**Kaggle Link:**

[**https://www.kaggle.com/c/ashrae-energy-prediction/data**](https://www.kaggle.com/c/ashrae-energy-prediction/data)

**Problem statement:**

Develop accurate models of metered building energy usage in the following areas: chilled water, electric, hot water, and steam meters.

**Data:**

The data comes from over 1,000 buildings over a three-year timeframe. With better estimates of these energy-saving investments, large scale investors and financial institutions will be more inclined to invest in this area to enable progress in building efficiencies.

**Background:**

Current methods to reduce costs and emissions of energy improve building efficiencies are not enough. Under pay-for-performance financing, the building owner makes payments based on the difference between their real energy consumption and what they would have used without any retrofits. The latter values have to come from a model. Methods in use of estimation are fragmented and do not scale well. Some assume a specific meter type or don’t work with different building types.

Replica of competition : <https://www.kaggle.com/c/great-energy-predictor-shootout-i/overview>

Data :

1. Train
2. Weather\_train
3. Building\_metadata
4. Test
5. Weather\_test

**5/11/2019**

1. **Download data**
2. **Understand data**
3. **Calculate missing values**
4. **Calculated missing value**
   1. Cloud\_coverage (35.98%)
   2. Precip\_depth\_1\_hr (49.49%)
   3. Year\_built (53.41%)
   4. Floor\_count (75%)

**10/11/2019**

**Missing value imputation**

1. Floor\_count (Method - Linear Algebra, KNN, MICE, DataWig)
2. Year\_built ( Map city to Site ID by looking at weather data (Code from Kaggle) Kaggle code: https://www.kaggle.com/patrick0302/locate-cities-according-weather-temperature
3. Weather data (In discussion)

**Plots**

1. Frequency plot for meter\_reading for meter type =0 (electricity), 1(chilled water), 2(steam), 3(hot water) - Discussions about their value

**Github Account setup**

**13/11/2019**

1. **Python code explanation for floor\_count**
2. Repository for Ashrae energy predictors (<https://github.com/grapestone5321/Great_Energy_Predictor_Shootout_1_2/tree/master/papers>)
3. Discussion about future meetings and set up

**17/11/2019**

1. KNN method explanation by Michael

Distances- 1. Euclidean distance 2. Manhattan distance 3. Humming distance

1. Weather data visualization (Not done)
2. Timestamp data preprocessing
   1. Day, date, hour, minutes, month, year
3. Weather data imputation
   1. For 7 variables from weather data by method rolling mean, median, max, min for - groupby site id for 3, 5, 7, 14, 30 days

**20/11/2019**

1. Download test data set and do some exploratory data analysis
2. Transform test set as train set (same imputation and preprocessing)
3. Combine them to do one-hot encoding
4. Understand bagging and boosting methods
5. Understand parameters in Python for Linear regression, Random forest regression and XGBoost regression

**10/12/2019**

1. Weather data imputation (Completed)
   1. Using a loop of 3 methods (3-day rolling mean, reverse 3-day rolling mean and average of column by timestamp)
2. Data preparation for model 1 (PCA)
   1. Merging data
   2. Feature engineering
3. Exploratory data analysis