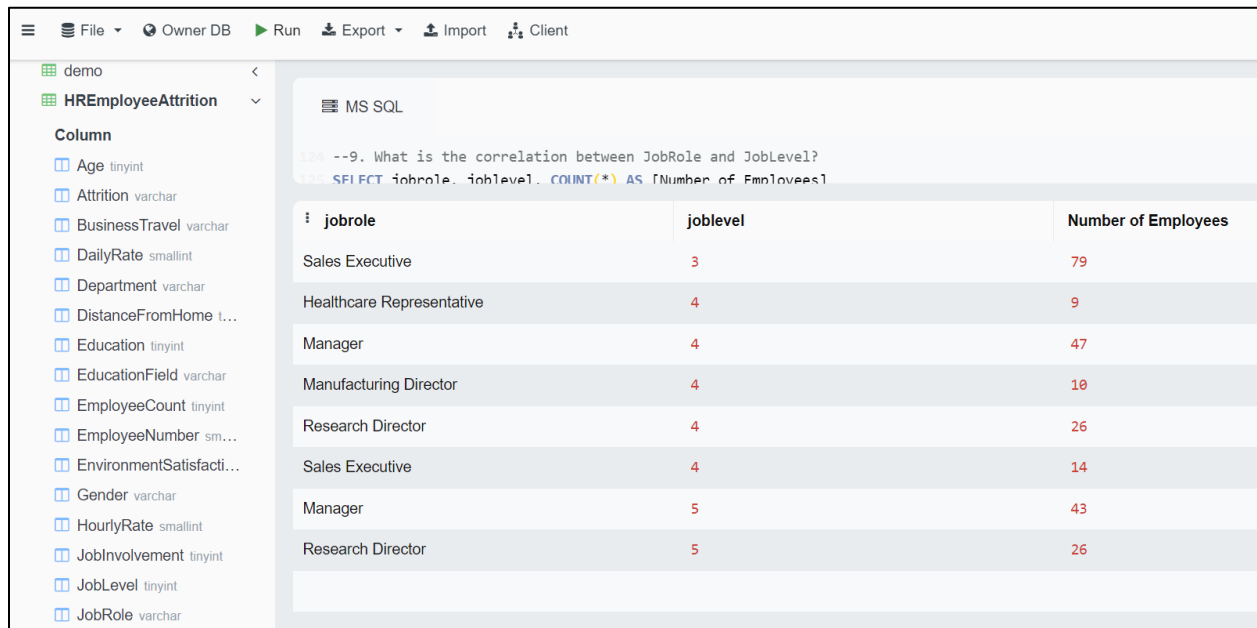


The screenshot shows a database client interface with a query window displaying the following SQL query:

```
--9. What is the correlation between JobRole and JobLevel?
SELECT jobrole, joblevel, COUNT(*) AS [Number of Employees]
```

The result set is a table with three columns: jobrole, joblevel, and Number of Employees. The data is as follows:

| jobrole                   | joblevel | Number of Employees |
|---------------------------|----------|---------------------|
| Research Scientist        | 2        | 57                  |
| Sales Executive           | 2        | 233                 |
| Sales Representative      | 2        | 7                   |
| Healthcare Representative | 3        | 44                  |
| Human Resources           | 3        | 6                   |
| Laboratory Technician     | 3        | 3                   |
| Manager                   | 3        | 12                  |
| Manufacturing Director    | 3        | 45                  |
| Research Director         | 3        | 28                  |
| Research Scientist        | 3        | 1                   |



The screenshot shows a database client interface with a query window displaying the following SQL query:

```
--9. What is the correlation between JobRole and JobLevel?
SELECT jobrole, joblevel, COUNT(*) AS [Number of Employees]
```

The result set is a table with three columns: jobrole, joblevel, and Number of Employees. The data is as follows:

| jobrole                   | joblevel | Number of Employees |
|---------------------------|----------|---------------------|
| Sales Executive           | 3        | 79                  |
| Healthcare Representative | 4        | 9                   |
| Manager                   | 4        | 47                  |
| Manufacturing Director    | 4        | 10                  |
| Research Director         | 4        | 26                  |
| Sales Executive           | 4        | 14                  |
| Manager                   | 5        | 43                  |
| Research Director         | 5        | 26                  |

