Predicting Building Energy Consumption

By: Melissa Barona

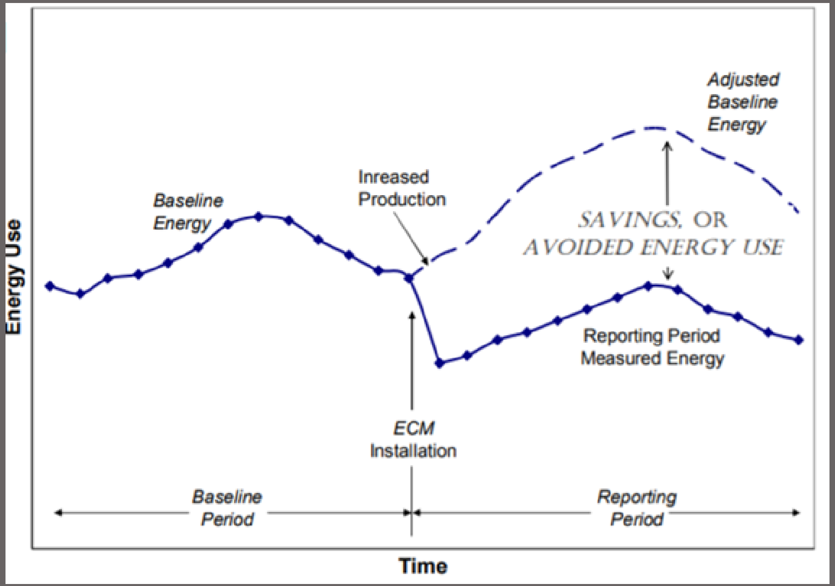
Springboard Capstone Project 2

Problem Identification

Problem → **Solution**

Facility owners must present accurate projected energy savings to enhance their chances of being financed by energy savings programs

Predict what the energy consumption baseline would be prior to applying an energy efficiency measure to estimate energy savings



Key Findings

- Dataset was obtained from the Great Energy Predictor III Kaggle competition by ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers)
- For this dataset, both linear regression and random forest models found that size of the building is the most important feature for predicting energy usage
- The random forest regression model predicts energy consumption the best
  - RMSE: 205 kWh
  - MAE: 85 kWh
- This is an improvement, compared to using the mean as a predictor
  - RMSE: 376 kWh
  - MAE: 179 kWh

Objectives

Details of the Dataset

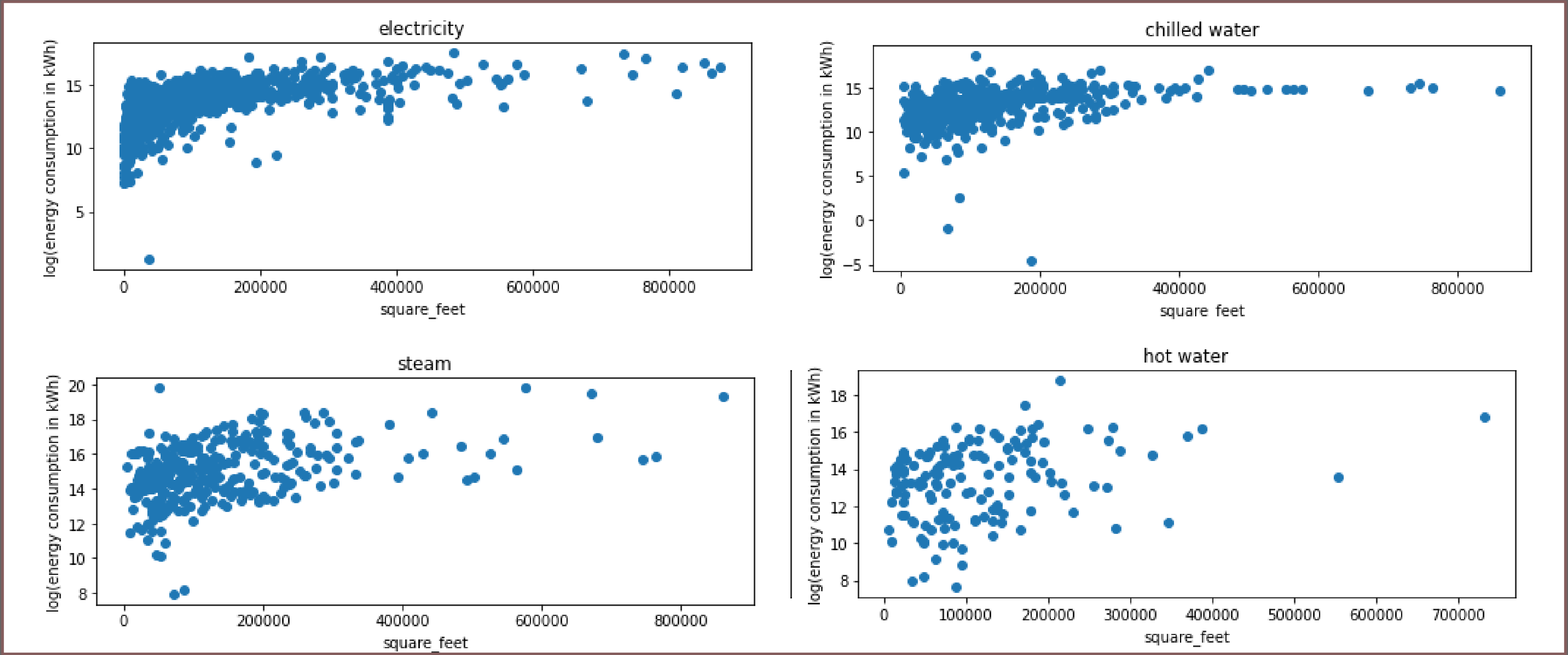
- **3 datasets from**
  - Details about the buildings
    - Size, floor count, year built, location
  - Weather features
    - Air and dew temperature
    - Sea level pressure
    - Wind speed and direction
    - Cloud coverage
    - Precipitation
  - Meter readings for each hour of year 2016 for each building
    - Electricity, hot water, steam, chilled water

