Case Study of Data Science Jobs

```
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

/* 1. You are a compensation anlyst employed by a multinational corporation. Your Assignment is to pinpoint Countries who give work fully remote work, for the title 'Managers' Paying salaries Exceeding $90,000 USD */

```
select * from salaries;
select distinct company_location from salaries
where job_title like '%Manager%' and salary_in_usd>90000 and
remote_ratio=100;
```

company_location
US
MX
AU
FR

/* 2. AS a remote work advocate Working for a progressive HR tech startup who place their ‘freshers’ clients IN large tech firms. you're tasked WITH Identifying top 5 Country Having greatest count of large (company size) number of companies. */

```
select company_location,count(*) from (select * from salaries
where experience_level= 'EN' and company_size='L') t
group by company_location
order by count(*) desc
limit 5;
```

company_location	count(*)
US	53
DE	10
CA	10
GB	8
IN	6

/* 3.Picture yourself AS a data scientist Working for a workforce management platform.
Your objective is to calculate the percentage of employees. Who enjoy fully remote roles WITH salaries Exceeding $100,000 USD, Shedding light ON the attractiveness of high-paying remote positions IN today's job market */

```
set @total = (select count(*) from salaries where
salary_in_usd>100000);
set @count = (select count(*) from salaries where
salary_in_usd>100000 and remote_ratio=100);
set @percentage =((select @count)/(select @total))*100;
select round(@percentage,2) as percentage;
```

percentage
32.24

/* 4. Imagine you're a data analyst Working for a global recruitment agency. Your Task is to identify the Locations where entry-level average salaries exceed the average salary for that job title IN market for entry level, helping your agency guide candidates towards lucrative opportunities. */

```
select t.job_title, company_location,Average,avg_per_country
from(select job_title, avg(salary_in_usd) as 'Average'
from salaries where experience_level='EN' group by job_title)t
inner join
(select company_location,job_title, avg(salary_in_usd)
'avg_per_country' from salaries
where experience_level='EN'
group by company_location,job_title)m
on t.job_title=m.job_title
where avg_per_country>Average;
```

job_title	company_location	Average	avg_per_country
Data Analyst	US	84808.6361	89800.3519
Analytics Engineer	US	96722.3000	110831.2500
Data Engineer	US	92713.4701	106791.2584
Research Analyst	US	107294.2143	110459.5385
Machine Learning Engineer	US	110718.3778	126188.8529
Business Intelligence Analyst	AU	76688.6333	91000.0000

/* 5. You've been hired by a big HR Consultancy to look at how much people get paid IN different Countries. Your job is to Find out for each job title which. Country pays the maximum average salary. This helps you to place your candidates IN those countries */

```
select * from (select *, dense_rank() over(partition by job_title order
by average desc) as rank_num
from (select company_location,job_title, avg(salary_in_usd) as 'average'
from salaries
group by company_location,job_title)t)m
where rank_num=1
order by average desc;
```

company_location	job_title	average	rank_num
CA	AI Architect	800000.0000	1
MX	Data Analyst	429950.0000	1
IL	AI Scientist	417937.0000	1
GB	Analytics Engineering Manager	399880.0000	1
US	Data Science Tech Lead	375000.0000	1
US	Head of Machine Learning	337000.0000	1

/* 6. AS a data-driven Business consultant, you've been hired by a multinational corporation to analyze salary trends across different company Locations. Your goal is to Pinpoint Locations WHERE the average salary Has consistently Increased over the Past few years (Countries WHERE data is available for 3 years Only(present year and past two years) providing Insights into Locations experiencing Sustained salary growth. */

```
with companies as
(
select * from salaries where company_location in
(
select company_location from
(
select company_location, avg(salary_in_usd) as avg_salary,count(distinct
work_year) as cnt from salaries
```