## Reasons for Attrition

### 1. What is the overall attrition rate in the company?



The screenshot displays a database management interface with a sidebar on the left showing a tree view of the database structure. The main area shows an SQL query being executed. The query is as follows:

```

131
132 --1. What is the overall attrition rate in the company?
133
134 SELECT
135     BusinessTravel VARCHAR(20) 1 No.of Rows],
136     SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) AS [Total Number of Attrition],
137     ((SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END)) / COUNT(*) * 100) AS AttritionRate
138 FROM AttritionTable

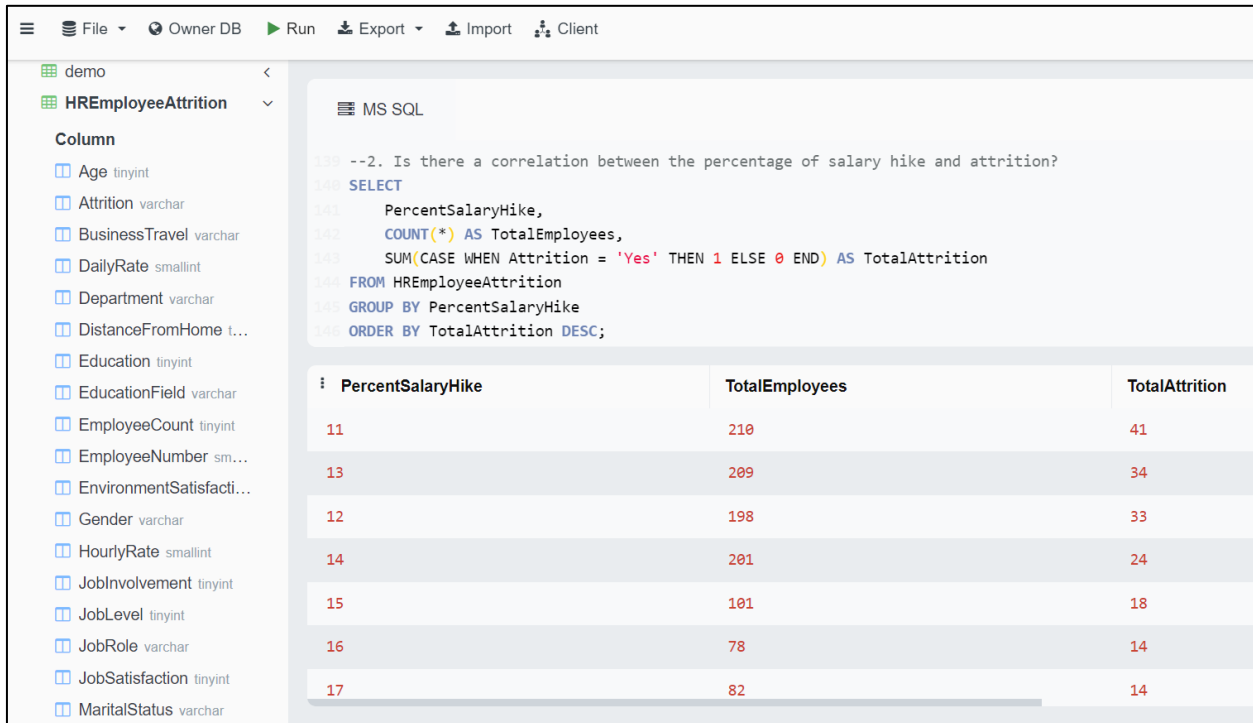
```

The results of the query are shown in a table with three columns: Total No.of Rows, Total Number of Attrition, and AttritionRate. The values are 1470, 237, and 0 respectively.

| Total No.of Rows | Total Number of Attrition | AttritionRate |
|------------------|---------------------------|---------------|
| 1470             | 237                       | 0             |

It provides a summary of attrition data from the AttritionTable. It counts the total number of employees, calculates the total number of attritions, and computes the attrition rate as a percentage. This information is useful for analyzing employee turnover, assessing HR performance, and making strategic decisions related to employee retention and workforce management.