Models

- The performance of the following models were evaluated
  - Linear Regression
  - Ridge Regression
  - **Random Forest Regression**

Exploratory Data Analysis

- **<u>Size of the building:</u>** Larger buildings consume more energy

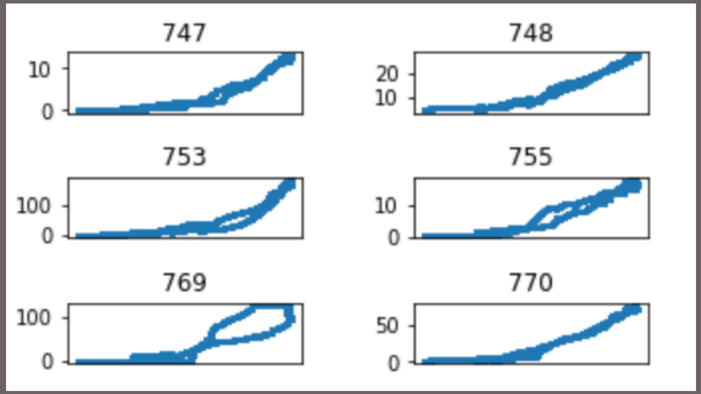


Exploratory Data Analysis

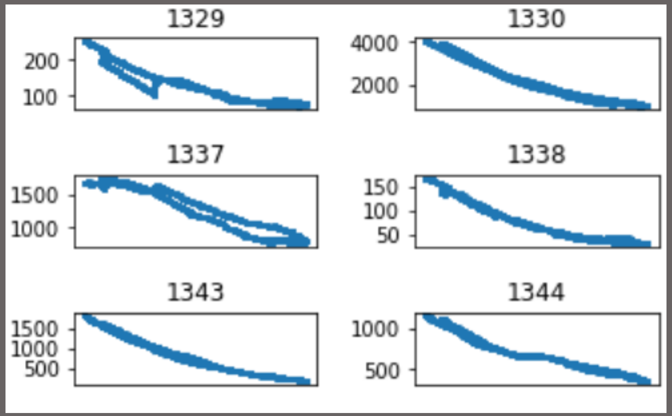
- **<u>Air and dew temperature:</u>** At warmer temperatures, AC will be used, driving up energy consumption. At colder temperatures, the heater will be used, driving up energy consumption

Energy Usage (kWh)

