# SQL_EmployeeRecords

Jas Ahuja

2024-05-14

## Contents

*css chunk to make tables fit*

*Hidden code chunk for mypassword*

*Make a MySQL connection from Rmarkdown*

```
library(DBI)
library(dplyr)
library(dbplyr)

## mypassword = {}!

conn <- dbConnect(RMySQL::MySQL(),
                  dbname = "",
                  Server = "localhost",
                  port = 3306,
                  user = "root",
                  password = mypassword)

knitr::opts_chunk$set(connection = "conn", echo = TRUE, comment = NA, message = FALSE, warning = FALSE)
```

# Course-end Project 1 - ScienceQtech Employee Performance Mapping. . .

## Description

ScienceQtech is a startup that works in the Data Science field. ScienceQtech has worked on fraud detection, market basket, self-driving cars, supply chain, algorithmic early detection of lung cancer, customer sentiment, and the drug discovery field. With the annual appraisal cycle around the corner, the HR department has asked you (Junior Database Administrator) to generate reports on employee details, their performance, and on the project that the employees have undertaken, to analyze the employee database and extract specific data based on different requirements.

## Objective:

To facilitate a better understanding, managers have provided ratings for each employee which will help the HR department to finalize the employee performance mapping. As a DBA, you should find the maximum salary of the employees and ensure that all jobs are meeting the organization's profile standard. You also need to calculate bonuses to find extra cost for expenses. This will raise the overall performance of the organization by ensuring that all required employees receive training.

**Note: You must download the dataset from the course resource section in LMS and create a table to perform the above objective.**

## Dataset description:

**emp_record_table: It contains the information of all the employees.**

- EMP_ID – ID of the employee

- FIRST_NAME – First name of the employee
- LAST_NAME – Last name of the employee
- GENDER – Gender of the employee
- ROLE – Post of the employee
- DEPT – Field of the employee
- EXP – Years of experience the employee has
- COUNTRY – Country in which the employee is presently living
- CONTINENT – Continent in which the country is
- SALARY – Salary of the employee
- EMP_RATING – Performance rating of the employee
- MANAGER_ID – The manager under which the employee is assigned
- PROJ_ID – The project on which the employee is working or has worked on

**Proj_table: It contains information about the projects.**

- PROJECT_ID – ID for the project
- PROJ_Name – Name of the project
- DOMAIN – Field of the project
- START_DATE – Day the project began
- CLOSURE_DATE – Day the project was or will be completed
- DEV_QTR – Quarter in which the project was scheduled
- STATUS – Status of the project currently

**Data_science_team: It contains information about all the employees in the Data Science team.**

- EMP_ID – ID of the employee
- FIRST_NAME – First name of the employee
- LAST_NAME – Last name of the employee
- GENDER – Gender of the employee
- ROLE – Post of the employee
- DEPT – Field of the employee
- EXP – Years of experience the employee has
- COUNTRY – Country in which the employee is presently living
- CONTINENT – Continent in which the country is

## The tasks to be performed:

**1. Create a database named employee, then import tables into the employee database from the given resources.**

```
- data_science_team.csv
- proj_table.csv and
- emp_record_table.csv
```

**Answer-1:** *Done: Using the import function of MySQLWorkbench.*

set nocount on
! This works with SQL database to run multiple queries from a code chunk. This does not work with MySQL and therefore I need to use one query per code chunk

1. Initiate use of our database.

```
use employee
```

2. Look at tables available in *employee*:

```
SHOW TABLES;
```

Table 1: 4 records

| Tables_in_employee |
|---|
| data_science_team |
| emp_record_table |
| proj_table |
| vw_country |

3. Fix *the proj_table*

- Alter table as text feild *project_id* cannot be **PRIMARY KEY**

```
ALTER TABLE proj_table
    MODIFY COLUMN project_id VARCHAR(10)
```

```
ALTER TABLE proj_table
    ADD PRIMARY KEY(project_id)
```

- Our altered table

```
SELECT * FROM proj_table
```

Table 2: 6 records

| project_id | PROJ_NAME | DOMAIN | START_DATE | CLOSURE_DATE | DEV_QTR | STATUS |
|---|---|---|---|---|---|---|
| P103 | Drug Discovery | HEALTHCARE | 04-06-2021 | 6/20/2021 | Q1 | DONE |
| P105 | Fraud Detection | FINANCE | 04-11-2021 | 6/25/2021 | Q1 | DONE |
| P109 | Market Basket Analysis | RETAIL | 04-12-2021 | 6/30/2021 | Q1 | DELAYED |
| P204 | Supply Chain Management | AUTOMOTIVE | 07/15/2021 | 9/28/2021 | Q2 | WIP |
| P302 | Early Detection of Lung Cancer | HEALTHCARE | 11-08-2021 | 12/18/2021 | Q3 | YTS |
| P406 | Customer Sentiment Analysis | RETAIL | 07-09-2021 | 9/24/2021 | Q2 | WIP |

