# Data Science Jobs
# SQL Case Study

Presented by: DISHA VARSHNEY.

# Info

- Table Name - salaries
- Total records - 13972
- Total columns - 11

| Column Name | Datatype |
|---|---|
| work_year | INT |
| experience_level | TEXT |
| employment_type | TEXT |
| job_title | TEXT |
| salary | INT |
| salary_currency | TEXT |
| salary_in_usd | INT |
| emplyoee_residence | TEXT |
| remote_ratio | INT |
| company_location | TEXT |
| company_size | TEXT |

# Data View

```
select * from salaries;
```

| work_year | experience_level | employment_type | job_title | salary | salary_currency | salary_in_usd | employee_residence | remote_ratio | company_location | company_size |
|---|---|---|---|---|---|---|---|---|---|---|
| 2024 | SE | FT | AI Engineer | 90000 | USD | 90000 | AE | 0 | AE | L |
| 2024 | SE | FT | Machine Learning Engineer | 180500 | USD | 180500 | US | 0 | US | M |
| 2024 | SE | FT | Machine Learning Engineer | 96200 | USD | 96200 | US | 0 | US | M |
| 2024 | SE | FT | Machine Learning Engineer | 235000 | USD | 235000 | AU | 0 | AU | M |
| 2024 | SE | FT | Machine Learning Engineer | 175000 | USD | 175000 | AU | 0 | AU | M |
| 2024 | MI | FT | Business Intelligence Developer | 95413 | USD | 95413 | US | 100 | US | M |
| 2024 | MI | FT | Business Intelligence Developer | 70692 | USD | 70692 | US | 100 | US | M |
| 2024 | SE | FT | Data Engineer | 110000 | USD | 110000 | UA | 0 | UA | M |
| 2024 | SE | FT | Data Engineer | 48000 | USD | 48000 | UA | 0 | UA | M |
| 2024 | SE | FT | Data Scientist | 222300 | USD | 222300 | US | 0 | US | M |
| 2024 | SE | FT | Data Scientist | 119700 | USD | 119700 | US | 0 | US | M |
| 2024 | SE | FT | Cloud Database Engineer | 170375 | USD | 170375 | US | 100 | US | M |
| 2024 | SE | FT | Cloud Database Engineer | 102500 | USD | 102500 | US | 100 | US | M |
| 2024 | SE | FT | Data Engineer | 270000 | USD | 270000 | AU | 0 | AU | M |
| 2024 | SE | FT | Data Engineer | 190000 | USD | 190000 | AU | 0 | AU | M |

# Insights

1 - As a market researcher, your job is to investigate the job market for a company that analyzes workforce data. Your task is to know how many people were employed in different types of companies as per their size in 2021.

```sql
select
    company_size,
    COUNT(company_size) as "people_employed"
from salaries
where work_year=2021
group by 1
order by 2 desc;
```

| company_size | people_employed |
|---|---|
| L | 124 |
| M | 52 |
| S | 42 |

# Insights

2- Imagine you are a Talent Acquisition specialist working for an International recruitment agency. Your task is to identify the top 3 job titles that command the highest average salary among part-time positions in the year 2023.

```sql
select
    job_title, ROUND(AVG(salary_in_usd),2) as "avg_salary"
from salaries
where work_year=2023 AND employment_type="PT"
group by 1
order by 2 desc
limit 3;
```

| job_title | avg_salary |
|---|---|
| Data Scientist | 95650.00 |
| Data Analyst | 18160.00 |

# Insights

3 - As a database analyst you have been assigned the task to select countries where average mid-level salary is higher than overall mid-level salary for the year 2023

```
select
    company_location,ROUND(AVG(salary_in_usd),2) as "avg_salary"
from salaries
where work_year=2023 and experience_level="MI"
group by 1
having AVG(salary_in_usd)> (
select
    avg(salary_in_usd) as "overall_salary"
from salaries
where work_year=2023 and experience_level="MI");
```

| company_location | avg_salary |
|---|---|
| US | 132783.25 |
| CA | 129789.97 |
| AU | 195984.36 |
| QA | 300000.00 |
| SA | 134999.00 |

