# Data Analyst Case Study Assignment

## **Context**

You are provided with two datasets from a community health program:

1. **Consultation Data (Excel)** – patient consultations (physical & online), doctors, prescriptions, timestamps.
2. **Vitals Data (Excel)** – patient vitals (blood pressure, glucose, pulse, etc.).

Your task is to analyze, clean, and extract insights that can guide doctors, program managers, CHWs, and policymakers in improving healthcare delivery.

## **Tasks**

### **1. Data Cleaning & Preparation**

* Identify and handle missing values & inconsistencies.
* Standardize timestamps into a consistent format.
* Merge Consultation & Vitals data on uhid and/or focusedGroupId.
* Convert vitals (BP, glucose, pulse) into numeric values for analysis.

### **2. Exploratory Analysis**

* Count unique patients (UHIDs).
* Analyze consultation distribution: **Physical vs Online**.
* Identify top doctors by consultation volume.
* Vitals analysis:
  + Distribution of BP, glucose, pulse.
  + % patients at risk: **Hypertension (≥140/90 mmHg)**, **Diabetes (glucose ≥126 mg/dL)**, abnormal pulse.

### **3. Stakeholder Insights**

* **Doctors:** Which vitals show highest risks?
* **CHWs:** Who contributed most consultations?
* **Program Managers:** Which facilities/districts are busiest?
* **Policymakers:** What gaps exist in vitals capture & online adoption?

(any other insight you can think of)

### **4. KPI Development & Visualization**

Define and calculate **4 additional KPIs**, ensuring they are aligned with industry priorities:

* **2 KPIs for the MedTech industry** (focused on vitals monitoring, device-enabled screenings, early detection).
* **2 KPIs for the Pharma industry** (focused on prescriptions, treatment patterns, therapy demand).

For each KPI: 1. Clearly define the KPI (in your own words). 2. Explain why it is relevant to the chosen industry. 3. Create a visualization that best represents it. 4. Provide a short commentary (2–3 lines) on its potential value to stakeholders.

### **5. Advanced Analysis**

#### **5A. Patient Risk Stratification (Predictive Insight)**

* Using vitals data (BP, glucose, pulse), classify patients into **Low, Moderate, and High risk** categories.
* Calculate % of patients in each risk category.
* Compare risk distribution across facilities or CHWs.
* *(Optional stretch goal)*: Build a simple predictive model (logistic regression / decision tree) to identify which factors most strongly predict “high risk.”

#### **5B. Consultation–Outcome Linkage Analysis**

* For patients with multiple consultations, compare their first vs most recent vitals.
* Quantify improvement, deterioration, or no change.
* Produce % of patients showing improvement vs deterioration.
* Compare doctor- or facility-level performance in driving better outcomes.
* *(Optional stretch goal)*: Suggest KPIs that could track treatment effectiveness over time.

## **Expected Deliverables**

* Cleaned & merged dataset (CSV/Excel).
* Analytical report with insights.
* Visual dashboards (Excel/Power BI/Tableau/Python).
* KPI definitions, calculations, and visualizations.
* Advanced analysis outputs (risk stratification, consultation–outcome linkage).
* 1-page executive summary for leadership.