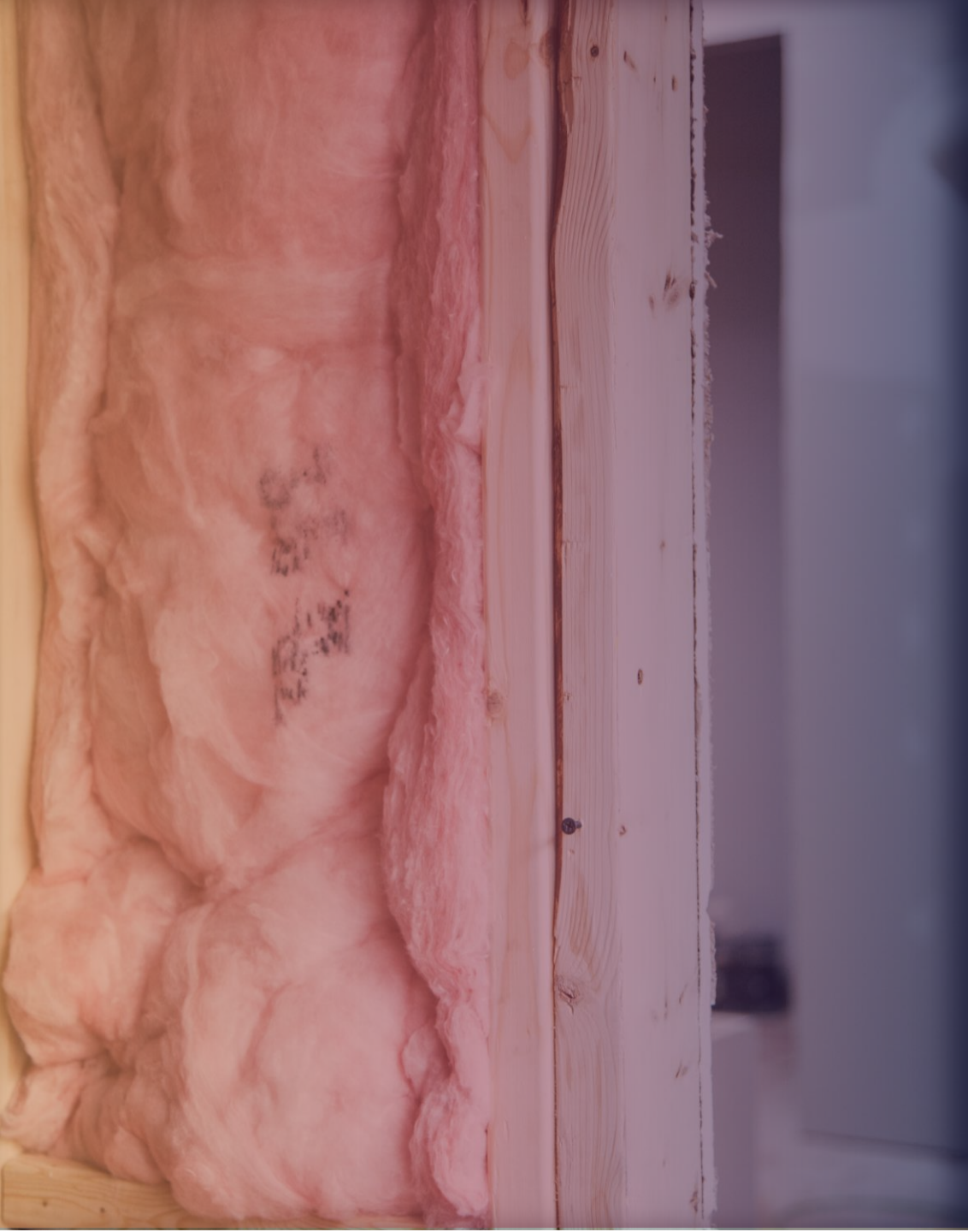




PREDICTING RESIDENTIAL ENERGY CONSUMPTION BASED ON ATTRIBUTES OF THE HOUSE

Julia Zozulia



POTENTIAL CLIENTS:

- homeowners thinking about home improvements or buying new appliances, answering the questions like “how much will I save if I insulate my attic”, or “should I really install smart thermostat”;
- contractors helping to persuade customers to renovation project;
- real estate brokers, helping make their customers more informed decisions by estimating heating and electric utilities cost.

The model will allow a variety of home improvement vendors/contractors to give the customer a very quick estimate of potential savings associated with updating of heating/cooling equipment, insulation, windows, appliances, etc.