4. Fix *emp_record_table*

- Lets look at the table

```
SELECT * FROM emp_record_table
```

Table 3: Displaying records 1 - 10

| emp_id | first_name | LAST_NAME | GENDER | ROLE | DEPT | EXP | COUNTRY | CONTINENT | SALARY | EMP_RATING | MANAGER_ID | Proj_ID | designation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E001 | Arthur | Black | M | PRESIDENT | ALL | 20 | USA | NORTH AMERICA | 16500 | 5 | NA | NA | NA |
| E005 | Eric | Hoffman | M | LEAD DATA SCIENTIST | FINANCE | 12 | USA | NORTH AMERICA | 8500 | 3 | E103 | P105 | NA |
| E010 | William | Butler | M | LEAD DATA SCIENTIST | AUTOMOTIVE | 12 | FRANCE | EUROPE | 9000 | 2 | E428 | P204 | NA |

| emp_id | first_name | LAST_NAME | GENDER | ROLE | DEPT | EXP | COUNTRY | CONTINENT | SALARY | EMP_RATING | MANAGER_ID | proj_id | designation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E052 | Dianna | Wilson | F | SENIOR DATA SCIENTIST | HEALTHCARE | 6 | CANADA | NORTH AMERICA | 5500 | 5 | E083 | P103 | NA |
| E057 | Dorothy | Wilson | F | SENIOR DATA SCIENTIST | HEALTHCARE | 9 | USA | NORTH AMERICA | 7700 | 1 | E083 | P302 | NA |
| E083 | Patrick | Voltz | M | MANAGER | HEALTHCARE | 15 | USA | NORTH AMERICA | 9500 | 5 | E001 | NA | NA |
| E103 | Emily | Grove | F | MANAGER | FINANCE | 14 | CANADA | NORTH AMERICA | 10500 | 4 | E001 | NA | NA |
| E204 | Karene | Nowak | F | SENIOR DATA SCIENTIST | AUTOMOTIVE | 8 | GERMANY | EUROPE | 7500 | 5 | E428 | P204 | NA |
| E245 | Nian | Zhen | M | SENIOR DATA SCIENTIST | RETAIL | 6 | CHINA | ASIA | 6500 | 2 | E583 | P109 | NA |
| E260 | Roy | Collins | M | SENIOR DATA SCIENTIST | RETAIL | 7 | INDIA | ASIA | 7000 | 3 | E583 | NA | NA |

- Alter the table and put *emp_id* as **PRIMARY_KEY**

```
ALTER TABLE emp_record_table
    MODIFY COLUMN emp_id VARCHAR(10)
```

```
ALTER TABLE emp_record_table
    ADD PRIMARY KEY(emp_id)
```

- *proj_id* needs to be set as **FOREIGN_KEY**
- it cannot be NA or Null! |
- feild type also needs to match in the two tables

```
SELECT * FROM emp_record_table
WHERE proj_id = 'NA' OR proj_id IS NULL
```

Table 4: 7 records

| emp_id | first_name | LAST_NAME | GENDER | ROLE | DEPT | EXP | COUNTRY | CONTINENT | SALARY | EMP_RATING | MANAGER_ID | proj_id | designation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E001 | Arthur | Black | M | PRESIDENT | ALL | 20 | USA | NORTH AMERICA | 16500 | 5 | NA | NA | NA |
| E083 | Patrick | Voltz | M | MANAGER | HEALTHCARE | 15 | USA | NORTH AMERICA | 9500 | 5 | E001 | NA | NA |
| E103 | Emily | Grove | F | MANAGER | FINANCE | 14 | CANADA | NORTH AMERICA | 10500 | 4 | E001 | NA | NA |

| emp_id | first_name | LAST_NAME | GENDER | ROLE | DEPT | EXP | COUNTRY | CONTINENT | SALARY | EMP_RATING | MANAGER_ID | proj_id | designation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E260 | Roy | Collins | M | SENIOR DATA SCIEN-TIST | RETAIL | 7 | INDIA | ASIA | 7000 | 3 | E583 | NA | NA |
| E428 | Pete | Allen | M | MANAGER | AUTOMOTIVE | 12 | GERMANY | EUROPE | 11000 | 4 | E001 | NA | NA |
| E583 | Janet | Hale | F | MANAGER | RETAIL | 14 | COLOMBIA | SOUTH AMER-ICA | 10000 | 2 | E001 | NA | NA |
| E612 | Tracy | Norris | F | MANAGER | RETAIL | 13 | INDIA | ASIA | 8500 | 4 | E001 | NA | NA |

- Fix the NA values in *proj_id*

  In case of Error: 1175- go to menu "MySQLWorkbench" > "Settings" > "SQL Editor" > uncheck "Safe Updates"_
  *on PC it is under "edit"*

```
UPDATE emp_record_table
    SET proj_id=NULL
    WHERE proj_id='NA'
```

- Now change type to match the proj_table.

```
ALTER TABLE emp_record_table
    MODIFY COLUMN proj_id VARCHAR(10)
```

```
ALTER TABLE emp_record_table
    ADD CONSTRAINT fk_proj
    FOREIGN KEY(proj_id) REFERENCES proj_table(PROJECT_ID)
```

- Let's see the altered table:

```
DESCRIBE emp_record_table
```

Table 5: Displaying records 1 - 10

| Field | Type | Null | Key | Default | Extra |
|---|---|---|---|---|---|
| emp_id | varchar(10) | NO | PRI | NA | |
| first_name | varchar(20) | YES | MUL | NA | |
| LAST_NAME | text | YES | | NA | |
| GENDER | text | YES | | NA | |
| ROLE | text | YES | | NA | |
| DEPT | text | YES | | NA | |
| EXP | int | YES | | NA | |
| COUNTRY | text | YES | | NA | |
| CONTINENT | text | YES | | NA | |
| SALARY | int | YES | | NA | |