```
where work_year>=year(current_date())-2 group by company_location having
cnt=3
)t
)
)
select company_location,
max(case when work_year=2022 then average end) as avg_salary_2022,
max(case when work_year=2023 then average end) as avg_salary_2023,
max(case when work_year=2024 then average end) as avg_salary_2024
from
(select company_location,work_year, avg(salary_in_usd) as average from
companies
group by company_location,work_year)q
group by company_location
having avg_salary_2024>avg_salary_2023 and
avg_salary_2023>avg_salary_2022 ;
```

company_location	avg_salary_2022	avg_salary_2023	avg_salary_2024
CA	126009.5526	150724.1414	153611.8077
ES	47997.3415	60327.9857	72184.6667
FI	63040.0000	71259.0000	77777.0000
FR	72684.4667	100411.1905	101370.1667
PT	48921.3750	51521.0000	53054.7500
AD	50000.0000	65000.0000	88500.0000

/* 7. Picture yourself AS a workforce strategist employed by a global HR tech startup.Your Mission is to Determine the percentage of fully remote work for each experience level IN 2021 and compare it WITH the corresponding figures for 2024, Highlighting any significant Increases or decreases IN remote work Adoption over the years */

```
select * from (select * , (count/total)*100 as 'remote_2021' from (select
a.experience_level,a.total,b.count from (select experience_level,count(*)
as total from salaries where work_year=2021
group by experience_level) a
inner join
(select experience_level,count(*) as count from salaries where
work_year=2021 and remote_ratio=100 group by experience_level) b
on a.experience_level=b.experience_level)t)m
inner join
(select * , (count/total)*100 as 'remote_2024' from (select
a.experience_level,a.total,b.count from (select experience_level,count(*)
as total from salaries where work_year=2021
group by experience_level) a
inner join
(select experience_level,count(*) as count from salaries where
work_year=2024 and remote_ratio=100 group by experience_level) b
on a.experience_level=b.experience_level)t)n
on m.experience_level=n.experience_level;
```

experience_level	total	count	remote_2021	experience_level	total	count	remote_2024
SE	75	44	58.6667	SE	75	483	644.0000
MI	87	45	51.7241	MI	87	227	260.9195
EN	46	22	47.8261	EN	46	87	189.1304
EX	10	5	50.0000	EX	10	35	350.0000

/* 8. AS a Compensation specialist at a Fortune 500 company, you're tasked WITH analyzing salary trends over time. Your objective is to calculate the average salary increase percentage for each experience level and job title between the years 2023 and 2024, helping the company stay competitive IN the talent market*/

```
with t as
      (select
job_title,experience_level,work_year,round(avg(salary_in_usd),2) as
average from salaries
      where work_year in (2023,2024)
      group by job_title,experience_level,work_year)

select *, round(((avg_salary_2024-avg_salary_2023)/avg_salary_2023),2) as
Changes
from (select job_title,experience_level,
           max(case when work_year=2023 then average end) as
avg_salary_2023,
           max(case when work_year=2024 then average end) as
avg_salary_2024
           from t group by job_title,experience_level)m
where round(((avg_salary_2024-avg_salary_2023)/avg_salary_2023),2)is not
null;
```

job_title	experience_level	avg_salary_2023	avg_salary_2024	Changes
AI Engineer	SE	172245.94	180068.57	0.05
Machine Learning Engineer	SE	196167.59	206863.44	0.05
Business Intelligence Developer	MI	84032.00	83385.63	-0.01
Data Engineer	SE	158309.32	161949.40	0.02
Data Scientist	SE	173480.98	160234.25	-0.08

/* 9. You're a database administrator tasked with role-based access control for a company's employee database.
Your goal is to implement a security measure where employees in different
experience level (e.g. Entry Level, Senior level etc.) can only access details relevant to their respective experience level, ensuring data confidentiality and minimizing the risk of unauthorized access.*/

