EMPLOYEES DATABASE

ABSTRACT

The purpose of this project was to provide a tutorial on executing diverse queries in a database. The project used the employees database which is a common database for those looking to practice SQL. Several queries were done, and these touched on aggregations, joins, case statements and subqueries.

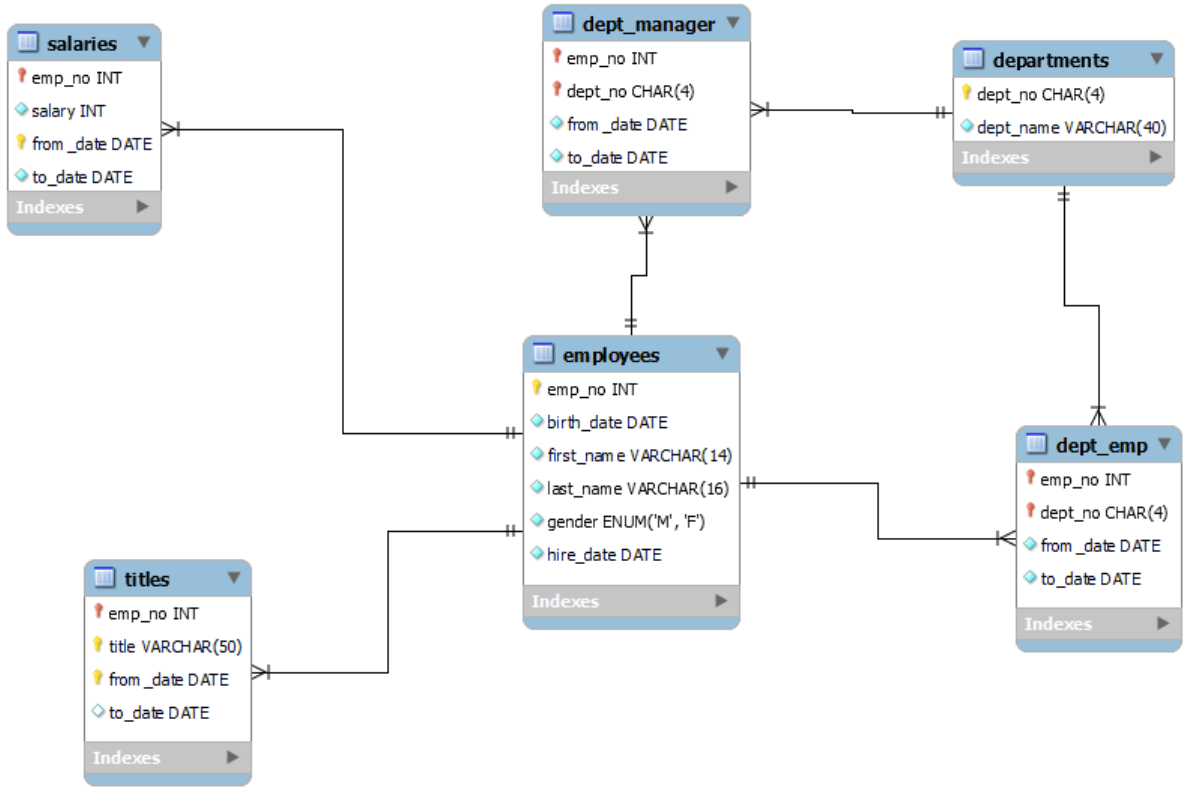
Nigel K. Gondo

Brief Note

For those looking to follow along and practice querying the database you may click on the hyperlink provided to download the database script – EMPLOYEE DATABASE

Entity Relationship Diagram (ERD)

An ERD is a type of flowchart that illustrates how different database objects (tables) are related to each other. The lines connecting the tables represent the cardinality - referring to the maximum number of times an instance in an object can relate to instances in another object. The ERD in question is made up of six tables and these tables have data about employees personal details, salary information, job titles, departments, and manager information related to those departments



For this database schema the cardinality is one-to-many where one side can only have one instance and the other side can have one or many instances. This type of cardinality is popular in relational databases as it uses 'Primary Keys' and 'Foreign Keys' to enforce the relationship. Viewing the connecting lines in the diagram the two vertical lines indicate that only one row is affected by this relationship. On the other hand, the crow's feet with the one vertical line indicates that there are many rows influenced by this relationship.