5. Similarly fix the *data_science_team*.

```
SELECT * FROM data_science_team
```

Table 6: Displaying records 1 - 10

| emp_id | FIRST_NAME | LAST_NAME | GENDER | ROLE | DEPT | EXP | COUNTRY | CONTINENT |
|---|---|---|---|---|---|---|---|---|
| E005 | Eric | Hoffman | M | LEAD DATA SCIENTIST | FINANCE | 11 | USA | NORTH AMERICA |
| E010 | William | Butler | M | LEAD DATA SCIENTIST | AUTOMOTIVE | 12 | FRANCE | EUROPE |
| E052 | Dianna | Wilson | F | SENIOR DATA SCIENTIST | HEALTHCARE | 10 | CANADA | NORTH AMERICA |
| E057 | Dorothy | Wilson | F | SENIOR DATA SCIENTIST | HEALTHCARE | 9 | USA | NORTH AMERICA |
| E204 | Karene | Nowak | F | SENIOR DATA SCIENTIST | AUTOMOTIVE | 8 | GERMANY | EUROPE |
| E245 | Nian | Zhen | M | SENIOR DATA SCIENTIST | RETAIL | 6 | CHINA | ASIA |
| E260 | Roy | Collins | M | SENIOR DATA SCIENTIST | RETAIL | 7 | INDIA | ASIA |
| E403 | Steve | Hoffman | M | ASSOCIATE DATA SCIENTIST | FINANCE | 4 | USA | NORTH AMERICA |
| E478 | David | Smith | M | ASSOCIATE DATA SCIENTIST | RETAIL | 3 | COLOMBIA | SOUTH AMERICA |
| E505 | Chad | Wilson | M | ASSOCIATE DATA SCIENTIST | HEALTHCARE | 5 | CANADA | NORTH AMERICA |

- match emp_id to emp_record_table

```
ALTER TABLE data_science_team
    MODIFY COLUMN emp_id VARCHAR(10)
```

```
ALTER TABLE data_science_team
    ADD CONSTRAINT fk_emp_record_table_emp_id
    FOREIGN KEY(emp_id) REFERENCES emp_record_table(emp_id)
```

**2. Create an ER diagram for the given employee database.**

**Answer-2:** goto "database" > "Reverse Engineer" and follow prompts to get **Figure-1**

**3. Write a query to fetch EMP_ID, FIRST_NAME, LAST_NAME, GENDER, and DEPARTMENT from the employee record table, and make a list of employees and details of their department.**

```
SELECT emp_id, first_name, last_name, gender, dept
    FROM emp_record_table
```

Table 7: Displaying records 1 - 10

| emp_id | first_name | last_name | gender | dept |
|---|---|---|---|---|
| E001 | Arthur | Black | M | ALL |
| E005 | Eric | Hoffman | M | FINANCE |
| E010 | William | Butler | M | AUTOMOTIVE |
| E052 | Dianna | Wilson | F | HEALTHCARE |
| E057 | Dorothy | Wilson | F | HEALTHCARE |
| E083 | Patrick | Voltz | M | HEALTHCARE |

| emp_id | first_name | last_name | gender | dept |
|---|---|---|---|---|
| E103 | Emily | Grove | F | FINANCE |
| E204 | Karene | Nowak | F | AUTOMOTIVE |
| E245 | Nian | Zhen | M | RETAIL |
| E260 | Roy | Collins | M | RETAIL |

**4. Write a query to fetch EMP_ID, FIRST_NAME, LAST_NAME, GENDER, DEPARTMENT, and EMP_RATING if the EMP_RATING is:**

**-4.1 *emp_rating* less than two**

```
SELECT emp_id, first_name, last_name, gender, dept
    FROM emp_record_table
    WHERE emp_rating < 2
```

Table 8: 3 records

| emp_id | first_name | last_name | gender | dept |
|---|---|---|---|---|
| E057 | Dorothy | Wilson | F | HEALTHCARE |
| E532 | Claire | Brennan | F | AUTOMOTIVE |
| E620 | Katrina | Allen | F | RETAIL |

**-4.2 *emp_rating* greater than four**

```
SELECT emp_id, first_name, last_name, gender, dept
    FROM emp_record_table
    WHERE emp_rating > 4
```

Table 9: 4 records

| emp_id | first_name | last_name | gender | dept |
|---|---|---|---|---|
| E001 | Arthur | Black | M | ALL |
| E052 | Dianna | Wilson | F | HEALTHCARE |
| E083 | Patrick | Voltz | M | HEALTHCARE |
| E204 | Karene | Nowak | F | AUTOMOTIVE |

**-4.3 *emp_rating* Between two and four**

```
SELECT emp_id, first_name, last_name, gender, dept
    FROM emp_record_table
    WHERE emp_rating BETWEEN 2 AND 4
```