# Insights

4-As a database analyst you have been assigned the task to Identify the company locations with the highest and lowest average salary for senior-level (SE) employees in 2023

```
CREATE DEFINER=`root`@`localhost` PROCEDURE `GetSeniorSalaryStats`()
BEGIN
    SELECT highest_location, highest_avg_salary, lowest_location, lowest_avg_salary FROM
    (SELECT company_location AS highest_location, AVG(salary_in_usd) AS highest_avg_salary
    FROM  salaries
    WHERE work_year = 2023 AND experience_level = 'SE'
    GROUP BY company_location
    ORDER BY highest_avg_salary DESC
    LIMIT 1) AS t1
    CROSS JOIN
    (SELECT company_location AS lowest_location, AVG(salary_in_usd) AS lowest_avg_salary
    FROM  salaries
    WHERE work_year = 2023 AND experience_level = 'SE'
    GROUP BY company_location
    ORDER BY lowest_avg_salary ASC
    LIMIT 1) AS t2;
END
```

```
call GetSeniorSalaryStats
```

| highest_location | highest_avg_salary | lowest_location | lowest_avg_salary |
|---|---|---|---|
| IL | 266468.5000 | TR | 18381.0000 |

# Insights

5 - You're a Financial analyst working for a leading HR Consultancy, and your task is to assess the annual salary growth rate for various job titles. By calculating the percentage increase in salary from previous year to this year you aim to provide valuable insights into salary trends within different job roles.

```sql
with cte as (
select job_title,
    avg(case when work_year=2020 then salary_in_usd end) as "2020_y",
    avg(case when work_year=2021 then salary_in_usd end) as "2021_y",
    avg(case when work_year=2022 then salary_in_usd end) as "2022_y",
    avg(case when work_year=2023 then salary_in_usd end) as "2023_y",
    avg(case when work_year=2024 then salary_in_usd end) as "2024_y"
    from salaries
group by 1  )

select job_title,
    round((2021_y-2020_y)/2020_y,2)*100 as "%_rate_2021",
    round((2022_y-2021_y)/2021_y,2)*100 as "%_rate_2022",
    round((2023_y-2022_y)/2022_y,2)*100 as "%_rate_2023",
    round((2024_y-2023_y)/2023_y,2)*100 as "%_rate_2024"
from cte;
```

| job_title | %_rate_2021 | %_rate_2022 | %_rate_2023 | %_rate_2024 |
|---|---|---|---|---|
| AI Engineer | NULL | NULL | NULL | 2.00 |
| Machine Learning Engineer | -49.00 | 102.00 | 27.00 | 2.00 |
| Business Intelligence Developer | NULL | NULL | 631.00 | -12.00 |
| Data Engineer | 13.00 | 45.00 | 8.00 | -1.00 |
| Data Scientist | -7.00 | 72.00 | 19.00 | -9.00 |
| Cloud Database Engineer | NULL | NULL | -19.00 | -4.00 |
| Research Engineer | NULL | NULL | 19.00 | 21.00 |
| Data Analyst | 28.00 | 39.00 | 0.00 | 1.00 |
| Machine Learning Scientist | -27.00 | -28.00 | 39.00 | 8.00 |
| Applied Scientist | NULL | NULL | 1.00 | 2.00 |
| Data Science Manager | -25.00 | 34.00 | 4.00 | -6.00 |
| Research Scientist | -66.00 | 71.00 | 33.00 | 17.00 |
| Prompt Engineer | NULL | NULL | NULL | 77.00 |
| Data Science | NULL | NULL | NULL | 9.00 |
| Data Science Consultant | -27.00 | 22.00 | 21.00 | -2.00 |
| Data Management Analyst | NULL | NULL | NULL | 45.00 |
| Research Analyst | NULL | NULL | NULL | 41.00 |
| Data Operations Analyst | NULL | NULL | 4.00 | -7.00 |
| Data Management Consultant | NULL | NULL | NULL | NULL |
| Business Intelligence Analyst | NULL | NULL | NULL | -20.00 |
| Analytics Engineer | NULL | NULL | 16.00 | 2.00 |
| Data Quality Analyst | NULL | NULL | NULL | -12.00 |
| Data Architect | NULL | 0.00 | -3.00 | 2.00 |

