

Energy Consumption Prediction

Climate change is a globally relevant, urgent, and multi-faceted issue heavily impacted by energy policy and infrastructure. Addressing climate change involves mitigation (i.e. mitigating greenhouse gas emissions) and adaptation (i.e. preparing for unavoidable consequences). Mitigation of GHG emissions requires changes to electricity systems, transportation, buildings, industry, and land use.

According to a report issued by the International Energy Agency (IEA), the lifecycle of buildings from construction to demolition were responsible for 37% of global energy-related CO2 emissions in 2020. Yet it is possible to drastically reduce the energy consumption of buildings by a combination of easy-to-implement fixes and state-of-the-art strategies. For example, retrofitted buildings can reduce heating and cooling energy requirements by 50-90 percent. Many of these energy efficiency measures also result in overall cost savings and yield other benefits, such as cleaner air for occupants. This potential can be achieved while maintaining the services that buildings provide.

Goal

The goal of this competition is to predict the energy consumption using building characteristics and climate and weather variables.

Data

The WiDS Datathon 2022 focuses on a prediction task involving roughly 100k observations of building energy usage records collected over 7 years and a number of states within the United States. The dataset consists of building characteristics (e.g. floor area, facility type etc), weather data for the location of the building (e.g. annual average temperature, annual total precipitation etc) as well as the energy usage for the building and the given year, measured as Site Energy Usage Intensity (Site EUI). Each row in the data corresponds to the a single building observed in a given year. Your task is to predict the Site EUI for each row, given the characteristics of the building and the weather data for the location of the building.

Setup

Load Data and Packages

```
library(gtExtras)
library(gtsummary)
library(tidyverse)
library(tidymodels)

data <- read_csv("data/buildings-data/train.csv")
```

Clean Data

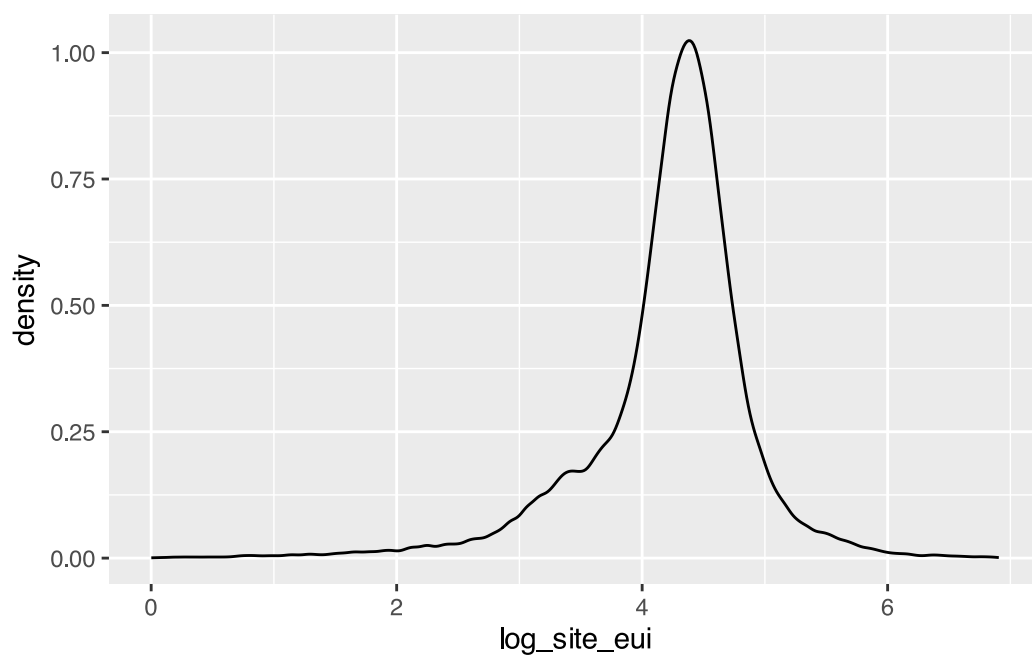
```
data <- data |>
  janitor::clean_names() |>
  mutate(
    state_factor = factor(state_factor),
    building_class = factor(building_class),
    facility_type = factor(facility_type),
    year_built = factor(year_built, ordered = TRUE),
    energy_star_rating = factor(energy_star_rating, ordered = TRUE),
    log_site_eui = log(site_eui)
  )

data |>
  gt_preview() |>
  tab_header(title = "Building Energy Dataset")
```

