

Strategic Energy Investment Prioritization

Challenge

Using energy meter data, weather conditions, and building metadata from the OSU Energy Research Data Hub, you are invited to develop an AI/M-enabled prototype that addresses a real capital-planning question:

If you had limited funding to invest in campus energy improvements, which buildings should be prioritized first - and how can data justify that decision?

Campus buildings experience broadly similar external conditions - weather, seasonal demand, and academic cycles- yet respond very differently in terms of energy consumption. In this challenge, you will treat **weather as an input** and **energy use as an output**, and analyze how buildings deviate from expected energy behavior under comparable conditions.

Rather than being given predefined “good” or “bad” buildings, **your task is to let the data itself reveal performance**. By modeling expected energy response to weather and comparing it to observed consumption across the full building portfolio, you will identify buildings that stand out, either as unusually efficient or persistently inefficient, and therefore merit deeper investigation or capital consideration.

You are expected to analyze **all available buildings** over approximately **60 days of data** using quantitative, scalable methods (not manual inspection). This challenge is not about diagnosing specific equipment faults or prescribing exact retrofits. Instead, it focuses on **screening and prioritization**: determining which buildings rise to the top when evaluated consistently, transparently, and comparatively.

Participants are strongly encouraged to use AI, machine learning, and data analytics techniques such as regression models, clustering, time-series analysis, or other statistical or ML approaches to:

- model expected energy response to weather,
- quantify deviations between expected and observed energy use, and
- rank buildings based on consistent, explainable performance signals.

OSU AI Hackathon Info Packet

Clear reasoning, methodological rigor, and interpretability are valued more than model complexity. The goal is to demonstrate how AI can support smarter, evidence-based investment decisions, especially when time, data, and budgets are constrained.

Imagine This!

Imagine being responsible for investing millions of dollars across a large campus—but instead of guessing, reacting to complaints, or relying on anecdote, you have a **data-backed shortlist** of buildings where investment is most likely to pay off.

You can see which buildings consume more energy than expected given their size, vintage, and weather exposure. You can identify facilities whose energy use is unusually sensitive to temperature or shows consistently elevated baseline load—signals that may indicate inefficiency or opportunity.

Now imagine scaling that insight across an entire campus, a university system, or a city.

Your work in this challenge explores how energy data, when analyzed as a **portfolio rather than in isolation**, can support smarter capital planning, equity-aware investment, and long-term sustainability goals. By allowing patterns, contrasts, and anomalies to emerge directly from the data, you build a framework that helps decision-makers focus limited dollars where they can have the greatest impact.

Let's explore how data can turn limited funding into maximum impact.

Expected Outcomes

Participants will deliver:

1. Decision-Support Artifact (Required)

Teams must build an **end-user artifact** that a non-technical decision-maker could realistically use. Acceptable formats include:

- An interactive web application or dashboard
- A lightweight decision tool (e.g., ranking interface or scenario simulator)
- A backend scoring service with a clear decision workflow

The artifact must:

OSU AI Hackathon Info Packet

- Produce a **ranked shortlist of buildings** (e.g., top 5–10) for potential energy investment
- Allow users to **inspect and understand why** buildings are prioritized
- Clearly separate **data inputs, signals, assumptions, and outputs**

2. AI / Model-Driven Reasoning Layer

Teams must implement **at least one quantitative or AI-assisted method** to infer expected energy behavior and support portfolio-level prioritization. Acceptable approaches include (but are not limited to):

- Regression or machine-learning models relating energy use to weather and building features
- Time-series, anomaly-detection, or representation-learning approaches
- Clustering or peer-group benchmarking to establish reference behavior

Models should be applied **consistently across all buildings** and support scalable reasoning.

Model sophistication is less important than **appropriateness, clarity, and robustness**.

3. Investment Signals (Required, Flexible)

Teams must identify and justify **multiple data-derived signals** that inform prioritization decisions. Signals may include, but are not limited to:

- Normalized energy intensity (e.g., energy per square foot)
- Sensitivity of energy use to weather
- Baseline or non-weather-dependent load
- Variability or instability in energy consumption
- Persistent excess energy use relative to peers

Teams must:

- Explain **why** each signal is relevant for capital prioritization
- Describe how signals are **combined, weighted, or traded off**
- Demonstrate how rankings change under **alternative assumptions or signal weightings**

OSU AI Hackathon Info Packet

There is no required or “correct” set of signals.