# Insights

6 -You've been hired by a global HR Consultancy to identify countries experiencing significant salary growth for entry-level roles. Your task is to list the top three countries with the highest salary growth rate from 2020 to 2023, helping multinational corporations identify  emerging talent markets

```
 with cte as (
select company_location,
    avg(case when work_year=2020 then salary_in_usd end) as "2020_y",
    avg(case when work_year=2021 then salary_in_usd end) as "2021_y",
    avg(case when work_year=2022 then salary_in_usd end) as "2022_y",
    avg(case when work_year=2023 then salary_in_usd end) as "2023_y"
from salaries
where experience_level="EN"
group by 1  )

select company_location,
    round((2021_y-2020_y)/2020_y,2)*100 as "%_rate_2021",
    round((2022_y-2021_y)/2021_y,2)*100 as "%_rate_2022",
    round((2023_y-2022_y)/2022_y,2)*100 as "%_rate_2023"
from cte;
```

| company_location | %_rate_2021 | %_rate_2022 | %_rate_2023 |
|---|---|---|---|
| US | -25.00 | 15.00 | -1.00 |
| LT | NULL | NULL | NULL |
| NL | NULL | NULL | -33.00 |
| AU | NULL | 86.00 | -32.00 |
| DE | 42.00 | -23.00 | 38.00 |
| GB | NULL | -45.00 | 42.00 |
| TR | NULL | NULL | NULL |
| CA | NULL | NULL | -13.00 |
| MX | NULL | NULL | NULL |
| FR | -9.00 | 44.00 | -45.00 |
| IT | NULL | NULL | 23.00 |
| AS | NULL | 177.00 | NULL |
| EE | NULL | NULL | NULL |
| MT | NULL | NULL | NULL |
| ES | NULL | NULL | 14.00 |
| LB | NULL | NULL | NULL |
| RO | NULL | NULL | NULL |
| ZA | NULL | NULL | NULL |
| HU | NULL | NULL | 143.00 |
| NG | NULL | NULL | NULL |
| PT | NULL | NULL | 11.00 |
| UA | NULL | NULL | NULL |
| IN | NULL | -17.00 | 34.00 |

# Insights

7- Picture yourself as a data architect responsible for database management. companies in US and AU(Australia) decided to create a hybrid model for employees they decided that employees earning salaries exceeding $90000 USD, will be given work from home. You now need to update the remote work ratio for eligible employees, ensuring efficient remote work management while implementing appropriate error handling mechanisms for invalid input parameters.

```
create table hybrid_model as select * from salaries;


update hybrid_model
set remote_ratio = 100
where company_location in ("US","AU") and salary_in_usd > 90000
```

```
select * from hybrid_model
```

| work_year | experience_level | employment_type | job_title | salary | salary_currency | salary_in_usd | employee_residence | remote_ratio | company_location | company_size |
|---|---|---|---|---|---|---|---|---|---|---|
| 2024 | SE | FT | AI Engineer | 90000 | USD | 133956 | AE | 0 | AE | L |
| 2024 | SE | FT | Machine Learning Engineer | 180500 | USD | 268656 | US | 100 | US | M |
| 2024 | SE | FT | Machine Learning Engineer | 96200 | USD | 143184 | US | 100 | US | M |
| 2024 | SE | FT | Machine Learning Engineer | 235000 | USD | 349774 | AU | 100 | AU | M |
| 2024 | SE | FT | Machine Learning Engineer | 175000 | USD | 260470 | AU | 100 | AU | M |
| 2024 | MI | FT | Business Intelligence Developer | 95413 | USD | 161248 | US | 100 | US | M |
| 2024 | MI | FT | Business Intelligence Developer | 70692 | USD | 119470 | US | 100 | US | M |
| 2024 | SE | FT | Data Engineer | 110000 | USD | 163724 | UA | 0 | UA | M |
| 2024 | SE | FT | Data Engineer | 48000 | USD | 71443 | UA | 0 | UA | M |
| 2024 | SE | FT | Data Scientist | 222300 | USD | 330871 | US | 100 | US | M |
| 2024 | SE | FT | Data Scientist | 119700 | USD | 178161 | US | 100 | US | M |
| 2024 | SE | FT | Cloud Database Engineer | 170375 | USD | 253587 | US | 100 | US | M |
| 2024 | SE | FT | Cloud Database Engineer | 102500 | USD | 152561 | US | 100 | US | M |

