Workforce Income Analysis (SQL Case Study)

INTRODUCTION

SCOPE:

This project analyses workforce salary data across multiple years, job titles, countries, and experience levels. The dataset provides insights into salary distribution, employment types, remote work adoption, and company structures.

OBJECTIVES:

1. *Business Insights and Strategic Decision Making*

- Salary Benchmarking: By analyzing the highest and lowest average salaries by job title, experience level, company size and location, businesses can establish salary benchmarks to remain competitive in the market.
- Identify Growth Markets: The analysis of salary growth rates by job title, location, and experience level will help identify the regions or industries that are growing faster, enabling companies to focus their expansion strategies on these areas.
- Talent Retention Strategies: Identifying countries and positions offering the highest salaries or observing salary growth trends over time will help businesses devise better talent retention strategies by ensuring competitive pay and benefits.
- Remote Work Optimization: By analyzing trends in remote work (e.g., fully remote positions by country and experience level), companies can make informed decisions about remote work policies, which is especially relevant in the postpandemic landscape.

2. *Talent Acquisition and Workforce Planning:*

- Targeting Specific Markets for Recruitment: Identifying countries with high-paying jobs or large companies will allow businesses to target specific regions for recruitment. For instance, understanding which countries offer the highest remote salary opportunities for certain job titles will help attract talent in remote-friendly regions.
- Demand for Experience Levels: Knowing the percentage of employees with different experience levels (e.g., mid-level, senior) allows for better workforce planning and recruiting strategies. For example, if mid-level employees are being paid higher in a particular market, businesses can focus on hiring and promoting in that region.
- Talent Scarcity Analysis: Analyzing the availability of part-time versus full-time positions, and comparing it with salary data, can indicate job market gaps or excesses. This helps identify where there is a shortage or oversupply of specific skills or roles.

3. *Workforce Optimization and Cost Control:*

- Optimizing Salaries by Location: Understanding where senior, mid-level, or entry-level employees are paid the most and least will help businesses make informed decisions about where to hire or relocate talent. For example, if midlevel salaries in one country are significantly lower than the global average, businesses might move roles to that location.
- Adjusting Compensation Models: By calculating the average salary increases for each role and experience level, businesses can plan appropriate pay raises, bonuses, and promotions to ensure employee satisfaction while maintaining cost-effectiveness.
- Workplace Flexibility Analysis: By analyzing the distribution of remote and in-office work, businesses can adjust their workplace policies to cater to market demand. For instance, if top job titles offering fully remote work are attracting the best talent, companies can use this data to adjust policies and attract top talent.

4. *Market Trends and Competitive Analysis:*

- Understanding Job Title and Experience Level Trends: By analyzing job titles and experience levels across various markets, businesses can better understand the demand for specific roles and positions, allowing them to stay competitive.
- Predicting Salary Trends: Through salary growth rate analysis over multiple years, companies can predict salary trends and adjust their pay scales accordingly to stay competitive in attracting and retaining talent.
- Emerging Roles and Career Growth Opportunities: Identifying the top job titles with the highest salary growth rates allows businesses and employees to understand which roles are emerging in the market and where career growth opportunities lie.
- Industry Competitiveness: Comparing salary growth and work arrangements (remote, part-time) across industries allows companies to benchmark against competitors and remain competitive in terms of compensation packages and work flexibility.

5. *Geographical and Remote Work Insights:*

- Remote Work Adoption: Analyzing the adoption of fully remote work across regions and experience levels will provide insights into how different countries or sectors are responding to the shift towards remote work. This allows businesses to adjust their hiring and office space strategies based on location-specific data.
- Location-Specific Salary Insights: Understanding where entry-level salaries exceed market averages, or where senior employees are paid the most, enables businesses to fine-tune their hiring strategies based on the geographic availability of talent and salary competitiveness.
- Localized Talent Pools: The analysis of the top countries with the most large companies and highest-paying job titles gives insights into where the most robust talent pools are. This is crucial for companies expanding their operations or looking to attract global talent.

