**CAPSTONE 01 GROUP E**

### Submitted By:

1. Farrah Ejaz [farrahnaveed0@gmail.com](mailto:farrahnaveed0@gmail.com)
2. LinkedIn www.linkedin.com/in/farrahejaz95

## HR ANALYTICS ATTRITION ANALYSIS

This dataset contains the details related to the HR department. It’s an annual data of employees from 03 departments. Some of the employees have left the company this year. Now, we have to analyze the reason behind Attrition. In this data, we have run multiple queries to analyze the relationship between Attrition and different attributes such as Age, marital status, Overtime, job role, salary, satisfaction and performance rating, etc. After running queries, we shared our insights in the presentation followed by recommendations.

# DATA ANALYSIS PROCESS

## Understanding Business Problems and Define Objectives

We started with a basic understanding of the problem and define objectives for the dataset.

## Data Preparation and Cleaning

Then, we prepare Data by making sure it’s clean, checking the data formats, finding null values (if any), check duplicates in the primary key.

## Data Processing and Analysis

Import data in SQL and run multiple queries to analyze different perspectives of data. Apply Attrition Rate Formula and divide it into different attributes to analyze the reasons behind it.

## Visualization, generating insights and Recommendations

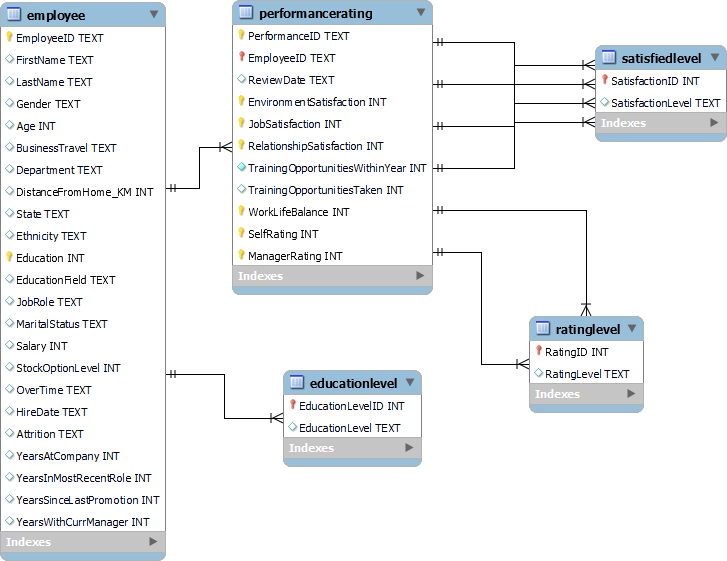
Based on the analysis, we created some visuals that elaborate on the findings. Based on insights, we gave some recommendations at the end of the presentations.

## Stakeholders

Stakeholders are the persons who have invested their time and efforts into projects

1. **The top management** team formulates policies and procedures in the long run.
2. **HR head/manager** is the person who is going to present the department's performance to upper management.
3. **Data Analytics Team** member whose results are important in making an effective data-driven decision.
4. **Employees** whose job roles and salaries are associated with the results

# ER DIAGRAM



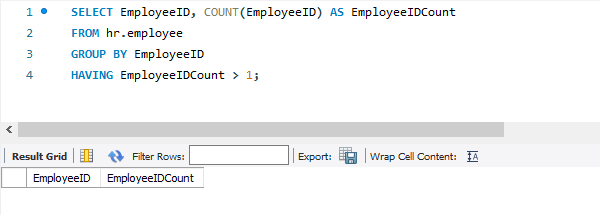
This is the ER Diagram of the dataset. It has 05 tables. Employee, Performance rating, education level, satisfaction level, and rating level.

* **Employee ID** is connected to Performance Rating in which different performance reviews have been taken of employees over a different period.
* **The education level ID** field of Educatio level table is connected to the Education field of the Employee table which gives the details of different education fields of employees
* **Rating ID of Rating Level table** is connected to Self Rating and Manager Rating of performance table which gives details of rating levels in the performance rating table
* **SatisfactionID of Satisfied level** is connected to Job Satisfaction, Environment Satisfaction, Relationship Satisfaction, and Worklife Balance in the performance rating table.

### Are there any duplicates in the data?

SELECT EmployeeID, COUNT(EmployeeID) AS EmployeeIDCount FROM hr.employee

GROUP BY EmployeeID HAVING EmployeeIDCount > 1;



# Attrition %

***Attrition Rate % = no. of employees who leave the company / Total Employees***

***Attrition Rate % of specific attribute= no. of employees leave in a given attribute / Total Attrition Employees***

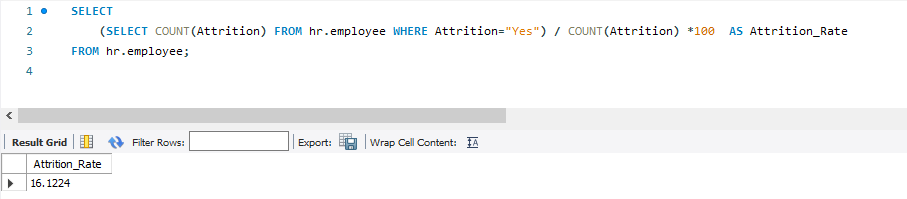
In this section, we calculated Attrition Rate and divided it into different attributes (Marital Status, Agegroup, Distance, Job role and salary, ethnicity, overtime, stock options, tenure etc) to see the reason of Attrition.

