Title

Poster Abstract: Leveraging Large Language Models to Reveal Interpretable Cooling Behaviors from Smart Thermostat Data

Journal

PROCEEDINGS OF THE 2025 THE 12TH ACM INTERNATIONAL CONFERENCE ON SYSTEMS FOR ENERGY-EFFICIENT BUILDINGS, CITIES, AND TRANSPORTATION, BUILDSYS 2025

Authors

Nihar, Kopal
Li, Han
Malik, Jeetika
et al.

Publication Date

2025-11-19

DOI

10.1145/3736425.3772110

Poster Abstract: Leveraging Large Language Models to Reveal Interpretable Cooling Behaviors from Smart Thermostat Data

Kopal Nihar†
Lawrence Berkeley National Lab
Berkeley, CA, USA
niharkopal@gmail.com

Han Li
Lawrence Berkeley National Lab
Berkeley, CA, USA
hanli@lbl.gov

Jeetika Malik
Lawrence Berkeley National Lab
Berkeley, CA, USA
jmalik@lbl.gov

Tianzhen Hong
Lawrence Berkeley National Lab
Berkeley, CA, USA
thong@lbl.gov

ABSTRACT

Frequent heatwaves and hot summers increasingly challenge occupant comfort, health, and energy grid stability. Addressing these challenges requires a detailed understanding of household cooling behaviors, such as thermostat adjustments and adaptive responses to extreme conditions. Traditional analyses often rely on aggregated numerical metrics that overlook subtle but important household-specific variations. In this study, we introduce a generalizable methodology that integrates large language models (LLMs) with vision capabilities to enable scalable and detailed analysis of residential thermostat data. Using Ecobee's Donate Your Data (DYD) dataset—which provides five-minute records of indoor temperatures, thermostat setpoints, and HVAC runtimes—we focus on two U.S. cities with contrasting summer climates: Austin (TX) and Phoenix (AZ). Because raw time-series data are not well suited for direct LLM analysis, we transform them into visual representations, such as daily indoor temperature trajectories and weekly runtime histograms, to better capture behavioral variations. Leveraging LLMs' visual interpretation, we extract descriptive behavioral features, including temperature preferences, time-of-day cooling orientation, anticipatory versus reactive heatwave responses, and behavioral consistency. These semantic features support unsupervised clustering to identify distinct occupant archetypes at scale, revealing differences—such as morning-centric anticipatory coolers versus households that shift toward warmer setpoints during heatwaves—that can inform demand response, resilience planning, and health-aware interventions. By converting raw numerical data into interpretable behavioral patterns, this methodology enables scalable and practical analysis of occupant behavior, supporting actionable insights for comfort, resilience, and energy management.

BUILDSYS '25, November 19–21, 2025, Golden, CO, USA

ACM ISBN 979-8-4007-1945-5/2025/11. https://doi.org/10.1145/3736425.3772110

CCS CONCEPTS

• Human-centered computing • Visualization • Empirical studies in visualization

KEYWORDS

Language models, Occupant behavior, Demand side management

ACM Reference format:

Kopal Nihar, Han Li, Jeetika Malik and Tianzhen Hong. 2025. Poster Abstract:Leveraging Large Language Models to Reveal Interpretable Cooling Behaviors from Smart Thermostat Data. In *12th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys 2025), November 19-21, 2025. Golden, Colorado, USA* https://doi.org/10.1145/3736425.3772110

1 Introduction

In light of increasing extreme heat events, understanding occupant behavior has become essential for ensuring comfort, health, and energy resilience [1]. However, most large-scale studies still rely on aggregated or survey-based data, which overlook day-to-day behavioral dynamics and the diversity of household adaptation strategies [2,3]. To bridge this gap, we explore the use of visualization and large language models (LLMs) to derive interpretable, scalable insights into how households adapt their cooling behavior during extreme heat.

2 Methodology

We analyzed data from 220 households each in Austin (TX) and Phoenix (AZ), covering the summer months from June 1 to September 30, 2023. We defined a heatwave as three or more consecutive days with maximum outdoor temperatures above 38 °C [4]. Using this definition, we found that Austin experienced heatwaves during 37% of the study period, while Phoenix experienced them during 69%. We divided the data into three sets per city, each spanning baseline, pre-heatwave, heatwave, and post-heatwave phases, identified from local weather data records.

We transformed daily indoor temperature trajectories into visualizations and used LLMs to interpret them. We designed prompts with clear definitions, labeling rules, and representative examples, and we explicitly referenced color-coded segments (grey = baseline, green = pre-heatwave, red = heatwave, yellow = post-heatwave). From these visuals, the LLMs extracted descriptors such as temperature preference, heatwave response, consistency, intra-day variation, time-of-day preference, anticipatory shift, and post-heatwave shift. Each run returned JSON-encoded labels in a fixed schema, which we standardized for clustering. We selected k-means for interpretability and consistency, and used silhouette scores to determine the final number of clusters. Validation involved two-dimensional t-SNE projections generated from both LLM-based and numeric clustering, providing a comparative view that highlighted where numeric features failed to capture nuanced occupant strategies such as anticipatory cooling or persistence of heatwave behavior. Figure 1 shows the full methodology framework.