## 2. Is there a correlation between the percentage of salary hike and attrition?



MS SQL

```

138 --2. Is there a correlation between the percentage of salary hike and attrition?
139
140 SELECT
141     PercentSalaryHike,
142     COUNT(*) AS TotalEmployees,
143     SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) AS TotalAttrition
144 FROM HREmployeeAttrition
145 GROUP BY PercentSalaryHike
146 ORDER BY TotalAttrition DESC;

```

| PercentSalaryHike | TotalEmployees | TotalAttrition |
|-------------------|----------------|----------------|
| 11                | 210            | 41             |
| 13                | 209            | 34             |
| 12                | 198            | 33             |
| 14                | 201            | 24             |
| 15                | 101            | 18             |
| 16                | 78             | 14             |
| 17                | 82             | 14             |

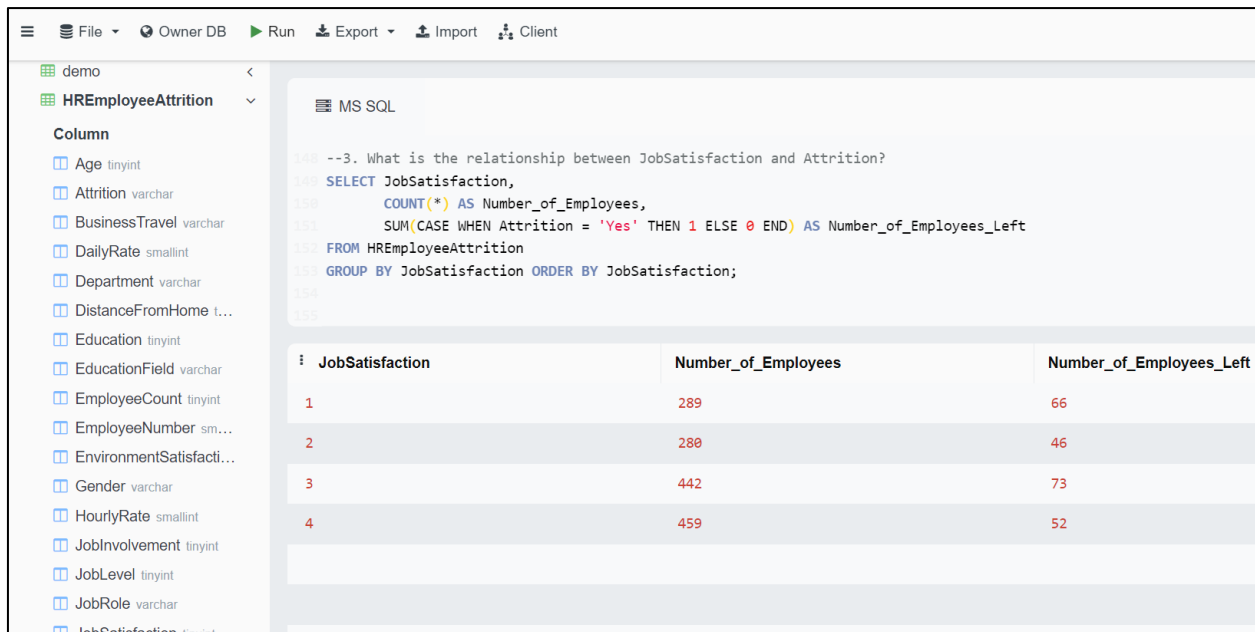
It analyzes the relationship between the **PercentSalaryHike** and employee attrition within the dataset **HREmployeeAttrition**.

The results can reveal trends such as whether employees with certain salary hike percentages are more likely to leave. For instance, if higher salary hikes correlate with higher attrition, this could indicate dissatisfaction with the company or other underlying issues.

## 3. What is the relationship between JobSatisfaction and Attrition?

This query helps to analyze how job satisfaction levels are associated with employee attrition. By grouping and counting employees based on job satisfaction and their attrition status, you can identify trends and patterns related to employee turnover.

Can observe if employees with lower or higher job satisfaction are more likely to leave the company. For instance, if employees with lower job satisfaction have higher attrition rates, this indicates a potential area for improvement in job satisfaction and retention strategies.