**What is the attrition % of the employees?**

SELECT

(SELECT COUNT(Attrition) FROM hr.employee WHERE Attrition="Yes") / COUNT(Attrition) \*100 AS Attrition\_Rate

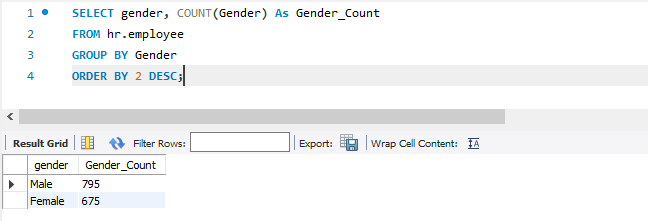
FROM hr.employee;



## Demographic Analysis and Attrition Rate What is the Gender Count?

SELECT gender, COUNT(Gender) As Gender\_Count

FROM hr.employee GROUP BY Gender ORDER BY 2 DESC;



## Gender Ratio in %

SELECT

((SELECT COUNT(Gender) FROM hr.employee WHERE Gender='Female') / COUNT(Gender) \*100) AS Female\_Employee\_perc,

((SELECT COUNT(Gender) FROM hr.employee WHERE Gender='Male') / COUNT(Gender) \* 100) AS Male\_Employee\_perc

FROM hr.employee;



## Age Group Analysis

WITH CTE AS (SELECT

CASE WHEN Age bETWEEN 0 AND 20 THEN '<20'

WHEN Age BETWEEN 20 AND 29 THEN '20-29'

WHEN Age BETWEEN 30 AND 39 THEN '30-40'

WHEN Age BETWEEN 40 AND 49 THEN '40-50'

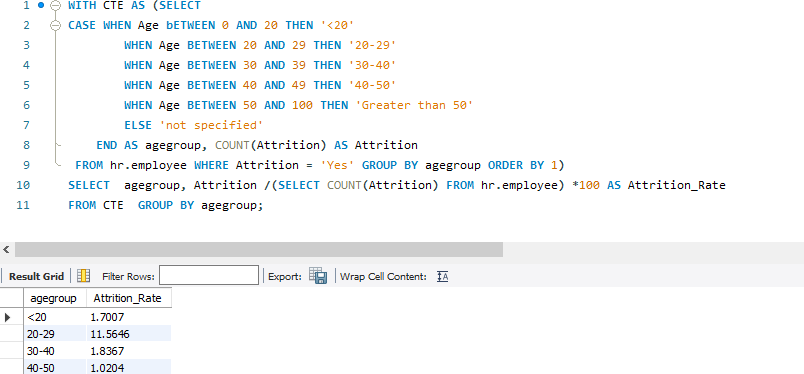
WHEN Age BETWEEN 50 AND 100 THEN 'Greater than 50'

ELSE 'not specified'

END AS agegroup, COUNT(Attrition) AS Attrition

FROM hr.employee WHERE Attrition = 'Yes' GROUP BY agegroup ORDER BY 1)

SELECT agegroup, Attrition /(SELECT COUNT(Attrition) FROM hr.employee) \*100 AS Attrition\_Rate FROM CTE GROUP BY agegroup;



## Distance and Attrition Rate

WITH CTE AS (SELECT

CASE when DistanceFromHome\_KM < 10 then '0-10'

when DistanceFromHome\_KM between 10 and 19 then '10-18'

when DistanceFromHome\_KM between 20 and 28 then '20-27'

when DistanceFromHome\_KM between 29 and 37 then '28-36'

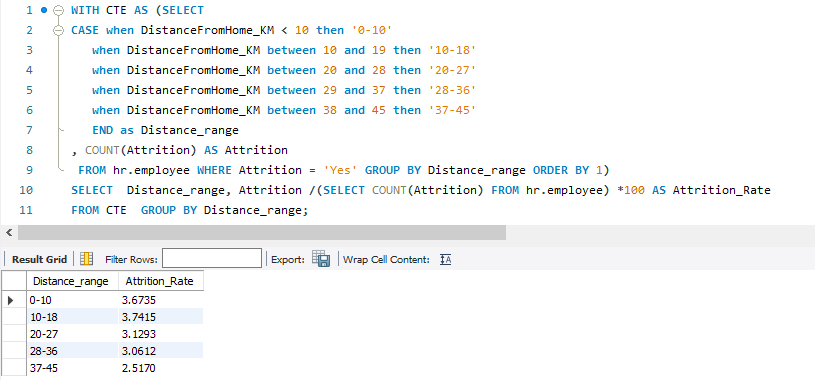
when DistanceFromHome\_KM between 38 and 45 then '37-45'

