SQL Case Study on Workforce Income Analysis

Introduction

The Workforce Income Analysis project aims to provide insights into salary trends and workforce dynamics based on various factors such as company size, job title, experience level, and geographical location. By querying a database with employee salary data, the project helps analyse different aspects of the labour market, offering valuable information for decision-making in workforce management, compensation planning, and career development.

Task 1

Investigating the job market based on company size in 2021:

Task: You need to count how many employees are working in different companies, categorized by size (S, M, L).

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

company_size	employees_working
L	124
M	52
S	42

Task 2

Top 3 job titles with the highest average salary for part-time positions in 2023:

Task: Identify the highest-paying job titles for part-time positions while ensuring you only include countries with more than 50 employees.

select top 3 job_title, AVG(salary) as avg_salary, count(*) as employee_count from salaries
where employment_type = 'PT' and
company_location in (select company_location from salaries group by company_location having count(*) > 50)
group by job_title
order by avg_salary desc

job_title	avg_salary	employee_count
Data Scientist	98540	5
Data Analyst	66687.3333333333	6
Data Engineer	62800	3

task 3

Countries where mid-level salary is higher than the overall mid level salary in 2023:
Task: Identify countries where the average salary for mid-level employees (MI) is greater than the overall average for that level.
select company_location, round(AVG(salary),1) as avg_salary from salaries
where experience_level = 'MI'
group by company_location
having AVG(salary) > (select avg(salary) from salaries)
order by avg_salary desc

company_location	avg_salary
CL	30400000
HU	5528500
JP	5191333.3
ID	4200000
IN	1994800
TH	840000
HK	510000
SG	437500
PH	343666.7
QA	300000
NZ	200000

task 4

Highest and lowest average salary locations for senior-level employees in 2023:

Task: Identify which countries pay seniorlevel (SE) employees the highest and lowest average salaries.

```
with salaryLocation as(
select company_location, avg(salary_in_usd) as avg_salary from salaries where experience_level = 'SE' and
work_year = 2023 group by company_location
)
select company_location, avg_salary from salaryLocation
where avg_salary = (select MIN(avg_salary) from salaryLocation) or avg_salary = (select MAX(avg_salary) from
salaryLocation) ;
```

	company_location	avg_salary
1	IL	266468.5
2	TR	18381

Task 5

Salary growth rates by job title: Task: Calculate the percentage increase in salaries for various job titles between two years (e.g., 2023 and 2024).

```
WITH salaryComparison AS (
select s1.job_title,
avg( s1.salary_in_usd) AS salary_2023,
avg(s2.salary_in_usd) AS salary_2024
from salaries s1
JOIN salaries s2 ON s1.job_title = s2.job_title
WHERE s1.work_year = 2023 and s2.work_year = 2024
group by s1.job_title
)
select job_title,salary_2023,salary_2024,
ROUND(((salary_2024 - salary_2023) / salary_2023) * 100, 2) AS salary_growth_percentage
from salaryComparison;
```

	job_title	salary_2023	salary_2024	salary_growth_percentage
1	Admin & Data Analyst	50000	54888.6	9.78
2	AI Architect	250328	318431.083333333	27.21
3	AI Developer	133266.823529412	43332.9	-67.48
4	AI Engineer	161487.829787234	204096.060476191	26.38
5	AI Product Manager	120000	198445	65.37
6	AI Programmer	72858.8	39000	-46.47
7	AI Research Engineer	70117	171166.45	144.12
8	AI Scientist	126842.4	213500	68.32
9	Analytics Engineer	162976.414414414	206993.308990827	27.01
10	Applied Scientist	189941.703571429	242558.441739129	27.7
11	BI Analyst	131315.592592593	116183.333333333	-11.52
12	BI Developer	112433.818181818	126225.954545455	12.27
13	Big Data Engineer	89292.5	20240	-77.33
14	Business Intelligence Analyst	118000.102564103	115611.786046512	-2.02
15	Business Intelligence Developer	109708.15	120636.613571429	9.96
16	Business Intelligence Engineer	149811.37804878	146546.622692306	-2.18
17	Business Intelligence Manager	126750	209430	65.23

task 6

Top three countries with the highest salary growth for entrylevel roles from 2020 to 2023:

```
with salarycomparison as (
select company_location,
avg(case when work_year = 2020 then salary_in_usd end) as salary_2020,
avg(case when work_year = 2023 then salary_in_usd end) as salary_2023,
count(*) as employee_count
from salaries where experience_level = 'EN'  and (work_year = 2020 or work_year = 2023)
group by company_location
having count(*) > 50
)
select top 3 company_location, salary_2020, salary_2023,
round(((salary_2023 - salary_2020) / salary_2020) * 100, 2) as salary_growth_percentage
from salarycomparison
order by salary_growth_percentage desc;
```


company_location	salary_2020	salary_2023	salary_growth_percentage
CA	44753	76852.75	71.73
DE	46120.2	69220	50.09
IT	21669	29691	37.02