Building Energy Dataset

| year_ | state_factor | building_class | facility_type | year_built | energy_star_rating | log_site_eui |
|-------|--------------|----------------|---------------|------------|--------------------|--------------|
| 2016 | CA | Commercial | Warehouse | 2016 | 1 | 1.57 |
| 2017 | CA | Commercial | Warehouse | 2017 | 1 | 1.57 |
| 2018 | CA | Commercial | Warehouse | 2018 | 1 | 1.57 |
| 2019 | CA | Commercial | Warehouse | 2019 | 1 | 1.57 |
| 2020 | CA | Commercial | Warehouse | 2020 | 1 | 1.57 |
| 2021 | CA | Commercial | Warehouse | 2021 | 1 | 1.57 |
| 2022 | CA | Commercial | Warehouse | 2022 | 1 | 1.57 |
| 2023 | CA | Commercial | Warehouse | 2023 | 1 | 1.57 |
| 2024 | CA | Commercial | Warehouse | 2024 | 1 | 1.57 |
| 2025 | CA | Commercial | Warehouse | 2025 | 1 | 1.57 |
| 2026 | CA | Commercial | Warehouse | 2026 | 1 | 1.57 |
| 2027 | CA | Commercial | Warehouse | 2027 | 1 | 1.57 |
| 2028 | CA | Commercial | Warehouse | 2028 | 1 | 1.57 |
| 2029 | CA | Commercial | Warehouse | 2029 | 1 | 1.57 |
| 2030 | CA | Commercial | Warehouse | 2030 | 1 | 1.57 |
| 2031 | CA | Commercial | Warehouse | 2031 | 1 | 1.57 |
| 2032 | CA | Commercial | Warehouse | 2032 | 1 | 1.57 |
| 2033 | CA | Commercial | Warehouse | 2033 | 1 | 1.57 |
| 2034 | CA | Commercial | Warehouse | 2034 | 1 | 1.57 |
| 2035 | CA | Commercial | Warehouse | 2035 | 1 | 1.57 |
| 2036 | CA | Commercial | Warehouse | 2036 | 1 | 1.57 |
| 2037 | CA | Commercial | Warehouse | 2037 | 1 | 1.57 |
| 2038 | CA | Commercial | Warehouse | 2038 | 1 | 1.57 |
| 2039 | CA | Commercial | Warehouse | 2039 | 1 | 1.57 |
| 2040 | CA | Commercial | Warehouse | 2040 | 1 | 1.57 |
| 2041 | CA | Commercial | Warehouse | 2041 | 1 | 1.57 |
| 2042 | CA | Commercial | Warehouse | 2042 | 1 | 1.57 |
| 2043 | CA | Commercial | Warehouse | 2043 | 1 | 1.57 |
| 2044 | CA | Commercial | Warehouse | 2044 | 1 | 1.57 |
| 2045 | CA | Commercial | Warehouse | 2045 | 1 | 1.57 |
| 2046 | CA | Commercial | Warehouse | 2046 | 1 | 1.57 |
| 2047 | CA | Commercial | Warehouse | 2047 | 1 | 1.57 |
| 2048 | CA | Commercial | Warehouse | 2048 | 1 | 1.57 |
| 2049 | CA | Commercial | Warehouse | 2049 | 1 | 1.57 |
| 2050 | CA | Commercial | Warehouse | 2050 | 1 | 1.57 |

```
data |>
  ggplot(mapping = aes(x = log_site_eui)) +
  geom_density()
```



```
res <- lm(site_eui ~ facility_type, data = data)
summary(res)
```