# Insights

8 -In year 2024, due to increase demand in data industry , there was  increase in salaries of data field employees.

Entry Level-35%  of the salary.
Mid junior – 30% of the salary.
Immediate senior level- 22% of the salary.
Expert level- 20% of the salary.
Director – 15% of the salary.

you have to update the salaries accordingly and update it back in the original database.

```
UPDATE hybrid_model
SET salary_in_usd=CASE
        WHEN experience_level="EN" THEN salary_in_usd*1.35
        WHEN experience_level="MI" THEN salary_in_usd*1.30
        WHEN experience_level="SE" THEN salary_in_usd*1.22
        WHEN experience_level="EX" THEN salary_in_usd*1.20
        WHEN experience_level="DX" THEN salary_in_usd*1.15
        ELSE salary_in_usd
        END
where work_year=2024;
SELECT * FROM hybrid model;
```

| work_year | experience_level | employment_type | job_title | salary | salary_currency | salary_in_usd | employee_residence | remote_ratio | company_location | company_size |
|---|---|---|---|---|---|---|---|---|---|---|
| 2024 | SE | FT | AI Engineer | 90000 | USD | 133956 | AE | 0 | AE | L |
| 2024 | SE | FT | Machine Learning Engineer | 180500 | USD | 268656 | US | 100 | US | M |
| 2024 | SE | FT | Machine Learning Engineer | 96200 | USD | 143184 | US | 100 | US | M |
| 2024 | SE | FT | Machine Learning Engineer | 235000 | USD | 349774 | AU | 100 | AU | M |
| 2024 | SE | FT | Machine Learning Engineer | 175000 | USD | 260470 | AU | 100 | AU | M |
| 2024 | MI | FT | Business Intelligence Developer | 95413 | USD | 161248 | US | 100 | US | M |
| 2024 | MI | FT | Business Intelligence Developer | 70692 | USD | 119470 | US | 100 | US | M |
| 2024 | SE | FT | Data Engineer | 110000 | USD | 163724 | UA | 0 | UA | M |
| 2024 | SE | FT | Data Engineer | 48000 | USD | 71443 | UA | 0 | UA | M |
| 2024 | SE | FT | Data Scientist | 222300 | USD | 330871 | US | 100 | US | M |
| 2024 | SE | FT | Data Scientist | 119700 | USD | 178161 | US | 100 | US | M |
| 2024 | SE | FT | Cloud Database Engineer | 170375 | USD | 253587 | US | 100 | US | M |
| 2024 | SE | FT | Cloud Database Engineer | 102500 | USD | 152561 | US | 100 | US | M |

# Insights

9 -You are a researcher and you have been assigned the task to Find the year with the highest average salary for each job title.

```sql
with cte as (
    select
        work_year,job_title,avg(salary_in_usd) as "avg_salary"
    from salaries
    group by 1,2)

select
    job_title,work_year,avg_salary
from
    (select
        job_title,work_year,avg_salary,
        RANK() OVER(PARTITION BY job_title ORDER BY avg_salary DESC) as "rank_sal"
    FROM cte) t
where rank_sal=1 ;
```