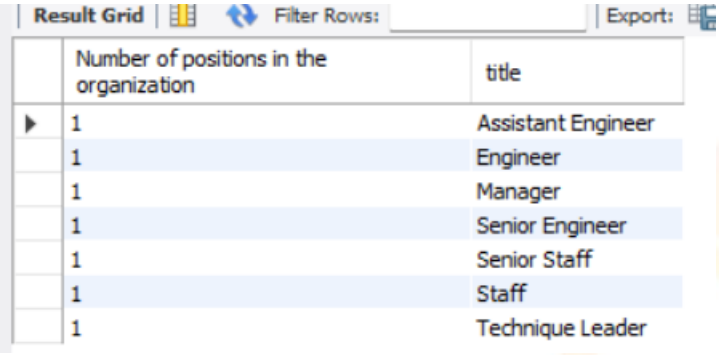
With all that said let's get down to writing some queries.

Happy analysing data warriors! 🕵🏾

Basic queries

```
-- Counting the number of unique positions in the company

SELECT

    COUNT(DISTINCT (title)) AS 'Number of positions in the organization',

    title

FROM

    titles

GROUP BY 2;
```


Output



Result Grid | Filter Rows: | Export:

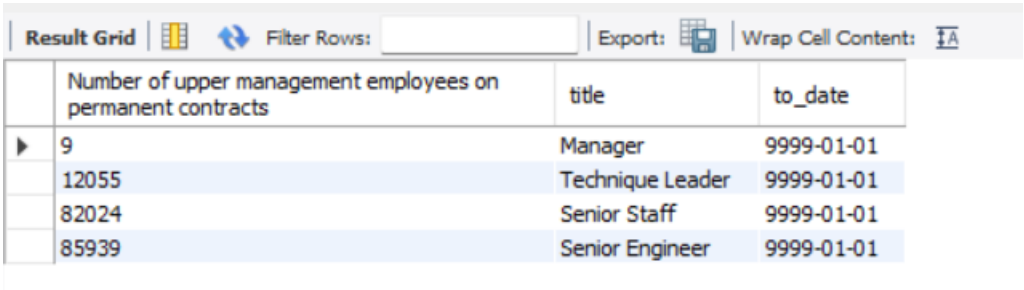
Number of positions in the organization	title
1	Assistant Engineer
1	Engineer
1	Manager
1	Senior Engineer
1	Senior Staff
1	Staff
1	Technique Leader

There are 7 different positions in the organisation and four of these positions are middle to upper management (Manager, Senior Engineer, Senior Staff and Technique Leader).


```
-- Counting the number of employees who are senior employees and on permanent contracts

SELECT

    COUNT(DISTINCT (emp_no)) AS 'Number of upper management employees on permanent contracts',

    title,

    to_date

FROM

    titles

WHERE

    (title LIKE '%Senior%'

        OR title LIKE 'Manager'

        OR title LIKE '%Leader')

        AND to_date = '9999-01-01'
```

```
GROUP BY 2
ORDER BY 1;
```

Output



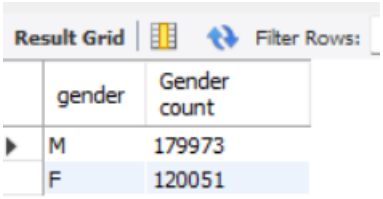
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Number of upper management employees on permanent contracts	title	to_date
9	Manager	9999-01-01
12055	Technique Leader	9999-01-01
82024	Senior Staff	9999-01-01
85939	Senior Engineer	9999-01-01

There are several employees in senior positions, however those in the top tier of upper management are only nine with permanent positions.

```
-- Number of male and female employees that have worked for the company
SELECT
    gender,
    COUNT(gender) AS 'Gender count'
FROM
    employees
GROUP BY 1;
```

