



# EMPLOYEE DATABASE ANALYSIS

Presented by: Arjun Domle



# Problem Statement:

The organization is facing challenges in effectively managing its workforce, and there is a need for a comprehensive data analysis project to gain insights into various aspects of employee demographics, job dynamics, and tenure. The lack of detailed information hinders the ability to make informed decisions regarding talent acquisition, employee engagement, and overall workforce optimization.



# Business Context:

- This data analysis project on the employee database serves to provide valuable insights for human resources and management teams within the organization.
- Understanding employee demographics, job dynamics, and tenure is crucial for making informed decisions to enhance workforce management, optimize departmental efficiency, and improve overall employee satisfaction.

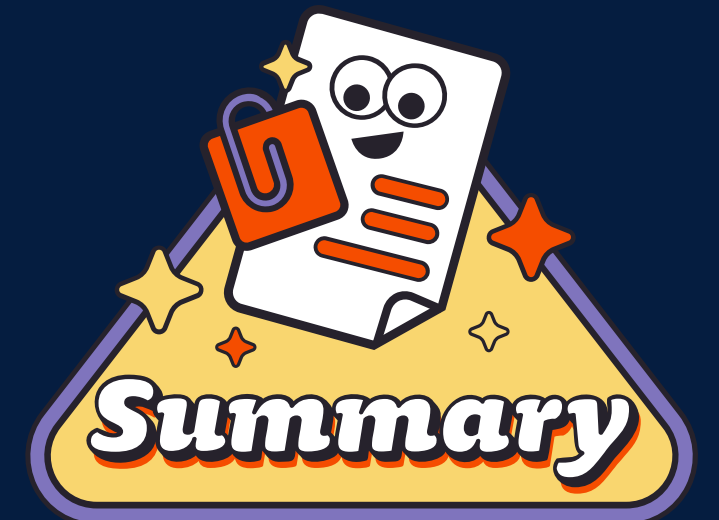


# Project Summary:

- The project involves conducting a comprehensive data analysis on an employee database.
- The primary objectives are to gain insights into : employee demographics, Salary Analysis, Employee Titles, Department Information, Managerial Analysis, Hiring Trends, Employee Longevity.

The database includes multiple tables such as :

