SQL Project Salaries Analysis

Presented by Saif Ullah

Data Set Overview

The provided dataset contains salary information for various full-time positions in the tech industry for the year 2024. The dataset includes senior and mid-level roles across different job titles such as AI Engineer, Machine Learning Engineer, Business Intelligence Developer, Data Engineer, Data Scientist, Cloud Database Engineer, and Research Engineer. Employees are located in the United States (US), United Arab Emirates (AE), Australia (AU), and Ukraine (UA). Most of the roles are based in medium-sized companies, except for a few in large companies in the UAE. Salaries in the dataset range from $48,000 to $283,800 USD, with the highest being for a fully remote Data Engineer in the US. The dataset highlights the presence of fully remote roles, particularly in the US, with significant salaries exceeding $100,000 USD. Senior positions generally command higher salaries compared to mid-level roles. The data reveals trends in remote work, salary distribution across different job titles and locations, and the impact of company size on compensation.

```sql
-- 1. Pinpoint countries with fully remote
-- 'Manager' titles paying salaries exceeding $90,000 USD
SELECT *
FROM salaries
WHERE remote_ratio = 100
  AND job_title LIKE '%Manager%'
  AND salary_in_usd > 90000;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio	company_location	compar
2024	MI	FT	Data Manager	98500	USD	98500	US	100	US	M
2024	MI	FT	Data Product Manager	171000	USD	171000	US	100	US	M
2024	MI	FT	Data Product Manager	136000	USD	136000	US	100	US	M
2024	SE	FT	Data Manager	131200	USD	131200	US	100	US	M
2024	SE	FT	Data Manager	95300	USD	95300	US	100	US	M
2024	MI	FT	Data Manager	127000	USD	127000	US	100	US	M
2024	SE	FT	Data Product Manager	170000	USD	170000	US	100	US	M
2024	SE	FT	Data Product Manager	145000	USD	145000	US	100	US	M
2024	SE	FT	Data Manager	131200	USD	131200	US	100	US	M
2024	SE	FT	Data Manager	95300	USD	95300	US	100	US	M
2024	SE	FT	Data Manager	131200	USD	131200	US	100	US	M

```sql
-- 2. Identify top 5 countries with the
-- greatest count of large companies hiring freshers
SELECT company_location, COUNT(company_size) AS count_size
FROM (
    SELECT *
    FROM salaries
    WHERE experience_level = 'EN' AND company_size = 'L'
) t
GROUP BY company_location
ORDER BY count_size DESC
LIMIT 5;
```

company_location	count_size
US	53
DE	10
CA	10
GB	8
IN	6

```sql
-- 3. Calculate the percentage of employees
-- in fully remote roles with salaries exceeding $100,000 USD
SET @count = (
    SELECT COUNT(*)
    FROM salaries
    WHERE remote_ratio = 100 AND salary_in_usd >= 100000
);

SET @total = (
    SELECT COUNT(*)
    FROM salaries
    WHERE salary_in_usd >= 100000
);

SET @percentage = ROUND(((@count / @total) * 100), 2);

SELECT @percentage AS '% Salary people work remotely with 100000 USD';
```

	% Salary people work remotely with 100000 USD
▶	32.24

```sql
-- 4. Identify locations where entry-level
-- average salaries exceed the market average for that job title
SELECT company_location, t.job_title, avg_salary_per_country, avg_salary
FROM (
    SELECT company_location, job_title, AVG(salary_in_usd) AS avg_salary_per_country
    FROM salaries
    WHERE experience_level = 'EN'
    GROUP BY company_location, job_title
) AS t
INNER JOIN (
    SELECT job_title, AVG(salary_in_usd) AS avg_salary
    FROM salaries
    WHERE experience_level = 'EN'
    GROUP BY job_title
) AS n ON t.job_title = n.job_title
WHERE avg_salary_per_country > avg_salary;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

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

```sql
-- 5. Find the country that pays
-- the maximum average salary for each job title
SELECT company_location, job_title, avg_salary
FROM (
    SELECT company_location, job_title, AVG(salary_in_usd) AS avg_salary,
           DENSE_RANK() OVER (PARTITION BY job_title ORDER BY AVG(salary_in_usd) DESC) AS rank_
    FROM salaries
    GROUP BY company_location, job_title
) AS ranked_salaries
WHERE rank_ = 1;
```

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

```sql
-- 6. Pinpoint locations where the average salary has consistently increased over the past three yea
WITH s AS (
    SELECT *
    FROM salaries
    WHERE company_location IN (
        SELECT company_location
        FROM (
            SELECT company_location, COUNT(DISTINCT(work_year)) AS num_years
            FROM salaries
            WHERE work_year >= YEAR(CURRENT_DATE) - 2
            GROUP BY company_location
            HAVING num_years = 3
        ) AS valid_locations
    )
)
SELECT company_location,
       MAX(CASE WHEN work_year = 2022 THEN avg_salary END) AS avg_salary_2022,
       MAX(CASE WHEN work_year = 2023 THEN avg_salary END) AS avg_salary_2023,
       MAX(CASE WHEN work_year = 2024 THEN avg_salary END) AS avg_salary_2024
FROM (
    SELECT company_location, work_year, AVG(salary_in_usd) AS avg_salary
    FROM s
    GROUP BY company_location, work_year
) AS yearly_salaries
GROUP BY company_location
HAVING avg_salary_2024 > avg_salary_2023 AND avg_salary_2023 > avg_salary_2022;
```

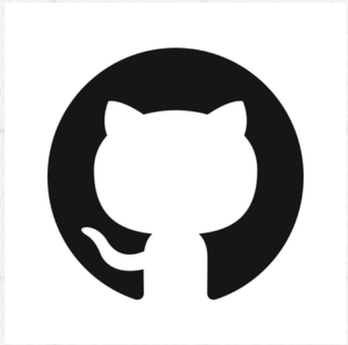
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

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
AR	50000.0000	65000.0000	88500.0000
IN	37328.3333	47777.5217	71538.3333

```sql
-- 7. Determine the percentage of fully remote work for each experience level in 2021 and 2024
WITH t1 AS (
    SELECT a.experience_level, total_remote, total_2021,
           ROUND((total_remote / total_2021) * 100, 2) AS '2021 remote %'
    FROM (
        SELECT experience_level, COUNT(*) AS total_remote
        FROM salaries
        WHERE work_year = 2021 AND remote_ratio = 100
        GROUP BY experience_level
    ) AS a
    INNER JOIN (
        SELECT experience_level, COUNT(*) AS total_2021
        FROM salaries
        WHERE work_year = 2021
        GROUP BY experience_level
    ) AS b ON a.experience_level = b.experience_level
    FROM (--
        SELECT experience_level, COUNT(*) AS total_remote
        FROM salaries
        WHERE work_year = 2024 AND remote_ratio = 100
        GROUP BY experience_level
    ) AS a
    INNER JOIN (
        SELECT experience_level, COUNT(*) AS total_2024
        FROM salaries
        WHERE work_year = 2024
        GROUP BY experience_level
    ) AS b ON a.experience_level = b.experience_level
)
SELECT t1.experience_level, t1.`2021 remote %`, t2.`2024 remot
FROM t1
INNER JOIN t2 ON t1.experience_level = t2.experience_level;
```

	experience_level	2021 remote %	2024 remote %
▶	SE	58.67	25.16
	MI	51.72	20.60
	EN	47.83	22.83
	EX	50.00	33.02

LinkedIn



GitHub

Project Link

Thank you !

Saif Ullah