

# **An Analysis of Career Trajectories - Salary & Experience Across Global Industries**

## Data Cleaning Process

Before we start analyzing the data, we first need to clean it up to ensure the information is accurate and useful. Here's how we approached this step:

### 1. Handling Missing Values

Some of the data in the dataset was missing. Instead of just deleting these rows, we decided to replace the missing values with a placeholder. For example:

- For **salary**, we chose to remove any extreme values (outliers) and replaced them with the **average salary**. This helps smooth out any drastic salary discrepancies that don't fit with the rest of the data.
- For **other missing information** like country, state, or city names, we decided to fill those gaps with the option "Other/Prefer not to answer". This lets us keep the data without making assumptions about what the missing values might be.

### 2. Standardizing Place Names

There were variations in the way some places were written—such as different spellings of the same city or country. We standardized these names so that each place only appears once in a consistent way. For example:

- "NYC" became "New York City"
- "California" was written the same way throughout

This step makes it easier to analyze salary patterns by location without confusion over name discrepancies.

### 3. Fixing Job Titles

We also fixed some inconsistencies in **job titles**. There were a few minor issues like capitalization (e.g., "Engineer" vs. "engineer") or slight spelling errors. We made sure that all the job titles were consistent across the dataset, ensuring that the same roles are categorized correctly.

### 4. Dealing with Outliers in Salaries

Outliers are data points that are very different from most others, like a salary much higher or lower than most people in the same job. In our dataset, we found some extreme outliers in

salary. Instead of removing them entirely, we decided to replace these extreme values with the **average salary** of the same job or category. This made sure that our analysis reflects more typical salaries.

## **Why These Changes Matter**

These changes might seem small, but they make a big difference. By cleaning and standardizing the data:

- We avoid biases that could arise from missing or inconsistent data.
- We ensure that the analysis will be more accurate because we are working with data that reflects the real patterns, not extreme anomalies or errors.
- We can make fair comparisons between different countries, industries, and job titles.

## Loading the Cleaned Data into MySQL (Using GUI)

### 1. Create a Database in MySQL

It seems like you've already created the database (survey\_salary), so we can move on to the next step.

#### Steps:

1. **Open MySQL Workbench.**
2. On the **Home** screen, in the **SCHEMAS** section, right-click on the survey\_salary database and select **Set as Default Schema**.

This ensures that all operations will be done within the survey\_salary database.

### 2. Create a Table (If Needed)

If you haven't already created a table within the survey\_salary database, you can use the following steps to do it manually through MySQL Workbench:

1. In the **SCHEMAS** section, right-click on the survey\_salary database and choose **Create Table**.
2. In the **Create Table** window, you can enter the appropriate **column names** and **data types** (e.g., INT for integers, VARCHAR for text, DECIMAL for salary data).
3. Once done, click **Apply** to create the table.

However, if your CSV has headers and you want to make the process easier, you can directly load the data and MySQL Workbench will try to match the columns for you.

### 3. Load Data Using the GUI (Right-Click Method)

This is the simple method you mentioned, and it involves using MySQL Workbench's **import tool**.

#### Steps:

1. **Right-click on the survey\_salary database** in the SCHEMAS panel.
2. Select **Table Data Import Wizard** from the dropdown.
3. In the **Table Data Import Wizard** window:

- Choose the **CSV file** you want to load (the cleaned dataset).
- Click **Next**.

#### 4. **Map the Columns:**

- MySQL Workbench will try to automatically map the CSV columns to the table columns.
- If it doesn't match exactly, you can manually map them by selecting the correct column from the dropdown list for each one.

#### 5. **Finish:**

- After ensuring the columns match, click **Next** and then **Finish**.
- This will automatically load the data from the CSV file into your table in MySQL.

## Query Data for Dashboard Creation

Once the data is cleaned and uploaded into MySQL, the next step is to run SQL queries to analyze the data and extract meaningful insights. These insights will form the foundation of your Tableau dashboard, where we will visualize various patterns and trends. Below are the SQL queries that will help in generating those insights.

### 1. Average Salary by Industry and Gender

**Objective:** Compare the average salary within each industry, split by gender.

sql

Copy

```
SELECT  
  
    Industry,  
  
    Gender,  
  
    AVG(Annual_Salary) AS Average_Salary  
  
FROM  
  
    salary_survey_2021  
  
GROUP BY  
  
    Industry, Gender  
  
ORDER BY  
  
    Industry, Gender;
```

- **Explanation:** This query helps compare the average salary between different industries and genders. It identifies any gender-based salary gaps within each industry by calculating the average salary for each gender in each industry.

