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

This project involves analyzing a salary survey dataset to explore career trajectories across global industries. The dataset provides insights into various factors that affect salary, such as age, job title, industry, experience, education, and compensation.

Dataset Attributes:

1. **Age Range**: The age group of the individual.
2. **Industry**: The sector in which the individual works.
3. **Job Title**: The official position held by the individual.
4. **Clarification of Job Title**: Additional details about the job title.
5. **Annual Salary**: The individual's annual salary or earnings.
6. **Additional Monetary Compensation**: Extra earnings beyond the base salary (e.g., bonuses, commissions).
7. **Currency**: The currency in which salary and compensation are reported.
8. **Other Currency**: A custom currency used when it does not match the standard options.
9. **Income Clarification**: Additional context regarding the income structure (e.g., commission-based).
10. **Country**: The country where the individual works.
11. **State**: The state or province of employment (if applicable).
12. **City**: The city where the individual works.
13. **Years of Professional Experience Overall**: The total number of years the individual has worked professionally.
14. **Years of Professional Experience in Field**: The number of years the individual has worked in their specific field.
15. **Highest Level of Education Completed**: The highest degree or educational level attained.
16. **Gender**: The gender identity of the individual.

Step-by-Step Instructions:

Step 1: Data Cleaning and Preprocessing

1. **Handle Missing Values**:
    - For numerical columns (e.g., Annual Salary), fill missing values with the mean or median, or remove rows with missing data if necessary.
    - For categorical columns (e.g., Industry, Job Title), replace missing values with "Unknown" or remove the rows.
2. **Standardize Data Types**:
    - Ensure all columns have appropriate data types (e.g., numeric for salaries, string for text values like Industry).
3. **Handle Inconsistent Values**:
    - Check for spelling variations (e.g., "Researcher" vs. "researcher") and ensure consistent capitalization for categorical values (e.g., all lowercase or proper case for job titles).
4. **Check for Outliers**:
    - Identify outliers in numerical columns (e.g., Annual Salary). Decide whether to remove or adjust them based on domain knowledge.
5. **Final Cleaned Dataset**:
    - Once cleaned, save the dataset in a CSV format for further processing.

Step 2: Incorporating the Cleaned Data into MySQL

1. **Create a Database in MySQL**:
    - Open MySQL Workbench and create a new database for the salary survey data.
2. **Create a Table**:
    - Design a table to hold the cleaned dataset, ensuring appropriate column names and data types.
3. **Upload the Data**:

- Use the LOAD DATA command or MySQL Workbench import tools to import the cleaned dataset into the MySQL table.

Step 3: Query Data for Dashboard Creation

1. **Execute the following queries**:

a. Average Salary by Industry and Gender

- Compare the average salary within each industry, split by gender. This helps identify potential salary discrepancies based on gender within industries.

b. Total Salary Compensation by Job Title

- Find the total monetary compensation (base salary + additional monetary compensation) for each job title. This can show which roles have the highest overall compensation.

c. Salary Distribution by Education Level

- Find the salary distribution (average salary, minimum, and maximum) for different education levels. This helps analyze the correlation between education and salary.

d. Number of Employees by Industry and Years of Experience

- Determine how many employees are in each industry, broken down by years of professional experience. This can show if certain industries employ more experienced professionals.

e. Median Salary by Age Range and Gender

- Calculate the median salary within different age ranges and genders. This can provide insights into salary trends across different age groups and gender.

f. Job Titles with the Highest Salary in Each Country

- Find the highest-paying job titles in each country. This can help understand salary trends across different countries and highlight high-paying positions.

g. Average Salary by City and Industry

- Calculate the average salary for each combination of city and industry. This shows which cities offer higher salaries within each industry.

h. Percentage of Employees with Additional Monetary Compensation by Gender

- Find the percentage of employees within each gender who receive additional monetary compensation, such as bonuses or stock options.

i. Total Compensation by Job Title and Years of Experience

- Determine the total compensation (salary + additional compensation) for each job title based on years of professional experience. This can help highlight compensation trends based on experience levels within specific job titles.

j. Average Salary by Industry, Gender, and Education Level

- Understand how salary varies by industry, gender, and education level. This query can provide a comprehensive view of how multiple factors influence salary.

2. **Export Query Results**:
    - Export the results of each query to CSV files. Each query's result should be saved as a separate CSV file.

Step 4: Create Dashboard in Tableau

1. **Import Data into Tableau**:
    - Import the CSV files containing the query results into Tableau.
2. **Create a Dashboard Story**:
    - Using Tableau, create an interactive dashboard that visualizes key insights from the data, such as salary distributions, gender comparisons, and the impact of education on salary.

3. **Write Insights**:
    - After completing the dashboard, write a detailed summary of the insights derived from the analysis. Include observations on salary trends, gender disparities, education impacts, and other meaningful patterns that emerge from the data.

Deliverables:

1. Cleaned dataset (CSV/Excel file).
2. MySQL database with the salaries table populated.
3. 10 SQL queries and their results exported to CSV files.
4. Tableau dashboard story with visualizations based on the exported query results.
5. A written summary of insights from the analysis, including documentation that outlines the project objective, dataset description, steps involved, and insights derived from the analysis.
6. A presentation file that contains the details of the project (10 slides).

Evaluation Rubric for M1:

Evaluation Criteria	Marks
1. Data Cleaning and Preparation	15
2. Data Import into MySQL	10
3. SQL Queries	20
4. Data Export to CSV files	10
5. Dashboard Creation in Tableau	15
6. Documentation	10
7. Presentation and Reporting	20
Total Marks	**100**