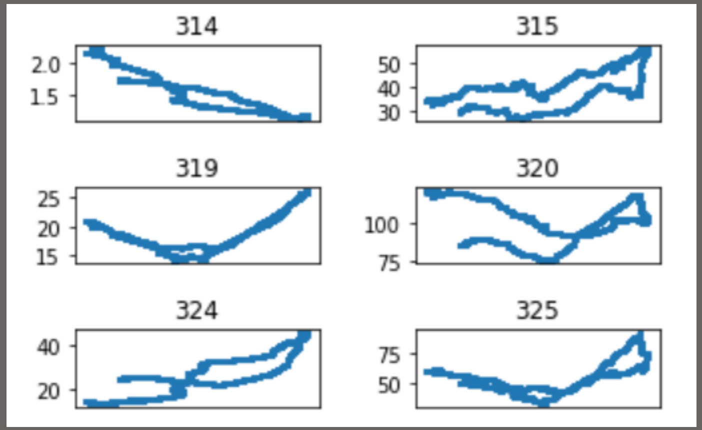
Chilled Water



Steam



Electricity



Air Temperature (°C)

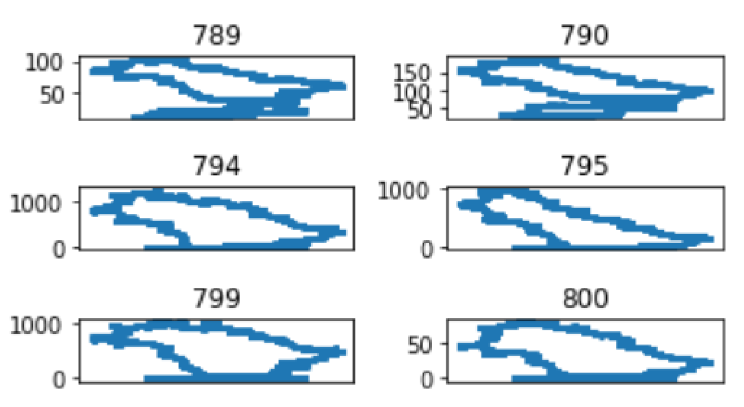
Exploratory Data Analysis

Sea level/air pressure: Air pressure is higher during winter because colder air is denser.

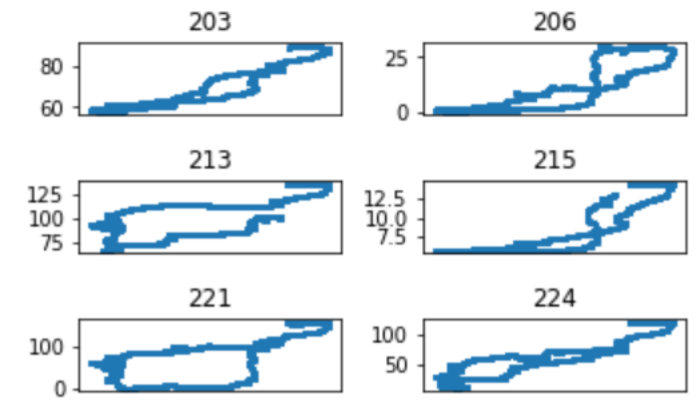
- Low pressure during summer leads to higher energy consumption in chilled water.
- High pressure during winter should lead to higher energy consumption in steam and heat.

Energy Usage (kWh)

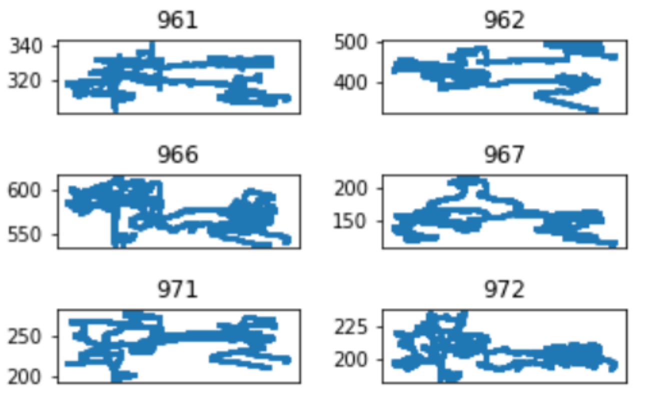
Chilled Water



Hot Water



Electricity



Sea Level Pressure (millibar)

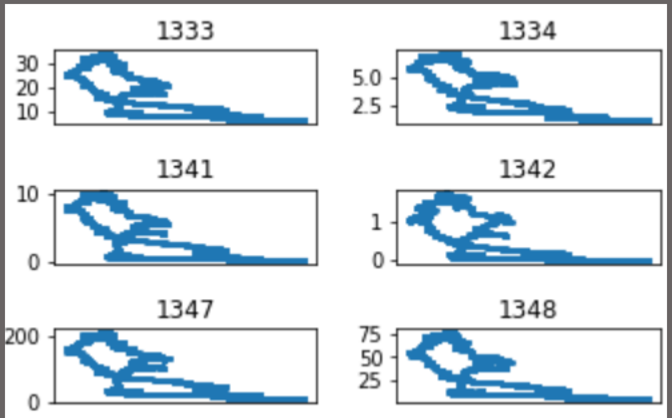
Exploratory Data Analysis

<u>Wind speed and wind direction:</u> Higher wind speed will cool the building.