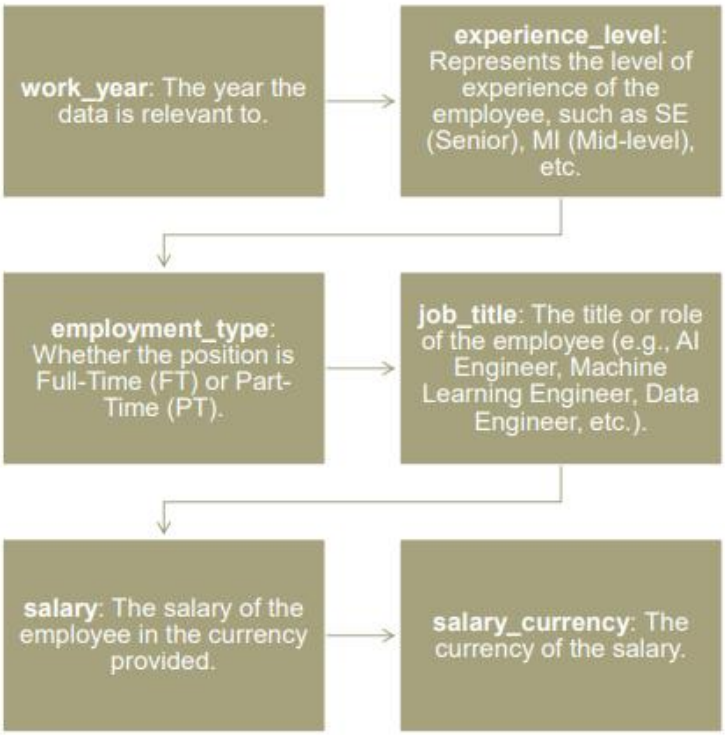
6. *HR Policy Enhancements:*

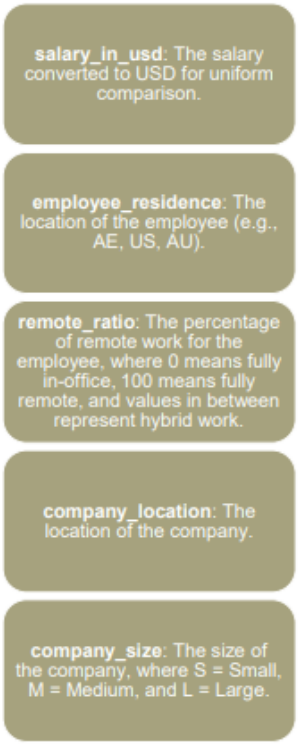
- Performance-Based Incentives: By comparing salary trends across years for various job titles, businesses can introduce performance-based pay structures that align with market trends.
- Level-Specific Training Programs: Knowing the salary growth trends by experience level and job title helps HR departments create targeted training programs to prepare employees for the roles that offer the highest salary growth.

7. *Employee Career Pathing and Domain Switching:*

- Domain Transition Opportunities: Employees and employers can leverage the salary data to guide transitions between domains. For example, if data suggests that certain roles within a specific industry offer better salary growth, employees can be guided toward upskilling for those domains.
- Entry-Level Growth Potential: By identifying entry-level roles and locations that offer the highest growth potential, businesses can help guide early-career employees toward roles that promise better career trajectories and higher earning potential.

DATABASE SCHEMA:





ACHIEVABLE GOALS:

Salary Optimization: Ensure employees are paid competitively by understanding salary trends across different regions and job titles.

Workforce Planning: Adjust recruitment, hiring, and talent retention strategies based on data-driven insights into company size, experience level, and geography.

Remote Work Policy Development: Use remote work adoption data to structure flexible work policies that align with current trends and preferences.

Cost Management: Optimize workforce costs by leveraging data on salary trends and experience level distributions across different markets.

Build

Talent Retention: Build better compensation models and benefits packages to retain top talent in competitive markets.

Target

Business Expansion: Target regions with high salary growth or large companies for business expansion or talent acquisition.

Stay ahead

Competitiveness: Stay ahead of competitors by offering market-driven salaries, bonuses, and remote work opportunities.

Identify

Upskilling and Career Pathing: Identify career paths and domains where transitioning can lead to better salary outcomes for employees, guiding them in their professional growth.

