



PREDICTING RESIDENTIAL ENERGY CONSUMPTION BASED ON ATTRIBUTES OF THE HOUSE

Julia Zozulia



POTENTIAL CLIENTS:

- homeowners thinking about home improvements;
- contractors persuading customers to renovation project;
- real estate brokers.

The model helps home improvement vendors/contractors to give the customer a very quick estimate of potential savings.

DATA WRANGLING

The data is provided by U.S. Energy Information Administration (residential energy consumption survey for the 2015 year).

Data has:

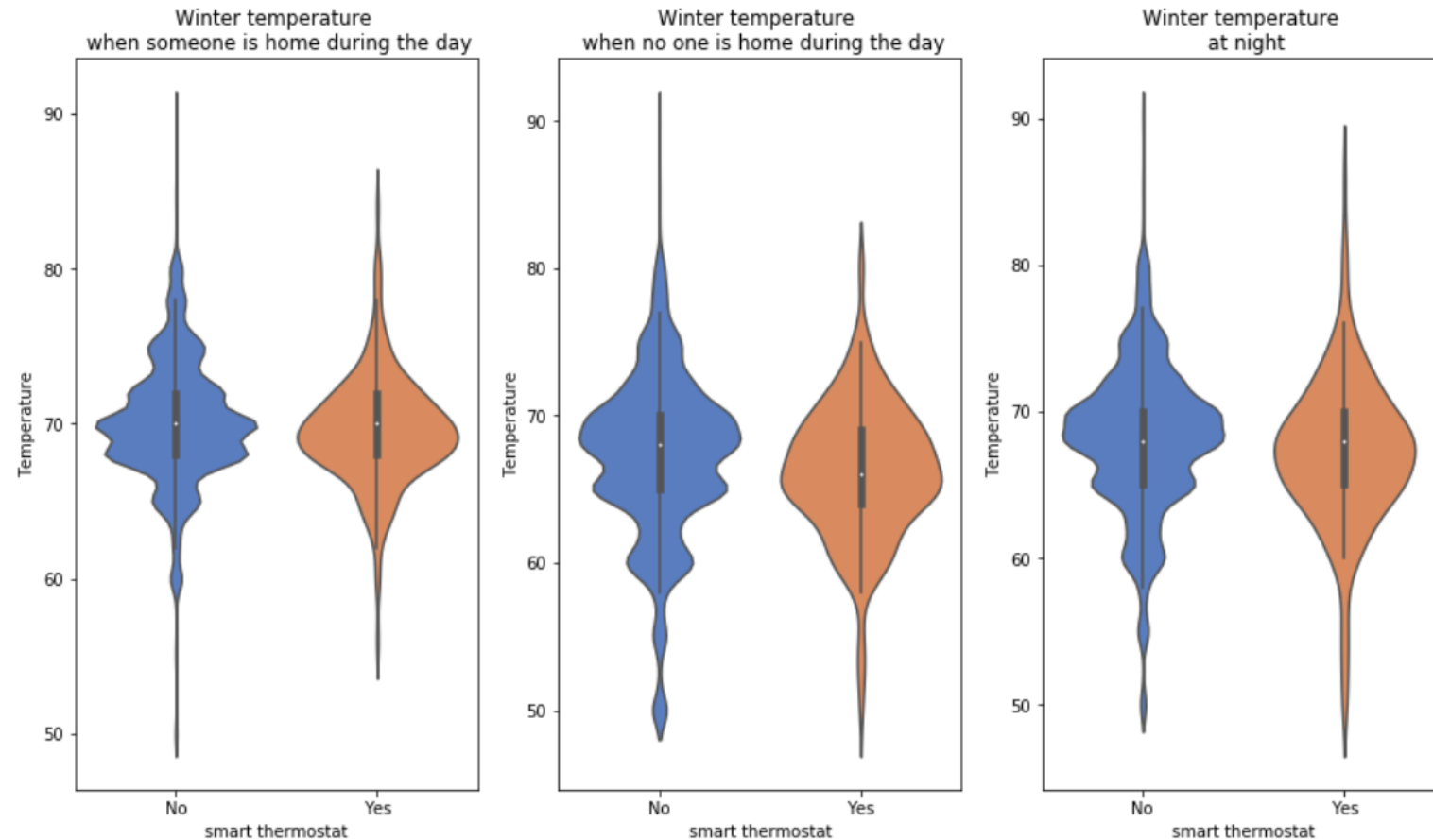
- 5686 households (rows);
- 759 attributes (columns).