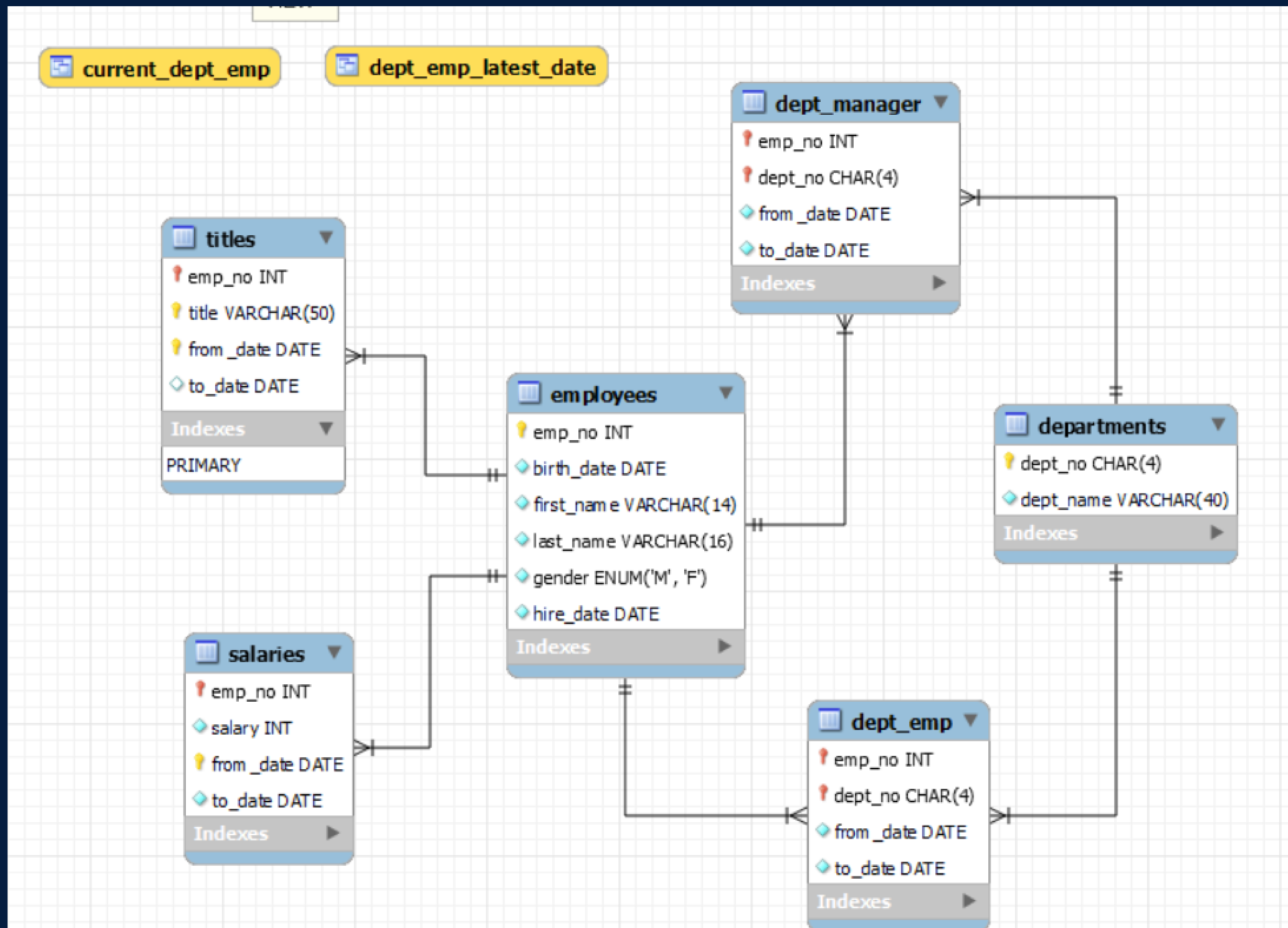
- employees
- departments
- dept\_emp
- dept\_manager
- salaries
- titles



# DATABASE SCHEMA

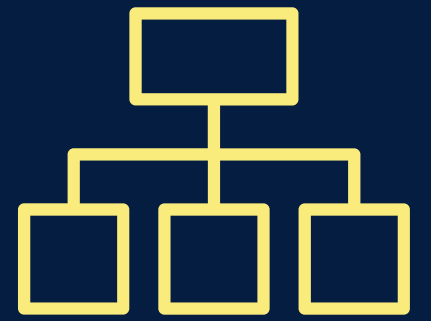
## TABLES

- employees
- departments
- dept\_emp
- dept\_manager
- salaries
- titles





# OVERVIEW OF ALL TABLES



SELECT \* FROM departments;

	dept_no	dept_name
▶	d009	Customer Service
	d005	Development
	d002	Finance
	d003	Human Resources
	d001	Marketing
	d004	Production
	d006	Quality Management
	d008	Research
	d007	Sales
⬤	NULL	NULL

SELECT \* FROM salaries;

	emp_no	salary	from_date	to_date
▶	10001	60117	1986-06-26	1987-06-26
	10001	62102	1987-06-26	1988-06-25
	10001	66074	1988-06-25	1989-06-25
	10001	66596	1989-06-25	1990-06-25
	10001	66961	1990-06-25	1991-06-25
	10001	71046	1991-06-25	1992-06-24
	10001	74333	1992-06-24	1993-06-24
	10001	75286	1993-06-24	1994-06-24
	10001	75994	1994-06-24	1995-06-24

