ANALYSIS OF

CAREER

TRAJECTORY-SALARY & EXPERIENCE

* INTRODUCTION:

In today's rapidly evolving job market, understanding career trajectories across industries is vital for both individuals and organizations. This project explores salary trends and professional experiences across various global industries, with a focus on the impact of factors such as age, job title, gender, education, experience, and geographic location. By analyzing a comprehensive salary survey dataset, this aims to identify meaningful patterns and insights that can help professionals make informed career decisions and organizations develop strategies to ensure fair compensation and support career development.

The dataset used in this analysis includes diverse attributes such as the individual’s age range, industry, job title, salary, additional monetary compensation, education level, and years of professional experience. By analyzing these attributes, we gain a deeper understanding of how various elements influence salary growth and professional progression across different sectors and regions. Through the use of SQL queries, data preprocessing, and visualization tools like Tableau, the goal of this project is to provide actionable insights that can address salary disparities.

**PROJECT OBJECTIVE:**

The objective of this project is to analyze a salary survey dataset to gain insights into the factors that influence career trajectories and salary growth across various global industries. The key goals of this project are:

1. **Exploring Salary Trends**: To understand how salary varies across different industries, job titles, experience levels, and education. This includes analyzing discrepancies in compensation across gender and age.
2. **Identifying Key Influencers of Salary**: To identify which factors such as professional experience, job title, industry, gender, and education have the most significant impact on salary.
3. **Highlighting Global Disparities**: By examining salary data across multiple countries, the project aims to provide a comprehensive view of salary trends, focusing on differences based on location and job market conditions.
4. **Facilitating Career Development**: To help individuals in the workforce better understand how their experience, education, and other factors can influence salary progression.
5. **Supporting Organizational Decision-Making**: To assist organizations in understanding current compensation trends and ensuring that their salary structures are competitive and fair across different sectors and regions.
6. Through a combination of data preprocessing, SQL queries, and Tableau visualizations, this project provides a comprehensive analysis that helps both professionals and organizations make informed decisions based on data-driven insights into career trajectories and salary distribution.

DATASET DESCRIPTION:

The dataset used in this project is a comprehensive salary survey that captures various attributes influencing career trajectories across global industries. The data includes information about individuals professional experience, compensation details, education level, and demographic factors, offering valuable insights into how these elements impact salaries across different sectors and regions. Below is a detailed description of the key attributes included in the dataset:

**Age Range**: This attribute represents the age group of the individual, categorized into various ranges to analyze salary trends based on different stages of a person’s career.

**Industry**: This indicates the sector in which the individual works, such as technology, healthcare, finance, or manufacturing.

**Job Title**: This attribute represents the official position or role held by the individual within their organization. It is used to identify how salaries differ across various job titles, from entry-level positions to senior leadership roles.

**Clarification of Job Title**: Additional details or a brief description of the individual’s job title, providing more context to understand the role.

**Annual Salary**: The annual base salary of the individual before any additional monetary compensation. It is the primary indicator used for salary analysis.

**Additional Monetary Compensation**: This includes any additional earnings beyond the base salary, such as bonuses, or stock options.

**Currency**: The currency in which the salary and compensation are reported, allowing for comparison across countries with different currencies.

**Other Currency**: In cases where the currency doesn’t match standard options, this attribute contains the custom currency used for reporting.

**Income Clarification**: This provides additional context regarding how the individual's income is structured, such as commission-based earnings, bonuses, or other forms of monetary compensation.

**Country**: The country in which the individual works. This allows for country-level comparisons and helps identify geographic variations in salary.

**State**: For countries with multiple states or provinces, this attribute indicates the specific state or province where the individual is employed, enabling location-based salary analysis.

**City**: The specific city where the individual works. City-level data is essential to understand salary differences across urban centers and rural areas within the same country.

**Years of Professional Experience Overall**: The total number of years the individual has worked in their professional career, providing insights into how experience correlates with salary growth.