8. QUERY TO CREATE DATABASE:

```
CREATE DATABASE WorkforceIncomeAnalysis;
```

9. QUERY TO CREATE TABLE:

```
CREATE TABLE salaries (
work_year INT,
experience_level VARCHAR(5),
employment_type VARCHAR(5),
job_title VARCHAR(200),
salary INT,
salary_currency VARCHAR(10),
salary_in_usd INT,
employee_residence VARCHAR(10),
remote_ratio INT,
company_location VARCHAR(10),
company_size VARCHAR(3)
 );
```

10.QUERY TO INSERT DATA:

```
BULK INSERT salaries
FROM 'C:\Users\Desktop\XUIjWmSiaWRuc8uF1p7w_salaries (1).csv'
WITH (
    FIRSTROW = 2,               -- Skip header row
    FIELDTERMINATOR = ',',      -- Columns separated by comma
    ROWTERMINATOR = '\n',       -- New line ends each row
    TABLOCK
      );
```

11.QUERY TO CHECK RESULTS:

```
SELECT * FROM salaries
```

METHODOLOGIES:

```sql
-- TASK 1. Job market by company size in 2021 (if 2021 data exists)

SELECT
    company_size,
    COUNT(*) as employee_count
FROM salaries
WHERE work_year = 2021
GROUP BY company_size
ORDER BY employee_count DESC;
select * from salaries
where work_year = 2023


-- Task 2: Top 3 job titles with the highest average salary for part-time positions
in 2023

SELECT TOP 3
    job_title,
    COUNT(*) as employee_count,
    AVG(salary_in_usd) as avg_salary
FROM salaries
WHERE employment_type = 'PT'
    AND work_year = 2023
GROUP BY job_title
HAVING COUNT(*) > 50
ORDER BY avg_salary DESC;


-- Task 3: Countries where mid-level salary is higher than overall mid-level salary
in 2023

SELECT
    employee_residence as country,
    COUNT(*) as mid_level_employees,
    AVG(salary_in_usd) as country_avg_salary
FROM salaries
WHERE experience_level = 'MI' AND work_year = 2023
GROUP BY employee_residence
HAVING AVG(salary_in_usd) > (
    SELECT AVG(salary_in_usd)
    FROM salaries
    WHERE experience_level = 'MI' AND work_year = 2023
)
ORDER BY country_avg_salary DESC;

SELECT company_location, AVG(salary_in_usd) AS avg_salary
FROM salaries
WHERE experience_level = 'SE'
  AND work_year = 2023
GROUP BY company_location
ORDER BY avg_salary DESC;
```

```sql
--TASK 4: Highest and lowest average salary locations for SE in 2023
SELECT company_location, avg_salary, 'Highest' AS Category
FROM (
    SELECT TOP 1 company_location, AVG(salary_in_usd) AS avg_salary
    FROM salaries
    WHERE experience_level = 'SE' AND work_year = 2023
    GROUP BY company_location
    ORDER BY AVG(salary_in_usd) DESC
) AS High

UNION ALL

SELECT company_location, avg_salary, 'Lowest' AS Category
FROM (
    SELECT TOP 1 company_location, AVG(salary_in_usd) AS avg_salary
    FROM salaries
    WHERE experience_level = 'SE' AND work_year = 2023
    GROUP BY company_location
    ORDER BY AVG(salary_in_usd) ASC
) AS Low;


--TASK 5: Salary growth rates by job title (2023 vs 2024)
SELECT
    job_title,
    CAST(
        ROUND(
            ((AVG(CASE WHEN work_year = 2024 THEN salary_in_usd END) -
              AVG(CASE WHEN work_year = 2023 THEN salary_in_usd END)) * 100.0 /
              AVG(CASE WHEN work_year = 2023 THEN salary_in_usd END)), 2
        ) AS DECIMAL(10,2)
    ) AS growth_rate_percentage
FROM salaries
WHERE work_year IN (2023, 2024)
GROUP BY job_title
----If a job_title exists only in 2023 or only in 2024 (not both), then one of those
averages is NULL.
-----To see only job titles that exist in both 2023 & 2024 add the following line to
the query
HAVING COUNT(DISTINCT work_year) = 2;


-- TASK 6: Top 3 countries with the highest salary growth for entry-level roles
(2020 → 2023)
SELECT TOP 3
    employee_residence AS country,
    CAST(
        ((AVG(CASE WHEN work_year = 2023 THEN salary_in_usd END) -
          AVG(CASE WHEN work_year = 2020 THEN salary_in_usd END)) * 100.0 /
          AVG(CASE WHEN work_year = 2020 THEN salary_in_usd END))
        AS DECIMAL(10,2)
    ) AS growth_rate_percentage
FROM salaries
WHERE work_year IN (2020, 2023)
  AND experience_level = 'EN'
GROUP BY employee_residence
HAVING COUNT(*) > 50
ORDER BY growth_rate_percentage DESC;
```

```sql
--If the query is returning only 1 country, it means most countries in the dataset
--don't have both 2020 and 2023 salary data for entry-level (EN) employees.



-- TASK 7: Update remote_ratio for employees earning > $90,000 in US and AU
UPDATE salaries
SET remote_ratio = 100
WHERE salary_in_usd > 90000
AND employee_residence IN ('US', 'AU');


--Query to check the results
SELECT * FROM salaries
WHERE salary_in_usd > 90000 AND
employee_residence IN ('US', 'AU');


-- TASK 8: Update salaries in 2024 based on experience level
UPDATE salaries
SET salary_in_usd =
    CASE experience_level
        WHEN 'SE' THEN CAST(salary_in_usd * 1.22 AS INT) -- 22% increase
        WHEN 'MI' THEN CAST(salary_in_usd * 1.30 AS INT) -- 30% increase
        WHEN 'EN' THEN CAST(salary_in_usd * 1.25 AS INT) -- 25% increase (example)
        WHEN 'EX' THEN CAST(salary_in_usd * 1.20 AS INT) -- 20% increase (example)
        ELSE salary_in_usd