SELECT \* FROM titles;

	emp_no	title	from_date	to_date
▶	10001	Senior Engineer	1986-06-26	9999-01-01
	10002	Staff	1996-08-03	9999-01-01
	10003	Senior Engineer	1995-12-03	9999-01-01
	10004	Engineer	1986-12-01	1995-12-01
	10004	Senior Engineer	1995-12-01	9999-01-01
	10005	Senior Staff	1996-09-12	9999-01-01
	10005	Staff	1989-09-12	1996-09-12
	10006	Senior Engineer	1990-08-05	9999-01-01
	10007	Senior Staff	1996-02-11	9999-01-01
	10007	Staff	1989-02-10	1996-02-11
	10008	Assistant Engi...	1998-03-11	2000-07-31
	10009	Assistant Engi...	1985-02-18	1990-02-18

SELECT \* FROM dept\_emp;

	emp_no	dept_no	from_date	to_date
▶	10001	d005	1986-06-26	9999-01-01
	10002	d007	1996-08-03	9999-01-01
	10003	d004	1995-12-03	9999-01-01
	10004	d004	1986-12-01	9999-01-01
	10005	d003	1989-09-12	9999-01-01
	10006	d005	1990-08-05	9999-01-01
	10007	d008	1989-02-10	9999-01-01
	10008	d005	1998-03-11	2000-07-31
	10009	d006	1985-02-18	9999-01-01

SELECT \* FROM dept\_manager;

	emp_no	dept_no	from_date	to_date
▶	110022	d001	1985-01-01	1991-10-01
	110039	d001	1991-10-01	9999-01-01
	110085	d002	1985-01-01	1989-12-17
	110114	d002	1989-12-17	9999-01-01
	110183	d003	1985-01-01	1992-03-21
	110228	d003	1992-03-21	9999-01-01
	110303	d004	1985-01-01	1988-09-09
	110344	d004	1988-09-09	1992-08-02

SELECT \* FROM employees;

emp_no	birth_date	first_name	last_name	gender	hire_date
10001	1953-09-02	Georgi	Facello	M	1986-06-26
10002	1964-06-02	Bezalel	Simmel	F	1985-11-21
10003	1959-12-03	Parto	Bamford	M	1986-08-28
10004	1954-05-01	Chirstian	Koblick	M	1986-12-01
10005	1955-01-21	Kyoichi	Maliniak	M	1989-09-12
10006	1953-04-20	Anneke	Preusig	F	1989-06-02
10007	1957-05-23	Tzvetan	Zielinski	F	1989-02-10
10008	1958-02-19	Saniya	Kalloufi	M	1994-09-15

# 1. Employee Demographics:



## A. What is the distribution of employees based on gender?

- Asking for an analysis of the number or percentage of employees in a given dataset categorized by their gender. The purpose is to understand how the workforce is distributed among different gender categories.

## B. What is the average age of employees in each department?

- Asking for an analysis of the average age of employees, categorized by their respective departments. The goal is to understand the distribution of the average age across different departments within the organization.

# A. What is the distribution of employees based on gender?

## QUERY

```
-- A. What is the distribution of employees based on gender
SELECT
    COUNT(emp_no) AS Total_Employees, gender
FROM
    employees
GROUP BY gender;
```

## OUTPUT

Total_Employees	gender
179973	M
120052	F



# B. What is the average age of employees in each department?

## QUERY

```
-- B. What is the average age of employees in each department?
SELECT
  d.dept_name,
  ROUND(AVG(YEAR(CURDATE()) - YEAR(e.birth_date)), 2) AS Average_age_of_employees
FROM
  departments d
  JOIN
  dept_emp de ON d.dept_no = de.dept_no
  JOIN
  employees e ON e.emp_no = de.emp_no
GROUP BY d.dept_name;
```

## OUTPUT

dept_name	Average_age_of_employees
Customer Service	65.92
Development	65.90
Finance	65.93
Human Resources	65.95
Marketing	65.92
Production	65.93
Quality Management	65.94
Research	65.90
Sales	65.92

## 2. Salary Analysis:



### **A. How has the average salary changed over the years?**

- Asking for an analysis of the trend or pattern in the average salary of employees over different time periods.