Output



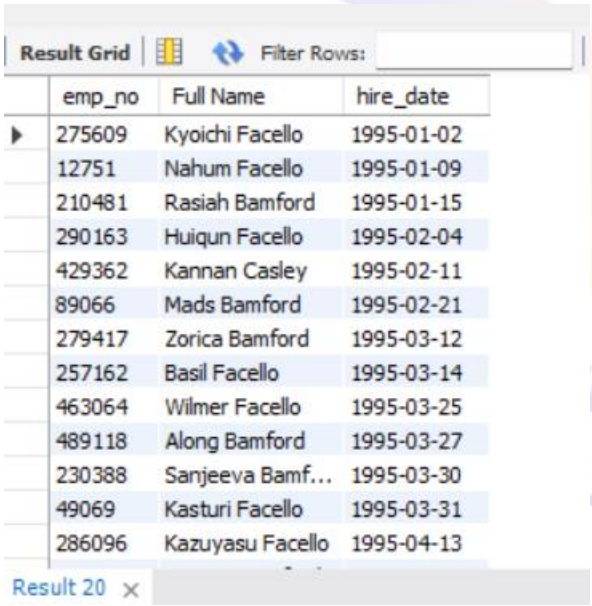
Result Grid | Filter Rows:

gender	Gender count
M	179973
F	120051

More men have worked for the company than women.

```
-- Looking for specific employees hired in a particular period
SELECT
    emp_no,
    CONCAT(first_name, ' ', last_name) AS 'Full Name',
    hire_date
FROM
    employees
WHERE
    last_name IN ('Bamford' , 'Facello', 'Casley')
        AND hire_date BETWEEN '1995-01-01' AND '1999-12-31'
ORDER BY 3;
```

Output



Result Grid | Filter Rows:

emp_no	Full Name	hire_date
275609	Kyoichi Facello	1995-01-02
12751	Nahum Facello	1995-01-09
210481	Rasiah Bamford	1995-01-15
290163	Huiqun Facello	1995-02-04
429362	Kannan Casley	1995-02-11
89066	Mads Bamford	1995-02-21
279417	Zorica Bamford	1995-03-12
257162	Basil Facello	1995-03-14
463064	Wilmer Facello	1995-03-25
489118	Along Bamford	1995-03-27
230388	Sanjeeva Bamf...	1995-03-30
49069	Kasturi Facello	1995-03-31
286096	Kazuyasu Facello	1995-04-13

Result 20

This is the list of specific employees hired between the period of 1995 and 1999.

```
-- the two most and two least hired jobs in the company – Using UNION

SELECT

  *

FROM

  ((SELECT

    COUNT(DISTINCT (emp_no)) AS 'The two most and two least hired jobs',

      title

  FROM

    titles

  GROUP BY 2

  ORDER BY 1 DESC

  LIMIT 2) UNION (SELECT

    COUNT(DISTINCT (emp_no)), title

  FROM

    titles

  GROUP BY 2

  ORDER BY 1 ASC

  LIMIT 2)) AS hiring_result;
```

Output

Result Grid | Filter Rows: | Export:

The two most and two least hired jobs	title
115003	Engineer
107391	Staff
24	Manager
15128	Assistant Engineer

The top two jobs in demand for the company are Engineer and Staff were over a hundred thousand each have been hired for this position, then the two least in demand jobs are Manager and Assistant Engineer.

Intermediate to advanced queries

```
-- The average salaries of each department
SELECT
    FORMAT(AVG(s.salary), '###,###') AS 'Average salary per department for the year 2000',
    d.dept_name
FROM
    salaries s
        JOIN
    employees e ON s.emp_no = e.emp_no
        JOIN
    dept_emp de ON e.emp_no = de.emp_no
        JOIN
    departments d ON de.dept_no = d.dept_no
WHERE
    s.from_date BETWEEN '2000-01-01' AND '2000-12-31'
GROUP BY 2
ORDER BY 1 DESC;
```

