Analysing building energy efficiency opportunities in California

Draft for EPA EnergyStar

Michael Bailey, Kumar H and Anthony Moubarak Environmental Impact Data Collaborative, Massive Data Institute, Georgetown University

Date published: August, 2023

1 Residential energy use and disadvantage at the zipcode level

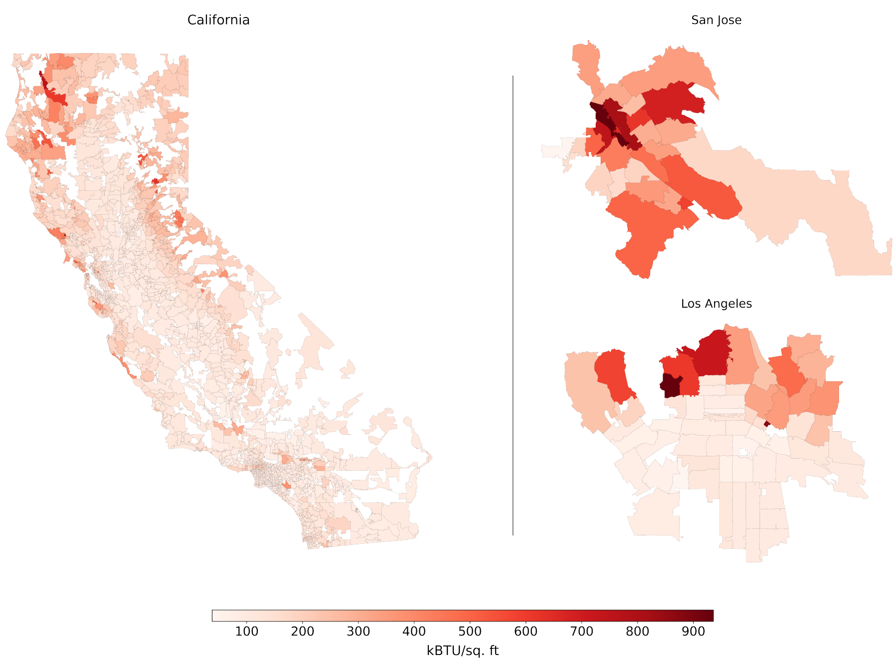

Figure 1: Avg. Energy Use Intensity, EUI, single-family building by zipcode in CA (kBTU/sq.ft)

Figure 1 shows the average zipcode level energy use intensity (EUI) of a single-family residential building, expressed in 1000 British thermal units per square foot (kBTU/sq.ft). The panel on the left shows all zipcodes in California. The one on the right illustrates how to zoom in to examine city-level patterns, as shown for San Jose and Los Angeles. The above figure is generated from BlocPower's building level data for California. The data is partially sourced from tax assessment records, including fields like built year, area and cooling system type. These fields then serve as inputs to an Automatic Building Energy Modeling (AutoBEM) software developed by Oak Ridge National Laboratory, to generate estimates of building energy use intensity. In contrast to the Residential Energy Consumption Survey (RECS) and Commercial Building Energy Consumption Survey (CBECS), the energy use field in the BlocPower data is not statistically sampled, but is simulated using an engineering model that takes in real data as inputs.

In its raw form, each row in the dataset represents a building. EIDC has developed a workflow to transform and spatially link a zipcode-level dataset, ready for analysis. The steps are:

- Subset the entire dataframe to single family buildings in California only
- Summarize building-level data to the zipcode level, averaging energy use intensity, finding the most common heating/cooling systems in a zipcode.
- Merge with aggregated zipcode-level LIHEAP [1] and CEJST [2] data. This shows how we can quickly merge in new datasets for added context.

The resulting dataset is used for the analyses and visualizations presented in this document. These can illustrate the potential uses of this data for policy analysis and decision support regarding energy efficiency of residential buildings.