### **B. Which department has the highest average salary?**

- Asking for an analysis of the average salary across different departments to identify the department with the highest average salary.

### **C. Identify employees with the highest and lowest salaries.**

- Asking for a list of employees who have either the highest or the lowest salaries within the organization.

# A. How has the average salary changed over the years?

## QUERY

```
-- A. How has the average salary changed over the years?  
SELECT  
    YEAR(from_date) AS years,  
    ROUND(AVG(salary), 2) AS Average_salary  
FROM  
    salaries  
GROUP BY years  
ORDER BY years ASC;
```

## OUTPUT

years	Average_salary
1985	53419.78
1986	54303.32
1987	55036.68
1988	55865.36
1989	56781.01
1990	57783.80
1991	58742.61
1992	59679.23
1993	60656.13
1994	61643.29
1995	62584.02
1996	63547.50

## B. Which department has the highest average salary?

### QUERY

```
-- B. Which department has the highest average salary?
SELECT
    d.dept_name, ROUND(AVG(s.salary), 2) AS Avg_salary
FROM
    departments d
    JOIN
    dept_emp de ON d.dept_no = de.dept_no
    JOIN
    salaries s ON s.emp_no = de.emp_no
GROUP BY d.dept_no
ORDER BY Avg_salary DESC
LIMIT 1;
```

### OUTPUT

	dept_name	Avg_salary
▶	Sales	80776.62

# C. Identify employees with the highest and lowest salaries.

## QUERY

```
-- C. Identify employees with the highest and lowest salaries.
WITH RankedSalaries AS (
  SELECT
    e.first_name,e.last_name, e.emp_no,
    s.salary,
    RANK() OVER (ORDER BY s.salary DESC) AS max_salary_rank,
    RANK() OVER (ORDER BY s.salary) AS min_salary_rank
  FROM
    employees e
    JOIN salaries s ON e.emp_no = s.emp_no
)
SELECT *
FROM RankedSalaries
WHERE max_salary_rank = 1 OR min_salary_rank = 1;
```

## OUTPUT

first_name	last_name	emp_no	salary	max_salary_rank	min_salary_rank
Fumiya	Unno	49239	38735	967330	1
Tokuyasu	Pesch	43624	158220	1	967330



# 3. Employee Titles:



## A. What are the most common job titles in the company?

- Asking for an analysis of the distribution of job titles among employees to identify which job titles are most prevalent or common within the organization.

## B. How often do employees change their job titles?

- Asking for an analysis of the frequency or rate at which employees undergo changes in their job titles within the company.

## C. Identify employees who have held the most number of titles.

- Asking for an analysis of employees who have experienced frequent changes in their job titles. The goal is to identify individuals who have had a significant number of different titles throughout their employment history within the company.

# A. What are the most common job titles in the company?

## QUERY

```
-- A. What are the most common job titles in the company?  
SELECT  
    title AS Most_common_job_title, COUNT(title) AS Total_jobs  
FROM  
    titles  
GROUP BY title  
ORDER BY Total_jobs DESC;
```

## OUTPUT

Most_common_job_title	Total_jobs
Engineer	115003
Staff	107391
Senior Engineer	97750
Senior Staff	92853
Technique Leader	15159
Assistant Engineer	15128
Manager	24

## B. How often do employees change their job titles?

### QUERY

```
-- B. How often do employees change their job titles?  
SELECT  
    emp_no,  
    COUNT(*) AS title_changes  
FROM  
    titles  
GROUP BY  
    emp_no  
HAVING  
    COUNT(*) > 1;
```

### OUTPUT

emp_no	title_changes
10004	2
10005	2
10007	2
10009	3
10012	2
10017	2
10018	2
10026	2