```
create user 'Entry_level'@'%' identified by 'EN';
create view entry_level as
(
select * from salaries where experience_level ='EN'
);
grant alter on salary.entry_level to 'Entry_level'@'%';
show privileges;
```

/* 10. You are working with a consultancy firm, your client comes to you with certain data and preferences such as
(their year of experience, their employment type, company location and company size) and want to make a transaction
into different domain in data industry (like a person is working as a data analyst and want to move to some other
domain such as data science or data engineering etc.) your work is to guide them to which domain they should switch to
base on the input they provided, so that they can now update their knowledge as per the suggestion/. The Suggestion should be based on average salary */

```
DELIMITER //
create PROCEDURE GetAverageSalary(IN exp_lev VARCHAR(2), IN emp_type
VARCHAR(3), IN comp_loc VARCHAR(2), IN comp_size VARCHAR(2))
BEGIN
    SELECT job_title, experience_level, company_location, company_size,
employment_type, ROUND(AVG(salary), 2) AS avg_salary
    FROM salaries
```

```
    WHERE experience_level = exp_lev AND company_location = comp_loc AND
company_size = comp_size AND employment_type = emp_type
    GROUP BY experience_level, employment_type, company_location,
company_size, job_title order by avg_salary desc ;
END//
DELIMITER ;
-- Deliminator  By doing this, you're telling MySQL that statements
within the block should be parsed as a single unit until the custom
delimiter is encountered.

call GetAverageSalary('EN','FT','AU','M');

drop procedure Getaveragesalary;
```

job_title	experience_level	company_location	company_size	employment_type	avg_salary
Data Scientist	EN	AU	M	FT	120000.00
Business Intelligence Analyst	EN	AU	M	FT	91000.00
AI Programmer	EN	AU	M	FT	40000.00
Machine Learning Developer	EN	AU	M	FT	40000.00
Data Analyst	EN	AU	M	FT	36276.50

/* 11. 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. */

```
select company_size,count(*) as total_employee from salaries
where work_year=2021
group by company_size;
```

company_size	total_employee
M	52
S	42
L	124

/* 12. 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. */

```
select job_title,avg(salary_in_usd) as average from salaries
where employment_type='PT'
group by job_title
order by average desc
limit 3;
```

job_title	average
Data Scientist	87767.0000
Data Engineer	65168.5000
Data Analyst	56916.6667

/* 13. 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. */

```
-- Overall mid-level salary for the year 2023
set @average = (select avg(salary_in_usd) from salaries
                      where experience_level='MI' and work_year=2023);
-- Countries where average mid-level salary is higher than overall mid-
level salary for the year 2023
```

```
select company_location,avg(salary_in_usd) as avg_salary from salaries
where experience_level='MI'
group by company_location
having avg(salary_in_usd) > @average;
```

company_location	avg_salary
US	135521.1182
CA	127300.0299
AU	144658.1304
EG	124642.8571
NZ	125000.0000
OA	300000.0000

/* 14. 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. */

```
-- Query to find the highest average salary for senior-level employees in
2023
    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;
```

company_location	Highest_avg_salary
IL	266468.5000

```
-- Query to find the lowest average salary for senior-level employees in
2023
    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;
```

lowest_location	lowest_avg_salary
TR	18381.0000

/* 15. 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. */

```
with Annual_salary as
                              (select
t.job_title,t.salary_2023,m.salary_2024 from
                              (select job_title, avg(salary_in_usd) as
'salary_2023' from salaries
                             where work_year=2023
                             group by job_title)t
                             inner join
                             (select job_title, avg(salary_in_usd) as
'salary_2024' from salaries
```