END as Distance\_range

, COUNT(Attrition) AS Attrition

FROM hr.employee WHERE Attrition = 'Yes' GROUP BY Distance\_range ORDER BY 1)

SELECT Distance\_range, Attrition /(SELECT COUNT(Attrition) FROM hr.employee) \*100 AS Attrition\_Rate FROM CTE GROUP BY Distance\_range;



## Attrition Rate (State Wise)

SELECT

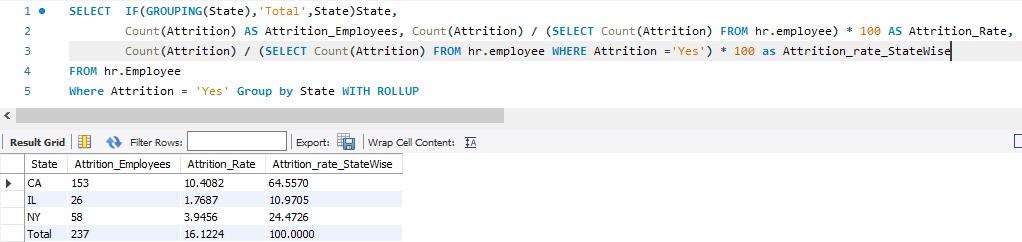
IF(GROUPING(State),'Total',State)State,

Count(Attrition) AS Attrition\_Employees, Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee) \* 100 AS Attrition\_Rate,

Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee WHERE Attrition ='Yes') \* 100 as Attrition\_rate\_StateWise

FROM hr.Employee

Where Attrition = 'Yes' Group by State WITH ROLLUP



## Ethnicity

SELECT

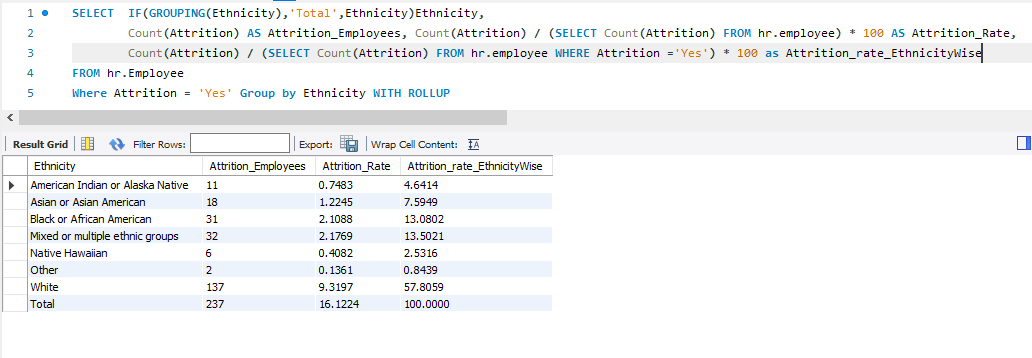
IF(GROUPING(Ethnicity),'Total',Ethnicity)Ethnicity,

Count(Attrition) AS Attrition\_Employees, Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee) \* 100 AS Attrition\_Rate,

Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee WHERE Attrition ='Yes') \* 100 as Attrition\_rate\_EthnicityWise

FROM hr.Employee

Where Attrition = 'Yes' Group by Ethnicity WITH ROLLUP



## State, Ethnicity Analysis along with their Average Salary and Attrition Rate

SELECT

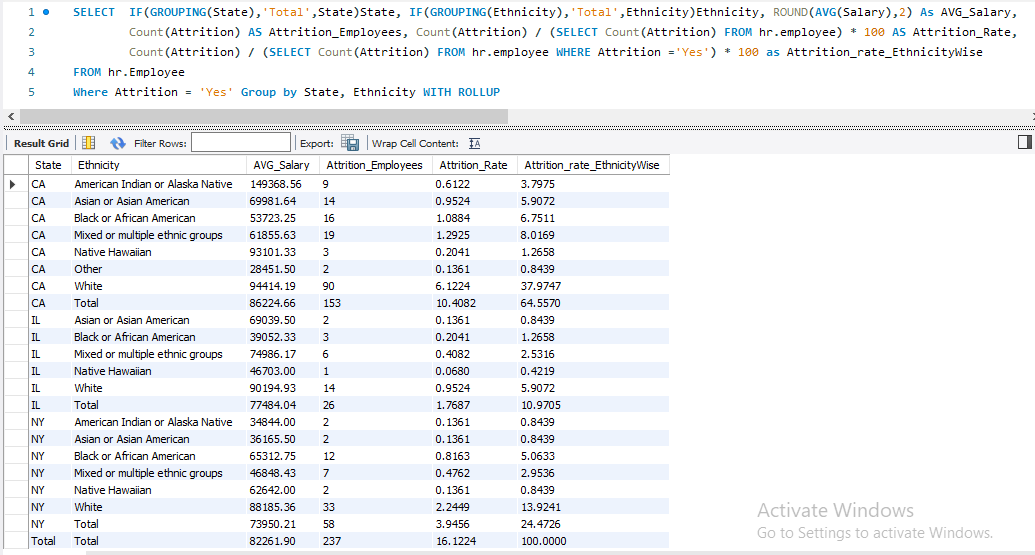
IF(GROUPING(State),'Total',State)State, IF(GROUPING(Ethnicity),'Total',Ethnicity)Ethnicity, ROUND(AVG(Salary),2) As AVG\_Salary,