Missing values:

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

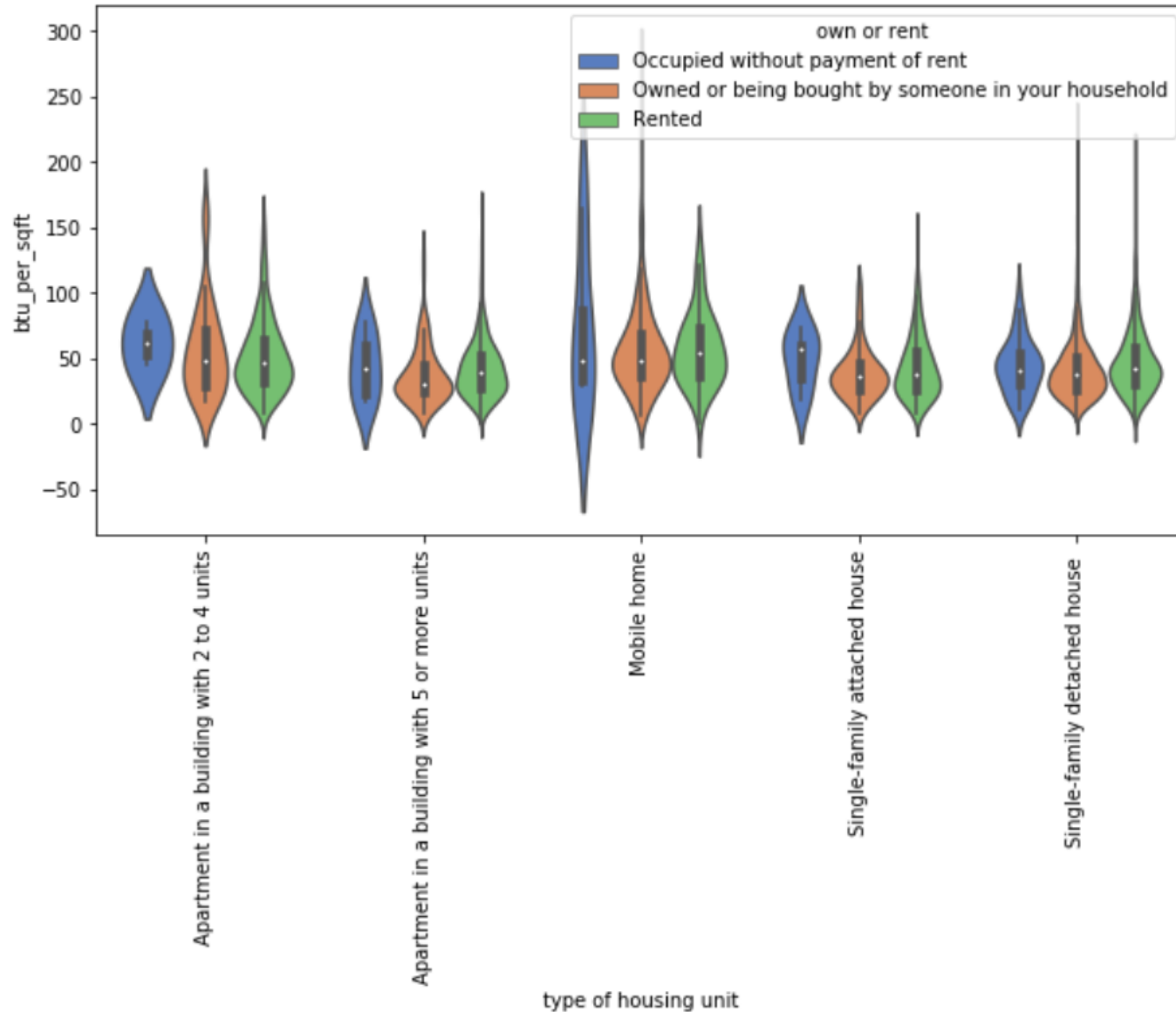
SMART THERMOSTATS:

Winter temperature vs. smart
thermostats



- smoother distribution of temperatures;
- easier to predict;
- temperature when someone at home is the same;
- temperature when no one at home is noticeably lower;
- helps to save.