Table 10: Displaying records 1 - 10

| emp_id | first_name | last_name | gender | dept |
|---|---|---|---|---|
| E005 | Eric | Hoffman | M | FINANCE |
| E010 | William | Butler | M | AUTOMOTIVE |
| E103 | Emily | Grove | F | FINANCE |
| E245 | Nian | Zhen | M | RETAIL |
| E260 | Roy | Collins | M | RETAIL |

| emp_id | first_name | last_name | gender | dept |
|---|---|---|---|---|
| E403 | Steve | Hoffman | M | FINANCE |
| E428 | Pete | Allen | M | AUTOMOTIVE |
| E478 | David | Smith | M | RETAIL |
| E505 | Chad | Wilson | M | HEALTHCARE |
| E583 | Janet | Hale | F | RETAIL |

**-4.4 *emp_rating* bins**

```
SELECT emp_id, first_name, last_name, gender, dept
    FROM emp_record_table
    WHERE emp_rating < 2 OR emp_rating BETWEEN 2 AND 4 OR emp_rating > 4
```

Table 11: Displaying records 1 - 10

| emp_id | first_name | last_name | gender | dept |
|---|---|---|---|---|
| E001 | Arthur | Black | M | ALL |
| E005 | Eric | Hoffman | M | FINANCE |
| E010 | William | Butler | M | AUTOMOTIVE |
| E052 | Dianna | Wilson | F | HEALTHCARE |
| E057 | Dorothy | Wilson | F | HEALTHCARE |
| E083 | Patrick | Voltz | M | HEALTHCARE |
| E103 | Emily | Grove | F | FINANCE |
| E204 | Karene | Nowak | F | AUTOMOTIVE |
| E245 | Nian | Zhen | M | RETAIL |
| E260 | Roy | Collins | M | RETAIL |

**5. Write a query to concatenate the FIRST_NAME and the LAST_NAME of employees in the Finance department from the employee table and then give the resultant column alias as NAME.**

```
SELECT CONCAT(first_name, ",", last_name) AS NAME FROM emp_record_table
    WHERE dept='FINANCE'
```

Table 12: 3 records

| NAME |
|---|
| Eric,Hoffman |
| Emily,Grove |
| Steve,Hoffman |

**6. Write a query to list only those employees who have someone reporting to them. Also, show the number of reporters (including the President).**

```
SELECT * FROM emp_record_table
    WHERE emp_id IN (SELECT DISTINCT manager_id from emp_record_table)
```