Output

Result Grid | Filter Rows: | Export: | Wrap

Average salary per department for the year 2000	dept_name
85,590	Sales
76,461	Marketing
74,847	Finance
64,657	Research
64,323	Production
64,284	Development
63,483	Customer Service
62,061	Quality Management
60,190	Human Resources

In the year 2000 the Sales department had the highest average salaries for the year 2000, by quite a significant margin. This could mean that the sales team could have exceeded their sales targets and got bonuses.

```
--Average salaries of managers who are on permanent contracts
SELECT
    de.emp_no,
    CONCAT(e.first_name, ' ', e.last_name) AS 'Full Name',
    FORMAT((AVG(s.salary)), '###,###') AS 'Average salary',
    t.title,
    d.dept_name
FROM
    departments d
        JOIN
    dept_manager de ON d.dept_no = de.dept_no
        JOIN
    employees e ON e.emp_no = de.emp_no
        JOIN
    salaries s ON s.emp_no = e.emp_no
        JOIN
    titles t ON e.emp_no = t.emp_no
WHERE
    de.to_date = '9999-01-01'
        AND t.title = 'Manager'
GROUP BY 3, 4
ORDER BY 2 DESC;
```


Output

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

emp_no	Full Name	Average salary	title	dept_name
110039	Vishwani Minakawa	87,571	Manager	Marketing
111133	Hauke Zhang	87,422	Manager	Sales
110114	Isamu Legleitner	68,809	Manager	Finance
111534	Hilary Kambil	65,917	Manager	Research
110854	Dung Pesch	59,734	Manager	Quality Management
110567	Leon DasSarma	56,384	Manager	Development
110228	Karsten Sigstam	53,582	Manager	Human Resources
111939	Yuchang Weedman	49,834	Manager	Customer Service
110420	Oscar Ghazalie	46,853	Manager	Production

These are the average salaries the permanent managers have been earning, and sales and marketing are the highest earners, exceeding finance in third by close to 20,000 dollars. Once again this could mean that sales and marketing managers are effectively running their campaigns and concomitant received huge bonuses and commissions.


```
-- Counting the number of male and female employees that started their contracts in the year 2000

SELECT

    COUNT(e.emp_no) AS 'Count of employees',

    e.gender,

    d.dept_name

FROM

    employees e

        JOIN

    dept_emp de ON e.emp_no = de.emp_no

        JOIN

    departments d ON d.dept_no = de.dept_no

WHERE

    de.from_date BETWEEN '2000-01-01' AND '2000-12-31'

GROUP BY 2 , 3

ORDER BY 1 DESC;
```

Output

Result Grid | Filter Rows: | Expo

Count of employees	gender	dept_name
718	M	Production
525	M	Customer Service
491	F	Production
476	M	Development
358	F	Customer Service
354	M	Research
334	F	Development
294	M	Sales
285	M	Quality Management
267	M	Marketing
227	F	Research
223	F	Marketing
198	F	Sales
196	F	Quality Management
123	M	Finance
118	M	Human Resources
84	F	Human Resources
79	F	Finance

Production had the highest number of hires in 2000 where most of them were male employees.

```
-- List of managers that where hired on or after 1987 - Subquery
SELECT
    *
FROM
    employees e
WHERE
    EXISTS( SELECT
        *
      FROM
        titles t
      WHERE
        t.emp_no = e.emp_no
          AND title = 'Manager'
          AND hire_date > '1987-12-31')
ORDER BY gender;
```

Output

Result Grid | Filter Rows: | Edit: | Export/

emp_no	birth_date	first_name	last_name	gender	hire_date
110386	1953-10-04	Shem	Kieras	M	1988-10-14
110420	1963-07-27	Oscar	Ghazalie	M	1992-02-05
110854	1960-08-19	Dung	Pesch	M	1989-06-09
111939	1960-03-25	Yuchang	Weedman	M	1989-07-10
110765	1954-05-22	Rutger	Hofmeyr	F	1989-01-07
111534	1952-06-27	Hilary	Kambil	F	1988-01-31
111784	1956-06-14	Marjo	Giarratana	F	1988-02-12
111877	1962-10-18	Xiaobin	Spinelli	F	1991-08-17
NULL	NULL	NULL	NULL	NULL	NULL