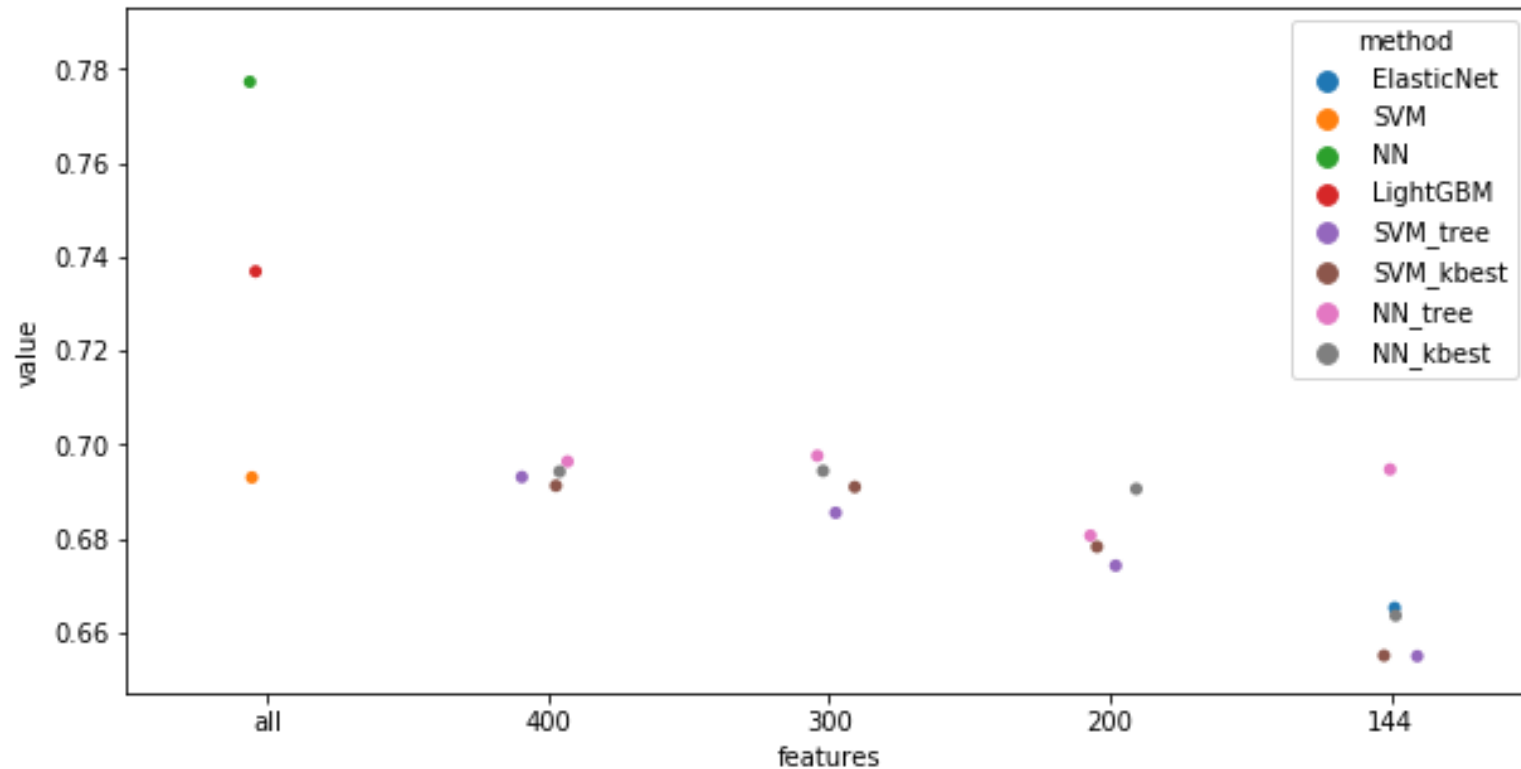
Energy usage per sq. ft. for housing
unit types per own or rent



OWNER OCCUPIED OR RENTED

- People who rent spend more per sq. ft. than those who own.
- People who occupied without payment of rent spend much more.
- Educational programs for people who don't pay rent needed.