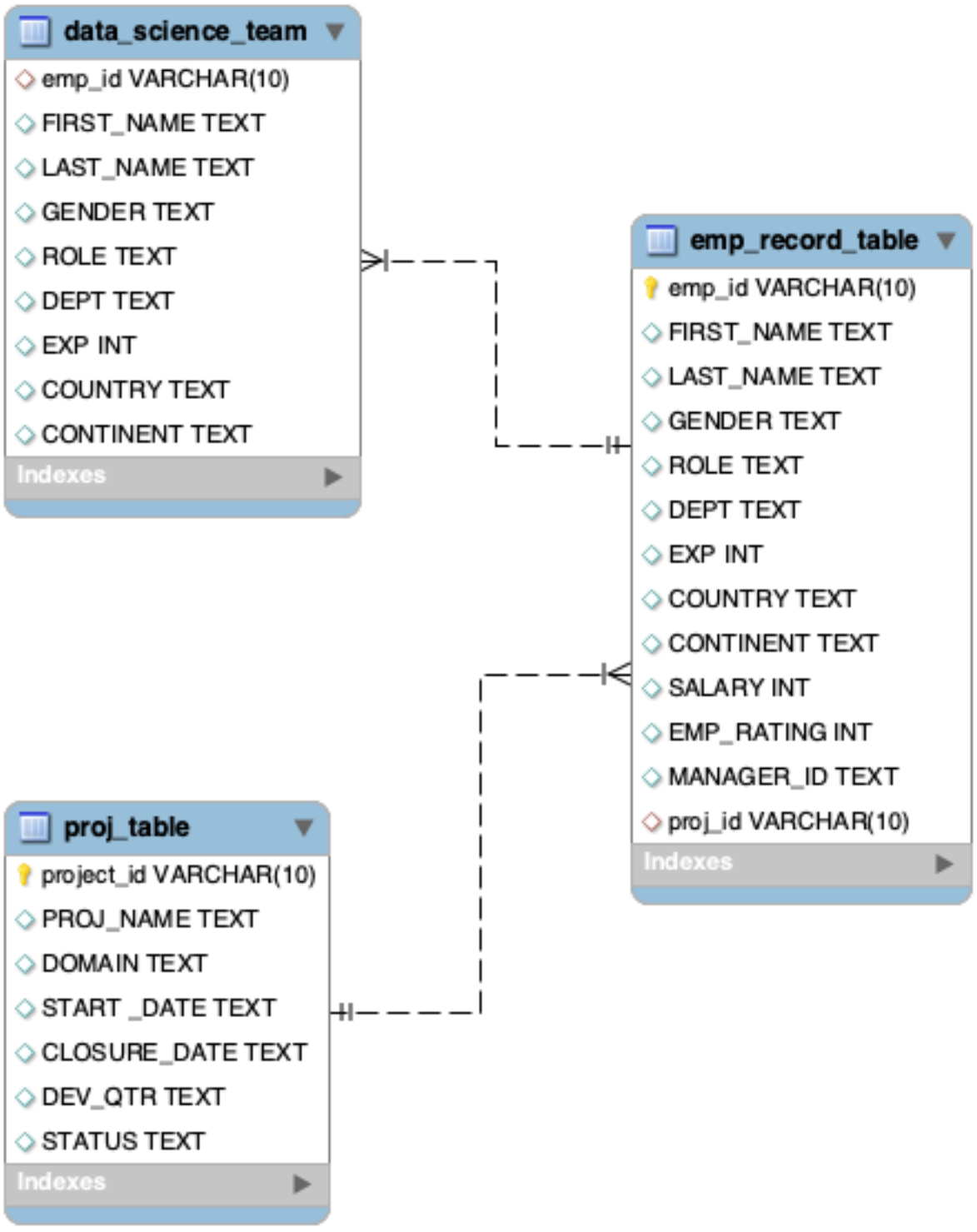


Figure 1: Figure-1: The Entity Relation diagram generated from MySQLWorkbench

Table 13: 6 records

| emp_id | first_name | LAST_NAME | GENDER | ROLE | DEPT | EXP | COUNTRY | CONTINENT | SALARY | EMP_RATING | MANAGER_ID | proj_id | designation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E103 | Emily | Grove | F | MANAGER | FINANCE | 14 | CANADA | NORTH AMERICA | 10500 | 4 | E001 | NA | NA |
| E428 | Pete | Allen | M | MANAGER | AUTOMOTIVE | 11 | GERMANY | EUROPE | 11000 | 4 | E001 | NA | NA |
| E083 | Patrick | Voltz | M | MANAGER | HEALTHCARE | 15 | USA | NORTH AMERICA | 9500 | 5 | E001 | NA | NA |
| E001 | Arthur | Black | M | PRESIDENT | MGT | 20 | USA | NORTH AMERICA | 16500 | 5 | NA | NA | NA |
| E583 | Janet | Hale | F | MANAGER | RETAIL | 14 | COLOMBIA | SOUTH AMERICA | 10000 | 2 | E001 | NA | NA |
| E612 | Tracy | Norris | F | MANAGER | RETAIL | 13 | INDIA | ASIA | 8500 | 4 | E001 | NA | NA |

**7. Write a query to list down all the employees from the healthcare and finance departments using union. Take data from the employee record table.**

```
SELECT * FROM emp_record_table
    WHERE dept = 'HEALTHCARE'
UNION
SELECT * FROM emp_record_table
    WHERE dept = 'FINANCE'
```

Table 14: 7 records

| emp_id | first_name | LAST_NAME | GENDER | ROLE | DEPT | EXP | COUNTRY | CONTINENT | SALARY | EMP_RATING | MANAGER_ID | proj_id | designation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E052 | Dianna | Wilson | F | SENIOR DATA SCIENTIST | HEALTHCARE | 16 | CANADA | NORTH AMERICA | 5500 | 5 | E083 | P103 | NA |
| E057 | Dorothy | Wilson | F | SENIOR DATA SCIENTIST | HEALTHCARE | 19 | USA | NORTH AMERICA | 7700 | 1 | E083 | P302 | NA |
| E083 | Patrick | Voltz | M | MANAGER | HEALTHCARE | 15 | USA | NORTH AMERICA | 9500 | 5 | E001 | NA | NA |
| E505 | Chad | Wilson | M | ASSOCIATE DATA SCIENTIST | HEALTHCARE | 15 | CANADA | NORTH AMERICA | 5000 | 2 | E083 | P103 | NA |
| E005 | Eric | Hoffman | M | LEAD DATA SCIENTIST | FINANCE | 11 | USA | NORTH AMERICA | 8500 | 3 | E103 | P105 | NA |
| E103 | Emily | Grove | F | MANAGER | FINANCE | 14 | CANADA | NORTH AMERICA | 10500 | 4 | E001 | NA | NA |
| E403 | Steve | Hoffman | M | ASSOCIATE DATA SCIENTIST | FINANCE | 11 | USA | NORTH AMERICA | 5000 | 3 | E103 | P105 | NA |

**8. Write a query to list down employee details such as EMP_ID, FIRST_NAME, LAST_NAME, ROLE, DEPARTMENT, and EMP_RATING grouped by dept. Also include the respective employee rating along with the max emp rating for the department.**

```
SELECT emp_id, first_name, last_name, role, dept, MAX(emp_rating) AS max_emp_rating
    FROM emp_record_table
-- this GROUP BY is used for max_emp_rating
    GROUP BY emp_id, first_name, last_name, role, dept
```

Table 15: Displaying records 1 - 10

| emp_id | first_name | last_name | role | dept | max_emp_rating |
|---|---|---|---|---|---|
| E001 | Arthur | Black | PRESIDENT | ALL | 5 |
| E005 | Eric | Hoffman | LEAD DATA SCIENTIST | FINANCE | 3 |
| E010 | William | Butler | LEAD DATA SCIENTIST | AUTOMOTIVE | 2 |
| E052 | Dianna | Wilson | SENIOR DATA SCIENTIST | HEALTHCARE | 5 |
| E057 | Dorothy | Wilson | SENIOR DATA SCIENTIST | HEALTHCARE | 1 |
| E083 | Patrick | Voltz | MANAGER | HEALTHCARE | 5 |
| E103 | Emily | Grove | MANAGER | FINANCE | 4 |
| E204 | Karene | Nowak | SENIOR DATA SCIENTIST | AUTOMOTIVE | 5 |
| E245 | Nian | Zhen | SENIOR DATA SCIENTIST | RETAIL | 2 |
| E260 | Roy | Collins | SENIOR DATA SCIENTIST | RETAIL | 3 |