4. Explainability & Uncertainty Communication (Required)

Solutions must go beyond ranking to demonstrate **how a human would interpret and trust the output**. Teams must:

- Explain *why* each shortlisted building rises to the top using data-derived evidence
- Communicate **uncertainty or confidence** in rankings (e.g., confidence tiers, scenario sensitivity, stability across methods)
- Clearly state limitations given the short data window and missing contextual variables
- Visual, textual, or interactive explanations are all acceptable.

5. Action Framing & Reflection

Teams must frame results in terms of **actionable next steps**, not diagnoses. Deliverables should include:

- A brief description of how the output could guide real decisions (e.g., audits, deeper investigation, phased investment)
- Reflection on how the framework could evolve with longer time horizons, additional campuses, or richer data (e.g., occupancy, schedules, controls)

Data Requirements

Solutions must incorporate:

- Energy meter data from `c3ntinel_meter_data_parquet` (~60 days)
- Weather data from `weather_data_parquet`
- Building metadata including gross square footage, number of floors, construction year (vintage), and geographic location

Evaluation Rubric

Teams will be evaluated using the following criteria:

1. Data Processing & Scale (25%)

- Analysis includes **all buildings**, not a small subset
- Automated, code-based workflows are used
- Data cleaning and aggregation are handled programmatically

2. Analytical Rigor & AI Usage (25%)

- Use of regression, machine learning, clustering, or other quantitative models
- Clear definition of expected vs. observed energy behavior
- Appropriate use of weather and building metadata

3. Explainability & Reasoning (20%)

- Clear explanation of *why* buildings rank highly based on data signals
- Visualizations directly support conclusions
- Assumptions and limitations are explicitly stated

4. Investment Prioritization Logic (15%)

- Rankings reflect consistent, defensible criteria
- Signals align with portfolio-level decision-making
- Bonus for partial alignment with facilities-identified high-performing references

5. Validation & Reflection (15%)

- Evidence of sanity checks or alternative comparisons
- Thoughtful discussion of uncertainty and next steps
- Reflection on how results could inform real investment decisions

Important Note on Ground Truth

There is **no single correct ranking**. Teams will be evaluated on methodological quality, appropriate use of AI, and clarity of reasoning—not on agreement with any predefined list. The emphasis is on demonstrating how data can surface meaningful performance differences across a portfolio and support smarter capital-planning decisions.

Documentation

Data Use Guidelines

Because portions of the datasets provided for this challenge are internal to The Ohio State University, all data access and use is limited to participation in this hackathon only.

By participating in this challenge, teams agree that:

- The provided data may be used solely for analysis, prototyping, and presentation within the scope and duration of the hackathon.
- The data may not be copied, redistributed, published, or used for external research, commercial purposes, or projects outside of this event.
- Any demos, notebooks, repositories, or presentations shared publicly must not include raw data or information that could reasonably be used to reconstruct the dataset.

Participants are encouraged to showcase methods, models, visualizations, and insights, but not the underlying data itself. The intent of this guideline is to support learning and experimentation while respecting institutional data stewardship responsibilities (<https://it.osu.edu/data/institutional-data-policy>).

Available Data Sets

Dataset Overview

- [Curated \(flattened\) smart meter data](#)
- [SIMS building data](#)
- [Historical weather data](#)

OSU AI Hackathon Info Packet

Together, these datasets provide a comprehensive view of building performance and utilization. When combined, they can reveal patterns that inform maintenance, facility planning, staffing, energy management, and cost-reduction decisions.

Data set	Source	Interval	Update frequency	Date range available
Smart meter - curated	Flattened smart meter data	1 hour	Daily	01/01/2025 – 12/31/2025
SIMS building (limited)	SIMS Space API	N/A	Once weekly	N/A
Weather	https://open-meteo.com/en/docs/historical-forecast-api	hourly	Daily	01/01/2025 – 12/31/2025

Data Set Documentation

Smart Meter Data

Source: Raw meter data

Format: CSV

Description: Each record represents a 15-minute interval energy meter reading, along with aggregated statistics over a rolling 24-hour window ending at that interval.

The `readingValue` reflects the energy consumed during a single 15-minute interval, measured in the units specified by `readingUnits` (e.g., kWh). These interval readings are not cumulative; total energy over longer periods must be computed by summing interval values.

In addition to the interval reading, each record includes summary statistics (`readingWindowSum`, `Mean`, `Min`, `Max`, etc.) calculated across the 96 expected 15-minute readings in the associated 24-hour window (`readingWindowStart` to `readingWindowEnd`). These fields provide contextual information about recent load behavior, variability, and extremes.