Call:

```
lm(formula = site_eui ~ facility_type, data = data)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|--------|
| -317.99 | -22.06 | -5.52 | 13.75 | 909.55 |

Coefficients:

| | Estimate | Std. Error |
|---|----------|------------|
| (Intercept) | 31.877 | 1.182 |
| facility_type5plus_Unit_Building | 4.861 | 1.864 |
| facility_typeCommercial_Other | 60.765 | 1.707 |
| facility_typeCommercial_Unknown | 81.275 | 5.406 |
| facility_typeData_Center | 307.858 | 9.965 |
| facility_typeEducation_College_or_university | 76.752 | 1.975 |
| facility_typeEducation_Other_classroom | 37.565 | 1.443 |
| facility_typeEducation_Preschool_or_daycare | 29.097 | 5.087 |
| facility_typeEducation_Uncategorized | 14.296 | 2.264 |
| facility_typeFood_Sales | 104.916 | 6.258 |
| facility_typeFood_Service_Other | -2.640 | 12.526 |
| facility_typeFood_Service_Restaurant_or_cafeteria | 163.717 | 6.535 |
| facility_typeFood_Service_Uncategorized | 96.925 | 12.176 |

| | | |
|--|---------|--------|
| facility_typeGrocery_store_or_food_market | 209.258 | 2.701 |
| facility_typeHealth_Care_Inpatient | 216.464 | 2.804 |
| facility_typeHealth_Care_Outpatient_Clinic | 71.736 | 6.972 |
| facility_typeHealth_Care_Outpatient_Uncategorized | 158.015 | 8.650 |
| facility_typeHealth_Care_Uncategorized | 152.068 | 7.296 |
| facility_typeIndustrial | 93.468 | 2.878 |
| facility_typeLaboratory | 297.572 | 5.109 |
| facility_typeLodging_Dormitory_or_fraternity_sorority | 49.719 | 2.313 |
| facility_typeLodging_Hotel | 73.058 | 1.630 |
| facility_typeLodging_Other | 89.081 | 6.053 |
| facility_typeLodging_Uncategorized | 34.719 | 23.024 |
| facility_typeMixed_Use_Commercial_and_Residential | 57.653 | 2.309 |
| facility_typeMixed_Use_Predominantly_Commercial | 37.256 | 3.424 |
| facility_typeMixed_Use_Predominantly_Residential | 49.921 | 17.179 |
| facility_typeMultifamily_Uncategorized | 52.002 | 1.210 |
| facility_typeNursing_Home | 99.437 | 2.196 |
| facility_typeOffice_Bank_or_other_financial | 58.019 | 4.084 |
| facility_typeOffice_Medical_non_diagnostic | 84.885 | 2.704 |
| facility_typeOffice_Mixed_use | 50.233 | 12.176 |
| facility_typeOffice_Uncategorized | 45.197 | 1.268 |
| facility_typeParking_Garage | 35.474 | 3.454 |
| facility_typePublic_Assembly_Drama_theater | 49.040 | 6.258 |
| facility_typePublic_Assembly_Entertainment_culture | 87.023 | 5.043 |
| facility_typePublic_Assembly_Library | 73.972 | 4.233 |
| facility_typePublic_Assembly_Movie_Theater | 71.218 | 8.317 |
| facility_typePublic_Assembly_Other | 94.827 | 4.474 |
| facility_typePublic_Assembly_Recreation | 83.301 | 6.174 |
| facility_typePublic_Assembly_Social_meeting | 47.045 | 5.607 |
| facility_typePublic_Assembly_Stadium | 125.165 | 17.179 |
| facility_typePublic_Assembly_Uncategorized | 30.996 | 10.351 |
| facility_typePublic_Safety_Courthouse | 71.305 | 8.424 |
| facility_typePublic_Safety_Fire_or_police_station | 99.246 | 4.270 |
| facility_typePublic_Safety_Penitentiary | 139.028 | 8.535 |
| facility_typePublic_Safety_Uncategorized | 51.832 | 7.929 |
| facility_typeReligious_worship | 12.684 | 2.832 |
| facility_typeRetail_Enclosed_mall | 69.089 | 4.840 |
| facility_typeRetail_Strip_shopping_mall | 78.542 | 4.979 |
| facility_typeRetail_Uncategorized | 49.026 | 1.933 |
| facility_typeRetail_Vehicle_dealership_showroom | 14.764 | 6.093 |
| facility_typeService_Drycleaning_or_Laundry | 10.236 | 17.179 |
| facility_typeService_Uncategorized | 81.689 | 6.346 |
| facility_typeService_Vehicle_service_repair_shop | 105.719 | 4.533 |
| facility_typeWarehouse_Distribution_or_Shipping_center | 7.683 | 2.403 |
| facility_typeWarehouse_Nonrefrigerated | 6.332 | 1.872 |
| facility_typeWarehouse_Refrigerated | 64.648 | 4.979 |
| facility_typeWarehouse_Selfstorage | -10.288 | 2.445 |
| facility_typeWarehouse_Uncategorized | 4.060 | 3.067 |
| t value Pr(> t) | | |

| | | | |
|---|--------|----------|-----|
| (Intercept) | 26.975 | < 2e-16 | *** |
| facility_type5plus_Unit_Building | 2.608 | 0.009101 | ** |
| facility_typeCommercial_Other | 35.607 | < 2e-16 | *** |