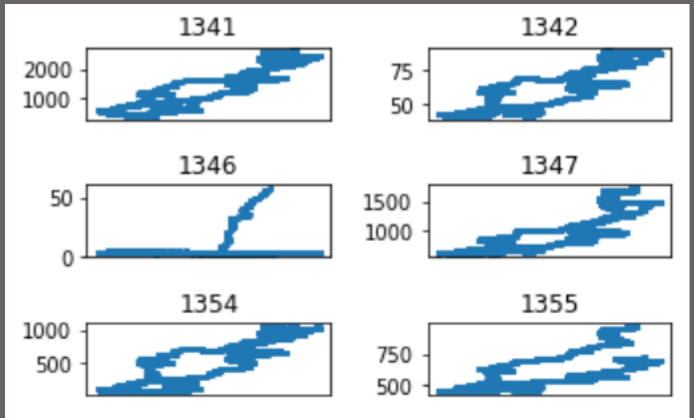
- In the summer it helps lower energy consumption, but in the winter, it may lead to higher energy consumption.
- If wind speed is low, wind direction won't matter, but if wind speed is high, wind direction will matter more and its effect on energy consumption will depend on the position of the building relative to the direction of the wind.

Energy Usage (kWh)

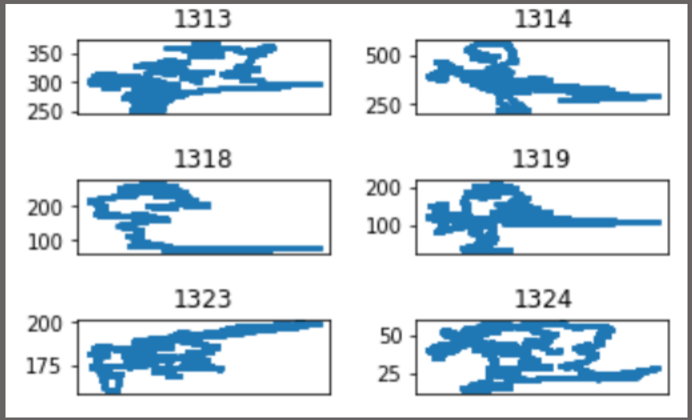
Chilled Water



Steam



Electricity