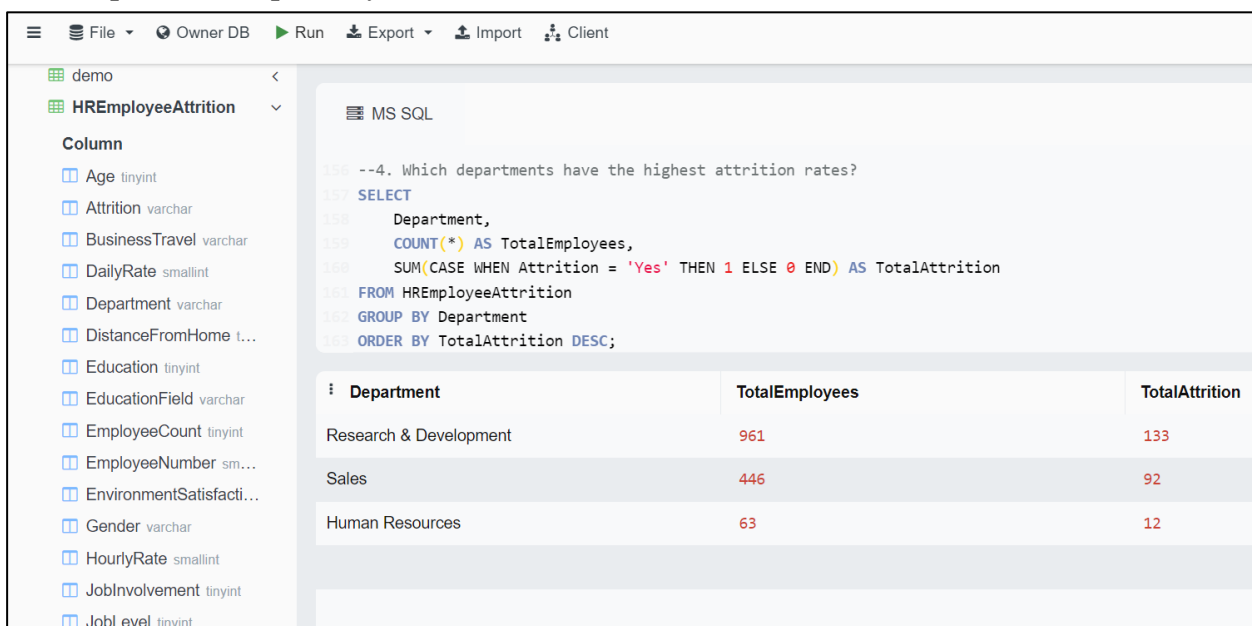
MS SQL

```
--3. What is the relationship between JobSatisfaction and Attrition?
SELECT JobSatisfaction,
       COUNT(*) AS Number_of_Employees,
       SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) AS Number_of_Employees_Left
FROM HREmployeeAttrition
GROUP BY JobSatisfaction ORDER BY JobSatisfaction;
```

| JobSatisfaction | Number_of_Employees | Number_of_Employees_Left |
|-----------------|---------------------|--------------------------|
| 1               | 289                 | 66                       |
| 2               | 280                 | 46                       |
| 3               | 442                 | 73                       |
| 4               | 459                 | 52                       |

#### 4. Which departments have the highest attrition rates?

The clause ensures that the aggregation functions (COUNT and SUM) are applied within each department separately, rather than across the entire dataset.



MS SQL

```
--4. Which departments have the highest attrition rates?
SELECT Department,
       COUNT(*) AS TotalEmployees,
       SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) AS TotalAttrition
FROM HREmployeeAttrition
GROUP BY Department
ORDER BY TotalAttrition DESC;
```

| Department             | TotalEmployees | TotalAttrition |
|------------------------|----------------|----------------|
| Research & Development | 961            | 133            |
| Sales                  | 446            | 92             |
| Human Resources        | 63             | 12             |

Departments with higher attrition rates are highlighted, allowing you to focus on these areas for further investigation. It may indicate potential issues within those departments, such as management problems, work environment concerns, or other factors affecting employee retention.

### 5. Employees from which designation falls attrition more?

Job roles with higher attrition counts are highlighted, indicating where there might be significant issues affecting employee retention. This can reveal potential problems specific to those roles.