Figure 2 depicts a simple framework to prioritize areas for energy efficiency investments. This is inspired in part by previous work that analyzed block-level EUI patterns in association with demographic and economic factors to balance energy efficiency with social equity priorities (Tong et al, 2021). Using more contemporary approaches, we overlay the latest version of the Climate and Economic Justice Screening Tool (CEJST) and calculate the proportion of the population in each zip code which lives in a tract classified as disadvantaged in the CEJST [3]. The other variable in the framework is the percent of buildings in each zipcode classified as having either high or medium energy efficiency potential (EEP) by BlocPower.

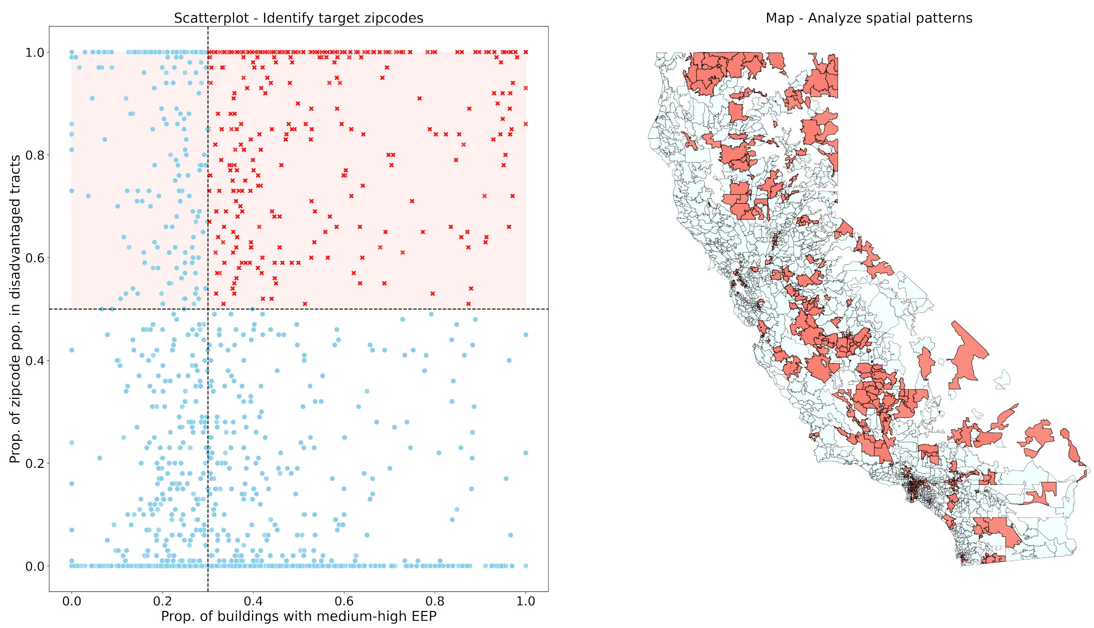

Figure 2: Overlaying building energy efficiency opportunities with environmental justice indicators

[1] US Department of Health and Human Services (HHS) Low Income Home Energy Assistance Assistance Program (LIHEAP) grants data, aggregated and anonymized to the zipcode level.

[2] Climate and Economic Justice Screening Tool

[3] We employ a raster-based method to map CEJST data from census tracts to ZIP Code Tabulation Areas (ZCTAs). Since these areas can overlap, we use a population density raster layer to proportionately map the population within these overlapping areas. This process produces a "crosswalking weight" that can be then used to reweight the percentage of disadvantaged populations from census tracts to matched ZCTA. This ensures a more accurate representation of the demographics across different geographic scales.

2 Future directions:

2.1 Utilizing a wider range of building attributes:

The EnergyStar certification process for residential buildings is different compared to the benchmarking for commercial buildings (see appendix). For new residential construction, "a home or apartment must meet strict program requirements for energy efficiency" based on building science and industry interaction (EPA, 2023), spanning four areas: Thermal Enclosure, Heating and Cooling, Water management and Lighting and Appliances [4]. A third–party Energy Rating Company then performs a rating to verify compliance with EPA requirements, set forth in the National Program Requirements, with a separate version for California. There are also specific checklists for raters, as in the National Rater Design checklist and the HVAC design checklist, shown below.