| job_title | work_year | avg_salary |
|---|---|---|
| Admin & Data Analyst | 2022 | 60000.0000 |
| AI Architect | 2024 | 256637.5000 |
| AI Developer | 2022 | 275000.0000 |
| AI Engineer | 2024 | 164314.7857 |
| AI Product Manager | 2024 | 152650.0000 |
| AI Programmer | 2023 | 72858.8000 |
| AI Research Engineer | 2024 | 131666.5000 |
| AI Research Scientist | 2024 | 88888.0000 |
| AI Scientist | 2024 | 175000.0000 |
| AI Software Engineer | 2024 | 174100.0000 |
| Analytics Engineer | 2024 | 166020.1284 |
| Analytics Engineerin… | 2023 | 399880.0000 |
| Applied Data Scientist | 2022 | 160800.0000 |
| Applied Machine Lear… | 2023 | 141726.3333 |
| Applied Machine Lear… | 2021 | 230700.0000 |
| Applied Scientist | 2024 | 193762.0870 |
| Autonomous Vehicle … | 2023 | 120000.0000 |
| AWS Data Architect | 2023 | 258000.0000 |
| Azure Data Engineer | 2020 | 100000.0000 |
| Azure Data Engineer | 2023 | 100000.0000 |
| BI Analyst | 2023 | 131315.5926 |
| BI Data Analyst | 2020 | 98000.0000 |
| BI Data Engineer | 2023 | 60000.0000 |

# Insights

10 - You have been hired by a market research agency where you been assigned the task to show the percentage of different employment type in different job roles, in the format where each row will be job title, each column will be type of employment type and  cell value  for that row and column will show the % value

```
select job_title,
    COUNT(case when employment_type="FT" then 1 end) *100 /COUNT(*) as "FT",
    COUNT(case when employment_type="PT" then 1 end) *100 /COUNT(*) as "PT",
    COUNT(case when employment_type="CT" then 1 end) *100 /COUNT(*) as "CT",
    COUNT(case when employment_type="FL" then 1 end) *100 /COUNT(*) as "FL"
    from salaries
    group by 1;
```

| job_title | FT | PT | CT | FL |
|---|---|---|---|---|
| AI Engineer | 98.8764 | 0.0000 | 1.1236 | 0.0000 |
| Machine Learning Engineer | 99.7954 | 0.0000 | 0.1364 | 0.0682 |
| Business Intelligence Developer | 100.0000 | 0.0000 | 0.0000 | 0.0000 |
| Data Engineer | 99.8343 | 0.1326 | 0.0000 | 0.0331 |
| Data Scientist | 99.6521 | 0.2436 | 0.0696 | 0.0348 |
| Cloud Database Engineer | 100.0000 | 0.0000 | 0.0000 | 0.0000 |
| Research Engineer | 100.0000 | 0.0000 | 0.0000 | 0.0000 |
| Data Analyst | 99.6633 | 0.2886 | 0.0481 | 0.0000 |
| Machine Learning Scientist | 100.0000 | 0.0000 | 0.0000 | 0.0000 |
| Applied Scientist | 100.0000 | 0.0000 | 0.0000 | 0.0000 |
| Data Science Manager | 100.0000 | 0.0000 | 0.0000 | 0.0000 |
| Research Scientist | 100.0000 | 0.0000 | 0.0000 | 0.0000 |
| Prompt Engineer | 100.0000 | 0.0000 | 0.0000 | 0.0000 |
| Data Science | 99.4186 | 0.5814 | 0.0000 | 0.0000 |
| Data Science Consultant | 98.6301 | 0.0000 | 0.0000 | 1.3699 |
| Data Management Analyst | 100.0000 | 0.0000 | 0.0000 | 0.0000 |
| Research Analyst | 100.0000 | 0.0000 | 0.0000 | 0.0000 |
| Data Operations Analyst | 100.0000 | 0.0000 | 0.0000 | 0.0000 |
| Data Management Consultant | 100.0000 | 0.0000 | 0.0000 | 0.0000 |
| Business Intelligence Analyst | 98.3471 | 0.0000 | 1.6529 | 0.0000 |
| Analytics Engineer | 100.0000 | 0.0000 | 0.0000 | 0.0000 |
| Data Quality Analyst | 100.0000 | 0.0000 | 0.0000 | 0.0000 |
| Data Architect | 100.0000 | 0.0000 | 0.0000 | 0.0000 |