It is designed to assess the candidate's competencies across several key domains in data science and business intelligence. It aims to test not only the technical skills in data transformation and machine learning but also his proficiency in natural language processing (NLP) and business intelligence (BI) tools. The tasks outlined demand a harmonious blend of analytical thinking, problem-solving acumen, and the capability to effectively communicate complex data insights. This case study particularly focuses on the candidate ability to utilize these skills in making strategic business decisions, ensuring he can transform intricate data into actionable, understandable formats for various users. The case study is divided into two parts:

1. **Part 1: Tailored Remarks and Instructions**: refining the work done in the first case study, with a focus on improving the accuracy of the machine learning model and adapt the analysis to provide more insightful data interpretations.
2. **Part 2 - New Business Problem Task for All Candidates**: This part presents a new challenge that all candidates will tackle. It involves a practical application of data integration skills in a business context. This task tests the candidate's ability to adapt and apply their skills to a real-world scenario, ensuring their contributions align with the company's operational needs.

# Part 1: Tailored Remarks and Instructions

## Objective

Refine the data cleaning process in Power BI and the Jupiter Notebook to improve the accuracy of the machine learning model, and adapt the analysis to provide more insightful data interpretations.

## Tasks

1. **Data Cleaning and Transformation:**
   * Modify the data extraction process to include a single 'Department' column for both clients and employees, categorizing as 'Home Care', 'Mental Health', or 'No Department assigned'.
   * Create a 'Zone' column in the Clients table, identifying client zones based on specified group IDs. Address scenarios with no or multiple zone assignments appropriately.
   * Develop tables for 'Departments' and 'Zones', and establish relevant relationships with the Clients and Employees tables.
2. **Analysis Adaptation in Power BI and Notebook:**
   * Redo the Notebook Analysis based on the above transformation process, since the Notebook given for the first round of interview was not giving any insight.
   * Redesign your analysis to focus on the new data structure, especially examining visits by client zone and department.
   * Ensure your analysis captures nuances and trends relevant to these categorizations.
3. **Notebook Presentation Refinement:**
   * Assume your notebook will be reviewed by non-technical executives, including the CEO.
   * Simplify explanations of complex machine learning models, making them accessible to a non-technical audience.

# Part 2: Opening a New Clinic - Strategic Business Analysis

## Objective

Use company data, along with external datasets, to analyze activities and identify suitable areas for opening new clinics. Develop insights and tools to assist company leadership in strategic decision-making.

## Tasks

1. **Data-Driven Location Analysis:**
   * Integrate external data (e.g., population, community age demographics) with company data to analyze activities in current service areas like Sudbury, North Bay, Elliot Lake, and Parry Sound as well as other Areas (Potential Areas).
   * Utilize any appropriate data analytics tool or programming language for this analysis.
2. **Insight Generation in Text Format:**
   * Focus on creating detailed insights from the data, presented in text format, ideally using Power BI. The goal is to narrate compelling stories about the company's activities, clients, and visit patterns.
   * Develop multiple tables to encapsulate these narratives. For instance, one table could summarize activities by community or city. Each row in this table might include:
     + **City/Community Column:** Name of the city or community.
     + **Summary Column:** A concise narrative encapsulating key data points.
       - Example Entry: “Azilda has 43 active clients, with 80% aged 70 or above. This community is serviced by 11 active personal support workers. Over the past three years, Azilda recorded an average of 123 visits per day in 2020, increasing to 145 in 2021 and 167 in 2022.”
   * This approach should be replicated across different communities or cities, providing a rich, data-driven story of the company's operations and client engagement in each area.
3. **Chatbot Development for Data Interaction:**
   * Use the generated text data to create a chatbot.
   * This chatbot should enable company leadership to interactively explore data, aiding in strategic discussions about new clinic locations.
4. **Utilizing Service Code Data:**
   * Incorporate the 'Visit Service Code Id' data to refine your model for identifying optimal clinic locations.
   * Focus on Nursing services (identified by keywords like Nursing, RN, RPN) while considering other service codes as relevant.

Deliverables

* A Power BI file showcasing your analysis.
* At least one Jupyter Notebook, along with a PDF or Word report generated from the notebook.
* Any additional tools or outputs you find relevant for this case study.