Knowing which job roles have the highest attrition can guide HR and management in focusing their efforts on improving conditions for those specific roles. This might include revising job responsibilities, improving job satisfaction, or addressing role-specific concerns.

The screenshot displays a data analysis interface with a menu on the left, a central SQL editor, and a results table on the right.

**Menu:**

- demo
- HREmployeeAttrition
- Column
- Age tinyint
- Attrition varchar
- BusinessTravel varchar
- DailyRate smallint
- Department varchar
- DistanceFromHome t...
- Education tinyint
- EducationField varchar
- EmployeeCount tinyint
- EmployeeNumber sm...
- EnvironmentSatisfacti...
- Gender varchar
- HourlyRate smallint
- JobInvolvement tinyint
- JobLevel tinyint
- JobRole varchar
- JobSatisfaction tinyint
- MaritalStatus varchar

**SQL Query:**

```
SELECT
  Jobrole,
  COUNT(Case when attrition = 'Yes' then 1 else 0 end) AS [Number of Employees Attrited]
FROM HREmployeeAttrition
GROUP BY jobrole
ORDER BY [Number of Employees Attrited] DESC;
```

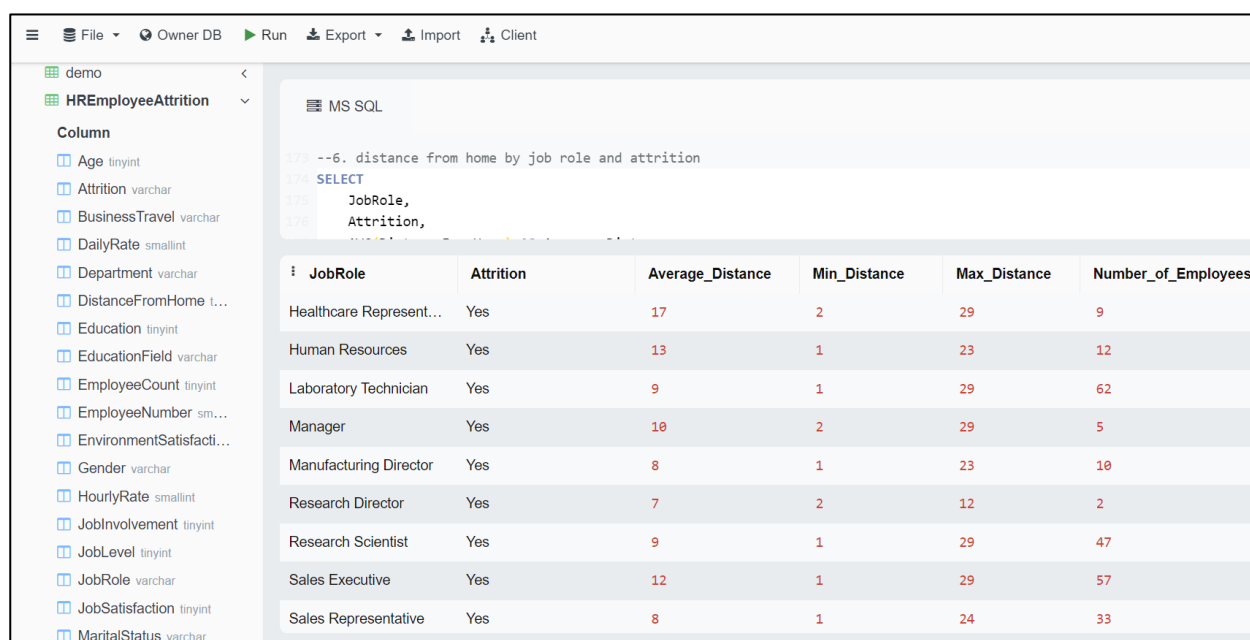
**Results Table:**

| Jobrole                   | Number of Employees Attrited |
|---------------------------|------------------------------|
| Sales Executive           | 326                          |
| Research Scientist        | 292                          |
| Laboratory Technician     | 259                          |
| Manufacturing Director    | 145                          |
| Healthcare Representative | 131                          |
| Manager                   | 102                          |
| Sales Representative      | 83                           |
| Research Director         | 80                           |

## 6. Distance from home by job role and attrition?

This SQL query is designed to analyze and summarize the distance employees live from home, segmented by their job role and attrition status within the **HREmployeeAttrition** dataset.

