Data Scientist Coding Challenge

Overview

This challenge assesses your ability to explore and analyse real-world energy data, generate insights, and prototype solutions to optimise flexibility in household energy use.

You'll work with simulated data from UK households and flexibility markets and optionally integrate live price data from the Octopus Agile API.

We're not expecting energy sector expertise - just demonstrate strong reasoning, modeling skills, and clear communication.

What We're Looking For

- Exploratory data analysis (EDA)
- Thoughtful problem framing
- Insightful visualisations and summaries
- A modeling approach (regression, classification, forecasting, etc.)
- Sensible assumptions, trade-offs, and communication

This is an offline task. Complete it at your own pace, push to a GitHub repository, and share the link with **Komodo-Rob** (rob@komodo.energy). We'll schedule a review session to discuss your work.

Provided Data Sources

Energy Prices

- A sample JSON file (agile_price_example.json) is available for convenience.
- Additional energy price data can also be retrieved from the Octopus Agile API

Household Energy Usage Data (household_usage.csv)

- Simulated half-hourly energy usage for three household types over 24 hours.

Flexibility Market Data (flexibility_market.json)

- Defines two flexibility events where energy shifting/exporting is valuable.

Your Task

Part 1: Exploratory Analysis & Insight Generation

Perform an EDA and answer questions like:

- How do different household types use energy throughout the day?
- When is usage typically highest and lowest?
- Which households have the most potential to shift load away from expensive periods?
- How would household energy costs change under different pricing schemes (e.g. Agile vs flat rate)?
- What might a flexibility provider learn from this data?

Use plots, tables, and narrative to communicate your insights.

Part 2: Model or Prototype (Choose One)

Pick one of the following directions (or suggest your own):

📈 Option 1: Forecasting

- Forecast short-term household energy demand (e.g., next 1–3 hours).
- Consider using basic time series models or ML techniques.

⚖️ Option 2: Cost Optimisation

- Build a simple model that recommends **when a household should shift energy use** to save money.
- Can be rule-based, linear program, or prototype ML.

🧩 Option 3: Classification

- Identify household types (or usage behavior patterns) using unsupervised or supervised learning.
- E.g. classify households into “flexible” vs “inflexible” based on usage patterns.

🔍 **Option 4: Flexibility Opportunity Scoring**

- Create a scoring system to rank flexibility events for each household (based on timing, cost savings, ability to shift, etc.).

What to Include

A **Jupyter Notebook** (or equivalent)

- Clear structure (load data → explore → model → insights)
- Well-commented code and clean outputs
- Inline markdown for explanation and context

A short **README**

- Problem framing
- Setup instructions
- Key assumptions
- How you'd take this further (e.g. productionisation, validation, next models) - especially if you had a larger dataset

(Optional) A dashboard or visualization summary (e.g. using Streamlit or Plotly Dash)

Submission Process

1. Complete the challenge in a GitHub repo
2. Share the link with us: **Komodo-Rob** (rob@komodo.energy).
3. We'll arrange a **review session** to discuss your approach and thinking

Final Notes

- You can use any tools you're comfortable with: pandas, scikit-learn, Prophet, PyTorch, etc. and any public libraries
- Please limit AI use to how you would use it day-to-day (e.g., debugging, refactoring) – not to generate the whole solution. We will expect you to explain your approach and reasoning in the review session.
- If you're unfamiliar with the energy flexibility domain, just document any assumptions - we'll discuss them in the review!