**9. Write a query to calculate the minimum and the maximum salary of the employees in each role. Take data from the employee record table.**

```
SELECT ROLE, MIN(salary), MAX(salary)
    FROM emp_record_table
    GROUP BY ROLE
```

Table 16: 6 records

| ROLE | MIN(salary) | MAX(salary) |
|---|---|---|
| PRESIDENT | 16500 | 16500 |
| LEAD DATA SCIENTIST | 8500 | 9000 |
| SENIOR DATA SCIENTIST | 5500 | 7700 |
| MANAGER | 8500 | 11000 |
| ASSOCIATE DATA SCIENTIST | 4000 | 5000 |
| JUNIOR DATA SCIENTIST | 2800 | 3000 |

**10. Write a query to assign ranks to each employee based on their experience. Take data from the employee record table.**

```
SELECT *, RANK() OVER (ORDER by exp DESC)
    FROM emp_record_table
```

Table 17: Displaying records 1 - 10

| emp_id | first_name | LAST_NAME | GENDER | ROLE | DEPT | EXP | COUNTRY | CONTINENT | SALARY | EMP_RATING | MANAGER_id | Proj_id | Designation | RANK() OVER (ORDER by exp DESC) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E001 | Arthur | Black | M | PRESIDENT | ALL | 20 | USA | NORTH AMERICA | 16500 | 5 | NA | NA | NA | 1 |
| E083 | Patrick | Voltz | M | MANAGER | HEALTHCARE | 15 | USA | NORTH AMERICA | 9500 | 5 | E001 | NA | NA | 2 |
| E103 | Emily | Grove | F | MANAGER | FINANCE | 12 | CANADA | NORTH AMERICA | 10500 | 4 | E001 | NA | NA | 3 |
| E428 | Pete | Allen | M | MANAGER | AUTOMOTIVE | 14 | GERMANY | EUROPE | 11000 | 4 | E001 | NA | NA | 3 |
| E583 | Janet | Hale | F | MANAGER | RETAIL | 14 | COLOMBIA | SOUTH AMERICA | 10000 | 2 | E001 | NA | NA | 3 |
| E612 | Tracy | Norris | F | MANAGER | RETAIL | 13 | INDIA | ASIA | 8500 | 4 | E001 | NA | NA | 6 |
| E010 | William | Butler | M | LEAD DATA SCIENTIST | AUTOMOTIVE | 12 | FRANCE | EUROPE | 9000 | 2 | E428 | P204 | NA | 7 |
| E005 | Eric | Hoffman | M | LEAD DATA SCIENTIST | FINANCE | 11 | USA | NORTH AMERICA | 8500 | 3 | E103 | P105 | NA | 8 |
| E057 | Dorothy | Wilson | F | SENIOR DATA SCIENTIST | HEALTHCARE | 9 | USA | NORTH AMERICA | 7700 | 1 | E083 | P302 | NA | 9 |
| E204 | Karene | Nowak | F | SENIOR DATA SCIENTIST | AUTOMOTIVE | 8 | GERMANY | EUROPE | 7500 | 5 | E428 | P204 | NA | 10 |

*NB: Lets understand this one-*
- _SELECT *: This selects all columns from the emp_record_table._
- *RANK() OVER (ORDER by exp DESC): This is a window function that calculates the rank of each record based on the exp column in descending order.*
- *The ORDER BY exp DESC specifies that the records should be ranked based on the exp column in descending order*

**11. Write a query to create a view that displays employees in various countries whose salary is more than six thousand. Take data from the employee record table.**

*Can't have two views by the same name!*

```
DROP VIEW IF EXISTS vw_country
```

```
CREATE VIEW vw_country
    AS
```

```
    SELECT first_name, last_name, country FROM emp_record_table
    WHERE salary > 6000
```

**12. Write a nested query to find employees with experience of more than ten years. Take data from the employee record table.**

```
SELECT * FROM
(
  SELECT * FROM emp_record_table
      WHERE exp > 10
) AS T
```

Table 18: 8 records

| emp_id | first_name | LAST_NAME | GENDER | ROLE | DEPT | EXP | COUNTRY | CONTINENT | SALARY | EMP_RATING | MANAGER_ID | PROJ_ID | designation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E001 | Arthur | Black | M | PRESIDENT | ALL | 20 | USA | NORTH AMERICA | 16500 | 5 | NA | NA | NA |
| E005 | Eric | Hoffman | M | LEAD DATA SCIENTIST | FINANCE | 11 | USA | NORTH AMERICA | 8500 | 3 | E103 | P105 | NA |
| E010 | William | Butler | M | LEAD DATA SCIENTIST | AUTOMOTIVE | 12 | FRANCE | EUROPE | 9000 | 2 | E428 | P204 | NA |
| E083 | Patrick | Voltz | M | MANAGER | HEALTHCARE | 16 | USA | NORTH AMERICA | 9500 | 5 | E001 | NA | NA |
| E103 | Emily | Grove | F | MANAGER | FINANCE | 14 | CANADA | NORTH AMERICA | 10500 | 4 | E001 | NA | NA |
| E428 | Pete | Allen | M | MANAGER | AUTOMOTIVE | 12 | GERMANY | EUROPE | 11000 | 4 | E001 | NA | NA |
| E583 | Janet | Hale | F | MANAGER | RETAIL | 14 | COLOMBIA | SOUTH AMERICA | 10000 | 2 | E001 | NA | NA |
| E612 | Tracy | Norris | F | MANAGER | RETAIL | 13 | INDIA | ASIA | 8500 | 4 | E001 | NA | NA |