### 2. Total Salary Compensation by Job Title

**Objective:** Find the total monetary compensation (base salary + additional monetary compensation) for each job title.

sql

Copy

SELECT

Job\_Title,

SUM(Annual\_Salary + Additional\_Monetary\_Compensation) AS Total\_Compensation

FROM

salary\_survey\_2021

GROUP BY

Job\_Title

ORDER BY

Total\_Compensation DESC;

- **Explanation:** This query calculates the total compensation for each job title by adding the Annual\_Salary and Additional\_Monetary\_Compensation (like bonuses and commissions). It helps identify which job titles offer the highest total compensation.

### 3. Salary Distribution by Education Level

**Objective:** Find the salary distribution (average salary, minimum, and maximum) for different education levels.

sql

Copy

SELECT

Highest\_Education,

AVG(Annual\_Salary) AS Average\_Salary,

```

MIN(Annual_Salary) AS Minimum_Salary,

MAX(Annual_Salary) AS Maximum_Salary

FROM

salary_survey_2021

GROUP BY

Highest_Education

ORDER BY

Highest_Education;

```

- **Explanation:** This query helps analyze the correlation between education and salary by calculating the **average**, **minimum**, and **maximum** salary for each education level.

#### 4. Number of Employees by Industry and Years of Experience

**Objective:** Determine how many employees are in each industry, broken down by years of professional experience.

sql

Copy

```

SELECT

Industry,

'Total_Years_of_Professional_Experience',

COUNT(*) AS Number_of_Employees

FROM

salary_survey_2021

GROUP BY

Industry, 'Total_Years_of_Professional_Experience'

ORDER BY

```



Industry, `Total\_Years\_of\_Professional\_Experience`;

- **Explanation:** This query groups the employees by Industry and Total\_Years\_of\_Professional\_Experience, and counts the number of employees in each group. This helps us identify industries with more experienced professionals.

## 5. Median Salary by Age Range and Gender

**Objective:** Calculate the median salary within different age ranges and genders.

sql

Copy

```
SELECT
    Age_Range,
    Gender,
    AVG(Annual_Salary) AS Median_Salary
FROM (
    SELECT
        Age_Range,
        Gender,
        Annual_Salary,
        @row_num := @row_num + 1 AS row_num,
        @total_rows := @row_num
    FROM
        salary_survey_2021,
        (SELECT @row_num := 0) AS init
    ORDER BY
        Age_Range, Gender, Annual_Salary
```

) AS sorted\_salaries

WHERE

row\_num IN (FLOOR(@total\_rows / 2), FLOOR(@total\_rows / 2) + 1)

GROUP BY

Age\_Range, Gender;

- **Explanation:** This query calculates the **median salary** by first ordering the data by Age\_Range, Gender, and Annual\_Salary. It then finds the middle salary values, which represent the median salary within each combination of Age\_Range and Gender.

## 6. Job Titles with the Highest Salary in Each Country

**Objective:** Find the highest-paying job titles in each country.

sql

Copy

SELECT

Country,

Job\_Title,

MAX(Annual\_Salary) AS Highest\_Salary

FROM

salary\_survey\_2021

GROUP BY

Country, Job\_Title

ORDER BY

Country, Highest\_Salary DESC;

- **Explanation:** This query helps us identify which job titles offer the highest salaries in each country by grouping the data by Country and Job\_Title.

## 7. Average Salary by City and Industry

**Objective:** Calculate the average salary for each combination of city and industry.

sql

Copy

SELECT

City,

Industry,

AVG(Annual\_Salary) AS Average\_Salary

FROM

salary\_survey\_2021

GROUP BY

City, Industry

ORDER BY

City, Average\_Salary DESC;

- **Explanation:** This query calculates the **average salary** within each combination of City and Industry. This can help identify cities that offer higher salaries for particular industries.