By examining the average, minimum, and maximum distances, you can assess whether distance from home might be a factor influencing employee attrition in different roles. For example, if certain roles have employees with significantly higher or lower average distances, it might suggest regional or logistical issues.



MS SQL

```
--6. distance from home by job role and attrition
SELECT
    JobRole,
    Attrition,
    Average_Distance,
    Min_Distance,
    Max_Distance,
    Number_of_Employees
```

| JobRole                 | Attrition | Average_Distance | Min_Distance | Max_Distance | Number_of_Employees |
|-------------------------|-----------|------------------|--------------|--------------|---------------------|
| Healthcare Represent... | Yes       | 17               | 2            | 29           | 9                   |
| Human Resources         | Yes       | 13               | 1            | 23           | 12                  |
| Laboratory Technician   | Yes       | 9                | 1            | 29           | 62                  |
| Manager                 | Yes       | 10               | 2            | 29           | 5                   |
| Manufacturing Director  | Yes       | 8                | 1            | 23           | 10                  |
| Research Director       | Yes       | 7                | 2            | 12           | 2                   |
| Research Scientist      | Yes       | 9                | 1            | 29           | 47                  |
| Sales Executive         | Yes       | 12               | 1            | 29           | 57                  |
| Sales Representative    | Yes       | 8                | 1            | 24           | 33                  |

## 7. How does attrition vary across different age groups?

It provides insights into how attrition rates vary across different age groups. By grouping data by age, it helps to understand if certain age groups are more prone to leaving the company.

By sorting the results in descending order of attrition, you can easily identify which age groups have the highest number of employees who have left. This can highlight potential trends or issues related to specific age groups.



File Owner DB Run Export Import Client

demo HREmployeeAttrition

Column

- Age tinyint
- Attrition varchar
- BusinessTravel varchar
- DailyRate smallint
- Department varchar
- DistanceFromHome t...
- Education tinyint
- EducationField varchar
- EmployeeCount tinyint
- EmployeeNumber sm...
- EnvironmentSatisfacti...
- Gender varchar
- HourlyRate smallint
- JobInvolvement tinyint
- JobLevel tinyint
- JobRole varchar
- JobSatisfaction tinyint

MS SQL

```

186 SELECT
187     Age,
188     COUNT(*) AS TotalEmployees,
189     SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) AS TotalAttrition
190 FROM HREmployeeAttrition
191 GROUP BY Age
192 ORDER BY TotalAttrition DESC;

```

| Age | TotalEmployees | TotalAttrition |
|-----|----------------|----------------|
| 29  | 68             | 18             |
| 31  | 69             | 18             |
| 28  | 48             | 14             |
| 33  | 58             | 12             |
| 26  | 39             | 12             |
| 32  | 61             | 11             |
| 35  | 78             | 10             |

| Age | TotalEmployees | TotalAttrition |
|-----|----------------|----------------|
| 30  | 60             | 9              |
| 24  | 26             | 7              |
| 44  | 33             | 6              |
| 21  | 13             | 6              |
| 41  | 40             | 6              |
| 19  | 9              | 6              |
| 25  | 26             | 6              |
| 36  | 69             | 6              |
| 39  | 42             | 6              |
| 20  | 11             | 6              |

## 8. Compare average monthly income by education and attrition?

Here, analyzed employee attrition based on their education level and field of education, focusing on how these factors relate to monthly income.

This query analyzes employee attrition by education level and field of education, calculating the average monthly income and counting the number of employees who have left the company within each group. It groups the data based on education and education field, filters to include only those who have attrited, and orders the results by education level. This analysis helps to understand how attrition rates and income are influenced by educational background, providing valuable insights for strategic HR decisions.

