

# Data Analyst Project: SQL Analysis Using GitHub's [DataCharmer](#) Dataset

## Project Overview:

In this project, I performed data analysis on a variety of business-related questions using SQL. The dataset used for this analysis was sourced from the [DataCharmer GitHub repository](#). I used SQL Workbench to execute the queries and extract valuable insights from the database.

## Tools and Technologies:

- **SQL Workbench:** To write and execute SQL queries for data analysis.
- **GitHub Data (DataCharmer Repository):** The dataset contains information on employees, salaries, departments, and more.

## Questions and SQL Queries:

### 1. List of employees by department

*Include employee number, first name, last name, department number, and department name.*

```
SELECT E.emp_no, E.first_name, E.last_name, D.dept_no, D.dept_name
FROM employees AS E
JOIN dept_emp AS DE ON E.emp_no = DE.emp_no
JOIN departments AS D ON DE.dept_no = D.dept_no;
```

	emp_no	first_name	last_name	dept_no	dept_name
▶	10011	Mary	Sluis	d009	Customer Service
	10038	Huan	Lortz	d009	Customer Service
	10049	Basil	Tramer	d009	Customer Service
	10060	Breannda	Billingsley	d009	Customer Service
	10088	Jungsoon	Syrzycki	d009	Customer Service
	10098	Sreekrishna	Servieres	d009	Customer Service
	10112	Yuichiro	Swick	d009	Customer Service
	10115	Chikara	Rissland	d009	Customer Service

## 2. Retrieve all the salary records of a given employee (by employee number)

*Include employee number, salary, from date, and to date*

```
SELECT *  
FROM salaries  
WHERE emp_no = '10001'  
ORDER BY salary DESC  
LIMIT 1; -- Reflects the highest salary
```

	emp_no	salary	from_date	to_date
►	10001	88958	2002-06-22	9999-01-01
*	NULL	NULL	NULL	NULL

## 3. Find all employees who have held a specific title (e.g., 'Engineer')

*Include employee number, first name, last name, and title.*

```
SELECT E.emp_no, E.first_name, E.last_name, T.title  
FROM employees AS E  
JOIN titles AS T ON E.emp_no = T.emp_no  
WHERE T.title = 'Engineer'  
ORDER BY 1;
```

	emp_no	first_name	last_name	title
►	10004	Chirstian	Koblick	Engineer
	10009	Sumant	Peac	Engineer
	10010	Duangkaew	Piveteau	Engineer
	10012	Patricio	Bridgland	Engineer
	10014	Berni	Genin	Engineer
	10018	Kazuhide	Peha	Engineer
	10020	Mayuko	Warwick	Engineer
	10022	Shahaf	Famili	Engineer

#### 4. List all departments along with their current managers

*Include department number, department name, manager's employee number, first name, and last name.*

```
SELECT DM.dept_no AS department_number, D.dept_name AS department_name,  
       E.emp_no AS managers_employee_number, E.first_name, E.last_name  
FROM dept_manager AS DM  
JOIN departments AS D ON DM.dept_no = D.dept_no  
JOIN employees AS E ON DM.emp_no = E.emp_no;
```

	department_number	department_name	managers_employee_number	first_name	last_name
►	d009	Customer Service	111692	Tonny	Butterworth
	d009	Customer Service	111784	Marjo	Giarratana
	d009	Customer Service	111877	Xiaobin	Spinelli
	d009	Customer Service	111939	Yuchang	Weedman
	d005	Development	110511	DeForest	Hagimont
	d005	Development	110567	Leon	DasSarma
	d002	Finance	110085	Ebru	Alpin
	d002	Finance	110114	Isamu	Legleitner

#### 5. Count the number of employees in each department

*Include department number, department name, and employee count.*

```
SELECT D.dept_no AS department_number, D.dept_name AS department_name,  
       COUNT(DE.emp_no) AS employee_count  
FROM dept_emp AS DE  
JOIN departments AS D ON DE.dept_no = D.dept_no  
GROUP BY department_number;
```

	department_number	department_name	employee_count
►	d001	Marketing	20211
	d002	Finance	17346
	d003	Human Resources	17786
	d004	Production	73485
	d005	Development	85707
	d006	Quality Management	20117
	d007	Sales	52245
	d008	Research	21126
	d009	Customer Service	23580