DATA

o b t a i n i n g a n d w r a n g l i n g

The data is provided by U.S. Energy Information Administration in their residential energy consumption survey ([RECS](#)) for the 2015 year.

Data has:

- 5686 households (rows);
- 759 attributes (columns).

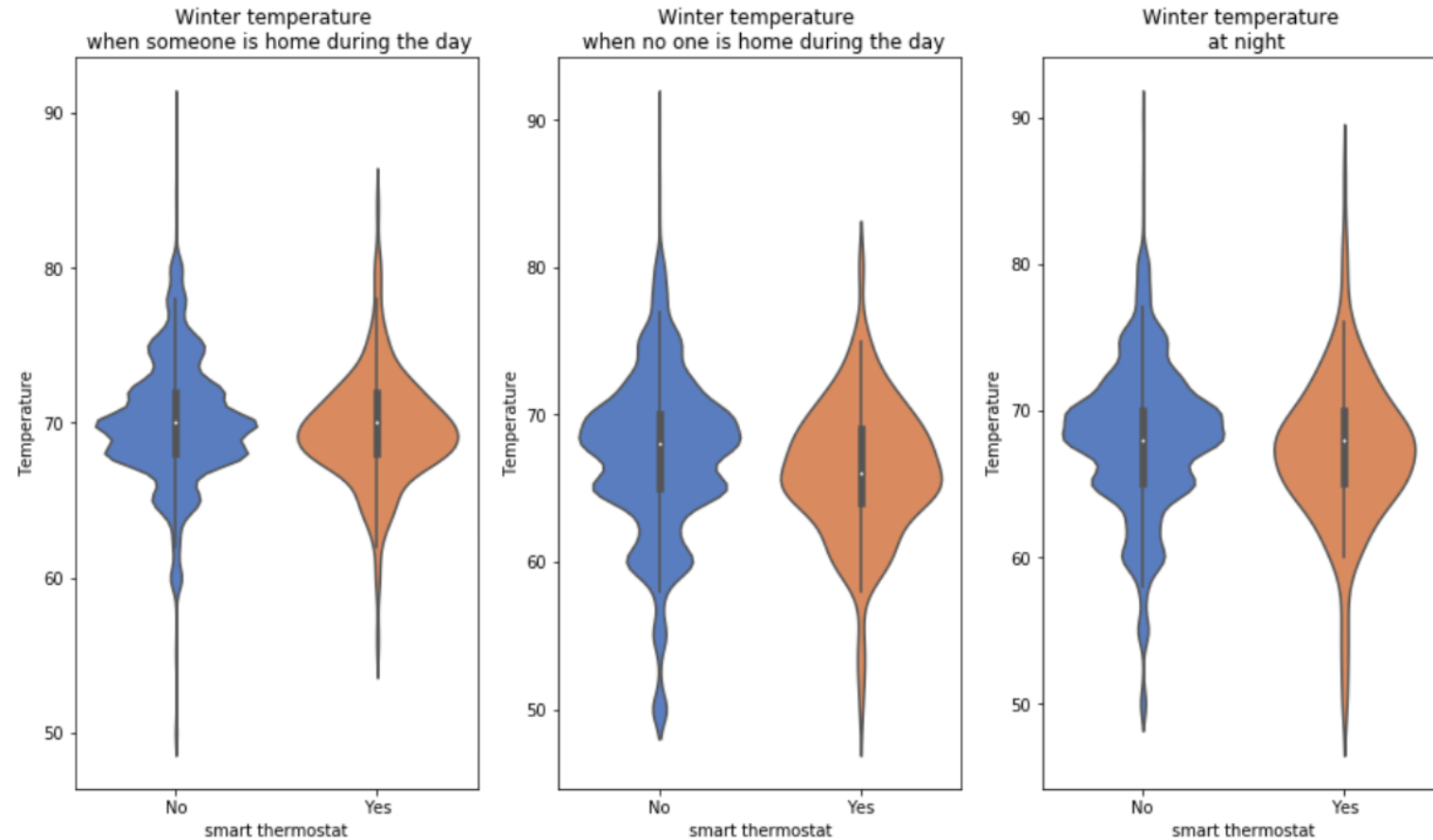
Missing values:

- Numerical: most replaced with 0, some with maximum or minimum of the column.
- Categorical: variables with “Yes”/”No” filled with “No”, some with default, and some with most common values.

