

Human Resources Analytics: Exploring Employee Data in SQL

Abstract

Businesses are increasingly counting on their human resources departments to answer employee engagement related questions. While, employee engagement has emerged as a critical driver of business success today. Engagement has significantly affect employee retention, productivity and loyalty. The research is based on the primary data developed by Patrick Crews and Giuseppe Maxia and provided a combination of a large base of data (approximately 160MB) spread over six separate tables and consisting of 4 million records in total. This study used SQL analysis and attempt to examine the level of employee engagement questions, such as the breakdown between male and female employees, managers, and the average salaries among different departments.

Tables

Table Names: Employees, Departments, Dept_emp, Dept_manager, Salaries

Employees table:

1	emp_no	birth_date	first_name	last_name	gender	hire_date
2	10008	1958/2/19	Saniya	Kalloufi	M	1994/9/15
3	10011	1953/11/7	Mary	Sluis	F	1990/1/22
4	10012	1960/10/4	Patricio	Bridgland	M	1992/12/18
5	10016	1961/5/2	Kazuhito	Cappelletti	M	1995/1/27

Departments table:

1	dept_no	dept_name
2	d009	Customer Service
3	d005	Development
4	d002	Finance

Dept_emp table:

1	emp_no	dept_no	from_date	to_date
2	10001	d005	1986/6/26	9999/1/1
3	10002	d007	1996/8/3	9999/1/1
4	10003	d004	1995/12/3	9999/1/1

Dept_manager table:

1	emp_no	dept_no	from_date	to_date
2	110022	d001	1995/12/30	1998/12/29
3	110039	d001	1997/4/9	9999/1/1
4	110040	d001	1991/12/31	1997/12/29
5	110041	d001	1993/8/2	1998/8/1

Salaries table:

1	emp_no	salary	from_date	to_date
2	10001	59461	1990/6/25	1991/6/25
3	10001	63546	1991/6/25	1992/6/24
4	10001	66833	1992/6/24	1993/6/24
5	10001	67786	1993/6/24	1994/6/24

Questions & Queries

1. What's employees' basic information, name, dept_name, from_date, to_date

```
1  # find out employees' name, dept_name, from_date, to_date
2  ●  SELECT
3      dm.emp_no,
4      e.first_name,
5      e.last_name,
6      d.dept_name,
7      dm.from_date,
8      dm.to_date
9  FROM
10     employees e
11  # inner join dept_manager and departments table
12  INNER JOIN
13     dept_manager dm
14  ON
15     dm.emp_no = e.emp_no
16  INNER JOIN
17     departments d
18  ON
19     dm.dept_no = d.dept_no
20  ORDER BY dm.emp_no;
```

Output:

	emp_no	first_name	last_name	dept_name	from_date	to_date
►	110022	Margareta	Markovitch	Marketing	1985-01-01	1991-10-01
	110039	Vishwani	Minakawa	Marketing	1991-10-01	9999-01-01
	110085	Ebru	Alpin	Finance	1985-01-01	1989-12-17
	110114	Isamu	Legleitner	Finance	1989-12-17	9999-01-01

This result of the query above gives us the general information about employees, employee number, first name, last name, department name, employment start date and employment end date.

2. What are the managers who hired between 1989 and 2002

```
1  # extract managers who were hired between 1989 and 2002
2  ● SELECT
3      *
4  FROM
5      dept_manager
6  WHERE
7      emp_no IN
8      (SELECT
9          emp_no
10         FROM
11             employees
12         WHERE hire_date BETWEEN '1989-01-01' AND '2002-01-01')
13 ORDER BY emp_no;
```

Output:

	emp_no	dept_no	from_date	to_date
►	110420	d004	1996-08-30	9999-01-01
	110765	d006	1989-05-06	1991-09-12
	110854	d006	1994-06-28	9999-01-01
	111877	d009	1992-09-08	1996-01-03

This result of the query above shows the managers who got hired from 1989 to 2002, represents their employment number, department number, employment start date and employment end date.

3. Who are the employees have salary higher than 145,000

```
1  # find out the employees who have salary higher than 145000
2  ● SELECT
3      e.emp_no,
4      e.first_name,
5      e.last_name,
6      s.salary
7  FROM
8      employees e
9  JOIN
10     salaries s
11  ON
12     e.emp_no = s.emp_no
13  WHERE s.salary > 145000
14  ORDER BY s.salary;
```

Output

	emp_no	first_name	last_name	salary
►	80823	Willard	Baca	145077
	44465	Brigham	Teitelbaum	145137
	18997	Basim	Tischendorf	145215
	37558	Juichirou	Thambidurai	145300

This result of the query above returns all the employees' employment number, first name, last name and salary whose salary is higher than 145,000.