BlocPower's building-energy model data simulates highly granular details, including end-use consumption estimates, HVAC airflow rates, fenestration and ventilation rates, and could be used to estimate EnergyStar compliance of existing buildings. A few possible example variables with good fill rates are given in table 1 below, with possibly analogous checklist items on the right.

BlocPower Variables	**HVAC Design Checklist V 3.2 (Rev 12)**
Coil_Sizing_Summary_Coil_Air_Volume_Flow_Rate_at_Ideal_Loads_Peak_m3s	Ventilation airflow design rate
EntireFacility_Conditioned_Window_Wall_Ratio	Window area used in loads
Cooling/Heating system type	AC/Heat Pump Equipment type

As an additional resource, the latest released version of the Residential Energy Consumption Survey (RECS 2020) contains a wide variety of building appliance and usage characteristics, available at the state level for the first time. The 2020 RECS microdata and state-level summary data have been cleaned and published by EIDC, potentially allowing them to be used for EnergyStar benchmarking purposes. The RECS data is not used for energy efficiency benchmarking for residential buildings in the same way the Commercial Building Energy Consumption Survey (CBECS) is used for commercial buildings. However, in terms of informing about general patterns and associations with appliance types, it could contain valuable insights combined with the BlocPower data.

2.2 Newer models for EUI prediction and benchmarking

Recent research has explored augmenting and enhancing the statistical tools used for EnergyStar benchmarking. Arjunan et al (2020) explore machine learning models that increase prediction accuracy over the base regression model (measured by the R-squared) by almost 24%. They also emphasize model interpretability, i.e. to understand which attributes of a building are driving its influencing its energy use prediction. Using similar techniques, we tried a basic nonlinear model of building energy use intensity (EUI) and used it to predict the EUI value for an unseen test observation. Calculating the Shapley value and plotting as shown in figure 3, it offers an account of which particular attributes of this building influenced its prediction away from the base value.

[4] https://www.energystar.gov/partner_resources/residential_new/about

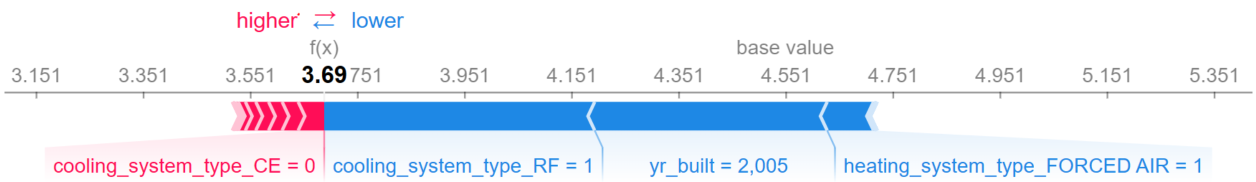

Figure 3: Example SHAP force plots for two office buildings with lower predicted EUI using an XGBoost model. Width of each bar denotes the SHAP value for that factor. Color indicates the attributes influencing the prediction to go higher (red) or lower (blue)

2.3 Geocoding

EIDC is also working on an easy-to-use geocoding application in R and Python for multiple use cases. It would likely be feasible to adapt it for the EnergyStar division's particular requirements for building benchmarking. For example, this could be integrated into a simple application where it takes user's self-reported data, geocodes the coordinates, assesses its energy use parameters and offers a comparison report with nearby/similar buildings.

---

3 Appendix

3.1 BlocPower-EIDC data

A majority of total U.S. emissions stem from four economic sectors: buildings, electricity, industry, and transportation. Reductions in all four areas are essential to reach the Paris agreement's Nationally Determined Contribution (NDC) of cutting emissions to 50% below 2005 levels by 2030. Princeton's Net Zero Initiative (Larson et al, 2021) sets out six pillars of economy-wide decarbonization, leading with improving end-use energy efficiency and electrification for buildings. Increasing electrification by itself is projected to reduce final energy use due to intrinsically more efficient technology like heat pumps.