## 6. Find all employees born in a specific year (e.g., 1954)

*Include employee number, first name, last name, and birth date.*

```
SELECT emp_no AS employee_number, first_name, last_name, birth_date
FROM employees
WHERE YEAR(birth_date) = 1954;
```

	employee_number	first_name	last_name	birth_date
►	10004	Chirstian	Koblick	1954-05-01
	10018	Kazuhide	Peha	1954-06-19
	10053	Sanjiv	Zschoche	1954-09-13
	10057	Ebbe	Callaway	1954-05-30
	10058	Berhard	McFarlin	1954-10-01
	10073	Shir	McClurg	1954-02-23
	10088	Jungsoon	Syrzycki	1954-02-25
	10096	Jayson	Mandell	1954-09-16

## 7. Find all employees hired in the last 50 years

*Include employee number, first name, last name, and hire date.*

```
SELECT emp_no AS employee_number, first_name, last_name, hire_date
FROM employees
WHERE hire_date >= DATE_SUB(CURDATE(), INTERVAL 50 YEAR);
```

	employee_number	first_name	last_name	hire_date
►	10001	Georgi	Facello	1986-06-26
	10002	Bezael	Simmel	1985-11-21
	10003	Parto	Bamford	1986-08-28
	10004	Chirstian	Koblick	1986-12-01
	10005	Kyoichi	Maliniak	1989-09-12
	10006	Anneke	Preusig	1989-06-02
	10007	Tzvetan	Zielinski	1989-02-10
	10008	Saniya	Kalloufi	1994-09-15

## 8. Calculate the average salary for each department

*Include department number, department name, and average salary.*

```
SELECT D.dept_no AS department_number, D.dept_name AS department_name,  
       AVG(S.salary) AS average_salary  
FROM dept_emp AS DE  
JOIN departments AS D ON DE.dept_no = D.dept_no  
JOIN salaries AS S ON DE.emp_no = S.emp_no  
GROUP BY department_name  
ORDER BY department_number;
```

	department_number	department_name	average_salary
▶	d001	Marketing	71913.2000
	d002	Finance	70489.3649
	d003	Human Resources	55574.8794
	d004	Production	59605.4825
	d005	Development	59478.9012
	d006	Quality Management	57251.2719
	d007	Sales	80667.6058
	d008	Research	59665.1817
	d009	Customer Service	58770.3665

## 9. Find the gender distribution (number of males and females) in each department

*Include department number, department name, count of males, and count of females.*

```
SELECT D.dept_no AS department_number, D.dept_name AS department_name,  
       SUM(E.gender = 'M') AS count_of_males, SUM(E.gender = 'F') AS count_of_females  
FROM departments AS D  
JOIN dept_emp AS DE ON D.dept_no = DE.dept_no  
JOIN employees AS E ON DE.emp_no = E.emp_no  
GROUP BY department_name  
ORDER BY department_number;
```

	department_number	department_name	count_of_males	count_of_females
▶	d001	Marketing	12174	8037
	d002	Finance	10331	7015
	d003	Human Resources	10711	7075
	d004	Production	43936	29549
	d005	Development	51449	34258
	d006	Quality Management	12039	8078
	d007	Sales	31391	20854
	d008	Research	12687	8439
	d009	Customer Service	14132	9448

# 10. Find the employees who have served the longest in the company

*Include employee number, first name, last name, and number of years served.*

```
SELECT emp_no AS employee_number, first_name, last_name,  
       TIMESTAMPDIFF(YEAR, hire_date, CURDATE()) AS number_of_years_served  
FROM employees  
ORDER BY number_of_years_served DESC  
LIMIT 5;
```

	employee_number	first_name	last_name	number_of_years_served
►	110022	Margareta	Markovitch	40
	110114	Isamu	Legleitner	40
	110303	Krassimir	Wegerle	40
	110085	Ebru	Alpin	40
	110183	Shirish	Ossenbruggen	40

## Analysis and Insights:

1. **Employee Distribution:** The SQL queries provided insights into employee distribution across departments, helping identify departments with more or fewer employees.
2. **Salary Analysis:** By retrieving salary records and calculating averages for departments, this project highlighted salary disparities and helped analyze trends in compensation across different departments.
3. **Gender Diversity:** The gender distribution query offered insights into the gender composition across departments, which could inform diversity initiatives.
4. **Employee Tenure:** The query that identifies the longest-serving employees provides key insights for workforce planning and recognizing employee loyalty.

## Conclusion:

This project provided valuable insights into employee demographics, department structures, salary trends, and tenure. The findings can guide HR, management, and business leaders in making data-driven decisions regarding employee retention, department growth, and salary planning.