**Years of Professional Experience in Field**: This indicates the number of years the individual has worked specifically within their field of expertise, differentiating between general professional experience and field-specific experience.

**Highest Level of Education Completed**: The highest degree or educational qualification attained by the individual, such as high school, bachelor's degree, master's degree, or doctorate. It helps analyze the correlation between education and salary.

**Gender**: The gender identity of the individual, enabling an analysis of salary disparities between different genders.

STEPS INVOLVED IN ANALYSIS:

**Data cleaning** is a fundamental step to ensure the dataset is accurate, consistent, and ready for analysis. In this project, the dataset underwent various cleaning procedures to handle missing values, standardize text data, remove duplicates, and detect and handle outliers. Below is an outline of the specific data cleaning steps:

* To ensure consistency across the dataset, we standardized **text data** such as job titles, industries, and gender by following proper capitalization rules. For example, job titles like "Software engineer" and "software engineer" were corrected to "Software Engineer" to maintain uniformity.
* This step helps avoid inconsistencies that could affect data analysis and querying, particularly when grouping or aggregating data based on these attributes.
* **Blank Fields**: For fields with missing or blank values, such as industry or job title, we replaced them with the value **"Unknown"**. This ensured that no data was left incomplete, allowing for smoother analysis without losing rows of data.
* **Numerical Fields**: For missing numerical values (such as annual salary or years of experience), we replaced them with the **mean** or **median** of the respective column, based on which method was most appropriate for maintaining data integrity.
* During the cleaning process, we identified and removed nearly **200 duplicated rows** in the dataset. Duplicate entries could distort analysis, especially when calculating averages, distributions, or other aggregations. Removing duplicates ensured that each row represented unique data and improved the accuracy of the analysis.
* We applied statistical techniques to detect potential **outliers** in numerical columns, particularly in the **Annual Salary** and **Years of Experience** columns.
* After identifying the outliers, we assessed whether these values were valid or erroneous. If they were deemed invalid or extreme without reasonable justification they were **removed** from the dataset to ensure they did not skew the analysis.
* After completing these steps, the dataset was thoroughly reviewed to ensure consistency and accuracy. The cleaned dataset was then saved in **CSV format** for use in further analysis and visualization.
* By applying these data cleaning techniques, we ensured that the dataset was free from inconsistencies, missing values, duplicates, and outliers, making it ready for deeper analysis and insightful decision-making.

**Loading data into Mysql** is a process of loading data into a MySQL database using Python. Before loading the data, create or use an existing MySQL database set up and create a table that matches the structure of your data.

* This table includes columns for attributes such as age range, industry, job title, clarification job title, annual salary, additional compensation, education level, experience and gender. This table should have columns that align with the fields in your pandas DataFrame. After loading the data into the DataFrame, it is important to check null values. This ensures that the dataset is complete and accurate before loading it into the MySQL database. **Converting DataFrame Rows into Text and Writing to a File.** The process involves iterating through each row of the DataFrame, converting the row into a tuple, and then writing the formatted content to a text file.
* For each row, the data is converted into a tuple. A tuple is a data structure that holds an ordered collection of items, making it a suitable format for storing the values of each row from the DataFrame. Each tuple contains all the column values for that specific row.
* Once all rows have been converted into tuples and stored in the list, we open a new text file in write mode. The file is opened with UTF-8 encoding to handle any special characters that may be present in the data. For each tuple in the list, it is converted into a string representation and written to the text file. Atlast, copy the text data into the table and perform operation.

***Sql query:***

**1. Compare the Average Salary within Each Industry, Split by Gender : This** query groups the data by gender and industry, then calculates the average salary for each group. The results are ordered by gender.

**2. Total Salary Compensation by Job Title :** It groups the data by job title and calculates the total compensation for each title. The result is ordered in descending order to show the highest-compensated roles first.

**3. Salary Distribution by Education Level :** It groups the data by education level and calculates the average, minimum, and maximum salary for each education level.