## C. Identify employees who have held the most number of titles.

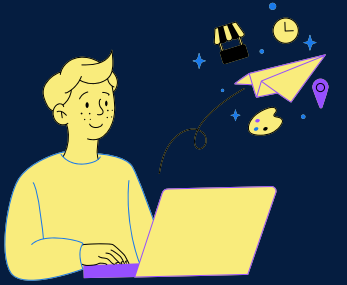
### QUERY

```
-- C. Identify employees who have held the most number of titles.  
SELECT  
    emp_no,  
    COUNT(DISTINCT title) AS most_titles  
FROM  
    titles  
GROUP BY  
    emp_no  
HAVING  
    COUNT(*) > 1  
ORDER BY most_titles DESC  
LIMIT 1;
```

### OUTPUT

emp_no	most_titles
486671	3

# 4. Department Information:



## **A. How many employees are currently assigned to each department?**

- Asking for a count of the number of employees currently associated with each department within the organization.

## **B. Which department has had the most managers over time?**

- Asking for an analysis of the historical managerial activity within each department.



# A. How many employees are currently assigned to each department?

## QUERY

```
-- A. How many employees are currently assigned to each department?
SELECT
    d.dept_name, COUNT(e.emp_no) AS Total_employees
FROM
    departments d
    JOIN
    dept_emp de ON d.dept_no = de.dept_no
    JOIN
    employees e ON e.emp_no = de.emp_no
GROUP BY d.dept_name
ORDER BY Total_employees;
```

## OUTPUT

dept_name	Total_employees
Finance	17346
Human Resources	17786
Quality Management	20117
Marketing	20211
Research	21126
Customer Service	23580
Sales	52245
Production	73485
Development	85707

## B. Which department has had the most managers over time?

### QUERY

```
-- B. Which department has had the most managers over time?  
SELECT  
    d.dept_name, COUNT(dm.emp_no) MOST_MANAGER  
FROM  
    departments d  
    INNER JOIN  
    dept_manager dm ON d.dept_no = dm.dept_no  
GROUP BY d.dept_name  
ORDER BY MOST_MANAGER DESC  
LIMIT 1;
```

### OUTPUT

dept_name	MOST_MANAGER
Customer Service	4

# 5. Managerial Analysis:



## A. Who are the current managers in each department?

- Asking for information about employees who are currently serving as managers in each department.

## B. How many times has each employee served as a manager?

- Asking for an analysis of the managerial roles held by each employee over time.

# A. Who are the current managers in each department?

## QUERY

```
-- A. Who are the current managers in each department?
SELECT
    dm.dept_no, d.dept_name, e.emp_no, e.first_name, e.last_name
FROM
    dept_manager dm
    JOIN
    employees e ON dm.emp_no = e.emp_no
    JOIN
    departments d ON dm.dept_no = d.dept_no
WHERE
    dm.to_date = '9999-01-01';
```

## OUTPUT

dept_no	dept_name	emp_no	first_name	last_name
d001	Marketing	110039	Vishwani	Minakawa
d002	Finance	110114	Isamu	Legleitner
d003	Human Resources	110228	Karsten	Sigstam
d004	Production	110420	Oscar	Ghazalie
d005	Development	110567	Leon	DasSarma
d006	Quality Management	110854	Dung	Pesch
d007	Sales	111133	Hauke	Zhang
d008	Research	111534	Hilary	Kambil

## B. How many times has each employee served as a manager?

### QUERY

```
-- B. How many times has each employee served as a manager?  
SELECT  
    emp_no, COUNT(*) AS manager_count  
FROM  
    dept_manager  
GROUP BY emp_no  
ORDER BY manager_count DESC;
```

### OUTPUT

emp_no	manager_count
110022	1
110039	1
110085	1
110114	1
110183	1
110228	1
110303	1
110344	1
110386	1
110420	1
110511	1



# 6. Hiring Trends:



## **A. How has the number of new hires changed over the years?**

- Asking for an analysis of the trends or patterns in the hiring activity of the organization over different years.