The `simsCode` field enables joining meter data to the Space Information and Management System (SIMS), allowing meter readings to be linked with building metadata such as square footage, number of floors, construction year, and location.

For more supporting information on utilizes, see section “[Utilities Overview](#)” below.

Sample JSON return

```
{  
  "meterId": 246014,
```

OSU AI Hackathon Info Packet

```
"siteName": "East Regional Chilled Water Plant",
"simsCode": "376",
"utility": "ELECTRICITY",
"readingTime": "2025-01-04T05:00:00",
"readingValue": 151.05446750165862,
"readingUnits": "kWh",
"readingUnitsDisplay": "Kilowatt hour",
"readingWindowStart": "2025-01-04T05:00:00",
"readingWindowEnd": "2025-01-05T04:45:00",
"expectedWindowReadings": 96,
"totalWindowReadings": 96,
"missingWindowReadings": 0,
"filteredWindowReadings": 0,
"readingWindowSum": 14142.396919778323,
"readingWindowMin": 88.3088609341783,
"readingWindowMinTime": "2025-01-05T03:30:00",
"readingWindowMax": 157.55109096750593,
"readingWindowMaxTime": "2025-01-05T01:45:00",
"readingWindowStandardDeviation": 10.561087028350984,
"readingWindowMean": 147.3166345810242
}
```

Data Dictionary

Key	Description	Data Type	Sample Value
meterId	Unique numeric identifier for the meter	integer	246073
siteName	Human-readable site name where the meter is located	string	"11th Ave, 33 W"
simsCode	Internal SIMS (Space Information and Management System) building code	integer	193
utility	Type of utility measured	string	"ELECTRICITY"
readingTimeEpoch	Timestamp of the reading in Unix epoch milliseconds	integer	1672574400000
readingValue	Recorded consumption for the 15 minute interval	float	8.186482615928398

OSU AI Hackathon Info Packet

readingUnits	Standardized unit of measurement	string	"kWh"
readingUnitsDisplay	Display-friendly unit name	string	"Kilowatt hour"
readingWindowStartEpoch	Start time of the reading window	integer	1451624400000
readingWindowEndEpoch	End time of the reading window	integer	1451624400000
expectedWindowReadings	Number of readings expected for the window	integer	96
totalWindowReadings	Number of readings actually received	integer	0
missingWindowReadings	Number of readings missing from the window	integer	96
filteredWindowReadings	Number of readings excluded due to quality filters	integer	0
sumWindowReadings	Sum of all readings in the window	float	0.0
minWindowReading	Minimum recorded value in the window	float	112
maxWindowReading	Maximum recorded value in the window	float	12
windowStandardDeviation	Standard deviation of readings in the window	float	12

SIMS Building

Source: SIMS Space API

Format: CSV

Description: Provides building-level metadata that is maintained by Facilities Information and Technology Services.

Sample JSON Return

```
{  
  "buildingNumber": "311",  
  "buildingName": "Mount Hall (0311)",  
  "campusName": "Columbus",  
  "address": "1050 Carmack Rd",  
  "city": "Columbus",  
  "state": "OH",
```

OSU AI Hackathon Info Packet

```
{
  "postalCode": "43210-1002",
  "county": "Franklin",
  "frameworkDistrict": "Western Lands",
  "geography": "Columbus Contiguous",
  "formalName": "Mount, John T. Hall",
  "alsoKnownAs": null,
  "schedulingAbbreviation": "MO",
  "grossArea": 75660.0,
  "floorsAboveGround": "2",
  "floorsBelowGround": "1",
  "constructionDate": "1974-07-01",
  "latitude": "40.00405648",
  "longitude": "-83.0367706"
}
```

Weather

Source: <https://open-meteo.com/en/docs/historical-forecast-api>

Format: CSV

Description: Weather attributes by date

Data Dictionary

Key	Description	Data Type	Units/format	Example Value
date	Timestamp of observation in UTC	datetime	ISO 8601	2025-08-10 04:00:00+00:00
latitude	Latitude coordinate for location of weather observation (will always be the same – location of Ohio State University – main campus)	float	lat/long	40.08
longitude	Longitude coordinate for location of weather observation (will always be the same – location of Ohio State University – main campus)	float	lat/long	-83.06