SMART THERMOSTATS

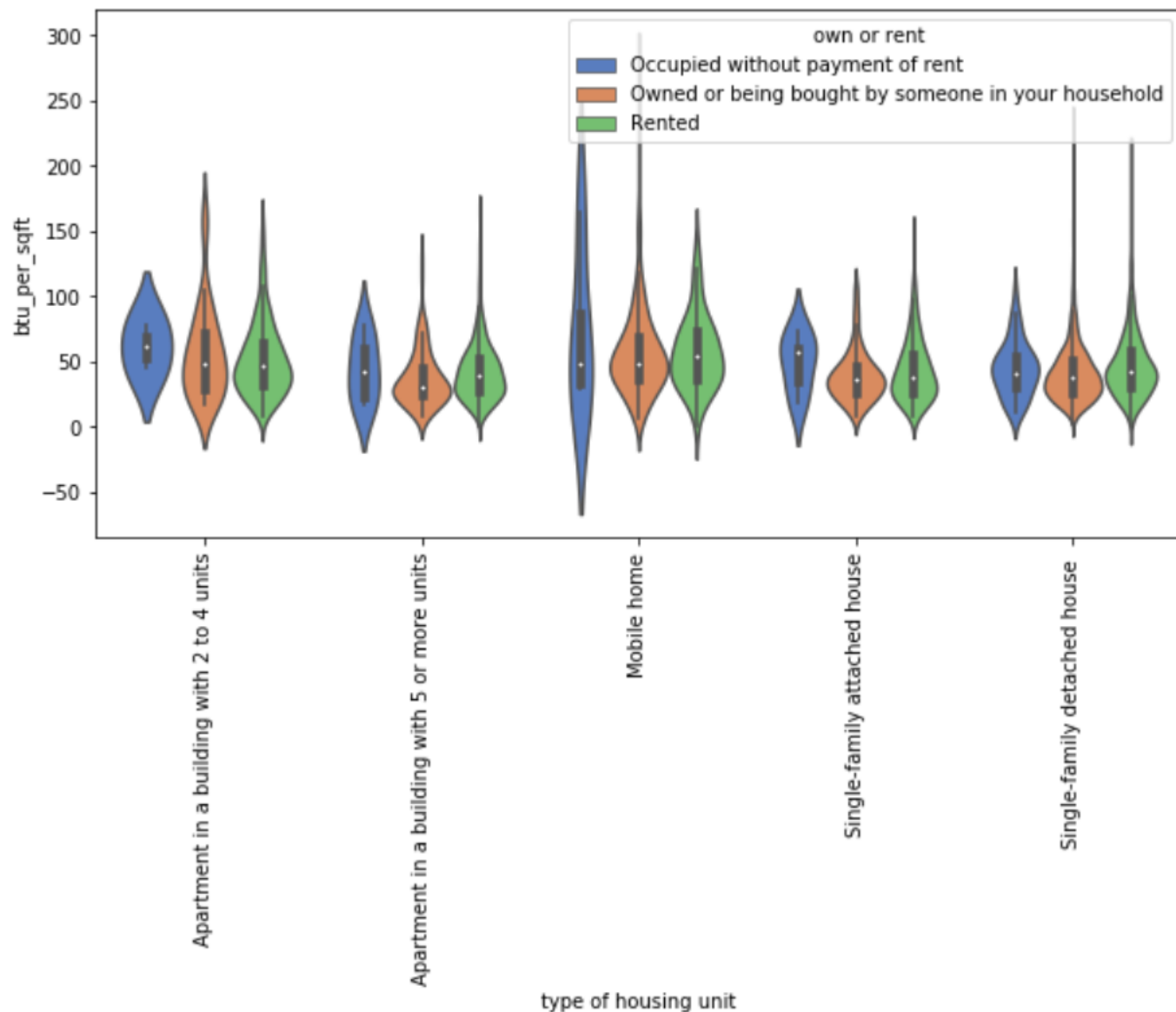
Do they help to lower the temperature when not needed?

Winter temperature vs. smart
thermostats



- Distribution of temperatures for homes with smart thermostats look smoother than those without, instead of uneven binomial distribution, we see almost perfect normal one.
- Smoother normal distribution makes households with smart thermostats easier to predict.
- 50% quantile for temperature when someone at home or at night quantile is the same for people with smart thermostats and without, while temperature when no one at home is noticeably lower.
- If our goal is to minimize consumption, it would make sense to promote using a smart thermostat.