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



MODELING

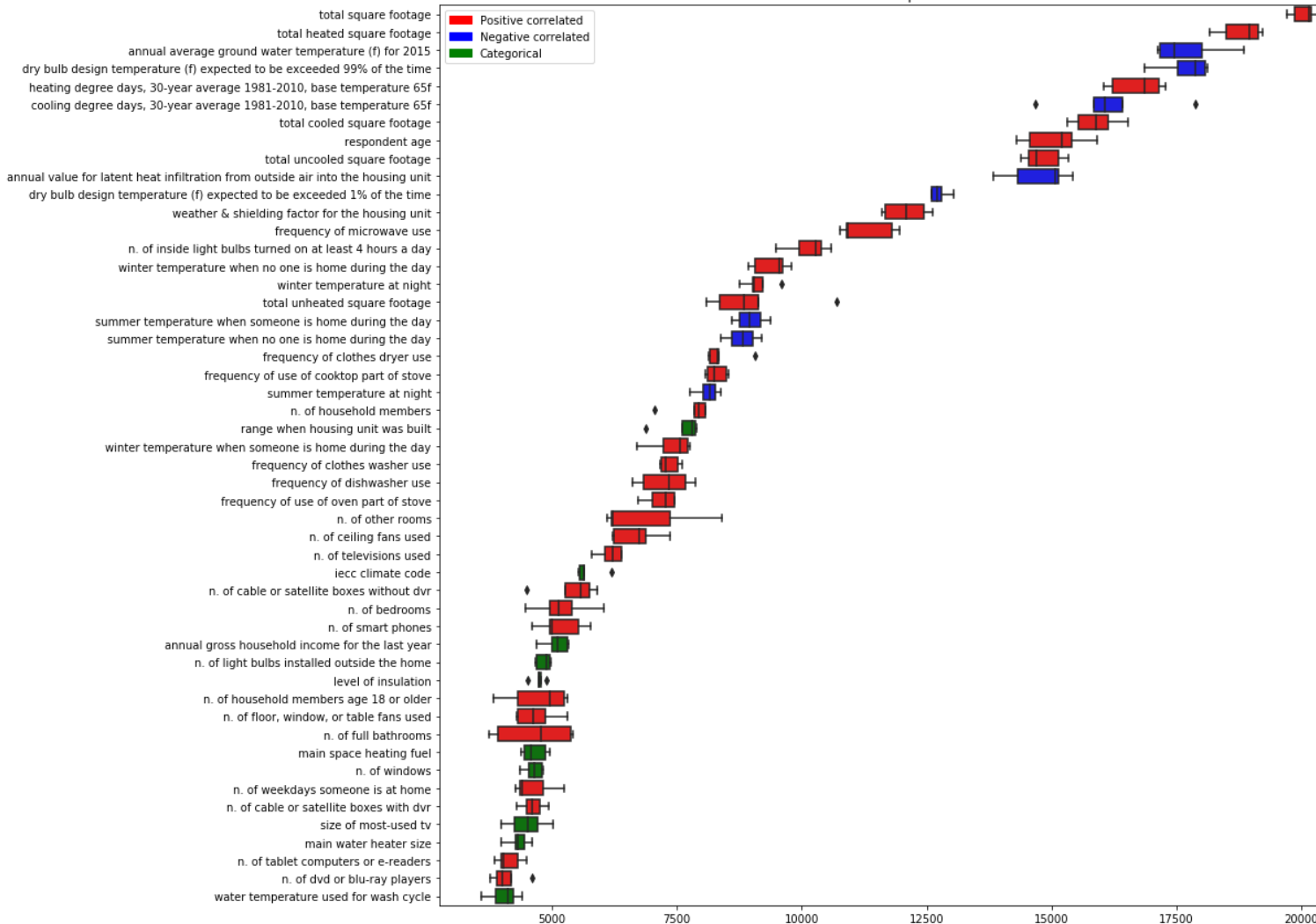
and feature engineering

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

FEATURE IMPORTANCE

based on LightGBM

Top 50 variables



- 5-fold cross-validation was used.
- Boxplot color:
 - Red - positively correlated with the outcome,
 - blue - negative,
 - and green - categorical feature.
- The spread of importance is small.



BUSINESS APPLICATION

- How much will I save per year?
- Is it worth the investment?
- What will allow to save more?

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

TO ESTIMATE SAVINGS:

1. Urban 2553 sq. ft. single family house located in very cold climate

Changing insulation from poor to good:

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

Changing insulation from adequately to good:

Absolute savings: 921.04 BTU

Percentage savings: **1.01 %**

Changing summer temperature when no one is at home from 65°F to 75°F
and installing smart thermostat:

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:

Model can be used by:

- **HVAC contractors;**
- **Companies selling/installing windows;**
- **Insulation contractors;**
- **Energy assessment companies.**

Allows making a very quick and inexpensive estimate, allowing to keep potential customers.

Traditional engineering calculations are time consuming and require a qualified engineer

TO REAL ESTATE BROKERS:

Real-estate based variables

- total sq. ft.
- total heated sq. ft.
- total cooled sq. ft.
- 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
- main space heating fuel

Real estate brokers can help customers to make more informed decisions.

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