🔥Heat Modelling: Technical Test

📚 Introduction and Problem Statement

At Octopus Energy, our mission is to decarbonise heat by accelerating the transition to electric heating, primarily through the deployment of heat pumps in homes across the UK. Electrifying heat is critical for reaching net-zero targets, and heat pumps are one of the most promising technologies to achieve this at scale.

Heat pumps are highly efficient, but their electricity demand can vary significantly depending on a range of factors. As we scale their deployment, it's essential that we understand and can accurately forecast this demand - both to ensure a smooth experience for customers and to help us buy the right amount of energy at the right times.

In this exercise, you'll work with real-world data to explore and model heat pump electricity usage. Your insights will help us better understand how heat pumps behave in the real world and how we might forecast their demand more effectively.

🎯 Objectives

Your goal is to analyse heat pump electricity consumption data and develop a model that can forecast demand. This task is designed to assess your ability to:

- Work with time series and structured data
- Identify patterns and trends in electricity usage
- Engineer meaningful features from raw data
- Build and evaluate a predictive model
- Communicate technical findings clearly and concisely

We’re interested in both your technical approach and your ability to draw actionable insights from the data.

📂 Data Description

For this exercise, you will be working with anonymised data collected from a sample of residential properties with heat pumps installed. The dataset has been designed to reflect real-world usage and contains the following components:

1. Heat Pump Electricity Consumption

- Granularity: Half-hourly
- Units: kWh
- Period: 2024-01-01 -> 2024-12-31
- Fields:
    - interval_start (TIMESTAMP UTC)
    - heat_pump_id (INT)
    - consumption_kwh (FLOAT)

2. External Temperature

- Granularity: Half-hourly
- Units: Celsius
- Period: 2024-01-01 -> 2024-12-31
- Fields:
    - interval_start (TIMESTAMP UTC)
    - location_id (INT)
    - external_temperature_c (FLOAT)

3. Property Metadata

- One row per heat pump
- Fields:
    - heat_pump_id (INT)
    - property_type (e.g. detached, semi-detached, flat) (STRING)
    - floor_area_m2 (INT)
    - number_of_bedrooms (INT)
    - decade_built (INT)
    - Insulation_level (STRING)
    - has_hot_water_storage (BOOLEAN)
    - location_id (INT)

🧪 Task

We're interested in both your technical approach and your ability to reason about the results. Please structure your work as you see fit, but the following steps provide a useful guide:

🔍 1. Explore the Data

- Perform exploratory data analysis (EDA) to understand usage patterns.
- How does heat pump consumption vary by temperature, time of day, and across different homes?
- Are there any data quality issues (e.g. missing values, outliers)?

🧠 2. Feature Engineering

- Create features that could help explain or predict consumption.

- This might include:
    - Degree days (e.g. how cold it is relative to a base temperature)
    - Time-based features (hour of day, day of week, etc.)
    - Property-level characteristics

📈 3. Build a Forecasting Model

- Train a model to predict half-hourly heat pump consumption.
- Focus on a single home or a group of homes - your choice, but explain your reasoning.
- You can use any modelling approach you think is suitable (e.g. linear regression, gradient boosting, time series models).

📊 4. Evaluate and Interpret

- Evaluate model performance using appropriate metrics (e.g. MAE, RMSE).
- How well does your model perform?
- What are the most important features?
- How would you improve it with more data?

📣 5. Communicate Your Findings

- Summarise your approach, key insights, and recommendations.
- If you had to explain this to a non-technical stakeholder (e.g. product manager or operations lead), what would you say?

✅ Deliverables

Please submit the following:

- A Jupyter notebook (or similar) that contains your code, analysis, and commentary.

- A short summary write-up (max 1 page or a few slides) that highlights your approach, findings, and any recommendations.

We're not looking for a polished report - clarity of thought, structured reasoning, and clear communication are what matter most.

You're welcome to use any tools or libraries you're comfortable with, but please make sure your work is clearly explained and reproducible.

📬 Submission & Time Guidance

We recommend spending 2–3 hours on this exercise. We're looking for clear reasoning, thoughtful analysis, and effective communication - not a perfect or production-ready solution.

Please submit your work as:

- A single file or folder containing your code and analysis (e.g., Jupyter notebook, Python script)

- A brief summary write-up (PDF or slides) highlighting your approach and key findings