SQL - Data Science Salaries Project

Dipti Soni

Introduction

This dataset, available at Kaggle, contains information about salaries for different years for jobs related to Data Science. Information about company location, experience level, remote ratio, and employment type is also included.
Dataset was analyzed using MySQL

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

100% Remote Work for Managers

You're a Compensation analyst employed by a multinational corporation. Your Assignment is to Pinpoint Countries who give work 100% remotely, for the title 'managers' Paying salaries Exceeding $90,000 USD

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

company_location
US
MX
AU
FR

Top Countries with Large Companies

As a remote work advocate working for a progressive HR tech startup who place their freshers' in large tech firms. You're tasked with Identifying top 5 Countries having greatest count of large(company size) number of companies.

```
select company_location, count(*) from salaries
where experience_level = 'EN' and company_size = 'L'
group by company_location order by count(*) desc limit 5;
```

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

High Paying Remote Positions

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

```
set @wfh = (select count(*) from salaries where remote_ratio = 100 and salary_in_usd > 100000);
set @total_work_force = (select count(*) from salaries);
set @percent = round(((select @wfh)/(select @total_work_force))*100,2);
select @percent as 'Percentage_of_remote_positions';
```

	Percentage_of_remote_positions
▶	24.37

Locations having greater Salary than average

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 countries.

```
select company_location, job_title, avg(salary) from salaries s1
where experience_level = 'EN' and
salary > (select avg(salary) from salaries s2 where experience_level = 'EN' and s2.job_title = s1.job_title)
group by company_location, job_title
order by company_location;
```

company_location	job_title	avg(salary)
AS	AI Scientist	1335000.0000
AT	Applied Data Scientist	50000.0000
AU	Business Intelligence Analyst	91000.0000
AU	Computer Vision Software Engineer	150000.0000
BA	AI Developer	120000.0000

Countries paying max average salary

You've been hired by a big HR Consultancy to look at how many 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.

```sql
with cte as
(select job_title, company_location, avg(salary) as 'avg_sal' from salaries
group by job_title, company_location
order by job_title)
select * from (select * ,dense_rank() over(partition by job_title order by avg_sal desc) as 'rnk' from cte) t
where t.rnk = 1;
```

job_title	company_location	avg_sal	rnk
Admin & Data Analyst	IN	60000.0000	1
AI Architect	CA	800000.0000	1
AI Developer	CA	275000.0000	1
AI Engineer	QA	300000.0000	1
AI Product Manager	US	152650.0000	1

Consistent increase in avg salary

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(this and past two years) providing Insights into Locations experiencing Sustained salary growth.

```
select company_location,
max(case when work_year = 2022 then avg end) as avg_2022,
max(case when work_year = 2023 then avg end) as avg_2023,
max(case when work_year = 2024 then avg end) as avg_2024
from (select company_location, work_year, avg(salary) as 'avg' from cte
        group by company_location, work_year
        order by company_location)t group by company_location having avg_2024 > avg_2023 and avg_2023 > avg_2022;
```

company_location	avg_2022	avg_2023	avg_2024
AR	50000.0000	65000.0000	88500.0000
CA	134865.7895	153553.9545	157440.5673
ES	45682.9268	56952.4714	64966.6667

Change in Remote Work over the years

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 increase or decrease in remote work adoption over the years.

```
select t1.experience_level, total_2021, wfh_2021, wfh_2021/total_2021*100 as 'percent_2021',
total_2024, wfh_2024, wfh_2024/total_2024*100 as 'percent_2024' from
(select experience_level, count(*) as 'total_2021' from salaries where work_year = 2021
group by experience_level order by experience_level)t1
join
(select experience_level, count(*) as 'wfh_2021' from salaries where remote_ratio = 100 and work_year = 2021
group by experience_level order by experience_level)t2
on t1.experience_level = t2.experience_level
join
(select experience_level, count(*) as 'total_2024' from salaries where work_year = 2024
group by experience_level order by experience_level)t3
on t1.experience_level = t3.experience_level
join
(select experience_level, count(*) as 'wfh_2024' from salaries where remote_ratio = 100 and work_year = 2024
group by experience_level order by experience_level)t4
on t1.experience_level = t4.experience_level;
```

experience_level	total_2021	wfh_2021	percent_2021	total_2024	wfh_2024	percent_2024
EN	46	22	47.8261	381	87	22.8346
EX	10	5	50.0000	106	35	33.0189
MI	87	45	51.7241	1102	227	20.5989
SE	75	44	58.6667	1920	483	25.1563