## 8. Percentage of Employees with Additional Monetary Compensation by Gender

**Objective:** Find the percentage of employees who receive additional monetary compensation (e.g., bonuses) by gender.

sql

Copy

SELECT

```

Gender,

ROUND(

    (SUM(CASE WHEN Additional_Monetary_Compensation > 0 THEN 1 ELSE 0 END)

    / COUNT(*)) * 100, 2

) AS Percentage_With_Additional_Compensation

FROM

    salary_survey_2021

GROUP BY

    Gender;

```

- **Explanation:** This query calculates the percentage of employees who receive **additional compensation** (e.g., bonuses, stock options) for each Gender. The query counts how many employees have non-zero Additional\_Monetary\_Compensation and divides it by the total number of employees to get the percentage.

## 9. Total Compensation by Job Title and Years of Experience

**Objective:** Determine the total compensation (salary + additional compensation) for each job title based on years of professional experience.

sql

Copy

```

SELECT

    Job_Title,

    `Total_Years_of_Professional_Experience`,

    SUM(Annual_Salary + Additional_Monetary_Compensation) AS Total_Compensation

FROM

    salary_survey_2021

```

GROUP BY

Job\_Title, `Total\_Years\_of\_Professional\_Experience`

ORDER BY

Total\_Compensation DESC;

- **Explanation:** This query calculates the **total compensation** (salary plus additional compensation) for each job title based on years of experience. This will show trends in compensation based on experience levels for each job title.

## 10. Average Salary by Industry, Gender, and Education Level

**Objective:** Understand how salary varies by industry, gender, and education level.

sql

Copy

SELECT

Industry,

Gender,

Highest\_Education,

AVG(Annual\_Salary) AS Average\_Salary

FROM

salary\_survey\_2021

GROUP BY

Industry, Gender, Highest\_Education

ORDER BY

Industry, Gender, Highest\_Education;

- **Explanation:** This query provides a **comprehensive view** of how salary varies by Industry, Gender, and Highest\_Education. This helps us understand how education level influences salary across different industries and genders.

## **Export Query Results**

After executing the queries, you can export the results from MySQL Workbench or using MySQL commands to CSV files. These CSV files will then be imported into Tableau for data visualization.

## **Create a Tableau Dashboard**

Once the data is exported into CSV format, we will import these files into Tableau. There, we will create visualizations based on the insights we have gathered from the queries. Tableau allows us to create interactive and engaging dashboards that display key trends such as salary differences by gender, education, job title, and location.

## Dashboard Insights: Visualizing Salary Trends and Disparities

### Gender Distribution of Additional Monetary Compensation

- Visualization: Pie Chart
- Key Insights:
  - Women receive the largest share of additional monetary compensation, at 76.72%.
  - Men follow with 20.48%.
  - Non-binary individuals represent only 1.10% of the total compensation.
- Analysis:

The data highlights a significant gender disparity in the allocation of additional monetary compensation, where women appear to benefit the most, but non-binary individuals are significantly underrepresented. This could point to either underreporting or underrepresentation, suggesting that further investigation into pay equity and bonus distribution is needed, especially with regards to non-binary employees. Understanding potential biases and industry-specific practices that influence this distribution will be crucial in driving more equitable compensation practices.

### Gender Pay Gap by Education Level: Salary Trends Across Genders

- Visualization: Bar Graph or Line Chart
- Key Insights:
  - Men consistently earn higher salaries than women and non-binary individuals across all experience levels and education categories.
  - However, the gap narrows slightly with increasing experience.
- Analysis:

Despite higher educational attainment (e.g., Master's, Ph.D., and Professional Degrees) leading to higher salaries, men continue to earn more than women and non-binary individuals at each education level. The gender pay gap is particularly evident among those with advanced degrees, indicating that gender bias persists, even in highly

educated fields. Non-binary individuals generally earn less across all educational levels, reflecting the need for targeted interventions to close these gaps.

### Impact of Education on Salary Trends

- Visualization: Heat Map or Scatter Plot
- Key Insights:
  - Higher education levels (Master's, Ph.D., and Professional Degrees) lead to significantly higher salaries.
  - Individuals with Ph.D. or Master's degrees tend to earn significantly higher salaries than those with only Some College or High School education.
  - Men tend to earn more than women and non-binary individuals, even at these higher education levels.
- Analysis:

The heat map clearly shows that higher education directly correlates with higher salaries, but it also highlights gender disparities at each level. Despite earning higher degrees, men still dominate the higher salary ranges, while women and non-binary individuals earn comparatively less. This trend suggests that gender inequality in salary persists across education levels, especially in fields requiring specialized skills or higher education.