Only eight managers have been hired from 1987 and onwards, with equal number of males and females.

```
-- Average salary of males and females in the company
SELECT
    e.gender,
    FORMAT(AVG(s.salary), '#,###,###') AS 'Average salary'
FROM
    salaries s
        JOIN
    employees e ON s.emp_no = e.emp_no
GROUP BY 1
ORDER BY 2 DESC;
```

Output

Result Grid | Filter Rows:

gender	Average salary
F	63,769
M	63,756

The difference of the average salaries between males and females is negligible and its roughly equal. This company is achieving its goal of gender equality in terms of pay. However, in the next query we need to drill down further by department and see if there are significant differences in salaries for males and females.


```
-- Average salary of males and females per department since company started
SELECT
    e.gender,
    d.dept_name,
    FORMAT(AVG(s.salary), '#,###,###') AS 'Average salary'
FROM
    salaries s
        JOIN
    employees e ON s.emp_no = e.emp_no
        JOIN
    dept_emp de ON e.emp_no = de.emp_no
```

```
    JOIN
    departments d ON de.dept_no = d.dept_no
GROUP BY 1 , 2
ORDER BY 3 DESC;
```

Output

Result Grid | Filter Rows:

gender	dept_name	Average salary
M	Sales	80,880
F	Sales	80,627
M	Marketing	72,198
F	Marketing	71,464
M	Finance	70,327
F	Finance	69,915
M	Research	59,966
F	Research	59,713
M	Production	59,596
M	Development	59,576
F	Production	59,456
F	Development	59,392
F	Customer S...	58,999
M	Customer S...	58,591
F	Quality Man...	57,423
M	Quality Man...	57,207
F	Human Reso...	55,596
M	Human Reso...	55,197

Once again, the salary difference by gender is negligible by a few hundred dollars.


```
-- Checking who is the highest paid permanent employee including their job title and department
SELECT
    e.emp_no,
    CONCAT(e.first_name, ' ', e.last_name) AS 'Full Name',
    e.gender,
    t.title,
    d.dept_name,
    MAX(s.salary) AS 'Highest Salary'
FROM
    employees e
```

```
    JOIN
    salaries s ON e.emp_no = s.emp_no
    JOIN
    titles t ON e.emp_no = t.emp_no
    JOIN
    dept_emp de ON e.emp_no = de.emp_no
    JOIN
    departments d ON d.dept_no = de.dept_no
WHERE
    de.to_date = '9999-01-01';
```

Output

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

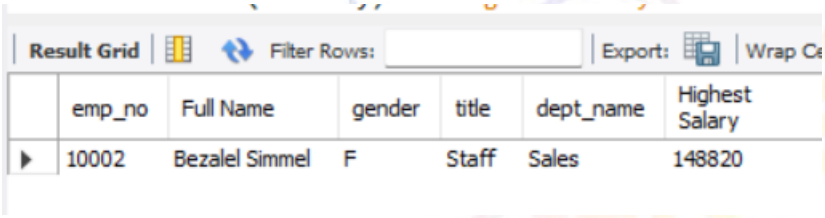
	emp_no	Full Name	gender	title	dept_name	Highest Salary
▶	10001	Georgi Facello	M	Senior Engineer	Development	158220

Georgi Facello is the highest paid male employee in the company, earning 158,220 dollars, who is a Senior Engineer and is under the Development department. Can also check who is the highest paid female employee for the next query.


```
-- Checking who is the highest paid permanent female employee, their job title and department
SELECT
    e.emp_no,
    CONCAT(e.first_name, ' ', e.last_name) AS 'Full Name',
    e.gender,
    t.title,
    d.dept_name,
    MAX(s.salary) AS 'Highest Salary'
FROM
    employees e
    JOIN
    salaries s ON e.emp_no = s.emp_no
```

```
    JOIN
titles t ON e.emp_no = t.emp_no
    JOIN
dept_emp de ON e.emp_no = de.emp_no
    JOIN
departments d ON d.dept_no = de.dept_no
WHERE
de.to_date = '9999-01-01'
    AND e.gender = 'F';
```