Count(Attrition) AS Attrition\_Employees, Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee) \* 100 AS Attrition\_Rate,

Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee WHERE Attrition ='Yes') \* 100 as Attrition\_rate\_EthnicityWise

FROM hr.Employee

Where Attrition = 'Yes' Group by State, Ethnicity WITH ROLLUP



## Marital Status and Attrtion Analysis

SELECT

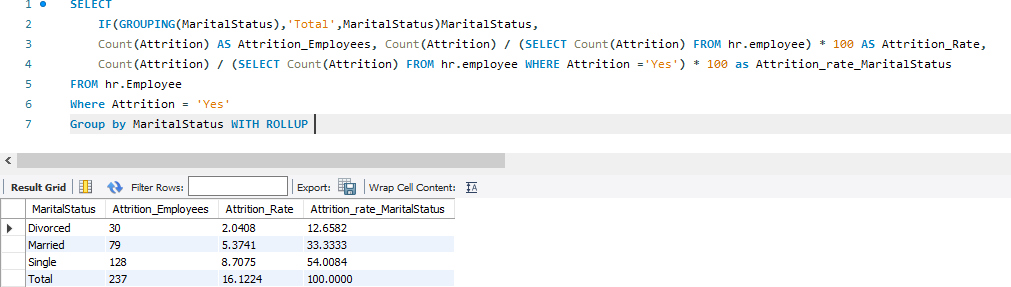
IF(GROUPING(MaritalStatus),'Total',MaritalStatus)MaritalStatus,

Count(Attrition) AS Attrition\_Employees, Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee) \* 100 AS Attrition\_Rate,

Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee WHERE Attrition ='Yes') \* 100 as Attrition\_rate\_MaritalStatus

FROM hr.Employee Where Attrition = 'Yes'

Group by MaritalStatus WITH ROLLUP



## How business travel affects the attrition of employees?

SELECT

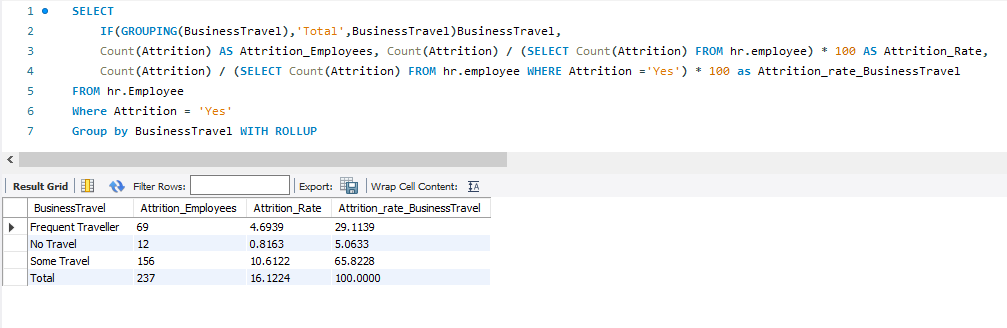
IF(GROUPING(BusinessTravel),'Total',BusinessTravel)BusinessTravel,

Count(Attrition) AS Attrition\_Employees, Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee) \* 100 AS Attrition\_Rate,

Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee WHERE Attrition ='Yes') \* 100 as Attrition\_rate\_BusinessTravel

FROM hr.Employee Where Attrition = 'Yes'