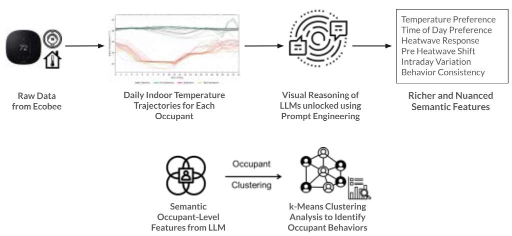

Figure 1: Methodology Framework for LLM-based Clustering

3 Preliminary Results

We focused first on Austin. To validate feature extraction, we manually labeled images for 25 households across each of the three sets and all selected features, consistently achieving accuracies above 70%. Using silhouette scores, we found that the optimal number of clusters was seven for both LLM-based clustering and numeric clustering. Figure 2 compares the t-SNE projections: the LLM-based analysis (left) shows well-separated, clearly defined clusters, while the numeric features (right) show overlapping groups, confirming the interpretability advantage of semantic features.

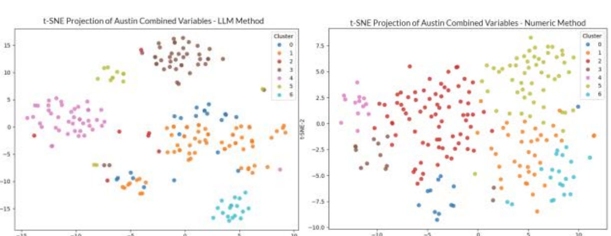

Figure 2: t-SNE Analysis Results for Austin using LLM-Based Features (Left) and Numeric Formula-based Features (Right)

Within the LLM-based results, the largest group (32%) consisted of morning cooling households that shifted to warmer setpoints during heatwaves. Another significant group (20%) had flat daily temperature profiles, also exhibiting a slight warm shift under heatwave conditions. We identified a small "wake-up" group with a warmer baseline preference and intensified cooling across the summer, often showing anticipatory responses to heatwaves. A smaller segment (~10%) represented aggressive morning coolers with consistent cooling profiles throughout the season. Figure 3 illustrates baseline heatwave responses and seasonal heatwave trends by cluster. Baseline behavior was polarized: C0 predominantly cooled more (~75% lowered setpoints), while C1, C4, and C6 shifted warmer, and C3/C5 showed little change—highlighting that the same heatwave can drive opposite load effects across households. Over the season, most clusters remained stable or reduced cooling, but C2 stood out, with ~70% intensifying cooling, signaling where load growth may occur.

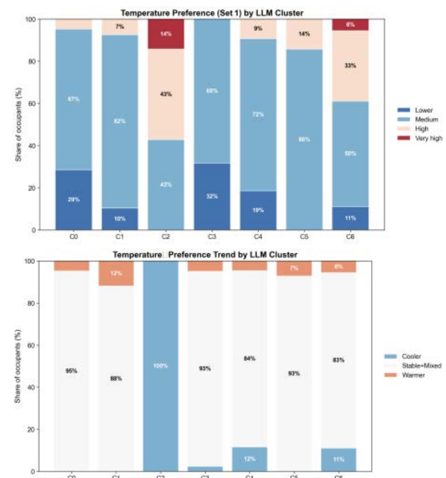

Figure 3: Results for Austin – Baseline Temperature Preference (Up) and Trend in Summer Season Across the Clusters (Down)

These findings underscore the value of LLM-derived semantic features. They capture richer behavioral nuances and yield clearer cluster separation than aggregated numeric features, with human-readable labels directly tied to comfort behaviors and energy adaptation strategies. This approach distinguishes *when* people act (morning, afternoon, or none), *how* they act (anticipatory vs. reactive; consistent vs. variable), and *how* they evolve (seasonal trends). By moving beyond one-size-fits-all averages, this framework enables utilities to design occupant-aware demand response strategies that align with both resilience and comfort.

REFERENCES

[1] Meier, Alan, Tsuyoshi Ueno, and Marco Pritoni. "Using data from connected thermostats to track large power outages in the United States." *Applied Energy* (2019): 113940. DOI: https://doi.org/10.1016/j.apenergy.2019.113940.

[2] Jung, Wooyoung, et al. "Smart thermostat data-driven US residential occupancy schedules and development of a US residential occupancy schedule simulator." *Building and Environment* 243 (2023): 110628. DOI: https://doi.org/10.1016/j.buildenv.2023.110628

[3] Li, Han, et al. "A Critical Review of Use Cases and Insights from A Large Dataset of Smart Thermostats." *Advances in Applied Energy* (2025): 100236. DOI: https://doi.org/10.1016/j.adapen.2025.100236

[4] Barnett, A. G., et al. "Cold and heat waves in the United States." *Environmental research* 112 (2012): 218-224. DOI: https://doi.org/10.1016/j.envres.2011.12.010