Output



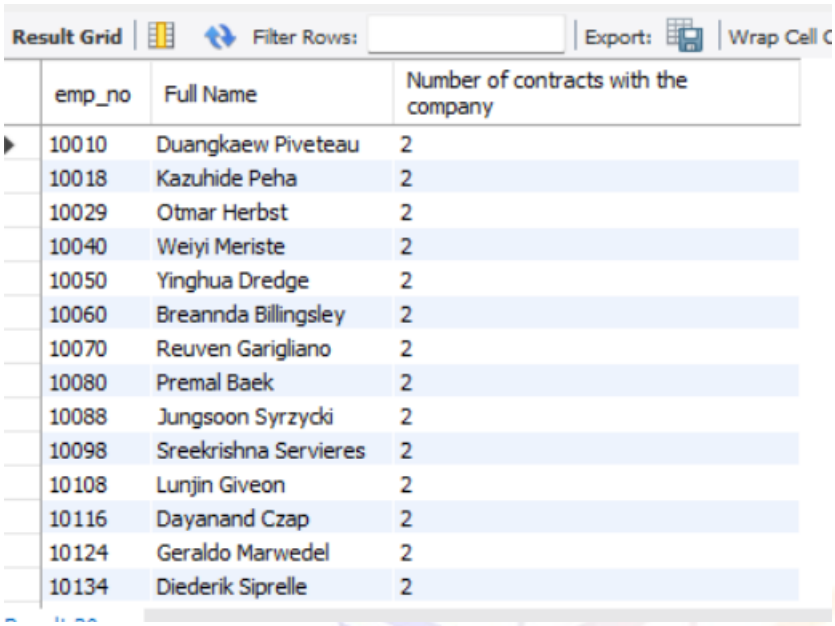
Result Grid | Filter Rows: | Export: | Wrap Ce

emp_no	Full Name	gender	title	dept_name	Highest Salary
10002	Bezalel Simmel	F	Staff	Sales	148820

Bezalel Simmel is the highest paid female employee on a permanent contract, earning 148,820 dollars, who is Staff and is under the Sales department.

```
-- List of employees who have renewed their contracts with the company
SELECT
    de.emp_no,
    CONCAT(e.first_name, ' ', e.last_name) AS 'Full Name',
    COUNT(de.emp_no) AS 'Number of contracts with the company'
FROM
    employees e
    JOIN
    dept_emp de ON e.emp_no = de.emp_no
GROUP BY 1
HAVING COUNT(de.emp_no) >= 2;
```

Output



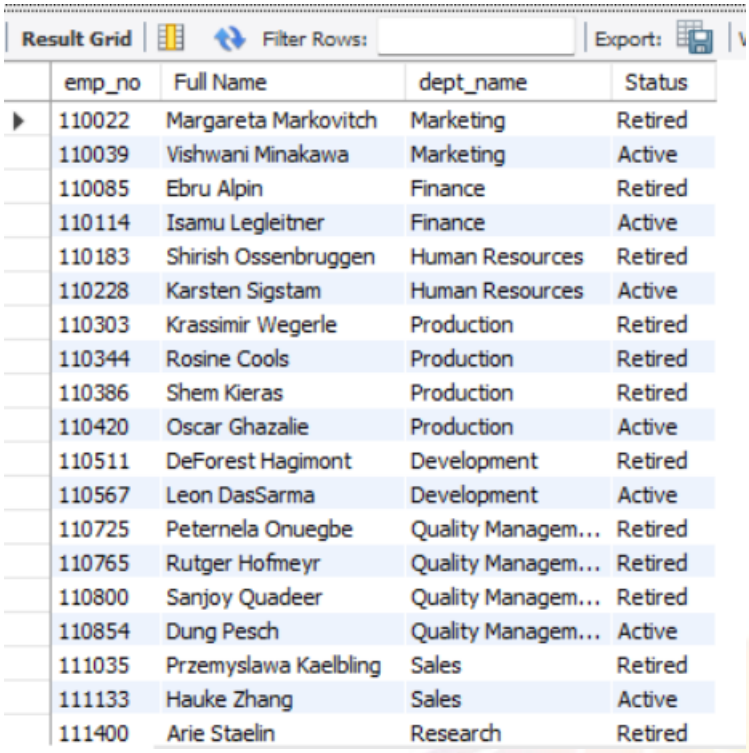
Result Grid | Filter Rows: | Export: | Wrap Cell C