Group by BusinessTravel WITH ROLLUP



## Which department is facing the highest Attrition?

SELECT

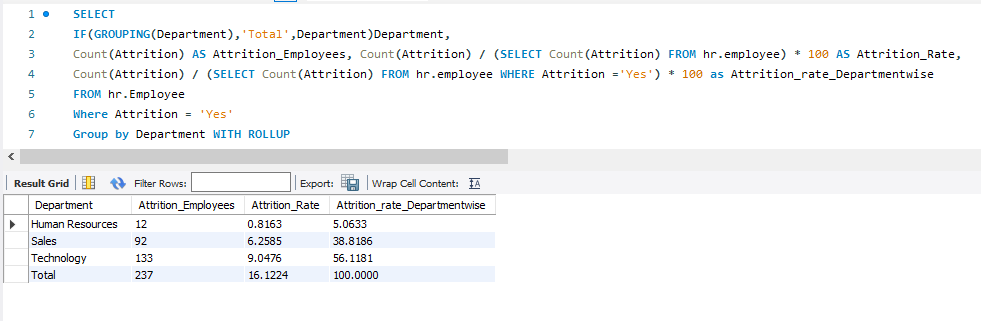
IF(GROUPING(Department),'Total',Department)Department,

Count(Attrition) AS Attrition\_Employees, Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee) \* 100 AS Attrition\_Rate,

Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee WHERE Attrition ='Yes') \* 100 as Attrition\_rate\_Departmentwise

FROM hr.Employee Where Attrition = 'Yes'

Group by Department WITH ROLLUP



## Which job role has the highest Attrition?

SELECT

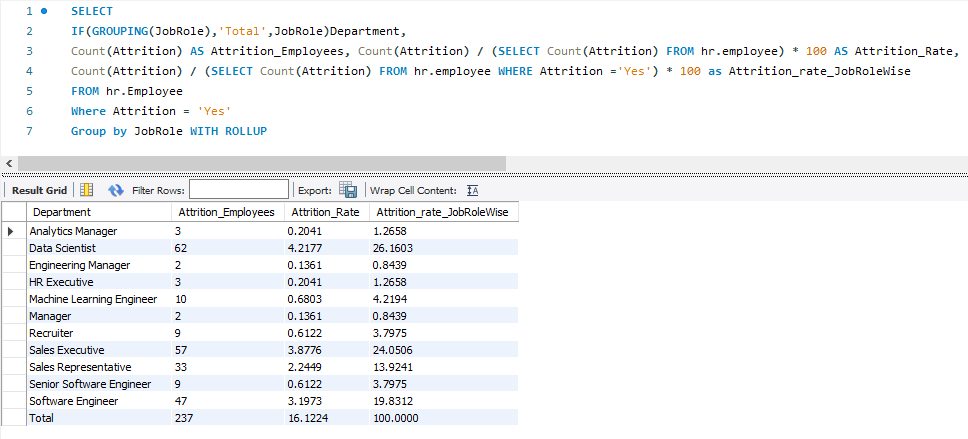
IF(GROUPING(JobRole),'Total',JobRole)Department,

Count(Attrition) AS Attrition\_Employees, Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee) \* 100 AS Attrition\_Rate,

Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee WHERE Attrition ='Yes') \* 100 as Attrition\_rate\_JobRoleWise

FROM hr.Employee Where Attrition = 'Yes'

Group by JobRole WITH ROLLUP



## Average Salary of Employees along with their Attrition Rate rates

SELECT

IF(GROUPING(Department),'Total',Department)Department,

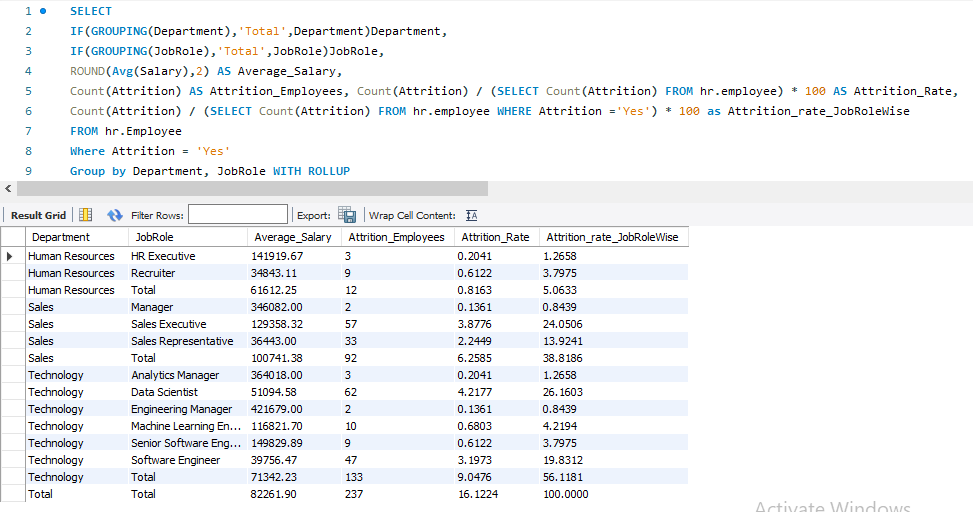
IF(GROUPING(JobRole),'Total',JobRole)JobRole, ROUND(Avg(Salary),2) AS Average\_Salary,

Count(Attrition) AS Attrition\_Employees, Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee) \* 100 AS Attrition\_Rate,

Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee WHERE Attrition ='Yes') \* 100 as Attrition\_rate\_JobRoleWise

FROM hr.Employee Where Attrition = 'Yes'

Group by Department, JobRole WITH ROLLUP



## Does overtime lead to attrition?

SELECT

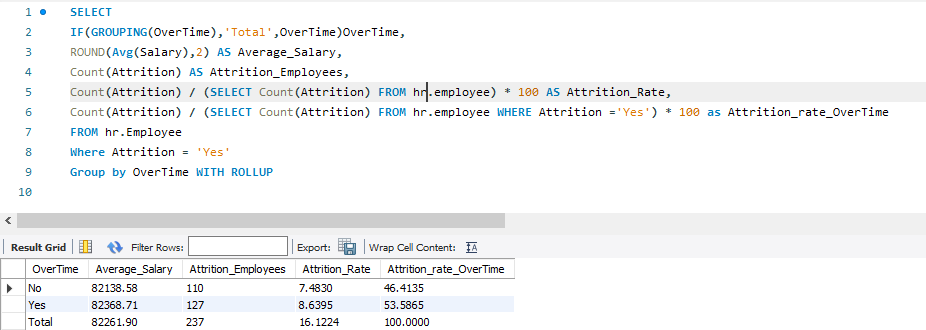
IF(GROUPING(OverTime),'Total',OverTime)OverTime, ROUND(Avg(Salary),2) AS Average\_Salary, Count(Attrition) AS Attrition\_Employees,

Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee) \* 100 AS Attrition\_Rate,

Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee WHERE Attrition ='Yes') \* 100 as Attrition\_rate\_OverTime

FROM hr.Employee Where Attrition = 'Yes'

Group by OverTime WITH ROLLUP



## Tenure Wise Analysis and Attrition Rate

SELECT

IF(GROUPING(YearsAtCompany),'Total',YearsAtCompany)YearsAtCompany,

ROUND(AVG(YearsSinceLastPromotion),2) AS AVG\_YearsSinceLastPromotion,

ROUND(AVG(YearsWithCurrManager),2) AS AVG\_YearsWithCurrManager,

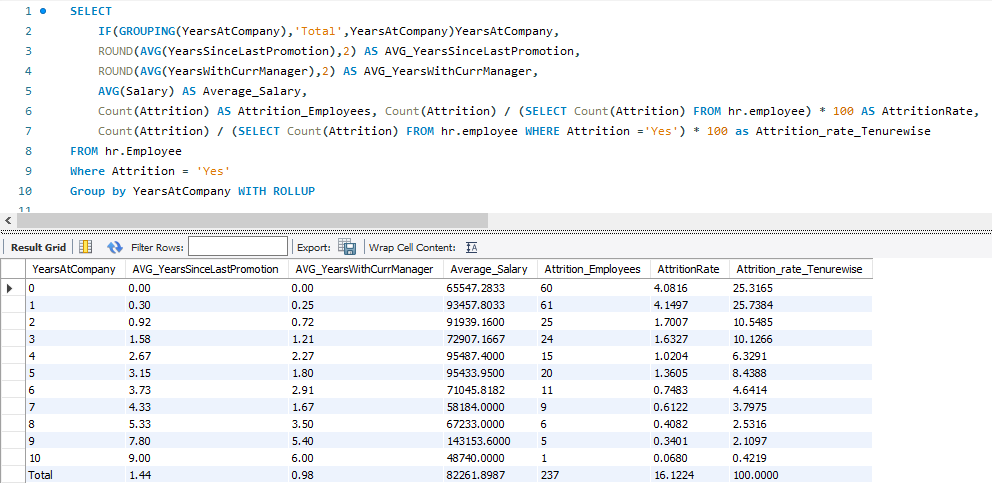
AVG(Salary) AS Average\_Salary,

Count(Attrition) AS Attrition\_Employees, Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee) \* 100 AS AttritionRate,

Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee WHERE Attrition ='Yes') \* 100 as Attrition\_rate\_Tenurewise

FROM hr.Employee Where Attrition = 'Yes'

Group by YearsAtCompany WITH ROLLUP



## Does Unavailability of stock level cause Attrition?

SELECT

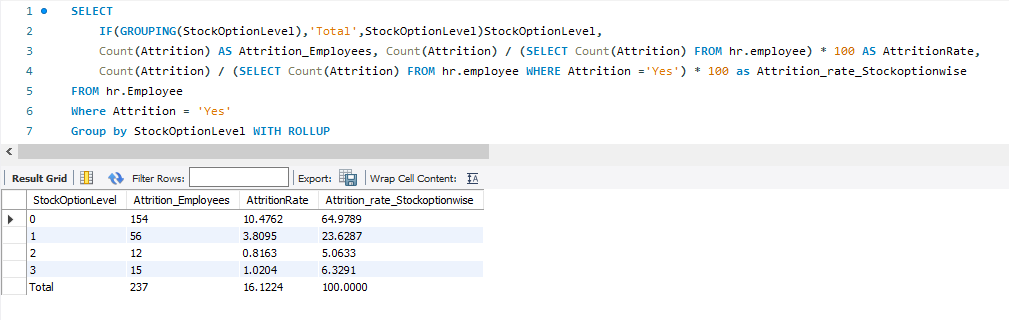
IF(GROUPING(StockOptionLevel),'Total',StockOptionLevel)StockOptionLevel,