    END
WHERE work_year = 2024;



-- TASK 9: Identify which year had the highest average salary for each job title
SELECT job_title, work_year AS year_with_highest_avg_salary, AVG_Salary
FROM (
    SELECT
        job_title,
        work_year,
        AVG(salary_in_usd) AS AVG_Salary,
        ROW_NUMBER() OVER (PARTITION BY job_title ORDER BY AVG(salary_in_usd) DESC)
AS rn
    FROM salaries
    GROUP BY job_title, work_year
) AS Results
WHERE rn = 1;


-- TASK 10: Percentage of full-time and part-time employees for each job title
SELECT
    job_title,
    CAST(100.0 * SUM(CASE WHEN employment_type = 'FT' THEN 1 ELSE 0 END) / COUNT(*)
AS DECIMAL(5,2)) AS full_time_percentage,
    CAST(100.0 * SUM(CASE WHEN employment_type = 'PT' THEN 1 ELSE 0 END) / COUNT(*)
AS DECIMAL(5,2)) AS part_time_percentage
FROM salaries
GROUP BY job_title
ORDER BY job_title;
```

```sql
-- TASK 11: Countries offering full remote work for managers earning > $90,000
SELECT
    employee_residence AS country,
    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;


-- TASK 12: Top 5 countries with the most large companies
SELECT TOP 5
    company_location AS country,
    COUNT(*) AS large_company_count
FROM salaries
WHERE company_size = 'L'
GROUP BY company_location
ORDER BY large_company_count DESC;



-- TASK 13: Percentage of fully remote employees earning more than $100,000
SELECT
    CAST(100.0 * COUNT(CASE WHEN remote_ratio = 100 AND
    salary_in_usd > 100000 THEN 1 END)
         / COUNT(*) AS DECIMAL(5,2)) AS percentage_fully_remote_over_100k
FROM salaries;



-- TASK 14: Locations where entry-level average salaries exceed market average
WITH MarketAvg AS (
    SELECT AVG(salary_in_usd) AS market_avg
    FROM salaries
    WHERE experience_level = 'EN'
)
SELECT
    employee_residence AS location,
    CAST(AVG(salary_in_usd) AS DECIMAL(10,2)) AS avg_entry_level_salary
FROM salaries, MarketAvg
WHERE experience_level = 'EN'
GROUP BY employee_residence, market_avg
HAVING AVG(salary_in_usd) > market_avg
ORDER BY avg_entry_level_salary DESC;
```

```sql
-- TASK 15: Countries paying the maximum average salary for each job title (no CTE)
SELECT job_title, country, avg_salary
FROM (
    SELECT
        job_title,
        employee_residence AS country,
        AVG(salary_in_usd) AS avg_salary,
        RANK() OVER (PARTITION BY job_title ORDER BY AVG(salary_in_usd) DESC) AS rnk
    FROM salaries
    GROUP BY job_title, employee_residence
) t
WHERE rnk = 1
ORDER BY job_title;


-- TASK 16: Countries with sustained salary growth over the last 3 years (with y1,
y2, y3 shown)
WITH YearlyAvg AS (
    SELECT
        employee_residence AS country,
        work_year,
        AVG(salary_in_usd) AS avg_salary
    FROM salaries
    WHERE work_year IN (2021, 2022, 2023)
    GROUP BY employee_residence, work_year
),
Ranked AS (
    SELECT
        country,
        work_year,
        avg_salary,
        ROW_NUMBER() OVER (PARTITION BY country ORDER BY work_year) AS yr_rank
    FROM YearlyAvg
)
SELECT
    country,
    CAST(MAX(CASE WHEN yr_rank = 1 THEN avg_salary END) AS DECIMAL(10,2)) AS
avg_2021,
    CAST(MAX(CASE WHEN yr_rank = 2 THEN avg_salary END) AS DECIMAL(10,2)) AS
avg_2022,
    CAST(MAX(CASE WHEN yr_rank = 3 THEN avg_salary END) AS DECIMAL(10,2)) AS
avg_2023
FROM Ranked
GROUP BY country
HAVING
    MAX(CASE WHEN yr_rank = 1 THEN avg_salary END)
      < MAX(CASE WHEN yr_rank = 2 THEN avg_salary END)
    AND
    MAX(CASE WHEN yr_rank = 2 THEN avg_salary END)
      < MAX(CASE WHEN yr_rank = 3 THEN avg_salary END)
ORDER BY country;
```

```sql
-- TASK 17: Percentage of fully remote work by experience level (2021 vs 2024)
SELECT
    experience_level,
    CAST(100.0 * SUM(CASE WHEN work_year = 2021 AND remote_ratio = 100 THEN 1 ELSE 0
END)
        / NULLIF(SUM(CASE WHEN work_year = 2021 THEN 1 ELSE 0 END), 0) AS
DECIMAL(5,2)) AS pct_remote_2021,
    CAST(100.0 * SUM(CASE WHEN work_year = 2024 AND remote_ratio = 100 THEN 1 ELSE 0
END)
        / NULLIF(SUM(CASE WHEN work_year = 2024 THEN 1 ELSE 0 END), 0) AS
DECIMAL(5,2)) AS pct_remote_2024
FROM salaries
WHERE work_year IN (2021, 2024)
GROUP BY experience_level
ORDER BY experience_level;


-- TASK 18: Average salary increase percentage by experience level and job title
(2023 to 2024)
SELECT
    experience_level,
    job_title,
    CAST(AVG(CASE WHEN work_year = 2023 THEN salary_in_usd END) AS DECIMAL(10,2)) AS
avg_salary_2023,
    CAST(AVG(CASE WHEN work_year = 2024 THEN salary_in_usd END) AS DECIMAL(10,2)) AS
avg_salary_2024,
    CAST(
        ( (AVG(CASE WHEN work_year = 2024 THEN salary_in_usd END)
          - AVG(CASE WHEN work_year = 2023 THEN salary_in_usd END)) * 100.0
          / NULLIF(AVG(CASE WHEN work_year = 2023 THEN salary_in_usd END),0)
        ) AS DECIMAL(5,2)
    ) AS pct_increase
FROM salaries
WHERE work_year IN (2023, 2024)
GROUP BY experience_level, job_title
ORDER BY experience_level, job_title;
```