Energy usage per sq. ft. for housing
unit types per own or rent



OWNER OCCUPIED OR RENTED

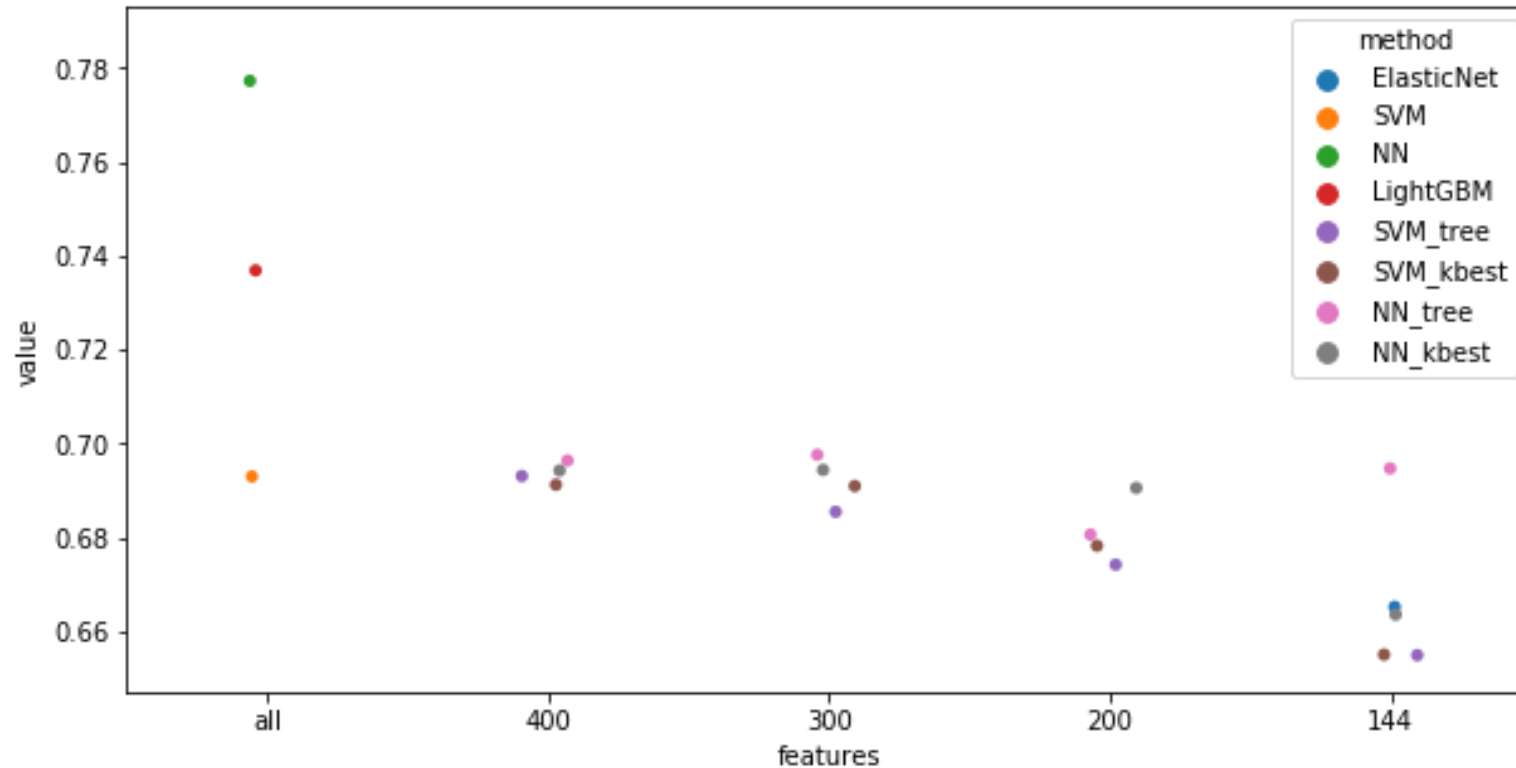
- People who rent spend more per sq. ft. than those who own for every unit type, except for 2-4 units apartments.
- Those, knowing if the unit is owner or renters occupied, we can expect higher consumption.
- People who occupied without payment of rent spend so much more. Perhaps, they are not financially responsible for utilities. Programs which will educate people about the environmental impact of wasteful use for such people could make a difference.

MODELING

and feature engineering

- Explored: ElasticNet, SVM, Neural Network, and LightGBM with different sets of features.
- Test/train split: 85%/15%.
- Hyper-parameters tuning: 5-fold cross-validation.
- The best performance: neural network on a full subset of variables (test score: 0.78), followed by LightGBM with 0.74.
- Using a subset of variables doesn't improve the overall result.

