NUPAT AI FELLOWSHIP - STAGE TWO

CASE STUDY ASSESSMENT

Your task is to analyze two datasets and deliver your findings in a Jupyter Notebook. We value clean code, clear reasoning, and actionable insights as much as the final model performance.

Timeline: Please submit your completed notebook and Github link on or before **11:59 PM, Friday 19TH DECEMBER, 2025.**

Datasets Provided:

- trades.csv

    o   timestamp: UNIX timestamp of the trade

    o   user_id: Unique identifier for the user

    o   pair: The trading pair (e.g., BTCNGN, ETHKES, USDTNGN)

    o   side: 'buy' or 'sell'

    o   price: Price in the quote currency (e.g., NGN for BTCNGN)

    o   volume: Amount of the base currency traded (e.g., amount of BTC)

- user_activity.csv

    o   timestamp: UNIX timestamp of the activity

    o   user_id: Unique identifier for the user

    o   activity_type: 'deposit', 'withdrawal'

    o   asset: The crypto asset involved (e.g., BTC, ETH, USDT)

Part 1: Exploratory Data Analysis & Market Insights

Your first task is to understand user behavior on our platform. Provide analysis and visualizations to answer the following:

1. **Market Dynamics:** What are the top 3 most traded pairs by total USD volume (use 1500 as the rate for all currency conversion to USD)?

2. **Volatility Analysis:** Calculate and plot the 7-day rolling average of daily price volatility for the **BTCNGN** pair.

3. **User Behavior:** Is there a specific day of the week or hour of the day when deposits are highest? Visualize your findings to show these peak times.

What we're testing: Your proficiency with **Pandas** for time-series manipulation, data aggregation, and your ability to extract basic business insights.

Part 2: Fraud Detection Model

One of our biggest priorities is platform security. **A common fraudulent pattern involves depositing funds, making minimal trades, and then quickly withdrawing everything.** Your task is to build a model to identify users exhibiting this behavior.

1. **Feature Engineering:** Using both datasets, create features for each user_id. Examples could include:

   o   Deposit/withdrawal frequency and ratio.

   o   Time between first deposit and first withdrawal.

   o   Trading volume vs. total deposited amount.

   o   Number of unique assets traded.

2. **Target Labeling:** Since there is no "is_fraud" column, you must define it. Create a rule-based label for users you considered suspicious based on the pattern described above. **Clearly state your logic for this rule.**

3. **Model Building:** Build a classification model (e.g., Logistic Regression, Random Forest, XGBoost) to predict your "suspicious" label.

4. **Evaluation:** How would you evaluate this model's performance? In the context of fraud detection, is **precision** or **recall** more important? Justify your choice and present the relevant score for your model.

What we're testing: Feature engineering, problem framing (creating a target variable), practical machine learning application, and a deep understanding of model evaluation metrics in a business context.

Part 3: Strategic Recommendation

Based on all your analysis, answer the following hypothetical question in a markdown cell.

Question: "The product team wants to launch a 'Low-Volume Trader' marketing campaign in Kenya. Using the data, how would you define the target audience for this campaign? Describe 2-3 data points you would use to create this user segment."

What we're testing: Your ability to translate data insights into a concrete, data-driven business strategy.

Submission Guidelines

- Submit a link to your Github repository containing a single Jupyter Notebook (.ipynb)

- Ensure your code is clean, commented, and can be run from top to bottom.

- Use markdown cells to explain your methodology, assumptions, and conclusions for each part.

- Visualizations should be clearly labeled and easy to interpret.