November 1, 2019

**Group 2**

Jinne Bak

Elliot Church

Adam Hubbell

Omar Malik

**STA 631: Research Proposal**

**Research Question:** How much energy would a building have used if it had not been overhauled with the aim of reducing its energy use?

**Context:** This is an open data science competition hosted on Kaggle. This competition is hosted by ASHRAE, which is a professional association that serves to advance knowledge in heating, ventilation, air-conditioning, refrigeration and associated domains.

Here is a link to the competition website:

<https://www.kaggle.com/c/ashrae-energy-prediction/overview>

**Data:** We are provided with data about the building and weather conditions. We have to use these variables to predict energy usage in kilowatt hours (kWh).

More accurate models can lead to better estimates of how much overhauls would save in terms of energy and lead to better incentives and lower costs for carrying out overhauls.

**Evaluation metric:** The competition evaluation metric is Root Mean Squared Logarithmic Error. This is just an extension of RMSE that we have discussed in class.

**Why is this a good project?** This is a good project for this class because:

1. It is a regression problem and clearly a predictive analytics problem.
2. It has real-world consequences in terms of our impact on the environment.
3. We can use all the tools learnt in this class: histograms, densities, correlations, regression modeling, model selection, model evaluation among others.
4. It is a focused problem and it can be addressed over the remainder of the semester.
5. It is a Big Data problem with more than 1,400 buildings. The dataset is more than 20 million rows and 22 columns.

**Our Goal:** Our goal for this class is to meet project requirements and also make a submission for this competition by December 12, 2019.