## **B. Identify the departments with the highest and lowest hiring rates.**

- Asking for an analysis of the relative hiring rates across different departments within the organization. The goal is to determine which departments have experienced higher or lower rates of new hires compared to their total workforce.

# A. How has the number of new hires changed over the years?

## QUERY

```
-- A. How has the number of new hires changed over the years?  
SELECT  
    YEAR(hire_date) AS hire_year, COUNT(*) AS new_hires_count  
FROM  
    employees  
GROUP BY hire_year  
ORDER BY hire_year;
```

## OUTPUT

hire_year	new_hires_count
1985	35316
1986	36150
1987	33501
1988	31436
1989	28394
1990	25610
1991	22568
1992	20402
1993	17772

## B. Identify the departments with the highest and lowest hiring rates.

### QUERY

```
-- B. Identify the departments with the highest and lowest hiring rates.
SELECT
  d.dept_no,
  d.dept_name,
  COUNT(DISTINCT e.emp_no) AS total_employees,
  COUNT(DISTINCT CASE
    WHEN YEAR(e.hire_date) = YEAR(CURDATE()) THEN e.emp_no
  END) AS new_hires,
  COUNT(DISTINCT CASE
    WHEN YEAR(e.hire_date) = YEAR(CURDATE()) THEN e.emp_no
  END) / COUNT(DISTINCT e.emp_no) AS hiring_rate
FROM
  departments d
  JOIN
  dept_emp de ON d.dept_no = de.dept_no
  JOIN
  employees e ON de.emp_no = e.emp_no
GROUP BY d.dept_no , d.dept_name
ORDER BY hiring_rate DESC;
```

### OUTPUT

dept_no	dept_name	total_employees	new_hires	hiring_rate
d001	Marketing	20211	0	0.0000
d002	Finance	17346	0	0.0000
d003	Human Resources	17786	0	0.0000
d004	Production	73485	0	0.0000
d005	Development	85707	0	0.0000
d006	Quality Management	20117	0	0.0000
d007	Sales	52245	0	0.0000

# 7. Employee Longevity:



## **A. What is the average tenure of employees in the company?**

- Asking for an analysis of the average length of time that employees have been working for the organization.

## **B. Identify employees who have been with the company the longest.**

- Asking for a list of employees who have the longest tenure or have been with the company for the greatest amount of time.

## **C. How many employees have retired or left the company in recent years?**

- Asking for an analysis of the number of employees who have either retired or voluntarily left the company within a specific recent time frame.

# A. What is the average tenure of employees in the company?

## QUERY

```
-- A. What is the average tenure of employees in the company?  
SELECT  
    ROUND(AVG(DATEDIFF(CURDATE(), hire_date)), 2) AS average_tenure  
FROM  
    employees;
```

## OUTPUT

average_tenure
12396.24



## B. Identify employees who have been with the company the longest.

### QUERY

```
-- B. Identify employees who have been with the company the longest.  
SELECT  
    emp_no, first_name, last_name, hire_date, DATEDIFF(CURDATE(), hire_date) AS tenure_days  
FROM  
    employees  
ORDER BY tenure_days DESC;
```

### OUTPUT

emp_no	first_name	last_name	hire_date	tenure_days
111035	Przemyslaw	Kaelbling	1985-01-01	14250
111400	Arie	Staelin	1985-01-01	14250
110303	Krassimir	Wegerle	1985-01-01	14250
110183	Shirish	Ossenbruggen	1985-01-01	14250
110725	Peternela	Onuegbue	1985-01-01	14250
111692	Tonny	Butterworth	1985-01-01	14250
110511	DeForest	Hagimont	1985-01-01	14250

# C. How many employees have retired or left the company in recent years?

## QUERY

```
-- C. How many employees have retired or left the company in recent years?  
SELECT  
    COUNT(*) AS employees_left  
FROM  
    dept_emp  
WHERE  
    to_date < CURDATE();
```

## OUTPUT

employees_left
91479



*Thank You*