emp_no	Full Name	Number of contracts with the company
10010	Duangkaew Piveteau	2
10018	Kazuhide Peha	2
10029	Otmar Herbst	2
10040	Weiyi Meriste	2
10050	Yinghua Dredge	2
10060	Breannda Billingsley	2
10070	Reuven Garigliano	2
10080	Premal Baek	2
10088	Jungsoon Syrzycki	2
10098	Sreekrishna Servieres	2
10108	Lunjin Giveon	2
10116	Dayanand Czap	2
10124	Geraldo Marwedel	2
10134	Diederik Siprelle	2

This is a list of employees who have had more than one contract with the company (or have renewed their contracts).

```
-- Checking to see whether a manager is retired or active - CASE statement – Will be treating as current we are in as the year 2000

SELECT
    dm.emp_no,
    CONCAT(e.first_name, ' ', e.last_name) AS 'Full Name',
    d.dept_name,
    CASE
        WHEN dm.to_date > '2000-01-01' THEN 'Active'
        ELSE 'Retired'
    END AS 'Status'
FROM
    employees e
        JOIN
    dept_manager dm ON e.emp_no = dm.emp_no
        JOIN
    departments d ON dm.dept_no = d.dept_no;
```

Output



Result Grid | Filter Rows: | Export:

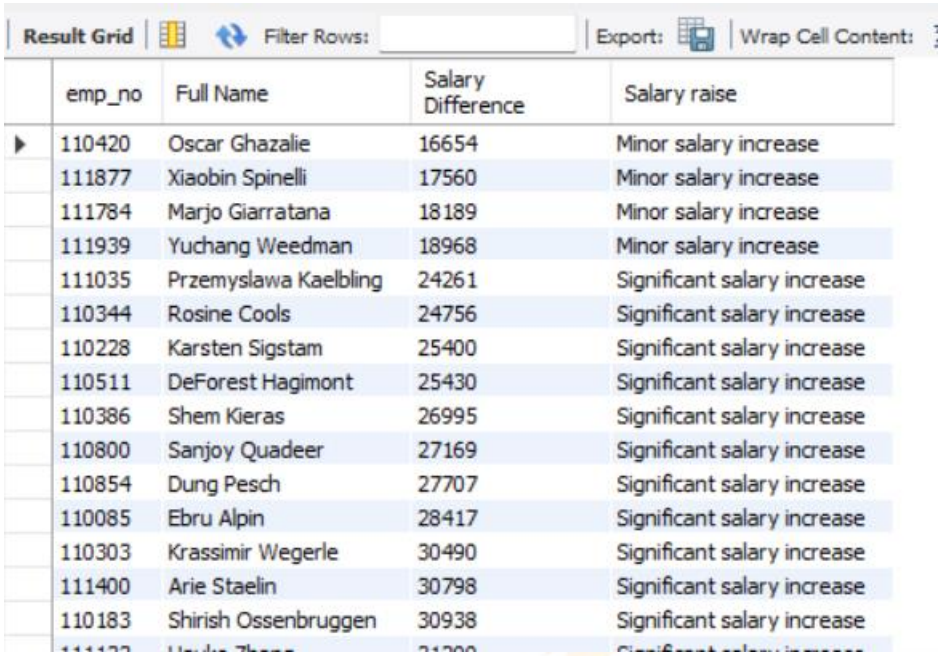
emp_no	Full Name	dept_name	Status
110022	Margareta Markovitch	Marketing	Retired
110039	Vishwani Minakawa	Marketing	Active
110085	Ebru Alpin	Finance	Retired
110114	Isamu Legleitner	Finance	Active
110183	Shirish Ossenbruggen	Human Resources	Retired
110228	Karsten Sigstam	Human Resources	Active
110303	Krassimir Wegerle	Production	Retired
110344	Rosine Cools	Production	Retired
110386	Shem Kieras	Production	Retired
110420	Oscar Ghazalie	Production	Active
110511	DeForest Hagimont	Development	Retired
110567	Leon DasSarma	Development	Active
110725	Peternela Onuegbe	Quality Managem...	Retired
110765	Rutger Hofmeyr	Quality Managem...	Retired
110800	Sanjoy Quadeer	Quality Managem...	Retired
110854	Dung Pesch	Quality Managem...	Active
111035	Przemyslawa Kaelbling	Sales	Retired
111133	Hauke Zhang	Sales	Active
111400	Arie Staelin	Research	Retired