Task 7

Updating remote work ratio for employees earning more than $90,000 in the US and AU:

Task: Update the remote_ratio for employees based on their salary and location.

```
update salaries
set remote_ratio = 100
where salary_in_usd > 90000
and employee_residence in ('US', 'AU');
```

Task 8

Salary updates based on percentage increases by level in 2024:

Task: Update the salaries for various experience levels (SE, MI, etc.) according to predefined percentage increases.

```
update salaries
set salary_in_usd = salary_in_usd *
case
when experience_level = 'SE' then 1.22
when experience_level = 'MI' then 1.30
when experience_level = 'EN' then 1.15
when experience_level = 'EX' then 1.10
else salary_in_usd   no change for other levels
end
where work_year = 2024;
```

Messages

(3509 rows affected)

Task 9

Year with the highest average salary for each job title:

Task: Identify which year had the highest average salary for each job title.

```
with avg_salary_per_year as (
select job_title,work_year,
avg(salary_in_usd) as avg_salary
from salaries
group by job_title, work_year
)
select a.job_title,a.work_year,a.avg_salary
from avg_salary_per_year a
inner join (
select job_title,max(avg_salary) as max_avg_salary from avg_salary_per_year group by job_title
) b on a.job_title = b.job_title and a.avg_salary = b.max_avg_salary;
```

Results   Messages

job_title	work_year	avg_salary
Staff Machine Learning Engineer	2021	185000
Staff Data Scientist	2020	164000
Staff Data Analyst	2023	179998
Software Data Engineer	2023	111627.666666667
Sales Data Analyst	2020	60000
Robotics Software Engineer	2024	310372.25
Robotics Engineer	2024	184811.5
Research Scientist	2024	339752.932765625
Research Engineer	2024	338500.597240909
Research Analyst	2024	186055.337
Quantitative Research Analyst	2024	67447.5
Prompt Engineer	2024	403320.3
Product Data Analyst	2022	120000
Principal Machine Learning Engineer	2022	190000
Principal Data Scientist	2021	239152.4
Principal Data Engineer	2021	192500
Principal Data Architect	2022	38154

task 10

Percentage of employment types for different job titles:

Task: Calculate the percentage of full-time and part-time employees for each job title.

```
with employment_counts as (
select job_title, employment_type, count(*) as employee_count
from salaries
group by  job_title, employment_type
)
select job_title,
round(
(cast(sum(case when employment_type = 'FT' then employee_count else 0 end) as float) * 100.0) /
sum(employee_count),
2) as full_time_percentage,
round(
(cast(sum(case when employment_type = 'PT' then employee_count else 0 end) as float)* 100.0) /
sum(employee_count),
2) as part_time_percentage
from employment_counts
group by job_title;
```

Results   Messages

job_title	full_time_percentage	part_time_percentage
Admin & Data Analyst	100	0
AI Architect	100	0
AI Developer	100	0
AI Engineer	98.88	0
AI Product Manager	100	0
AI Programmer	100	0
AI Research Engineer	85.71	0
AI Research Scientist	0	0
AI Scientist	100	0
AI Software Engineer	100	0
Analytics Engineer	100	0
Analytics Engineering Manager	100	0
Applied Data Scientist	100	0
Applied Machine Learning Engineer	100	0
Applied Machine Learning Scientist	78.57	0
Applied Scientist	100	0
Autonomous Vehicle Technician	50	0

task 11

COUNTRIES OFFERING FULL REMOTE WORK FOR MANAGERS WITH SALARIES OVER $90,000:

TASK: FIND COUNTRIES WHERE MANAGERS EARN MORE THAN $90,000 AND WORK FULLY REMOTELY.

```
select employee_residence, count(*) as manager_count
from salaries
where job_title like '%Manager%' and salary_in_usd > 90000  and remote_ratio = 100
group by employee_residence
order by manager_count desc;
```

employee_residence	manager_count
US	311
AU	1
DE	1
MX	1

task 12

Top 5 countries with the largest companies:

Task: Identify which countries have the highest number of large companies.

select top 5 company_location, count(*) as large_company_count
from salaries
where company_size = 'L'
group by company_location
order by large_company_count desc;

company_location	large_company_count
US	708
IN	37
DE	29
CA	28
GB	26

task 13

Percentage of employees with fully remote roles earning more than $100,000:

Task: Calculate the percentage of fully remote employees earning more than $100,000.

