## Project: Energy Efficiency

Student: Huynh Vo

## One page summary:

My topic, and this dataset, are related to Energy Efficiency, particularly the energy efficiency of housing. The dataset was obtained from [http://archive.ics.uci.edu/ml/datasets/energy+efficiency#](http://archive.ics.uci.edu/ml/datasets/energy+efficiency). The dataset relates how the size, shape, and other features of a building can help or hurt the energy required to keep the internal temperature of a building, such as a home, at a desired temperature. This is an important real world problem that our energy hungry society is only now beginning to look at in detail using data science tools. Some of the tools used to explore the data presented were different visualization graphs provided through RStudio tools and Azure Machine Learning for a more in depth analysis.

The machine learning tools were required to predict the responses of the heating and cooling loads for different building parameters such as relative compactness, surface area, glazing area, etc. Azure Machine Learning and R are the main tools for this analysis. Through exploration of the data using these tools I have concluded that many of the factors in the data help to determine the energy required to heat and cool a building, but not all of them do. I have used rcorr to see the correlation between given features. The Rcorr function showed that there is a positive correlation between heating load and cooling load, meaning if the house does not require much heat to warm up, it does not require much A/C to cool down either. Through histograms and box plots, created using the ggplot package, I also see that Relative Compactness, Wall Area, the number of stories of the building, and glazing area play a major role in determining the heating load and cooling load.

Besides visualization and after cleaning up the data, I tested out Azure Machine Learning using a Neural Network Regression model and a Boosted Decision Tree regression model to predict the outcome of heating load and cooling loads. The models were evaluated with a high Coefficient of Determination ( ~ 0.98). Azure Machine Learning is one of the more powerful tools for model prediction. Azure Machine is easy to use, the models are loaded automatically, are highly visualized, and most importantly, it requires minimal to no installation. Azure Studio is an online platform owned by Microsoft. They do very well in terms of organization and has a very gentle learning curve. It is open source. The only con is that for a free account, there is a limit for storage and speed when training the data. If you would like to analyze data that is bigger than 10MB, you will have to purchase the premium account and your data will be processed with faster speed.

Youtube URLs:

2 minute (short): <https://youtu.be/bIL9AZTkGKU>

15 minutes (long):  <https://youtu.be/2sjkyZTjbrs>