Wind Speed

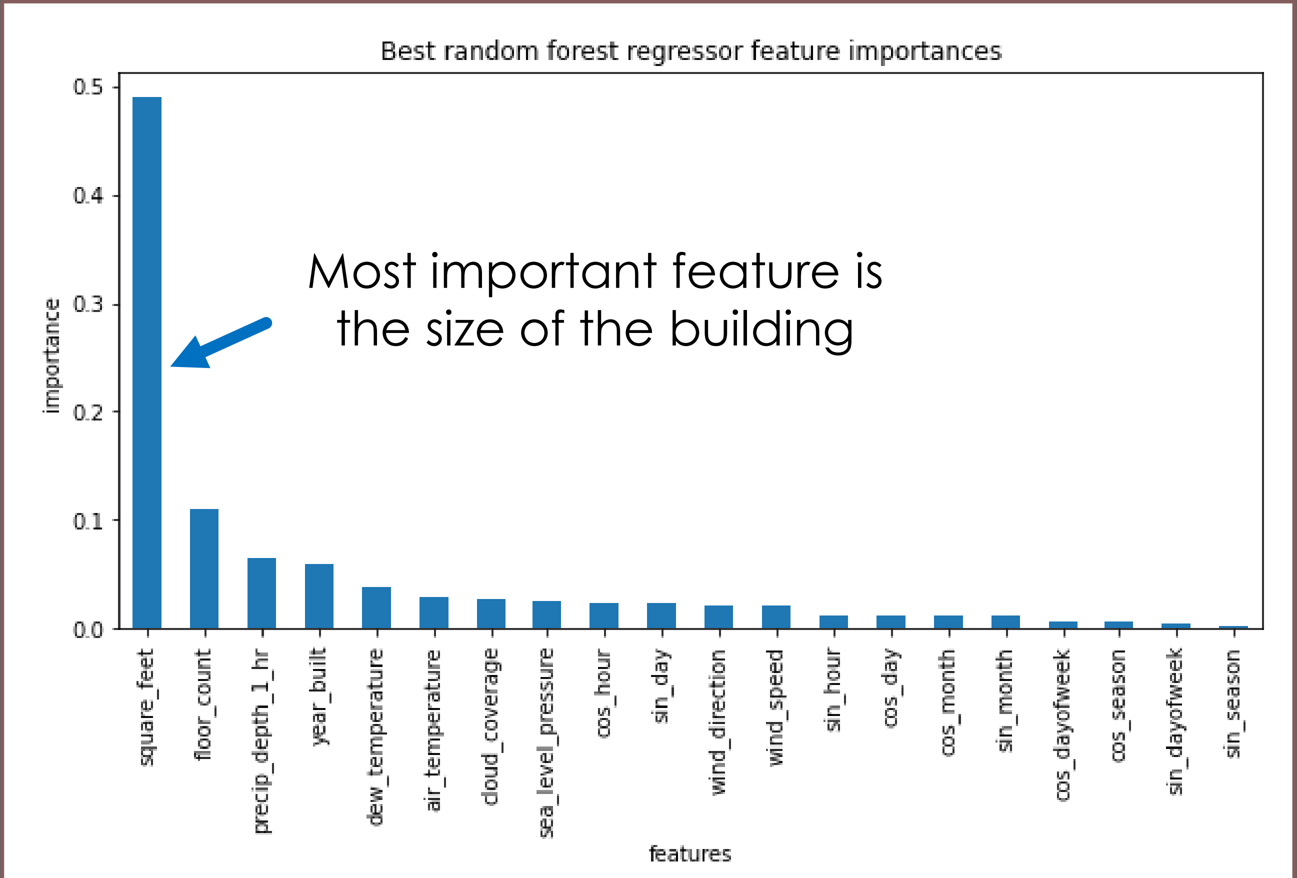
Preprocessing and Feature Engineering

- Missing data in the weather features were filled in by averaging out the data over time using the moving average.
- Season, month, day of week, and hour were added as features and transformed into cyclic features using sine and cosine functions.
- This dataset contains ~ 20 M rows. Subsample of a 100,000 rows for the linear regression models, and 10,000 rows for the random forest regression model.