select round(
(cast(count(case when remote_ratio = 100 and salary_in_usd > 100000 then 1 end) as float) * 100) / cast(count(*) as float), 2
) as percentage_remote_over_100k
from salaries;

percentage_remote_over_100k
76.88

task 14

Locations where entry-level average salaries exceed market average for entry level:

Task: Identify locations where entrylevel salaries surpass the market average.

with market_avg as (
select avg(salary_in_usd) as market_average
from salaries
where experience_level = 'EN'
),
location_avg as (
select company_location,avg(salary_in_usd) as location_average

```
from salaries
where experience_level = 'EN'
group by company_location
)
select la.company_location, la.location_average
from location_avg la
join market_avg ma on la.location_average > ma.market_average;
```

company_location	location_average
BA	120000
MX	295941.77625
SE	105000
US	117053.287650721

task 15

Countries paying the maximum average salary for each job title:

Task: For each job title, identify which country pays the highest average salary.

```
with job_title_avg as (
select job_title, employee_residence, round(avg(salary_in_usd),2) as avg_salary
from salaries
group by job_title, employee_residence
)
select job_title, employee_residence, avg_salary
from job_title_avg j1
where avg_salary = (select max(avg_salary) from job_title_avg j2 where j2.job_title = j1.job_title);
```

Results | Messages | Client Statistics

job_title	employee_residence	avg_salary
Admin & Data Analyst	ES	71355.18
AI Architect	CA	1352000
AI Developer	CA	275000
AI Engineer	QA	300000
AI Product Manager	US	257978.5
AI Programmer	CA	97043.5
AI Research Engineer	US	338000
AI Research Scientist	DE	117554.38
AI Scientist	IL	417937
AI Software Engineer	EG	259130.44
Analytics Engineer	FR	279819.2
Analytics Engineering Manager	GB	399880
Applied Data Scientist	US	238000
Applied Machine Learning Engineer	US	177500
Applied Machine Learning Scientist	US	188800
Applied Scientist	US	212314.65

task 16

Countries with sustained salary growth over three years:

Task: Identify countries with consistent salary growth over the past three years.

```
with yearly_avg_salaries as (
select employee_residence, work_year, ROUND( avg(salary_in_usd),2) as avg_salary
from salaries
where work_year in (2021, 2022, 2023)
group by employee_residence, work_year
```

```
),
salary_growth as (
select a.employee_residence, a.avg_salary as salary_2021, b.avg_salary as salary_2022,  c.avg_salary as salary_2023
from yearly_avg_salaries a
join yearly_avg_salaries b on a.employee_residence = b.employee_residence and a.work_year = 2021 and
b.work_year = 2022
join yearly_avg_salaries c on a.employee_residence = c.employee_residence and a.work_year = 2021 and c.work_year
= 2023
)
select employee_residence, salary_2021, salary_2022, salary_2023 from salary_growth
where salary_2022 > salary_2021 and salary_2023 > salary_2022;
```

Results | Messages | Client Statistics

employee_residence	salary_2021	salary_2022	salary_2023
AU	58539	84048.75	163288.82
CA	101475.33	124209.07	151545.16
ES	49383.2	52159.14	60604.72
FR	53539.5	70402.88	110120.95
GB	80495.08	84265.02	105723.28
IN	39913.79	47756.8	50047.55
PL	37536.5	38987	60448.83
TR	24093.5	25000	43326.25
US	148133.99	148237.3	159752.25

task 17

PERCENTAGE OF FULLY REMOTE WORK BY EXPERIENCE LEVEL (2021 VS 2024):

TASK: COMPARE THE ADOPTION OF FULLY REMOTE WORK ACROSS EXPERIENCE LEVELS BETWEEN 2021 AND 2024.

```
with remote_percentage as (
select experience_level, work_year,
round(cast(count(case when remote_ratio = 100 then 1 end) as float) / count(*) * 100,2) as remote_percentage
from salaries
where work_year in (2021, 2024)
group by experience_level, work_year
)
select experience_level,
max(case when work_year = 2021 then remote_percentage end) as remote_percentage_2021,
max(case when work_year = 2024 then remote_percentage end) as remote_percentage_2024
from remote_percentage
group by experience_level;
```

Results | Messages | Client Statistics

experience_level	remote_percentage_2021	remote_percentage_2024
EN	50	66.93
EX	60	92.45
MI	60.92	88.02
SE	73.33	91.51

task 18

Average salary increase percentage by experience level and job title (2023 to 2024):