Count(Attrition) AS Attrition\_Employees, Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee) \* 100 AS AttritionRate,

Count(Attrition) / (SELECT Count(Attrition) FROM hr.employee WHERE Attrition ='Yes') \* 100 as Attrition\_rate\_Stockoptionwise

FROM hr.Employee Where Attrition = 'Yes'

Group by StockOptionLevel WITH ROLLUP



### PERFORMANCE RATING

In this section, We analyze the average of satisfaction indicators based on their jobroles.

This is the satisfaction survey in which employees rate different perspective of their job according to their job role. The results are above Average which indicates that employees are Neutral or Satisfied with the overall job. The answer to all satisfaction indicator appears above average which is why we suggest that the attrition rate does not relate that much with the ratings.

## Organization’s Performance Indicators according to job Role

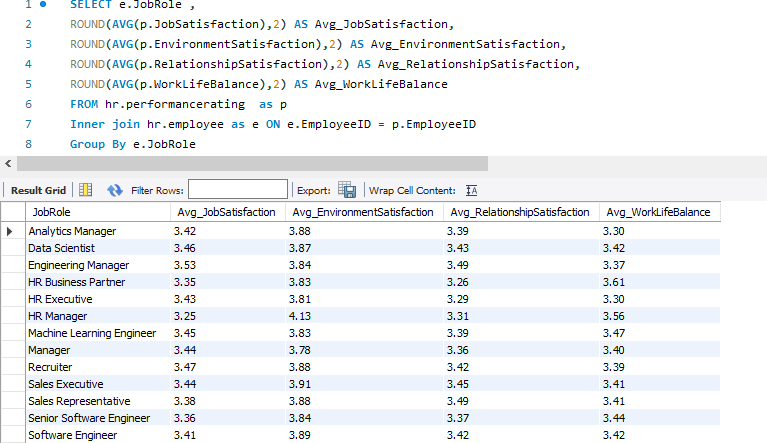
SELECT e.JobRole ,

ROUND(AVG(p.JobSatisfaction),2) AS Avg\_JobSatisfaction, ROUND(AVG(p.EnvironmentSatisfaction),2) AS Avg\_EnvironmentSatisfaction, ROUND(AVG(p.RelationshipSatisfaction),2) AS Avg\_RelationshipSatisfaction, ROUND(AVG(p.WorkLifeBalance),2) AS Avg\_WorkLifeBalance

FROM hr.performancerating as p

Inner join hr.employee as e ON e.EmployeeID = p.EmployeeID Group By e.JobRole

HAVING COUNT(DISTINCT p.EmployeeID) ORDER BY 1;



## Organization’s Performance Indicators according to Over Time

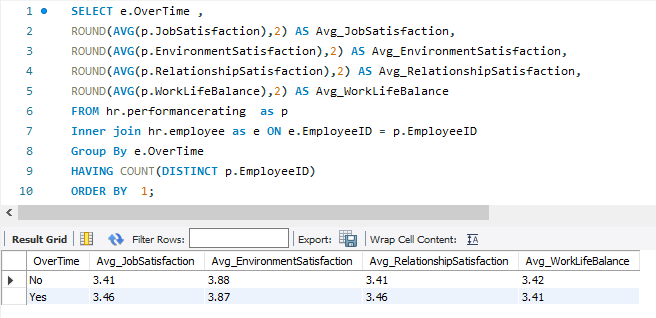
SELECT e.OverTime ,

ROUND(AVG(p.JobSatisfaction),2) AS Avg\_JobSatisfaction, ROUND(AVG(p.EnvironmentSatisfaction),2) AS Avg\_EnvironmentSatisfaction, ROUND(AVG(p.RelationshipSatisfaction),2) AS Avg\_RelationshipSatisfaction, ROUND(AVG(p.WorkLifeBalance),2) AS Avg\_WorkLifeBalance

FROM hr.performancerating as p

Inner join hr.employee as e ON e.EmployeeID = p.EmployeeID Group By e.OverTime

HAVING COUNT(DISTINCT p.EmployeeID) ORDER BY 1;



