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

Attributes of the given Salary Survey dataset:

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

Step-by-step Instructions:

Step 1: Data Cleaning and Preprocessing

Clean the dataset by handling missing values, standardizing data types, handling inconsistent values, and checking for outliers.

1. **Handle Missing Values:**
    - Identify columns with missing data. Decide how to handle them:
        - For numerical columns (like Annual Salary), you can either fill the missing values with the mean or median, or you can remove rows with missing values if necessary.
        - For categorical columns (like Industry, Job Title), you may replace missing values with a placeholder like "Unknown" or remove the rows.
2. **Standardize Data Types:**
    - Ensure all columns have appropriate data types.
3. **Handle Inconsistent Values:**
    - Review each column for inconsistencies, such as:
        - Variations in spelling for job titles (e.g., "Researcher" vs. "researcher").
        - Different formats for location (e.g., USA vs. United States).
        - Ensure consistent capitalization (e.g., all lowercase or proper case for categorical values).
4. **Check for Outliers:**
    - Identify outliers in numerical columns like Annual Salary.
    - For any outliers, decide whether to remove them or adjust them based on domain knowledge.
5. **Final Cleaned Dataset:**
    - Once the dataset is cleaned, save the cleaned version so that it can be further incorporated into MySQL.

Step 2: Incorporating the Cleaned Data into MySQL

Upload the cleaned dataset into a MySQL database and create a single table.

1. **Create a Database in MySQL:**
    - Open MySQL Workbench.
    - Create a new database.
2. **Create a Table:**
    - Design a table to hold the dataset with appropriate column names and data types.
3. **Upload the Data:**
    - Import the cleaned data from the CSV or Excel file into the MySQL table using the LOAD DATA command or the MySQL Workbench import tools.

Step 3: Query the data required for dashboard creation.

1. **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.
2. **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.
3. **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.
4. **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.

5. **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.
6. **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.
7. **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.
8. **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.
9. **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.
10. **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.

Step 4: Create Tables in Excel

Export the data from MySQL and create separate sheets in Excel for each query result.

1. **Export Data from MySQL:**
    - After running each SQL query, export the result to a CSV file.
2. **Import into Excel:**
    - Open Excel and create a new workbook.
    - For each query result create a new sheet.

- Import the respective CSV file into each sheet.
- You should have one sheet for each query result.

Step 5: Dashboard Creation

Create pivot tables and a dashboard to visualize the results.

1. **Create Pivot Tables:**
   - In Excel, create pivot tables from the data in each sheet to summarize and analyze the key metrics.
   - Use appropriate fields for rows, columns, values, and filters to get meaningful insights from the data.
2. **Create a Dashboard:**
   - Use various Excel charts (bar charts, pie charts, line charts, etc.) to visualize the following insights:
     - Salary distribution by industry.
     - Gender distribution across different job titles.
     - Top-paying job titles.
   - Arrange the charts and tables on a dedicated dashboard sheet in an easy-to-read format.
3. **Final Touches:**
   - Ensure the dashboard is clear and visually appealing, with proper labels, titles, and legends.
   - Add any filters or slicers for interactive analysis.

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. Excel workbook with separate sheets for each query result.
5. A final dashboard sheet that visualizes key insights.
6. A written summary of insights from the analysis, including documentation that outlines the project objective, dataset description, steps involved, screenshots of results obtained at each phase, insights derived from the analysis, and a conclusion for the project.
7. A presentation file that contains the details of the project (10 slides).

Evaluation Rubric for M1:

Evaluation Criteria	Marks
1. Data Cleaning and Preparation	10
2. Data Import into MySQL	10
3. SQL Queries	15
4. Data Export and Table Setup in Excel	10
5. Pivot Tables and Pivot Charts	10
6. Dashboard Creation	15
7. Documentation	10
8. Presentation and Reporting	20
Total Marks	**100**