Problem Statement

1. Background

The competitive landscape of the tech industry demands that companies offer competitive salaries to attract and retain top talent. An international HR consultancy aims to leverage data-driven insights to analyze salary trends globally. By examining key factors such as job role, experience level, remote work, and company size, the consultancy seeks to provide clients with actionable recommendations to enhance their hiring and compensation strategies.

To achieve this goal, the team will explore a dataset containing salary information from thousands of employees worldwide. This dataset will be analyzed at three levels:

1. **Exploratory Data Analysis (EDA)** – Understanding the structure and key statistics.
2. **Visualization & Trend Analysis** – Identifying patterns and salary trends.
3. **Predictive Modeling** – Building models to predict salaries and optimize future data collection.

2. Challenges & Analytical Objectives

Level 1: Exploratory Data Analysis (EDA)

- Assess the dataset's structure by identifying the total number of records and the range of years covered.
- Determine the average salaries of Data Scientists and Data Engineers and compare which role earns more.
- Identify the number of full-time employees based in the US who work 100% remotely.

Level 2: Visualization & Trend Analysis

- Create a bar chart displaying the top 5 job titles with the highest average salary.
- Compare the average salaries for employees working remotely at different levels (100%, 50%, 0%) to identify emerging trends.
- Visualize salary distribution across different company sizes (small, medium, large) to determine which company size offers the highest average salary.

Level 3: Predictive Modeling & Advanced Analysis

- Analyze how country, experience level, and remote work ratio impact salaries for Data Analysts, Data Scientists, and Machine Learning Engineers.
- Develop a predictive model to estimate an employee’s salary based on experience level, company location, and remote ratio, identifying the strongest predictors.
- Expand the model by incorporating additional features (e.g., company size and employment type) and evaluate its performance.
- Propose new features that could enhance the accuracy of future salary predictions.

3. Expected Outcomes

- **Actionable Insights:** Clear understanding of how different factors influence salaries, enabling clients to adjust compensation strategies.

- **Data-Driven Decision Making:** HR professionals can use the findings to attract and retain talent more effectively.
- **Predictive Capabilities:** A robust model that provides salary estimates, assisting businesses in setting competitive salaries.
- **Enhanced Future Data Collection:** Recommendations on additional features or data points that can further improve salary predictions.

By systematically addressing these challenges, the consultancy will empower clients with precise, data-backed salary insights to maintain a competitive edge in talent acquisition and retention.