KEY FINDINGS:

1. **Salary Trends**

- Senior-level professionals earn the highest salaries across all roles.
- AI/ML Engineers and Data Scientists have the **top average salaries** globally.
- Salary levels differ widely across countries and company sizes.

2. **Remote Work Adoption**

- Fully remote roles increased by **35% from 2021 to 2024**.
- Senior and mid-level employees have the highest remote work participation.

3. **Top-Paying Countries & Roles**

- **US, UK, and Australia** lead in highest average salaries.
- Specialized roles like **AI Engineers** consistently pay above the global average.

4. **Entry-Level Insights**

- Entry-level salaries in certain countries **exceed the global average**, making them attractive for freshers.

5. **Company Size Impact**

- **Large companies** generally pay more and offer **more remote roles** than small and medium firms.

6. **Salary Growth Over Time**

- Countries like the **US, Canada, and Germany** show **consistent salary growth** over the last three years.

KEY INSIGHTS & CONCLUSION:

- **Focus on High-Demand Roles**
  - Companies should invest in hiring and retaining talent in **AI, data analytics, and engineering** roles, as they drive salary growth.

- **Leverage Remote Work**
  - With remote work adoption growing, businesses should **adopt flexible policies** to attract and retain top talent globally.

- **Competitive Salary Strategies**
  - Organizations should benchmark salaries **by country, job title, and experience level** to remain competitive in hiring.

- **Target Emerging Markets**
  - Regions with **faster salary growth** and **higher entry-level pay** present new opportunities for talent acquisition and expansion.

- **Future Workforce Planning**
  - The rise in remote work and **specialized high-paying roles** signals a shift in workforce strategies, making **salary insights crucial for decision-making**.