High-resolution data on building equipment and energy consumption is essential to plan and measure the impact of decarbonization interventions. Existing data sources for this purpose have limited coverage.The Environmental Impact Data Collaborative (EIDC) has partnered with BlocPower - a Brooklyn-based climate technology company that uses building-level energy use data to guide community decarbonization projects. As a key output of this partnership, EIDC has worked with BlocPower to add a large building-level dataset to our data lake, containing energy equipment and consumption data for more than 121 million buildings. The data is partially sourced from tax assessment records, which provide data on building system types and attributes like built year and area. This data then serves as inputs to an Automatic Building Energy Modeling (AutoBEM) developed by Oak Ridge National Laboratory, to generate modeled estimates of building energy use [5]. Cloud-based building-level data pipelines are used by BlocPower to deploy their software-as-a-service (SaaS) solution - BlocMaps - to provide building decarbonization insights at city and

[5]"BlocPower collects data from over 100 million buildings from external sources, such as the Department of Energy's National Laboratories. These laboratories store their data using intermediate data format files, which require BlocPower to use EnergyPlus to process and render simulations of individual buildings."

local levels [6].

Apart from the transforming and aggregating, EIDC conducted quality and consistency checks of the data by comparing estimates with other US building and energy datasets (please contact us for data and results). To look at the active subset of the raw BlocPower dataset, visit this page. For other analyses created with this data, check out this interactive website.

3.2 EPA EnergyStar ratings - Commercial Buildings

The U.S. Department of Energy's Energy Information Administration conducts a national survey to gather data on building characteristics and energy use from thousands of buildings across the United States. This Commercial Building Energy Consumption Survey (CBECS) is the only national-level source of data on the characteristics and energy use of commercial buildings in the United States. For most property types, a building's peer group for comparison consists of those buildings in the CBECS survey that are similar. For each type of building for which there is an ENERGY STAR score, EPA ensures that the data is suitable to support development of an ENERGY STAR score, and creates a statistical regression model that adjusts for the key drivers of energy use. Based on the information you enter about your building, such as its size, location, number of occupants, number of PCs, etc., the regression estimates how much energy the building would use if it were the best performing, the worst performing, and every level in between. It then compares the actual energy data entered to the estimate to determine where the building ranks relative to its peers [7].

Arjunan et al (2020) describe how the EnergyStar methodology uses the CBECS:

- "Energy Star system normalizes building's energy use for differences in building operations by fitting linear regression models between building attributes and energy use. After empirically removing statistically insignificant attributes, the final model for each building type has 5–7 factor variables. Energy Efficiency Ratio (EER) of a building is calculated as a ratio between actual source EUI and normalized source EUI, as predicted by the respective peer group model. A lower EER indicates higher energy performance relative to the peer group. EER values are translated into a 1–100 percentile ranking by using a Score Lookup Table. This Score Lookup Table is created by using the parameters of a gamma distribution function (shape and scale), which was fit to the cumulative percentage of sorted EER values. A score of 75 or higher is eligible for Energy Star certification, and it can be interpreted as: this building performs better than 75 percent of similar buildings nationwide."

The EPA EnergyStar adopts a regression approach towards building energy consumption, which looks at measured energy data and determines statistically significant correlations between key indicators of business activity and source EUI. This is distinct from an engineered model, which attempts to predict or simulate the exact kWh of specific activities (lighting, HVAC, etc.) and sum these together into a building total. The BlocPower dataset was generated by feeding building attributes as inputs into ORNL's Building Energy modeling toolkit, which fits this second description.

[6] "BlocPower collects data from over 100 million buildings from external sources, such as the Department of Energy's National Laboratories. These laboratories store their data using intermediate data format files, which require BlocPower to use EnergyPlus to process and render simulations of individual buildings."

[7] Guide to computing an EnergyStar Score - https://portfoliomanager.energystar.gov/pdf/reference/ENERGY%20STAR%20Score.pdf