**4. Number of Employees by Industry and Years of Experience :** The data is grouped by industry and years of experience, and the number of employees in each category is counted.

**5. Median Salary by Age Range and Gender** : The query calculates the median salary by partitioning the data by age range and gender, and then calculating the middle salary.

**6. Job Titles with the Highest Salary in Each Country** : The query finds the highest salary in each country and selects the corresponding job title. It uses a subquery to identify the maximum salary for each country.

**7. Average Salary by City and Industry :** It groups the data by city and industry, calculating the average salary in each city-industry combination. The results are ordered by salary in descending order.

**8. Percentage of Employees with Additional Monetary Compensation by Gender** : It calculates the percentage of employees who receive additional salary for each gender by using a case statement to count those with additional salary greater than zero.

**9. Total Compensation by Job Title and Years of Experience** : It groups the data by years of experience and job title, calculating the total compensation (salary + additional salary) for each group.

**10. Average Salary by Industry, Gender, and Education Level** : It groups the data by industry, gender, and education level, calculating the average salary for each combination. The results are ordered by the average salary in descending order.

* TABLEAU:

The process began with exporting the SQL query output and saving it in CSV format. The main dataset and the SQL output were then linked by establishing a connection, allowing for seamless integration. Using Tableau, various charts and graphs were created to visualize the data, helping to uncover key trends and insights. Dashboards were designed with Key Performance Indicators (KPIs) to monitor and analyze important metrics. The gathered insights were thoroughly examined, and the findings were presented through storytelling in Tableau, where the data was showcased in a narrative format to provide a clear and impactful understanding of the results. This process effectively combined data visualization and analysis, enabling informed decision-making based on the insights extracted from the data.

* STORY & KEY INSIGHTS:
* Bar graph illustrates that the Computing/Tech industry tends to employ more experienced professionals compared to other industries.
* Map helps to visualize that different level of educated people are from which country. It shows that PhD holders are predominantly located in Russia, while individuals with college degrees are scattered across various states.
* Area chart visualize that which education earns high annual salry based on their age. People with a college degree in the 18-24 age group earn more than 1000 million annually.
* Donut chart visualize that among the different gender how the age group people are performing in earning annual salary The donut chart shows that:

1. Men: The 45-54 age group earns the most, while those aged 65+ earn the least.
2. Women: The 25-44 age group is the highest earning, while those aged 65+ again earn the least.
3. Non-Binary: The 25-34 age group has the highest earnings, with the lowest earnings seen in those aged 55-64.
4. Prefer Not to Answer: The 35-44 age group has the highest earnings.

* The bar chart highlights that the USA ranks highest in providing additional compensation, such as stock options and bonuses, while New Zealand offers minimal supplementary financial rewards.
* Pie chart illustrates that the highest-earning age group is between 25 and 34 years old, reflecting their peak earning potential.
* Software Engineers tend to earn the highest compensation as their experience level increases.
* Histogram visualize the salary distribution by age groups. It reveals that the maximum annual salary earned by various age groups typically falls within the range of 89 million to 105 million dollars.
* Woman earns potential salary discrepancies based on gender within industries.
* US Dollars is the currency used by most of the people among different states.
* CONCLUSION:

The process began with exporting the SQL query results and saving them in CSV format. Both the main data and the SQL output were linked by establishing a connection, allowing for effective integration and analysis. Once the data was connected, various charts and graphs were created in Tableau to visualize the information, helping to uncover patterns and trends. Dashboards were built around Key Performance Indicators (KPIs) to monitor and analyze critical metrics. These visualizations and KPIs provided valuable insights into the data, enabling deeper analysis. Finally, Tableau’s storytelling feature was used to present the gathered insights in a compelling narrative format, guiding stakeholders through the findings in a clear and structured way. This process of linking data, visualizing it through graphs and dashboards, and storytelling ultimately facilitated the extraction of meaningful insights that can inform strategic decisions.