Iqana’s Quant / ML Engineer Take-Home Challenge

A multi-asset machine learning meta-strategy

You are provided with a real multi-asset baseline trading strategy (long/cash) plus market and on-chain data. Your task is to **design**, **implement**, and **evaluate** an **ML-driven meta-strategy** that improves the baseline across assets under realistic backtesting constraints (timing, costs, leakage prevention).

This challenge mirrors the day-to-day workflow of a quant researcher:

data reconstruction → feature engineering → label design → walk-forward modeling → strategy evaluation → interpretation and failure analysis.

Provided data

1. *Trade_log.csv*
    - Matrix of **timestamps x assets**
    - Values: 0 = cash, 1 = long
2. *Prices.csv*
    - Matrix of **timestamps x assets**
    - Asset prices used to compute returns and portfolio performance
3. *Glassnode.csv*
    - Timestamped **Bitcoin on-chain features** (single set of features aligned to time).Cumulative returns by regime
    - Can be used as global regime/context features for all assets

Objective

Build a machine-learning filter / meta-policy that improves the baseline strategy, e.g.:

$$signal_{ml}[t,a] = signal_{base}[t,a] \times \mathbf{1}[\hat{p}(t,a) > \tau]$$

Where $\hat{p}(t,a)$ is your model's predicted probability that taking the baseline long at time t on asset a is profitable after costs, and τ is a threshold you select.

Your goal is to show improvement in risk-adjusted performance robustness.

What you need to do

1. **Data & Baseline**
    a. Align signals and prices (no lookahead)
    b. Sanitize data
    c. Evaluate baseline strategy

2. **Feature engineering**
    a. Build time-t features only

3. **Labels**
    a. Justify horizon and cost treatment

4. **Modeling**
    a. Train a model
    b. Use walk-forward validation
    c. Prevent leakage and imbalances

5. **Evaluation**
    a. Compare baseline vs improved strategy

Deliverables

1. Jupyter notebook (clean, reproducible) covering methodology + results

2. Short PDF report summarizing

3. Clear explanation of expectations, limitations and when the approach is expected to fail and next steps

Additional Guidance

You may assume the data comes from a process with latent regime dynamics that influence the behavior of the features and returns. The regimes may be persistent to a some degree. There is no need to know what the features are and there is no labeled regime variable — part of your challenge is to uncover it. You may use any public libraries and tools, but document clearly.

Why this matter?

At Iqana, we're not just optimizing signals — we're building the next generation of AI-driven strategy systems in crypto markets. This challenge reflects real problems we work on every day: uncovering hidden structure, building smarter signals, and pushing strategies toward robustness and performance.

We're excited to see your thinking — there are no perfect answers, only strong reasoning. Think like a quant, build like an engineer, and challenge assumptions.

If you enjoy solving this, you'll love working here.

iqana