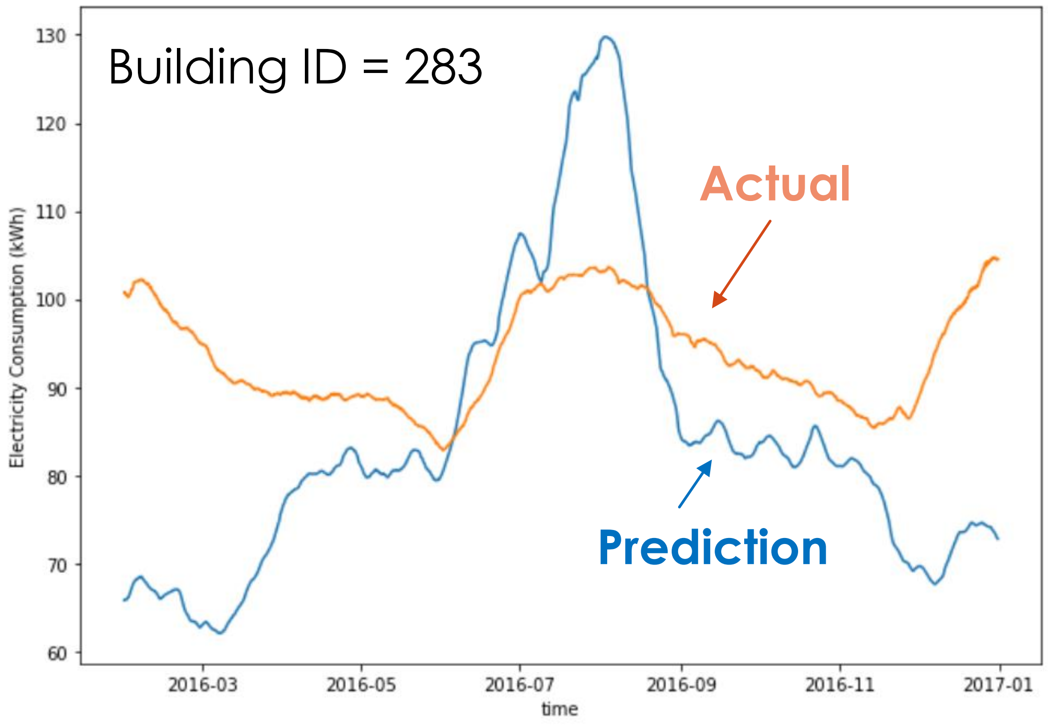
Modeling

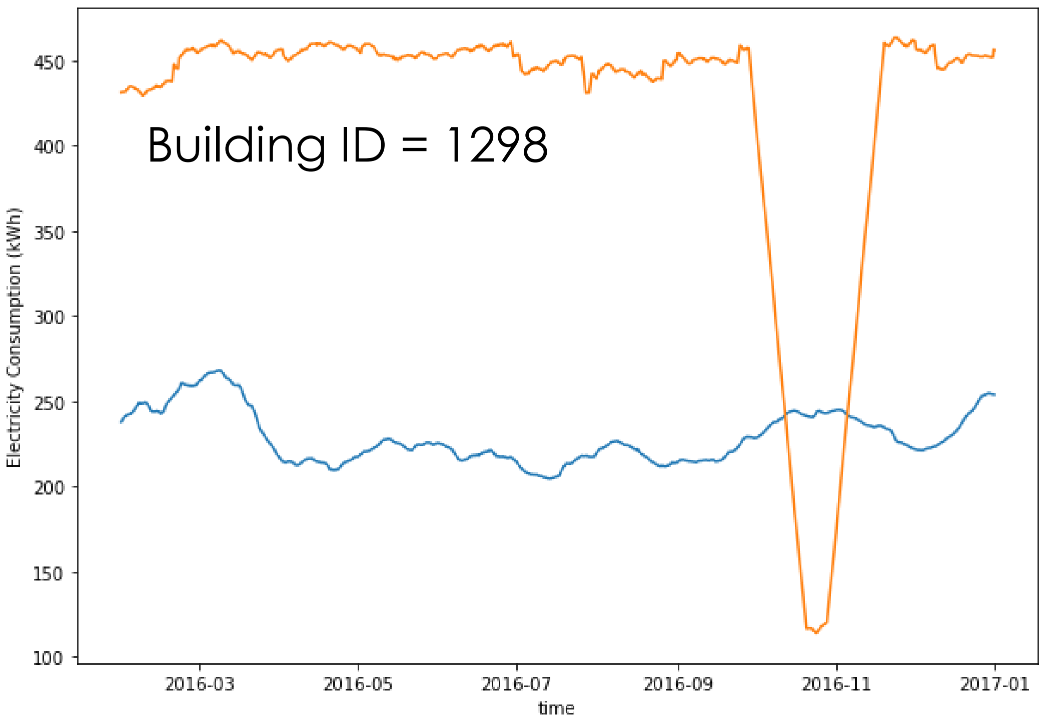
Random Forest Model was found to be the best model

Model	RMSE (kWh)		MAE (kWh)	
	Training Set	Testing Set	Training Set	Testing Set
Dummy Regressor	367	374	176	179
Linear Regression	304	307	128	129
Ridge Regression	304	307	128	129
Random Forest Regression	82	205	32	85



Prediction


Building ID = 283
Actual
Prediction
Electricity Consumption (kWh)
130
120
110
100
90
80
70
60
2016-03
2016-05
2016-07
2016-09
2016-11
2017-01
time


Building ID = 1298
Electricity Consumption (kWh)
450
400
350
300
250
200
150
100
2016-03
2016-05
2016-07
2016-09
2016-11
2017-01
time

Future Improvements

- **Detect outliers in the meter readings and address them appropriately**
  - Use the hampel identifier
    - uses moving estimates such as the rolling average to detect outliers in a time series and replace them with the average
- **Need to reproduce the seasonality in the energy consumption**
  - Better feature engineering is needed
    - fit a model for a single building to get an idea of the most important weather features to predict energy consumption
  - Use a model that regulates the size of the building as a feature and gives more importance to the weather features