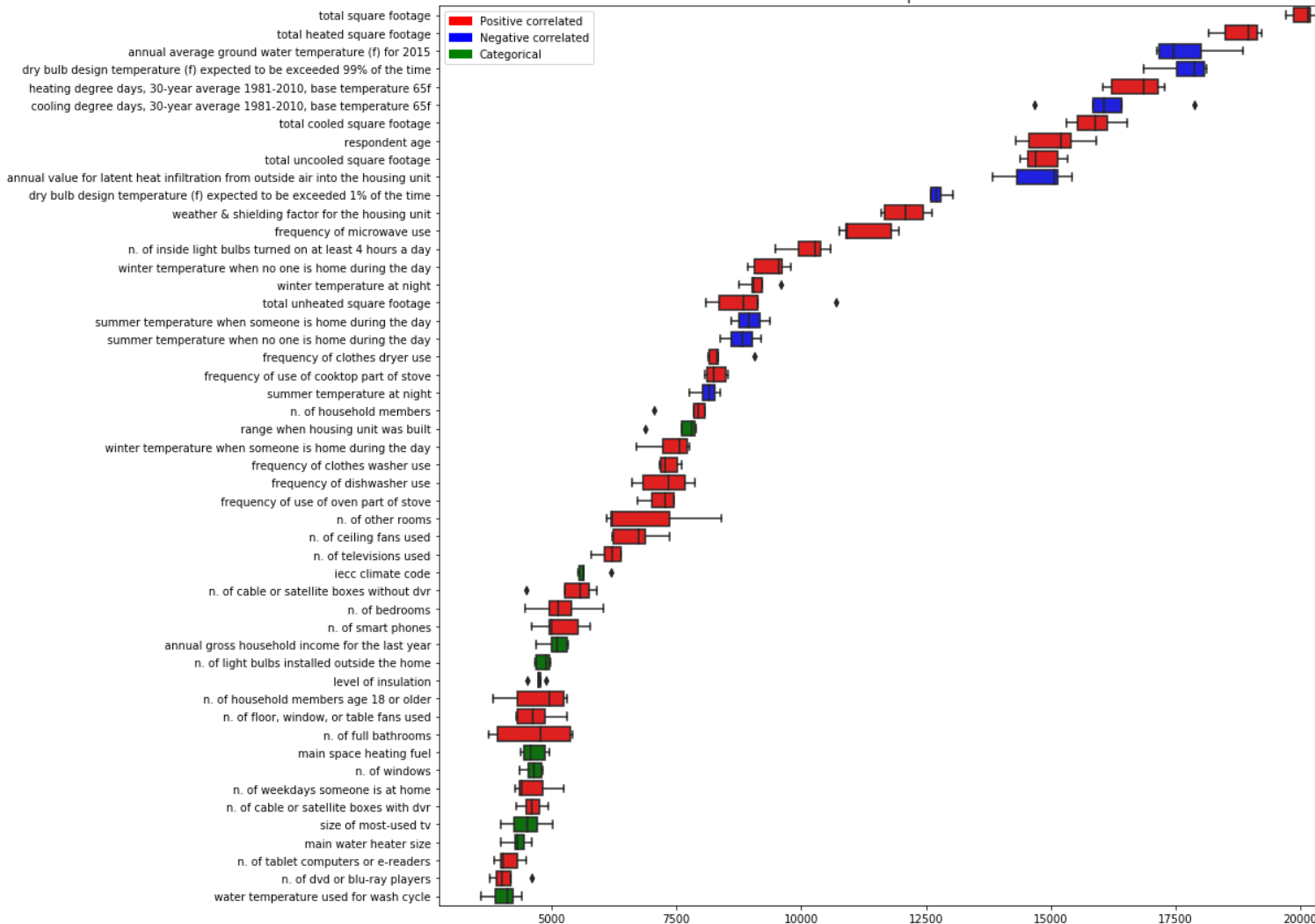
Test r^2 scores for different methods
and subset of features



FEATURE IMPORTANCE

based on LightGBM

Top 50 variables



- To be sure that feature importance we observe is not just a property of a certain run of the model, 5-fold cross-validation has been used.
- Boxplot color:
 - Red - the feature is positively correlated with the outcome,
 - blue - negative,
 - and green - categorical feature.
- The spread of importance is fairly small, so we can assume that the results are stable.