4. What's the total number breakdown between male and female employees?

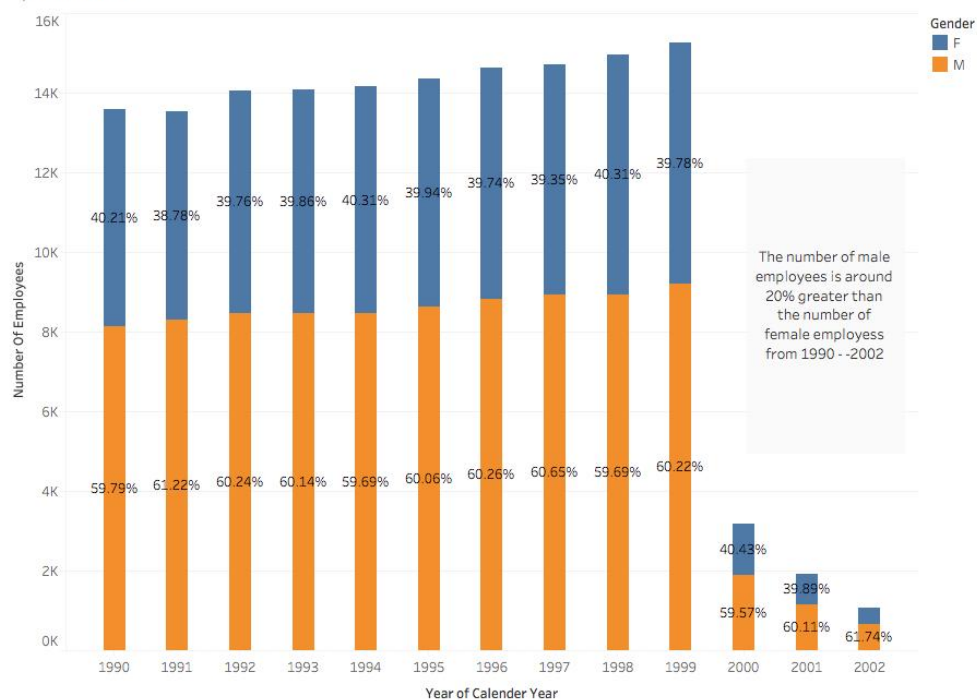
```
2      #Select year, gender, emp_no from t_employees and t_dept_emp tables
3      SELECT
4          YEAR(d.from_date) AS calender_year,
5          e.gender,
6          COUNT(e.emp_no) AS number_of_employees
7      FROM employees AS e
8      JOIN dept_emp AS d
9      ON e.emp_no = d.emp_no
10     GROUP BY calender_year, e.gender
11     HAVING calender_year >= 1990
12     ORDER BY calender_year;
```

Output:

▶ 1990	M	12568
1990	F	8471
1991	M	12727
1991	F	8257
1992	M	12955
1992	F	8593

Visualization:

m/f_employees



The plot of sum of Number Of Employees for Calendar Year Year. Color shows details about Gender. The marks are labeled by % of Total Number Of Employees.

The output and visualization show the breakdown between male and female employees for each year from year 1990 to year 2002. The result shows that the number of male employees is around 20% greater than the number of female employees for each year from 1990 – 2002. I created this visualization by Tableau with stacked bars.