```
                                where work_year=2024
                                group by job_title)m
                                on t.job_title=m.job_title)
select *,round(((salary_2024-salary_2023)/(salary_2023))*100,2) as
'Pct_change_in_salary' from Annual_salary;
```

job_title	salary_2023	salary_2024	Pct_change_in_salary
AI Engineer	161487.8298	164314.7857	1.75
Machine Learning Engineer	190703.5093	194716.1351	2.10
Business Intelligence Developer	109708.1500	96794.2500	-11.77
Data Engineer	150154.7500	147970.1100	-1.45
Data Scientist	162890.9357	148394.2106	-8.90
Cloud Database Engineer	141666.6667	136437.5000	-3.69

/* 16. 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 t as (select company_location, work_year, avg(salary_in_usd) as
average from salaries
                where  experience_level='EN' and (work_year=2021 or
work_year=2023)
                group by company_location, work_year)

select *, (((AVG_salary_2023 - AVG_salary_2021) / AVG_salary_2021) * 100)
AS changes
from (select company_location,
           max(case when work_year=2021 then average end) as
'AVG_Salary_2021',
           max(case when work_year=2023 then average end) as
'AVG_Salary_2023'
           from t
           group by company_location)a
           where (((AVG_salary_2023 - AVG_salary_2021) /
AVG_salary_2021) * 100) is not null
           order by (((AVG_salary_2023 - AVG_salary_2021) /
AVG_salary_2021) * 100) desc
           limit 3;
```

company_location	AVG_Salary_2021	AVG_Salary_2023	changes
AU	42028.0000	53089.3333	26.31896188
US	88617.6471	101592.8575	14.64179069
IN	24407.1667	27344.1667	12.03335084

/* 17. 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. */

```
-- creating temporary table so that changes are not made in actual table
```

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

update temp_salaries
set remote_ratio=100
where (company_location='AU' or company_location='AS') and
salary_in_usd>90000;
select * from temp_salaries
where (company_location='AU' or company_location='AS') and
salary_in_usd>90000;
```

/* 18. In the year 2024, due to increased demand in the data industry, there was an increase in salaries of data field
employees.

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

You must update the salaries accordingly and update them back in the original database. */

```
update temp_salaries
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 temp_salaries
where work_year=2024;
```

/* 19. You are a researcher and you have been assigned the task to Find the year with the highest average salary for
each job title. */

```
WITH avg_salary_per_year AS
(
    -- Calculate the average salary for each job title in each year
    SELECT work_year, job_title, AVG(salary_in_usd) AS avg_salary
    FROM salaries
    GROUP BY work_year, job_title
)

SELECT job_title, work_year, avg_salary FROM
    (
       -- Rank the average salaries for each job title in each year
       SELECT job_title, work_year, avg_salary,
       RANK() OVER (PARTITION BY job_title ORDER BY avg_salary DESC) AS
rank_by_salary
         FROM avg_salary_per_year
    ) AS ranked_salary
WHERE
    rank_by_salary = 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

/* 20. You have been hired by a market research agency where you been assigned the task to show the percentage of
different employment type (full time, part time) 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,
    ROUND((SUM(CASE WHEN employment_type = 'PT' THEN 1 ELSE 0 END) /
COUNT(*)) * 100, 2) AS PT_percentage, -- Calculate percentage of part-
time employment
    ROUND((SUM(CASE WHEN employment_type = 'FT' THEN 1 ELSE 0 END) /
COUNT(*)) * 100, 2) AS FT_percentage, -- Calculate percentage of full-
time employment
    ROUND((SUM(CASE WHEN employment_type = 'CT' THEN 1 ELSE 0 END) /
COUNT(*)) * 100, 2) AS CT_percentage, -- Calculate percentage of contract
employment
    ROUND((SUM(CASE WHEN employment_type = 'FL' THEN 1 ELSE 0 END) /
COUNT(*)) * 100, 2) AS FL_percentage -- Calculate percentage of freelance
employment
FROM
    salaries
GROUP BY
    job_title;
```

job_title	PT_percentage	FT_percentage	CT_percentage	FL_percentage
AI Engineer	0.00	98.88	1.12	0.00
Machine Learning Engineer	0.00	99.80	0.14	0.07
Business Intelligence Developer	0.00	100.00	0.00	0.00
Data Engineer	0.13	99.83	0.00	0.03
Data Scientist	0.24	99.65	0.07	0.03
Cloud Database Engineer	0.00	100.00	0.00	0.00