BUSINESS APPLICATION

Let's imagine you are a homeowner trying to minimize your utility cost. Everyone knows that, for example, if you insulate your home well, your energy consumption will go down. But by how much? **How much will I save per year? Is it worth the investment?** What will allow to save more: installing new windows or changing heating equipment?

The developed model allows us to estimate how much more/less will a household consume under new parameters.



TO ESTIMATE SAVINGS:

1. One-story single-family detached house, 2553 sq. ft., located in an urban area in Cold/Very Cold climate zone

House now is poorly insulated. If the owner will insulate the house, his estimated saving will be as follows:

Original consumption: 127095.96 BTU
Modified consumption: 125607.74 BTU
Absolute savings: 1488.22 BTU
Percentage savings: **1.17 %**

Changing windows to energy star qualified:

Absolute savings: 1301.38 BTU
Percentage savings: **1.02 %**

Changing main space heating equipment to newer one (same type):

Absolute savings: 1926.4 BTU
Percentage savings: **1.52 %**

Changing main space heating equipment from central furnace to heat :

Absolute savings: 8179.43 BTU
Percentage savings: **6.44 %**



TO ESTIMATE SAVINGS:

2. Rural 1676 sq. ft. single family house located in hot-humid climate

It is already adequately insulated, but, if the owner insulates the house better, he would save as follows:

Original consumption:	90945.05 BTU
Modified consumption:	90024.01 BTU
Absolute savings:	921.04 BTU
Percentage savings:	1.01 %

The occupants keep summer temperature when no one is home during the day at 65°F, which is unusually low. By keeping it at 75°F and installing smart thermostat they would save 1.41%:

Original consumption:	90945.05 BTU
Modified consumption:	89658.33 BTU
Absolute savings:	1286.72 BTU
Percentage savings:	1.41 %

Not everyone would benefit from installing smart thermostat equally: our first example home would save only 0.14%.



TO CONTRACTORS:

Apart from the end consumer, this model can be used by a variety of businesses:

- **HVAC contractors**, who would be able to give the customer a very quick estimate of potential savings associated with updating of heating/cooling equipment;
- **Companies selling/installing windows**;
- **Insulation contractors**;
- **Energy assessment companies**, who would be able to recommend customers which parts of the house would yield to the highest savings.

Of course, each of these constructors would be able to estimate potential savings based on detailed engineering calculations, but those are usually very time consuming and require a qualified engineer, while this model allows *making a very quick and inexpensive estimate*, allowing to *keep potential customers*.

TO REAL ESTATE BROKERS:

40 variables for unseen real estate

Locations based	Family based	Real estate based
<ul style="list-style-type: none">▪ iecc climate code▪ dry bulb design temperature▪ annual average ground water temperature▪ heating degree days▪ cooling degree days	<ul style="list-style-type: none">▪ respondent age▪ annual gross household income for the last year▪ n. of weekdays someone is at home▪ n. of inside light bulbs turned on at least 4 h▪ winter temperature when no one is home▪ winter temperature when someone is home▪ winter temperature at night▪ summer temperature when no one is home▪ summer temperature when someone is home▪ summer temperature at night▪ n. of household members▪ frequency of clothes dryer use▪ frequency of clothes washer use▪ frequency of dishwasher use▪ frequency of microwave use▪ frequency of use of cooktop part of stove▪ frequency of use of oven part of stove▪ size of most-used tv▪ n. of televisions used▪ n. of cable or satellite boxes without dvr▪ n. of smart phones	<ul style="list-style-type: none">▪ total square footage▪ total heated square footage▪ total cooled square footage▪ range when housing unit was built▪ n. of bedrooms▪ n. of full bathrooms▪ n. of other rooms▪ n. of windows▪ n. of ceiling fans used▪ n. of light bulbs installed outside the home▪ main space heating fuel

Real estate brokers can help customers to make more informed decisions by energy utilities cost for different homes.

Using subset of only 40 variables gives test score of 0.71.

Variables can be split into 3 subsets:

- locations based, can be assessed using zip code;
- family based, can be obtained by creating simple questionnaire for the clients;
- real estate based, are easily accessible in the listings.

New home buyers would be able to predict their utilities bills with high precision. This is one unknown factor less when choosing monthly mortgage payment they can afford.

THANK YOU

juliazozulia.github.io



Julia Zozulia



[github.com/JuliaZozulia/
Predicting_utilitiy_cost](https://github.com/JuliaZozulia/Predicting_utilitiy_cost)



y.s.zozulia@gmail.com