The screenshot displays a database management interface with a menu bar (File, Owner DB, Run, Export, Import, Client) and a sidebar showing a tree view with 'demo' and 'HREmployeeAttrition'. The main area shows an MS SQL query and its results.

```

195 SELECT
196     Education,
197     Educationfield,
198     AVG(monthlyincome) AS [Monthly Income],
199     COUNT(*) AS [Number of Employees Attrited]
200 FROM HREmployeeAttrition
201 GROUP BY education, educationfield, attrition
202 HAVING attrition = 'Yes'
203 ORDER BY education;
  
```

| Education | Educationfield   | Monthly Income | Number of Employees Attrited |
|-----------|------------------|----------------|------------------------------|
| 1         | Human Resources  | 1555           | 1                            |
| 1         | Life Sciences    | 4279           | 8                            |
| 1         | Marketing        | 6810           | 4                            |
| 1         | Medical          | 4350           | 10                           |
| 1         | Other            | 5123           | 2                            |
| 1         | Technical Degree | 3064           | 6                            |



File Owner DB Run Export Import Client

demo

HREmployeeAttrition

Column

- Age tinyint
- Attrition varchar
- BusinessTravel varchar
- DailyRate smallint
- Department varchar
- DistanceFromHome t...
- Education tinyint
- EducationField varchar
- EmployeeCount tinyint
- EmployeeNumber sm...
- EnvironmentSatisfacti...
- Gender varchar
- HourlyRate smallint
- JobInvolvement tinyint
- JobLevel tinyint
- JobRole varchar
- JobSatisfaction tinyint
- MaritalStatus varchar

MS SQL

195 SELECT

Education

| Education | Educationfield   | Monthly Income | Number of Employees Attrited |
|-----------|------------------|----------------|------------------------------|
| 2         | Life Sciences    | 3900           | 18                           |
| 2         | Marketing        | 4470           | 6                            |
| 2         | Medical          | 4592           | 15                           |
| 2         | Other            | 2090           | 1                            |
| 2         | Technical Degree | 5105           | 4                            |
| 3         | Human Resources  | 4332           | 4                            |
| 3         | Life Sciences    | 4709           | 37                           |
| 3         | Marketing        | 6516           | 15                           |
| 3         | Medical          | 4269           | 25                           |
| 3         | Other            | 3477           | 2                            |

| Education | Educationfield   | Monthly Income | Number of Employees Attrited |
|-----------|------------------|----------------|------------------------------|
| 3         | Technical Degree | 4327           | 16                           |
| 4         | Human Resources  | 2073           | 1                            |
| 4         | Life Sciences    | 5108           | 25                           |
| 4         | Marketing        | 7981           | 9                            |
| 4         | Medical          | 5723           | 13                           |
| 4         | Other            | 3760           | 6                            |
| 4         | Technical Degree | 2713           | 4                            |
| 5         | Human Resources  | 2956           | 1                            |
| 5         | Life Sciences    | 7446           | 1                            |
| 5         | Marketing        | 6134           | 1                            |

## Conclusion

In this SQL project, I thoroughly explored and analyzed the **HREmployeeAttrition** dataset to gain insights into employee attrition patterns and other key metrics. This analysis enhanced my SQL skills, particularly in data querying, aggregation, and interpretation. The findings provided a deeper understanding of how various factors, such as department, job role, salary, and demographic attributes, influence employee retention and turnover.

## Recommendations to the Company:

1. **Targeted Retention Strategies:** Develop specific retention strategies for departments and job roles with high attrition rates. This may include career development programs, competitive compensation packages, and enhanced workplace support.
2. **Compensation Review:** Conduct a comprehensive review of salary structures, particularly focusing on roles with lower satisfaction levels and higher turnover. Ensuring competitive and fair compensation can enhance employee satisfaction and reduce attrition.
3. **Employee Engagement Initiatives:** Implement programs to boost engagement, such as regular feedback sessions, recognition programs, and opportunities for professional growth.
4. **Demographic-Specific Policies:** Recognize and address the diverse needs of different employee demographics. For example, younger employees may value career development opportunities, while older employees might prioritize job stability and benefits.

## Reference

1. SQL Tutorial - [https://www.w3schools.com/sql/sql\\_syntax.asp](https://www.w3schools.com/sql/sql_syntax.asp)
2. IBM HR Analytics Employee Attrition & Performance - <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>
3. SQL Tuning - <https://www.tutorialspoint.com/sql/sql-syntax.htm>
4. Database Management System - <https://www.geeksforgeeks.org/sql-tutorial/>