Analyzing Salary Trends (2023 to 2024)

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.

```
select t1.experience_level, t1.job_title, avg_sal_2023, avg_sal_2024, (avg_sal_2024 - avg_sal_2023)*100/avg_sal_2023 as 'percent_change' from
        (select experience_level, job_title, avg(salary) as 'avg_sal_2023' from salaries where work_year = 2023
        group by experience_level, job_title
        order by experience_level, job_title) t1
        join
        (select experience_level, job_title, avg(salary) as 'avg_sal_2024' from salaries where work_year = 2024
        group by experience_level, job_title
        order by experience_level, job_title) t2
        on t1.experience_level = t2.experience_level and t1.job_title = t2.job_title;
```

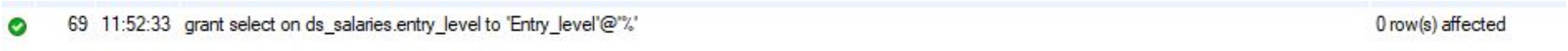
experience_level	job_title	avg_sal_2023	avg_sal_2024	percent_change
EN	AI Engineer	27500.0000	40000.0000	45.45454545
EN	Analytics Engineer	87908.3333	102333.3333	16.40913831
EN	Applied Scientist	178367.5000	135000.0000	-24.31356609

Role based authorized access

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'
);

show privileges;
grant select on ds_salaries.entry_level to 'Entry_level'@'%';
```



Switch Job based on data

You are working with an 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 transition 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 based 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 ;
```

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

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

Job Market for 2021

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(*), avg(salary) from salaries where work_year = 2021
group by company_size;
```

company_size	count(*)	avg(salary)
M	52	206337.3846
S	42	518526.1905
L	124	721216.7097

Highest Salary for Part Time in 2023

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) from salaries where work_year = 2023 and employment_type = 'PT'
group by job_title order by avg(salary) desc limit 3;
```

job_title	avg(salary)
Data Scientist	95650.0000
Data Analyst	78000.0000

Higher than average mid level salary

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, avg(salary) from salaries
where work_year = 2023 and experience_level = 'MI'
group by company_location
having avg(salary) > (select avg(salary) from salaries where work_year = 2023 and experience_level = 'MI');
```

company_location	avg(salary)
AU	199544.5455
PH	343666.6667
IN	2208500.0000
BR	140400.0000
QA	300000.0000
ZA	627000.0000
HK	510000.0000
SG	770000.0000
TH	840000.0000

Salary Analysis for Senior Level

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.

```
(select company_location, avg(salary) from salaries where work_year = 2023 and experience_level = 'SE' group by company_location order by avg(salary) limit 1)
UNION
(select company_location, avg(salary) from salaries where work_year = 2023 and experience_level = 'SE' group by company_location order by avg(salary) desc limit 1);
```

company_location	avg(salary)
RO	36000.0000
IN	1205085.7143

Salary Trends for different job roles

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.

```
select *, lag(avg_sal) over() as 'percent_change' from
(
    select job_title, work_year, avg(salary) as 'avg_sal' from salaries group by job_title, work_year order by job_title
)t;
```

job_title	work_year	avg_sal	percent_change
Admin & Data Analyst	2022	60000.0000	NULL
Admin & Data Analyst	2023	50000.0000	60000.0000
Admin & Data Analyst	2024	38000.0000	50000.0000
AI Architect	2023	249000.0000	38000.0000
AI Architect	2024	256637.5000	249000.0000
AI Developer	2022	275000.0000	256637.5000
AI Developer	2023	132588.2353	275000.0000
AI Developer	2024	30000.0000	132588.2353