Task: Calculate the average salary increase for each experience level and job title.

```
with salary_increase as (
select a.experience_level, a.job_title, a.salary_in_usd as salary_2023, b.salary_in_usd as salary_2024,
```

```
cast((b.salary_in_usd - a.salary_in_usd) as float) / a.salary_in_usd * 100 as salary_increase_percentage
from salaries a
join salaries b on a.experience_level = b.experience_level and a.job_title = b.job_title and a.work_year = 2023 and
b.work_year = 2024
)
select experience_level, job_title, round(avg(salary_increase_percentage),2) as avg_salary_increase_percentage
from salary_increase
group by experience_level, job_title
order by avg_salary_increase_percentage desc;
```

J % ▼ ◄

Results 📄 **Messages** 📊 **Client Statistics**

experience_level	job_title	avg_salary_increase_percentage
MI	Prompt Engineer	498.69
MI	MLOps Engineer	431.71
MI	ML Engineer	320.47
MI	Data Science Consultant	277.77
SE	Data Analytics Lead	246.12
MI	Research Engineer	240.48
MI	Data Operations Specialist	229.65
MI	Data Science Manager	229.09
SE	Data Operations Specialist	204.47
SE	Machine Learning Researcher	194.56
SE	Prompt Engineer	187.41
MI	Machine Learning Scientist	181.71
MI	Business Intelligence Engineer	177.52
MI	Data Infrastructure Engineer	174.74
MI	Data Science Engineer	165.58
EN	Research Analyst	165.43

task 19

Role-based access control for employees based on experience level:

Task: Implement security to restrict access based on an employee's experience level.

View for Entry-Level Employees (EL)

```
create procedure get_employee_data_by_experience_level
@user_experience_level varchar(5)
as
begin
select experience_level, salary_in_usd, company_location, work_year,employment_type, job_title
from salaries
where experience_level = @user_experience_level;
end;

exec get_employee_data_by_experience_level 'MI';
```

experience_level	salary_in_usd	company_location	work_year	employment_type	job_title
MI	161247.97	US	2024	FT	Business Intelligence Developer
MI	119469.48	US	2024	FT	Business Intelligence Developer
MI	165620	US	2024	FT	Data Engineer
MI	115427	US	2024	FT	Data Engineer
MI	377039	US	2024	FT	Machine Learning Engineer
MI	213616	US	2024	FT	Machine Learning Engineer
MI	280540	US	2024	FT	Research Engineer
MI	123370	US	2024	FT	Research Engineer
MI	228150	US	2024	FT	Data Scientist
MI	185900	US	2024	FT	Data Scientist
MI	354900	US	2024	FT	Data Engineer
MI	219700	US	2024	FT	Data Engineer
MI	229840	US	2024	FT	Data Scientist
MI	180323	US	2024	FT	Data Scientist
MI	277160	US	2024	FT	Data Scientist
MI	221390	US	2024	FT	Data Scientist

task 20

Guiding clients in switching domains based on salary insights:

Task: Based on an employee's data (experience, job title, location), suggest new domains they can transition to, based on salary trends.

with salary_growth_cte as (
select job_title, company_location, experience_level ,
round(avg(case when work_year = 2023 then salary_in_usd end),2) as salary_2023,
round(avg(case when work_year = 2024 then salary_in_usd end),2) as salary_2024
from salaries
where work_year in (2023, 2024)
group by job_title, company_location,experience_level
)
select job_title, company_location,experience_level, salary_2023, salary_2024,
round(((salary_2024 - salary_2023) / salary_2023) * 100,2) as salary_growth_percentage
from salary_growth_cte
where ((salary_2024 - salary_2023) / salary_2023) * 100 > 10
order by salary_growth_percentage desc;

job_title	company_location	experience_level	salary_2023	salary_2024	salary_growth_percentage
Machine Learning Engineer	ES	MI	38868	196884.16	406.55
Lead Data Analyst	IN	MI	18241	66949.35	267.03
Data Analytics Lead	US	SE	99960	337159.81	237.29
ML Engineer	US	MI	130775	432772.38	230.93
Data Operations Specialist	US	MI	57500	186322.5	224.04
Data Operations Specialist	US	SE	53210	161647.68	203.79
MLOps Engineer	US	MI	124275	371800	199.18
Data Infrastructure Engineer	US	MI	180827.5	494325	173.37
Research Engineer	US	MI	159959.33	433319.17	170.89
Prompt Engineer	US	SE	166500	444659.5	167.06
Data Science Engineer	US	MI	121500	321100	164.28
Business Intelligence Analyst	CA	MI	77584	197138.5	154.1
Data Scientist	GB	MI	79475.88	195578.91	146.09
Machine Learning Scientist	US	MI	246250	602766.67	144.78
Data Analyst	CA	EN	40747	98561.3	141.89
AI Research Engineer	US	MI	150000	338000	125.33
Data Science Consultant	US	MI	86666.67	194350	124.25