OSU AI Hackathon Info Packet

temperature_2m	Air temperature at 2 meters above ground	float	°F	72.8798
dew_point_2m	Dew point temperature at 2 meters above ground	float	°F	64.871765
relative_humidity_2m	Relative humidity at 2 meters above ground	integer	%	76
precipitation	Total precipitation during the hour	float	mm or inches*	0
direct_radiation	Direct solar radiation	float	W/m ²	0
wind_speed_10m	Wind speed at 10 meters above ground	float	mph	4.297137
wind_speed_80m	Wind speed at 80 meters above ground	float	mph	12.164427
wind_direction_10m	Wind direction at 10 meters above ground	float	degrees	141.34016
wind_direction_80m	Wind direction at 80 meters above ground	float	degrees	147.77127
cloud_cover	Fraction of the sky covered by clouds	integer	%	0
apparent_temperature	Feels-like temperature based on air temp, humidity, and wind speed	float	°F	76.52037
shortwave_radiation		Big int		77
direct_radiation		double		36.0
diffuse_radiation		double		158
direct_normal_irradiance		double		16.9898

Utilities Overview

The smart meter dataset includes measurements for multiple utility types, each representing a different form of energy or thermal service delivered to campus buildings or systems. Utilities should generally be analyzed separately, unless teams explicitly convert or normalize values using appropriate engineering assumptions.

Not all buildings have all utilities, and the presence of a meter does not imply exclusive service to a single building or piece of equipment.

OSU AI Hackathon Info Packet

Utility Descriptions

Utility	Description	Typical Interpretation
ELECTRICITY	Electrical energy consumption	Lighting, plug loads, motors, equipment, and building systems
ELECTRICAL_POWER	Instantaneous electrical demand	Real-time power draw (kW), often used for peak demand analysis
GAS	Natural gas consumption	Heating, hot water, cooking, or process loads
HEAT	Thermal energy delivered for heating	Hydronic or district heating systems
STEAM	Thermal energy delivered as steam	Space heating, hot water, or process uses
STEAMRATE	Steam flow rate	Instantaneous steam delivery rather than total energy
COOLING	Thermal energy delivered for cooling	Chilled water used for space cooling
COOLING_POWER	Instantaneous cooling demand	Real-time cooling load rather than total cooling energy
OIL28SEC	Fuel oil consumption	Legacy or backup heating systems

Energy vs Power

Energy represents total consumption over time (e.g., kilowatt-hours of electricity or ton-hours of cooling). This applies to electricity, gas, heat, steam, and cooling. You can look at a 15 min. Interval and understand, for example, “how much cooling is being delivered in this 15-minute interval?”

Power represents the rate at which energy is being used at a given moment or over a short interval (e.g., kilowatts or tons of cooling). This applies to cooling power and electrical power. You can look at a 15 min. Interval and understand, for example, “what is the rate of cooling in this 15-minute interval?” Or “how hard is the cooling system working in this 15-minute interval?”

Important Interpretation Notes

- In this dataset, meter readings are recorded at 15-minute intervals and represent the amount of energy consumed during each interval, not a cumulative total. To compute daily or monthly energy use, interval values should be summed. To analyze demand or intensity, interval energy values can be converted to average power over the interval.
- Utilities are not directly comparable without appropriate unit conversion or normalization. For example, electricity (kWh) should not be summed with steam or cooling energy without justification.

OSU AI Hackathon Info Packet

- Some utilities represent energy over time (in this case, within a 15 minute period) (e.g., ELECTRICITY, STEAM, HEAT), while others represent instantaneous demand or flow (e.g., ELECTRICAL_POWER, STEAMRATE, COOLING_POWER).
- Meter data may reflect delivered energy from centralized campus systems, not on-site generation or individual equipment performance.
- Buildings may be served by district energy systems, and a single meter may represent aggregated service rather than a single end use.

Teams are encouraged to clearly state assumptions when comparing or combining utilities and to focus on analyses on consistent utility types unless conversions are explicitly justified.

Recommended Best Practices (Optional Guidance)

- Analyze one utility at a time when exploring trends or comparisons.
- Normalize energy-based utilities (e.g., ELECTRICITY, HEAT) by square footage or time where appropriate.
- Use power-based utilities (e.g., ELECTRICAL_POWER, COOLING_POWER) for peak demand or variability analysis, not total consumption.
- Clearly distinguish between energy and power in visualizations and interpretations

Schema joins

Join Category	Table (db.table) A	Join key (A)	Join Type	Athena table (db.table) B	Join key (B)	Transform Required
Building metadata → Meter readings	Building_metadata	buildingNumber	direct	meter_data	simrcode	None
Weather → Meter readings by time	weather_data	date	direct/partial	meter_data	readingtime	None if full date; SPLIT if partial date (e.g. day)