## Organization’s Performance Indicators according to Marital Status

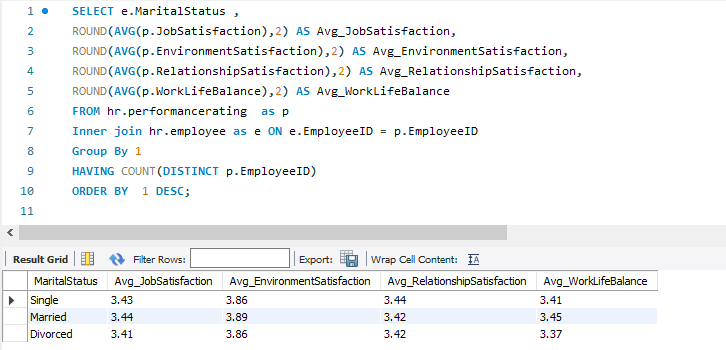
SELECT e.MaritalStatus , ROUND(AVG(p.JobSatisfaction),2) AS Avg\_JobSatisfaction,

ROUND(AVG(p.EnvironmentSatisfaction),2) AS Avg\_EnvironmentSatisfaction, ROUND(AVG(p.RelationshipSatisfaction),2) AS Avg\_RelationshipSatisfaction, ROUND(AVG(p.WorkLifeBalance),2) AS Avg\_WorkLifeBalance

FROM hr.performancerating as p

Inner join hr.employee as e ON e.EmployeeID = p.EmployeeID Group By 1

HAVING COUNT(DISTINCT p.EmployeeID) ORDER BY 1 DESC;



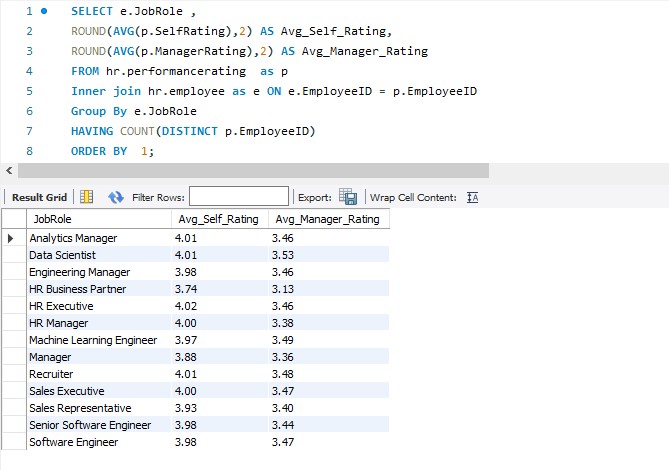
## Manager and Self Rating according to Job Role

SELECT e.JobRole ,

ROUND(AVG(p.SelfRating),2) AS Avg\_Self\_Rating, ROUND(AVG(p.ManagerRating),2) AS Avg\_Manager\_Rating FROM hr.performancerating as p

Inner join hr.employee as e ON e.EmployeeID = p.EmployeeID Group By e.JobRole

HAVING COUNT(DISTINCT p.EmployeeID) ORDER BY 1;



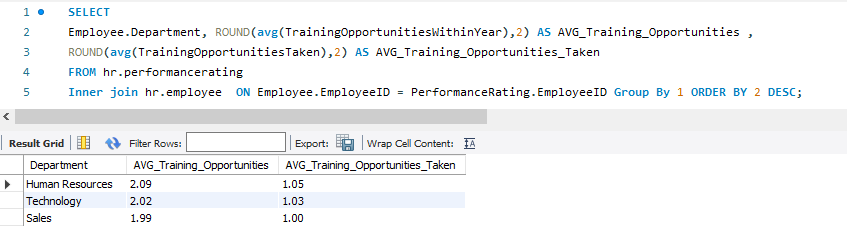
## Average Availability of Training Opportunities and Avg Training Opportunities Taken Chart

SELECT

Employee.Department, ROUND(avg(TrainingOpportunitiesWithinYear),2) AS AVG\_Training\_Opportunities , ROUND(avg(TrainingOpportunitiesTaken),2) AS AVG\_Training\_Opportunities\_Taken

FROM hr.performancerating

Inner join hr.employee ON Employee.EmployeeID = PerformanceRating.EmployeeID Group By 1 ORDER BY 2 DESC;



# Conclusion:

After running all the worries, we summarized our results and presented them in the presentation. Some of the insights are as follows:

* + From the insights gathered, we understand OVERTIME is the main reason for the attrition rate as 30% of employees are those who do overtime, and still do not receive Stock options. Likely, these individuals do not get enough return for their hard work.
  + Employees tend to leave early who have Traveled in their job as compares to those who don’t

have travel

* + Through findings, we analyzed that job roles like sales executives, sales representatives, and software engineers have greater attrition rates and the reason could be the average salary.
  + These roles have less average salary than others and these people do overtime more often and

don’t receive valuable salary packages.

* + The average rating of job satisfaction, Performance satisfaction, environment satisfaction, and work-life balance seems fine, and the rating above average means that employees are satisfied with their job, performance, and environment.
  + Training opportunities in the technology and sales department are far less than required. This may impact the attrition rate.