| facility_typeCommercial_Unknown | 15.035 | < 2e-16 | *** |
| facility_typeData_Center | 30.893 | < 2e-16 | *** |
| facility_typeEducation_College_or_university | 38.866 | < 2e-16 | *** |
| facility_typeEducation_Other_classroom | 26.038 | < 2e-16 | *** |
| facility_typeEducation_Preschool_or_daycare | 5.720 | 1.07e-08 | *** |
| facility_typeEducation_Uncategorized | 6.315 | 2.72e-10 | *** |
| facility_typeFood_Sales | 16.765 | < 2e-16 | *** |
| facility_typeFood_Service_Other | -0.211 | 0.833054 | |
| facility_typeFood_Service_Restaurant_or_cafeteria | 25.054 | < 2e-16 | *** |
| facility_typeFood_Service_Uncategorized | 7.960 | 1.74e-15 | *** |
| facility_typeGrocery_store_or_food_market | 77.465 | < 2e-16 | *** |
| facility_typeHealth_Care_Inpatient | 77.211 | < 2e-16 | *** |
| facility_typeHealth_Care_Outpatient_Clinic | 10.290 | < 2e-16 | *** |
| facility_typeHealth_Care_Outpatient_Uncategorized | 18.267 | < 2e-16 | *** |
| facility_typeHealth_Care_Uncategorized | 20.843 | < 2e-16 | *** |
| facility_typeIndustrial | 32.481 | < 2e-16 | *** |
| facility_typeLaboratory | 58.244 | < 2e-16 | *** |
| facility_typeLodging_Dormitory_or_fraternity_sorority | 21.499 | < 2e-16 | *** |
| facility_typeLodging_Hotel | 44.824 | < 2e-16 | *** |
| facility_typeLodging_Other | 14.716 | < 2e-16 | *** |
| facility_typeLodging_Uncategorized | 1.508 | 0.131577 | |
| facility_typeMixed_Use_Commercial_and_Residential | 24.971 | < 2e-16 | *** |
| facility_typeMixed_Use_Predominantly_Commercial | 10.881 | < 2e-16 | *** |
| facility_typeMixed_Use_Predominantly_Residential | 2.906 | 0.003663 | ** |
| facility_typeMultifamily_Uncategorized | 42.986 | < 2e-16 | *** |
| facility_typeNursing_Home | 45.289 | < 2e-16 | *** |
| facility_typeOffice_Bank_or_other_financial | 14.207 | < 2e-16 | *** |
| facility_typeOffice_Medical_non_diagnostic | 31.395 | < 2e-16 | *** |
| facility_typeOffice_Mixed_use | 4.126 | 3.70e-05 | *** |
| facility_typeOffice_Uncategorized | 35.645 | < 2e-16 | *** |
| facility_typeParking_Garage | 10.271 | < 2e-16 | *** |
| facility_typePublic_Assembly_Drama_theater | 7.836 | 4.69e-15 | *** |
| facility_typePublic_Assembly_Entertainment_culture | 17.257 | < 2e-16 | *** |
| facility_typePublic_Assembly_Library | 17.475 | < 2e-16 | *** |
| facility_typePublic_Assembly_Movie_Theater | 8.563 | < 2e-16 | *** |
| facility_typePublic_Assembly_Other | 21.197 | < 2e-16 | *** |
| facility_typePublic_Assembly_Recreation | 13.493 | < 2e-16 | *** |
| facility_typePublic_Assembly_Social_meeting | 8.391 | < 2e-16 | *** |
| facility_typePublic_Assembly_Stadium | 7.286 | 3.23e-13 | *** |
| facility_typePublic_Assembly_Uncategorized | 2.995 | 0.002750 | ** |
| facility_typePublic_Safety_Courthouse | 8.465 | < 2e-16 | *** |
| facility_typePublic_Safety_Fire_or_police_station | 23.242 | < 2e-16 | *** |
| facility_typePublic_Safety_Penitentiary | 16.289 | < 2e-16 | *** |
| facility_typePublic_Safety_Uncategorized | 6.537 | 6.33e-11 | *** |
| facility_typeReligious_worship | 4.478 | 7.54e-06 | *** |

| | | | |
|--|--------|----------|-----|
| facility_typeRetail_Enclosed_mall | 14.274 | < 2e-16 | *** |
| facility_typeRetail_Strip_shopping_mall | 15.775 | < 2e-16 | *** |
| facility_typeRetail_Uncategorized | 25.365 | < 2e-16 | *** |
| facility_typeRetail_Vehicle_dealership_showroom | 2.423 | 0.015384 | * |
| facility_typeService_Drycleaning_or_Laundry | 0.596 | 0.551299 | |
| facility_typeService_Uncategorized | 12.872 | < 2e-16 | *** |
| facility_typeService_Vehicle_service_repair_shop | 23.320 | < 2e-16 | *** |
| facility_typeWarehouse_Distribution_or_Shipping_center | 3.197 | 0.001387 | ** |
| facility_typeWarehouse_Nonrefrigerated | 3.383 | 0.000716 | *** |
| facility_typeWarehouse_Refrigerated | 12.984 | < 2e-16 | *** |
| facility_typeWarehouse_Selfstorage | -4.208 | 2.58e-05 | *** |
| facility_typeWarehouse_Uncategorized | 1.324 | 0.185652 | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 51.42 on 75697 degrees of freedom

Multiple R-squared: 0.2217, Adjusted R-squared: 0.221

F-statistic: 365.4 on 59 and 75697 DF, p-value: < 2.2e-16

Bibliography