The result of the managers who are retired and those still working for the company.

```
-- Checking on department managers who had salary increases – CASE statement
SELECT
    dm.emp_no,
    CONCAT(e.first_name, ' ', e.last_name) AS 'Full Name',
    MAX(s.salary) - MIN(s.salary) AS 'Salary Difference',
    CASE
        WHEN MAX(s.salary) - MIN(s.salary) > 20000 THEN 'Significant salary increase'
        ELSE 'Minor salary increase'
    END AS 'Salary raise'
FROM
    dept_manager dm
        JOIN
    employees e ON e.emp_no = dm.emp_no
        JOIN
```

```
    salaries s ON s.emp_no = dm.emp_no
GROUP BY 1
ORDER BY 3;
```

Output



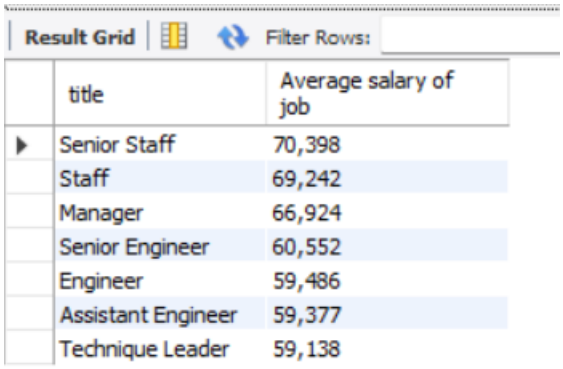
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

emp_no	Full Name	Salary Difference	Salary raise
110420	Oscar Ghazalie	16654	Minor salary increase
111877	Xiaobin Spinelli	17560	Minor salary increase
111784	Marjo Giarratana	18189	Minor salary increase
111939	Yuchang Weedman	18968	Minor salary increase
111035	Przemyslawa Kaelbling	24261	Significant salary increase
110344	Rosine Cools	24756	Significant salary increase
110228	Karsten Sigstam	25400	Significant salary increase
110511	DeForest Hagimont	25430	Significant salary increase
110386	Shem Kieras	26995	Significant salary increase
110800	Sanjoy Quadeer	27169	Significant salary increase
110854	Dung Pesch	27707	Significant salary increase
110085	Ebru Alpin	28417	Significant salary increase
110303	Krassimir Wegerle	30490	Significant salary increase
111400	Arie Staelin	30798	Significant salary increase
110183	Shirish Ossenbruggen	30938	Significant salary increase

Quite a number of department managers had significant salary increases, this could be dude to excellent performance in executing their duties.

```
-- Calculating average salary per job
SELECT
    t.title,
    FORMAT(AVG(s.salary), '###,###') AS 'Average salary of job'
FROM
    employees e
        JOIN
    titles t ON e.emp_no = t.emp_no
        JOIN
    salaries s ON e.emp_no = s.emp_no
GROUP BY 1
ORDER BY 2 DESC;
```

Output



Result Grid | Filter Rows:

	title	Average salary of job
▶	Senior Staff	70,398
	Staff	69,242
	Manager	66,924
	Senior Engineer	60,552
	Engineer	59,486
	Assistant Engineer	59,377
	Technique Leader	59,138

On average senior staff employees are the top earners in the company.