**Case Study: Employment Agreement Data Analysis**

#### **Scenario:**

Gorai Technology Solutions is developing an Employment Agreement Analytics Tool that analyzes and provides insights from employment agreements of various companies. As a part of this project, you are required to work on a dataset containing information extracted from a sample of . The dataset includes the following fields:

* Employee\_ID
* Name
* Department
* Position
* Agreement\_Date
* Contract\_Term (in months)
* Base\_Salary
* Bonus
* Benefits
* Cloud\_Service\_Used (e.g., AWS, Azure, GCP)
* Tools\_Used (e.g., Python, R, SQL)

#### **Tasks:**

1. **Data Cleaning and Preprocessing:**
   * Load the dataset and perform data cleaning to handle any missing or inconsistent data.
   * Standardize the format of dates and numerical values.
2. **Exploratory Data Analysis (EDA):**
   * Provide a summary of the data, including descriptive statistics for numerical columns.
   * Visualize the distribution of key variables such as Base\_Salary, Contract\_Term, and Bonus.
   * Identify any trends or patterns in the use of cloud services and tools.
3. **Technical Analysis:**
   * Use SQL queries to extract insights from the data. For example, find the average Base\_Salary for different Positions and Departments.
   * Write Python scripts to perform data analysis and visualization. Use libraries like Pandas, Matplotlib, and Seaborn.
4. **Cloud Services Integration:**
   * Demonstrate how you would store and process this data using a cloud service (AWS, Azure, or GCP).
   * Explain the benefits of using cloud services for this type of analysis.
   * Show how to set up a simple data pipeline using a cloud service for automated data processing and reporting.
5. **Advanced Analysis (Bonus):**
   * Use machine learning techniques to predict the likelihood of an employee leaving the company based on the agreement terms.
   * Build a simple model using tools like Scikit-Learn or TensorFlow.

#### **Submission Requirements:**

* A Jupyter notebook containing your data cleaning, EDA, and technical analysis.
* SQL queries used for data extraction.
* Python scripts for data analysis and visualization.
* A short report (2-3 pages) summarizing your findings and explaining the cloud service integration.
* (Optional) Code and results for the machine learning model.

#### **Evaluation Criteria:**

1. **Technical Knowledge:**
   * Understanding of data science concepts and techniques.
   * Proficiency in SQL and Python.
   * Knowledge of cloud services and their application in data analysis.
2. **Problem-Solving Skills:**
   * Ability to clean and preprocess data effectively.
   * Capability to derive insights from data and visualize them appropriately.
3. **Practical Application:**
   * Demonstration of how cloud services can be leveraged for data storage, processing, and analysis.
   * Application of machine learning for predictive analysis (if included).
4. **Communication:**
   * Clarity and coherence in the report summarizing your findings.
   * Proper documentation and presentation of your code and results.

By completing this case study, you will demonstrate your technical skills and problem-solving abilities, which are crucial for a Data Scientist role at Gorai Technology Solutions. We look forward to your innovative solutions and insights.