### Experience vs. Salary Growth

- Visualization: Line Graph or Box Plot
- Key Insights:
  - As experience increases, the salary gap between men and women widens, especially in the 10+ years experience bracket.
  - Non-binary individuals show a notable disparity, earning significantly less in the 1-5 years and 10+ years categories. However, they earn more in the 6-10 years experience range compared to women.
- Analysis:

The experience vs. salary relationship shows a clear positive correlation—the more years of experience, the higher the salary. However, the gender salary gap is more



pronounced as experience increases. For men, salary growth appears to be more substantial as experience grows, particularly in the 11-20 years range. Non-binary employees, while showing a slight improvement in earnings with experience, still trail behind men and women. This trend indicates potential career progression disparities based on gender identity.

#### Top 10 Industries by Average Salary

- Visualization: Bar Chart or Table
- Key Insights:
  - The highest-paying industries are predominantly in technical and specialized fields such as Restoration, Oil & Gas, and Scientific Research.
  - Other high-paying industries include Technology, Healthcare, and Pharmaceuticals.
- Analysis:

The industries at the top of the salary list reflect sectors where high levels of expertise and innovation are valued, such as technology and oil & gas. These fields generally offer above-average compensation due to the specialized knowledge required. However, it's also important to note that while some medical and real estate industries offer competitive salaries, they typically fall behind in comparison to the more technical and resource-driven fields.

#### Industry-Specific Salary Trends

- Visualization: Bar Graph or Box Plot
- Key Insights:
  - Technology and Finance industries show the most significant gender pay gap, with men consistently earning more than women across all roles.
- Analysis:

Industry disparities in salary are especially evident in technology and finance, where men typically earn significantly more than women. This raises concerns about gender bias within these industries, where traditionally, men have dominated leadership and

high-paying roles. Exploring these disparities can shed light on potential barriers to career progression for women and non-binary employees.

#### Annual Salary Growth by Experience and Gender

- Visualization: Line Graph
- Key Insights:
  - Men consistently have higher salaries across all experience levels.
  - The salary growth for men is more significant in the 11-20 years experience range, while non-binary individuals appear to earn lower salaries in the 1-5 years and 10+ years categories.
- Analysis:

The data shows that gender plays a significant role in salary growth over the years. Men see greater earnings growth with increased experience, while non-binary individuals earn less, even as their experience increases. This highlights the need for further analysis into how gender identity impacts long-term career earnings, especially in terms of both salary and additional compensation.

# Conclusion

This project provided a detailed analysis of salary data, shedding light on key trends across global industries and the influence of factors like gender, education, and experience. By leveraging data cleaning, MySQL queries, and advanced visualizations, we gained valuable insights into how compensation varies across various demographics and industries.

Key findings include:

1. **Gender Pay Gaps:** The data clearly reveals that men, on average, earn higher salaries than women and non-binary individuals across all levels of experience and education. This disparity is especially pronounced in certain industries like technology and finance, highlighting a potential need for more equitable pay practices.
2. **Impact of Education:** Higher levels of education (e.g., Master's, Ph.D.) are associated with higher salaries, but men tend to earn more at every education level. Non-binary individuals generally earn less across all education levels compared to both men and women.
3. **Experience and Salary Growth:** As professional experience increases, salaries generally grow, particularly in higher experience brackets. However, the growth is more significant for men, and non-binary individuals continue to earn lower salaries as experience rises.
4. **Industry-Specific Trends:** Certain industries, particularly in technical, research-based, and specialized fields like Oil & Gas, Restoration, and Scientific Research, offer the highest salaries. In contrast, industries like Real Estate and Medical Technology provide competitive compensation but not as high as those in resource-based or highly specialized sectors.
5. **Dashboard Insights:** The interactive Tableau dashboard provided a comprehensive, visual understanding of these trends. It highlighted gender-based pay gaps, the correlation between education and salary, and how industries differ in compensation levels. With filters for gender, age range, country, and education level, the dashboard allowed for a deeper dive into the data, making it easier to identify and explore patterns.

6. Further Investigations Needed: The disparities revealed in this analysis, especially concerning gender and non-binary employees, call for further investigation into potential biases in salary distribution. There is also a need to explore ways to address the gender pay gap, particularly in high-paying industries.

In conclusion, this project not only highlighted the key factors influencing salaries but also emphasized the need for continued work in improving pay equity, particularly for underrepresented groups. The data-driven insights presented in the dashboard can serve as a foundation for further research, policy discussions, and industry-wide improvements in compensation practices.