**13. Write a query to create a *stored procedure* to retrieve the details of the employees whose experience is more than three years. Take data from the employee record table.**

- Need to drop procedure if exists

```
DROP PROCEDURE IF EXISTS sp_exp;
```

```
CREATE PROCEDURE sp_exp()
BEGIN
    SELECT * FROM emp_record_table WHERE exp > 3;
END;
```

**In MySQLWorkbench we need a delimiter change. The code would therefore be-**

```
DELIMITER //

CREATE PROCEDURE sp_exp()
BEGIN
SELECT * FROM emp_record_table WHERE exp > 3;
END//

DELIMITER ;
```

*Here, // is used as the delimiter. It's changed back to ; after the stored procedure definition.*

*This ensures that the SQL interpreter doesn't interpret the semicolons within the stored procedure as statement terminators until the entire procedure is defined.*

**14. Write a query using stored functions in the project table to check whether the job profile assigned to each employee in the data science team matches the organization's set standard.**

The standard being:
- For an employee with experience less than or equal to 2 years assign 'JUNIOR DATA SCIENTIST',
- For an employee with the experience of 2 to 5 years assign 'ASSOCIATE DATA SCIENTIST',
- For an employee with the experience of 5 to 10 years assign 'SENIOR DATA SCIENTIST',
- For an employee with the experience of 10 to 12 years assign 'LEAD DATA SCIENTIST',
- For an employee with the experience of 12 to 16 years assign 'MANAGER'.

```
SELECT * ,
    CASE WHEN exp <= 2 THEN 'JUNIOR DATA SCIENTIST'
    WHEN exp > 2 AND EXP  <= 5 THEN 'ASSOCIATE DATA SCIENTIST'
    WHEN exp > 5 AND EXP  <= 10 THEN 'SENIOR DATA SCIENTIST'
    WHEN exp > 10 AND EXP  <= 12 THEN 'LEAD DATA SCIENTIST'
    WHEN exp > 12 AND EXP  <= 16 THEN 'MANAGER'
    END AS designation
FROM emp_record_table
```

Table 19: Displaying records 1 - 10

| emp_id | first_name | LAST_NAME | GENDER | ROLE | DEPT | EXP | COUNTRY | CONTINENT | SALARY | EMP_MANAGER_id | MANAGER_Proj_id | designation | designation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E001 | Arthur | Black | M | PRESIDENT | ALL | 20 | USA | NORTH AMERICA | 16500 | 5 | NA | NA | NA | NA |
| E005 | Eric | Hoffman | M | LEAD DATA SCIENTIST | FINANCE | 11 | USA | NORTH AMERICA | 8500 | 3 | E103 | P105 | NA | LEAD DATA SCIENTIST |
| E010 | William | Butler | M | LEAD DATA SCIENTIST | AUTOMOTIVE | 12 | FRANCE | EUROPE | 9000 | 2 | E428 | P204 | NA | LEAD DATA SCIENTIST |
| E052 | Dianna | Wilson | F | SENIOR DATA SCIENTIST | HEALTHCARE | 6 | CANADA | NORTH AMERICA | 5500 | 5 | E083 | P103 | NA | SENIOR DATA SCIENTIST |
| E057 | Dorothy | Wilson | F | SENIOR DATA SCIENTIST | HEALTHCARE | 8 | USA | NORTH AMERICA | 7700 | 1 | E083 | P302 | NA | SENIOR DATA SCIENTIST |

| emp_id | first_name | LAST_NAME | GENDER | ROLE | DEPT | EXP | COUNTRY | CONTINENT | SALARY | EMP_RATING | MANAGER_id | Proj_id | designation | designation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E083 | Patrick | Voltz | M | MANAGER | HEALTH | 15 | USA | NORTH AMERICA | 9500 | 5 | E001 | NA | NA | MANAGER |
| E103 | Emily | Grove | F | MANAGER | FINANCE | 14 | CANADA | NORTH AMERICA | 10500 | 4 | E001 | NA | NA | MANAGER |
| E204 | Karene | Nowak | F | SENIOR DATA SCIENTIST | AUTOMOTIVE | 8 | GERMANY | EUROPE | 7500 | 5 | E428 | P204 | NA | SENIOR DATA SCIENTIST |
| E245 | Nian | Zhen | M | SENIOR DATA SCIENTIST | RETAIL | 6 | CHINA | ASIA | 6500 | 2 | E583 | P109 | NA | SENIOR DATA SCIENTIST |
| E260 | Roy | Collins | M | SENIOR DATA SCIENTIST | RETAIL | 7 | INDIA | ASIA | 7000 | 3 | E583 | NA | NA | SENIOR DATA SCIENTIST |