5. How many managers were active from 1990 to 2000

```

1  #select detp_name, gender,emp_no,from_date,to_date, calendar_year
2  ●  SELECT
3      d.dept_name,
4      ee.gender,
5      ee.emp_no,
6      dt.from_date,
7      dt.to_date,
8      e.calendar_year ,
9      # active count 1 to calendar_year between from_date and to_date
10     CASE
11         WHEN e.calendar_year >= YEAR(dt.from_date) AND e.calendar_year <= YEAR (dt.to_date) THEN 1
12         ELSE 0
13     END AS active
14     #subquery e with calendar_year
15 FROM
16     (SELECT
17         YEAR (hire_date) AS calendar_year
18     FROM employees
19     GROUP BY calendar_year) AS e
20 #join t_departments, t_dept_manager, t_employees tables
21 CROSS JOIN
22     departments AS d
23 JOIN
24     dept_manager AS dt
25 ON
26     d.dept_no = dt.dept_no
27 JOIN
28     employees AS ee
29 ON
30     ee.emp_no = dt.emp_no
31 ORDER BY ee.emp_no, calendar_year;

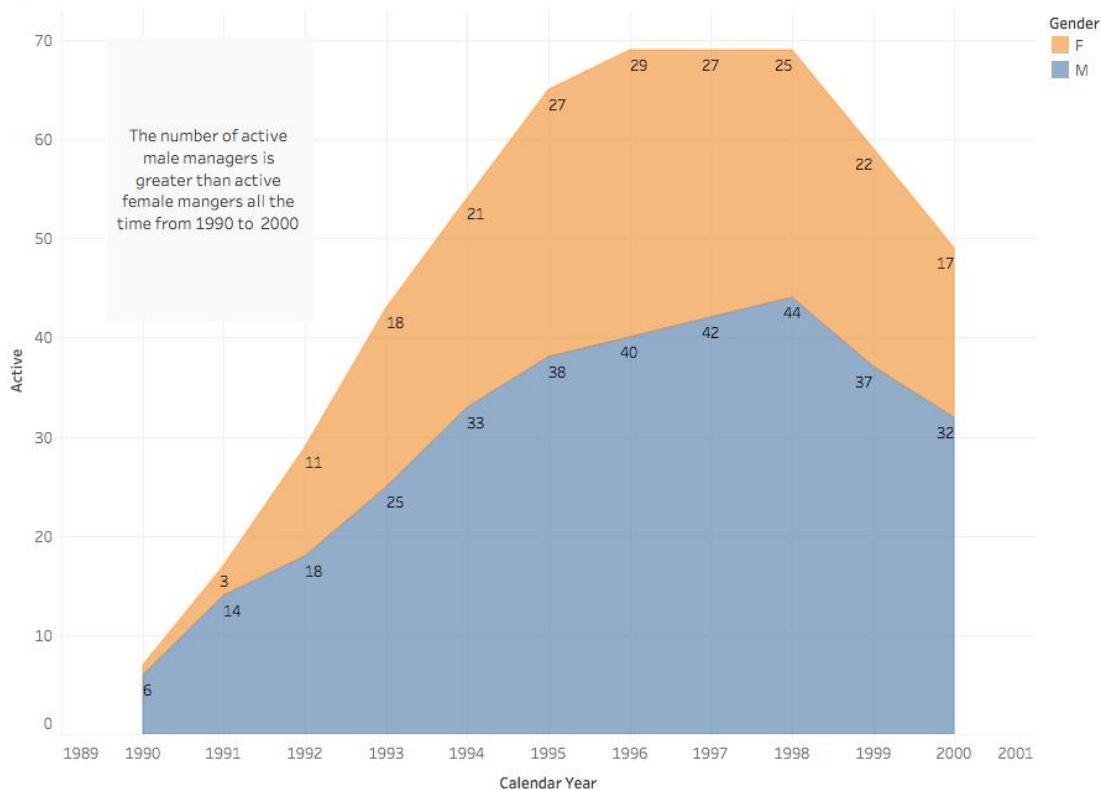
```

Output:

	dept_name	gend...	emp_no	from_date	to_date	calendar_y...	active
►	Marketing	M	110022	1985-01-01	1991-10-01	1985	1
	Marketing	M	110022	1985-01-01	1991-10-01	1986	1
	Marketing	M	110022	1985-01-01	1991-10-01	1987	1
	Marketing	M	110022	1985-01-01	1991-10-01	1988	1
	Marketing	M	110022	1985-01-01	1991-10-01	1989	1
	Marketing	M	110022	1985-01-01	1991-10-01	1990	1
	Marketing	M	110022	1985-01-01	1991-10-01	1991	1
	Marketing	M	110022	1985-01-01	1991-10-01	1992	0

Visualization:

m/f_managers



The plot of sum of Active for Calendar Year. Color shows details about Gender. The data is filtered on Dept Name, which keeps 7 of 7 members.

This output and visualization show the breakdown numbers of active managers between male and female from 1990 to 2000. The result shows the active male managers have greater amount of numbers than active female managers from all the time.

