**PROJECT TITLE: Fake Drug Detection via Pharmacy Sales Analytics**

**Case Study: Counterfeit Drug Crisis in Nigeria**

**Background**

In Nigeria, the problem of counterfeit and substandard medicines is deeply rooted and dangerous.  
According to the **National Agency for Food and Drug Administration and Control (NAFDAC)**, an estimated **15-30% of medicines in circulation** are counterfeit, especially in rural and semi-urban markets.

A tragic example occurred in **1990** when **109 children died** after consuming paracetamol syrup contaminated with diethylene glycol — a toxic industrial chemical.  
Even decades later, counterfeit antimalarials, antibiotics, and anti-hypertensive drugs remain widespread.

**Example**

In 2015, an investigative report in **Lagos and Kano states** uncovered fake brands of **Coartem (a key anti-malaria drug)** being sold in open markets.  
Symptoms of the counterfeit issue included:

* Suspiciously low pricing compared to verified pharmacies.
* Availability of "new brands" that were not registered by NAFDAC.
* Sudden spikes in the volume of sales for certain brands during malaria season.

**Key Observations:**

* Rural pharmacies and informal drug sellers (market stalls) were hotspots for fake drug distribution.
* Suppliers using mobile vans were harder to trace and often distributed unregistered brands.
* Expiry dates on fake drugs were often forged or missing.

**Problem Statement**

Counterfeit and substandard drugs pose a serious health threat across African markets, particularly in rural and semi-urban areas where regulatory oversight is weaker.  
Traditional manual inspections are slow and ineffective at scale.  
This project aims to **analyze pharmacy sales patterns** to **detect anomalies** that could indicate **distribution of fake drugs**.

**Project Goal**

* Develop an analytics system that flags pharmacies, drugs, and suppliers with suspicious sales behavior.
* Detect anomalies such as:
  + Unusual price drops.
  + Surge in unknown brands.
  + Sudden spikes in drug quantity sales.
  + Suspicious supplier activity.
* Generate insights to assist regulatory authorities and pharmacy chains in mitigating the fake drug crisis.

**Data Dictionary**

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| Date | Date | Date when the sale occurred. |
| Pharmacy | String | Name of the pharmacy where the drug was sold. |
| Location | String | Whether the pharmacy is Urban or Rural. |
| Drug | String | Name of the drug (e.g., Amoxicillin, Artemisinin). |
| Brand | String | Brand name of the drug (may include unknown/fake brands). |
| Price | Float | Price per unit sold (in Naira ₦). |
| Quantity | Integer | Number of units sold in the transaction. |
| Supplier\_Name | String | Supplier or distributor that provided the drug batch. |
| Batch\_Number | String | Unique batch identifier assigned by supplier. |
| Expiry\_Date | Date | Expiration date of the drug batch. |
| Sales\_Channel | String | Channel where the sale was made (In-Store, Mobile Van, Market Stall). |
| Drug\_Form | String | Form of the drug (Tablet, Syrup, Injection, etc.). |

**Project Workflow**

1. **Data Exploration**
   * Analyze price distribution, sales volume, and brand diversity.
   * Investigate differences between Urban vs Rural pharmacies.
2. **Data Preprocessing**
   * Handle missing values.
   * Engineer useful features (e.g., Price per Quantity, Days to Expiry).
   * Encode categorical variables where necessary.
3. **Anomaly Detection**
   * Detect unusual pricing or sales spikes.
   * Flag unfamiliar or rare brands.
   * Spot unreliable suppliers based on anomaly frequency.
4. **Risk Scoring**
   * Assign a **risk score** to pharmacies, suppliers, and brands based on suspicious activity patterns.
5. **Visualization & Reporting**
   * Build graphs and dashboards to summarize findings.
   * Report high-risk entities and key insights.

**Evaluation Criteria**

* Insight Clarity: Are your findings actionable and clear?
* Visualizations: Do your graphs support your insights effectively?
* Anomaly Detection Accuracy: Are you catching reasonable and interpretable anomalies?
* Risk Prioritization: Can you rank pharmacies/suppliers by threat level?
* Presentation: Is the final report easy to follow for non-technical stakeholders?

**Research Questions to be answered**

* Which pharmacies sell drugs with significantly lower prices than average?
* Are there brands with disproportionately high sales but unknown origins?
* Are there specific suppliers often associated with anomalies?
* How do sales patterns differ between urban and rural settings?
* Are expiry dates of drugs being manipulated or suspiciously close?