- If we were to add a column like this -

```
ALTER TABLE emp_record_table
    ADD designation VARCHAR(50)
```

```
UPDATE emp_record_table
    SET designation=CASE WHEN exp <= 2 THEN 'JUNIOR DATA SCIENTIST'
        WHEN exp > 2 AND EXP  <= 5 THEN 'ASSOCIATE DATA SCIENTIST'
        WHEN exp > 5 AND EXP  <= 10 THEN 'SENIOR DATA SCIENTIST'
        WHEN exp > 10 AND EXP  <= 12 THEN 'LEAD DATA SCIENTIST'
        WHEN exp > 12 AND EXP  <= 16 THEN 'MANAGER'
    END
```

*lets remove this **designation** column before moving on*

```
ALTER TABLE emp_record_table
    DROP COLUMN designation
```

**15. Create an index to improve the cost and performance of the query to find the employee whose FIRST_NAME is 'Eric' in the employee table after checking the execution plan.**

- index helps in speeding up queries, its performance is quite observable in large datasets.
- index actually creates a backend table that would keep a ~binary record of these values.
- text was not compatible with indexing so we need to change to **VARCHAR**

```
ALTER TABLE emp_record_table
    MODIFY COLUMN first_name VARCHAR(20)
```

```
CREATE INDEX ix_firstname ON emp_record_table(FIRST_NAME)
```

- This table is too short to tell the difference in query time

```
SELECT * FROM emp_record_table WHERE first_name = "eric"
```

Table 20: 1 records

| emp_id | first_name | LAST_NAME | GENDER | ROLE | DEPT | EXP | COUNTRY | CONTINENT | SALARY | EMP_RATING | MANAGER_ID | PROJ_ID | designation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E005 | Eric | Hoffman | M | LEAD DATA SCIEN-TIST | FINANCE | 11 | USA | NORTH AMER-ICA | 8500 | 3 | E103 | P105 | NA |

**16. Write a query to calculate the bonus for all the employees, based on their ratings and salaries (Use the formula: 5% of salary * employee rating).**

```
SELECT * , salary*0.05*emp_rating as bonus
    FROM emp_record_table
```

Table 21: Displaying records 1 - 10

| emp_id | first_name | LAST_NAME | GENDER | ROLE | DEPT | EXP | COUNTRY | CONTINENT | SALARY | EMP_RATING | MANAGER_ID | PROJ_ID | designation | bonus |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E001 | Arthur | Black | M | PRESIDENT | ALL | 20 | USA | NORTH AMER-ICA | 16500 | 5 | NA | NA | NA | 4125 |
| E005 | Eric | Hoffman | M | LEAD DATA SCIEN-TIST | FINANCE | 11 | USA | NORTH AMER-ICA | 8500 | 3 | E103 | P105 | NA | 1275 |
| E010 | William | Butler | M | LEAD DATA SCIEN-TIST | AUTOMOTIVE | 12 | FRANCE | EUROPE | 9000 | 2 | E428 | P204 | NA | 900 |
| E052 | Dianna | Wilson | F | SENIOR DATA SCIEN-TIST | HEALTHCARE | 6 | CANADA | NORTH AMER-ICA | 5500 | 5 | E083 | P103 | NA | 1375 |
| E057 | Dorothy | Wilson | F | SENIOR DATA SCIEN-TIST | HEALTHCARE | 9 | USA | NORTH AMER-ICA | 7700 | 1 | E083 | P302 | NA | 385 |
| E083 | Patrick | Voltz | M | MANAGER | HEALTHCARE | 15 | USA | NORTH AMER-ICA | 9500 | 5 | E001 | NA | NA | 2375 |
| E103 | Emily | Grove | F | MANAGER | FINANCE | 14 | CANADA | NORTH AMER-ICA | 10500 | 4 | E001 | NA | NA | 2100 |
| E204 | Karene | Nowak | F | SENIOR DATA SCIEN-TIST | AUTOMOTIVE | 8 | GERMANY | EUROPE | 7500 | 5 | E428 | P204 | NA | 1875 |
| E245 | Nian | Zhen | M | SENIOR DATA SCIEN-TIST | RETAIL | 6 | CHINA | ASIA | 6500 | 2 | E583 | P109 | NA | 650 |

| emp_id | first_name | LAST_NAME | GENDER | ROLE | DEPT | EXP | COUNTRY | CONTINENT | SALARY | EMP_RATING | MANAGER_ID | PROJ_ID | designation | bonus |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E260 | Roy | Collins | M | SENIOR DATA SCIEN-TIST | RETAIL | 7 | INDIA | ASIA | 7000 | 3 | E583 | NA | NA | 1050 |

**17. Write a query to calculate the average salary distribution based on the continent and country. Take data from the employee record table.**

```
SELECT CONTINENT, COUNTRY, AVG(SALARY) AS average_salary
FROM
    emp_record_table
GROUP BY
    continent, country;
```

Table 22: 7 records

| CONTINENT | COUNTRY | average_salary |
|---|---|---|
| NORTH AMERICA | USA | 9440.000 |
| EUROPE | FRANCE | 9000.000 |
| NORTH AMERICA | CANADA | 7000.000 |
| EUROPE | GERMANY | 7600.000 |
| ASIA | CHINA | 6500.000 |
| ASIA | INDIA | 6166.667 |
| SOUTH AMERICA | COLOMBIA | 5600.000 |