6. What's the average salary of male versus male employees in the entire company until year 2002?

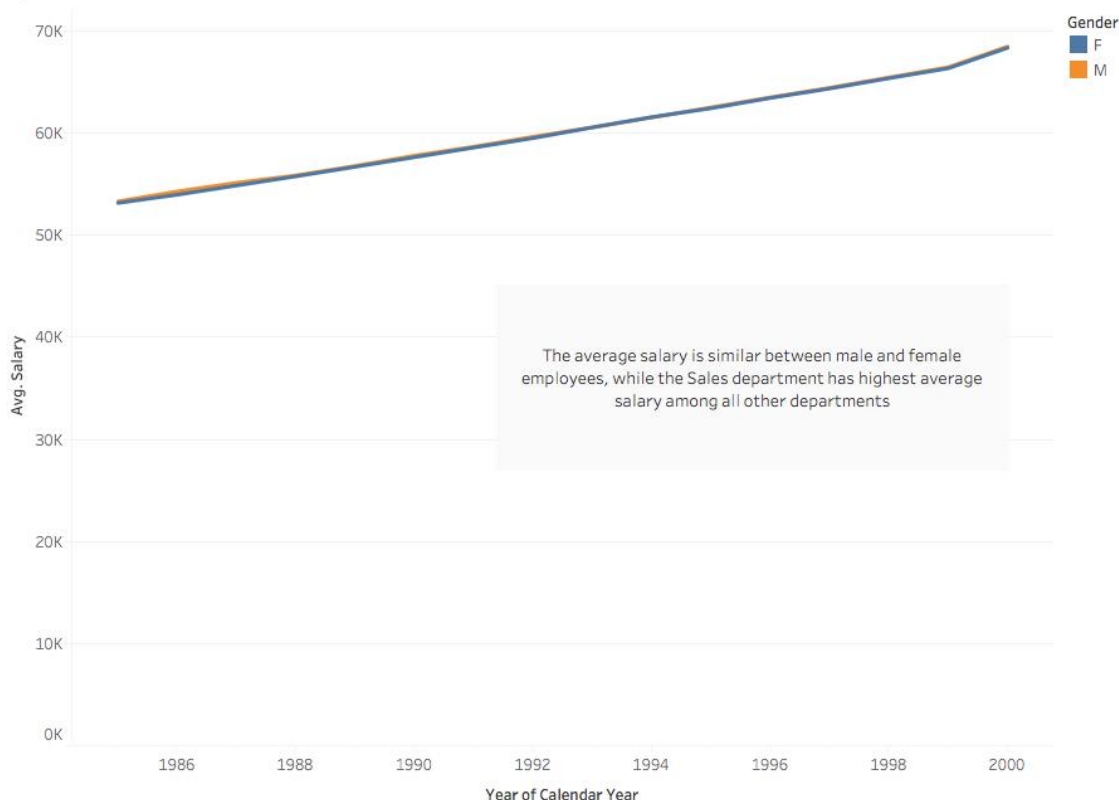
```
1 • USE employees;
2 # select column gender, dept_name, avg(salary), calendar_year
3 • SELECT
4     e.gender, d.dept_name, ROUND(AVG(s.salary),2) AS salary, YEAR(s.from_date) AS calendar_year
5 FROM
6     salaries s
7 # join salaries, employees, dept_emp, departments together
8 JOIN
9     employees e ON s.emp_no = e.emp_no
10 JOIN
11     dept_emp de ON de.emp_no = e.emp_no
12 JOIN
13     departments d ON d.dept_no = de.dept_no
14 GROUP BY d.dept_no, e.gender, calendar_year
15 # condition, calendar_year less than year 2000
16 HAVING calendar_year <= 2000
17 ORDER BY d.dept_no;
```

Output:

	gend...	dept_name	salary	calendar_y...
►	M	Marketing	62799.09	1985
	M	Marketing	63475.08	1986
	M	Marketing	64287.73	1987
	M	Marketing	64904.15	1988
	M	Marketing	65691.94	1989

Visualization:

m/f_avg_salary



The trend of average of Salary for Calendar Year Year. Color shows details about Gender. Details are shown for Gender. The data is filtered on Dept Name, which keeps 9 of 9 members.

The output and visualization show the comparison of average salary between male and female employees and the average salary among the different department. The result shows that the average salary is similar between male and female employees, while the Sales department has highest average salary among all other departments.

Conclusion:

The study revealed that the level of employee engagement used SQL analysis and Tableau visualization. This paper provided the insights that employers may hire, transfer, promote or terminate employees at any time for any cause, and the relationship between the employer and the employee. The employees how recognize this paper usefully could proactively use strategies to promote employee involvement.