Task: Data Analysis and Insights Generation for Employee Dataset

1. **Data Exploration and Cleaning**:

* Load the dataset and perform basic data exploration to understand the data structure.
* Identify and handle any missing values, duplicates, or inconsistencies.
* Convert any necessary columns to appropriate data types (e.g., salary to numeric, work\_year to integer).

1. **Exploratory Data Analysis (EDA)**:

* **Salary Analysis**:
* Calculate average and median salary both in salary\_currency and salary\_in\_usd.
* Compare salaries across different experience\_levels and job\_titles.
* **Remote Work Trends**:
* Analyze the distribution of remote\_ratio to identify the prevalence of remote work.
* Identify any trends in remote work by job\_title and company\_location.
* **Experience and Employment Insights**:
* Create a breakdown of employment\_type by experience\_level to see which employment types are most common at each level.
* Generate counts or percentages of different experience\_levels across various company\_sizes.

1. **Data Visualization**:

* Visualize the salary distribution across different job roles and experience levels using box plots or violin plots.
* Create a bar chart to show the average remote\_ratio by job\_title.
* Use a heatmap to show the correlation between remote\_ratio, salary\_in\_usd, and experience\_level.
* Plot a bar chart for company\_size distribution and how it correlates with employment\_type and experience\_level.

1. **Insights and Reporting**:

* Summarize key insights found from the analysis, such as:
* Which job titles offer the highest remote work ratios.
* The most common employment\_type and experience\_level combinations.
* Any interesting trends in salary based on experience\_level, job\_title, and company\_size.
* Prepare a report with visualizations and written insights, discussing patterns, trends, and any recommendations.

1. **Storing Results in Oracle Database**:

* Design an Oracle database table schema to store the cleaned and analyzed data.
* Insert the analyzed data (e.g., average salary by job\_title, remote work trends) into the Oracle database.
* Verify the data insertion by querying the database and comparing sample records to ensure consistency.

Expected Deliverables:

* **Python Script**: Script for data cleaning, analysis, and Oracle database storage.
* **Visualizations**: Graphs and charts summarizing insights.
* **Database Table**: Oracle table with stored results.
* **Report**: Documented insights and observations with visualizations.