

Modelling Building Energy Efficiency in New York

07/23/2017





1. Introduction

The following report will evaluate data from the IBM Data Science Experience: “*Modeling energy usage in New York City*” and New York State public data. The purpose of the analysis will be to classify 103 buildings based on the efficient consumption of energy and to create a model to identify inefficient energy buildings. Inefficient buildings identified may be able to take corrective actions which translate to annual energy savings.

This report will also analyze energy use Intensity for buildings required to benchmark their energy consumption in accordance to a city law placed in 2012, designed to cut emissions. Based on the findings of this report, recommendations are provided to the City of New York as well as building managers.

2. Data

Data comes from four sources in csv and excel files: BlocPower, public heating and cooling data for 103 buildings in New York City for 2016, NYC OpenData, and *freegeocoding.com* for coordinate locations of buildings. BlocPower is a New York company that analyzes building data and connects investors to green building projects to save energy. The following tables describe the data coming from csv and xlsx files. It is important to note that the buildings data provided by BlocPower are not drawn from a random sample, but will be treated as such for the purpose of this analysis. Buildings from NYC Open Data are all buildings required to benchmark energy data per law.

Source	CSV	Fields
 BlocPower	BlocPower_T.csv	UTSUM_Electricity_Usage, INFO_Year of Construction, INFO_Number of Stories, INFO_Total Square Feet, PLEI_1_Quantity, PLEI_3_Quantity
 BlocPower	clusterEnergyLocation.csv	AddressID, property_name, Address, Zipcode, Long, Lat, Annual Energy Bill (USD)
Public Heating & Cooling	CDD-HDD-Features.csv	Property Name, plug_load_consumption, ac_consumption, domestic_gas, heating_gas
 NYC OpenData	NYC_Municipal_Building_Energy_Benchmarking_Results__2014_.csv	Borough, Block, Lot, BIN, Building, Agency, 2010 Score *, 2010Source EUI (kBtu/ft ²)*, 2010 GHG Emissions Intensity (kgCO ₂ e/ft ²)*, 2014 Score *, 2014 Source EUI (kBtu/ft ²)*, 2014 GHG Emissions Intensity (kgCO ₂ e/ft ²)*
 NYC OpenData	2012_nyc_cy2011_1184_disclosure_data.xlsx	BBL, Street Number, Street Name, Borough, Zip, Benchmarking Submission, Entry Number, Site EUI, Weather

		Normalized Source EUI, Water per Square Foot, ENERGY STAR Score GHG, Reported Building Square Footage, Reported Facility Type, Number of Buildings, Reported BINs
NYC OpenData	2013_nyc_1184_disclosure.xlsx	Borough, Zip, Benchmarking Submission, Entry Number, "Site EUI (kBtu/ft2)", "Weather Normalized Source EUI (kBtu/ft2)", "Indoor Water Intensity (All Water Sources) (gal/ft2)", Reported Water Method, ENERGY STAR Score, "Total GHG Emissions (MtCO2e)", "Property Floor Area (Buildngs and Parking) (ft2)", Primary Property Type - Self Selected, Number of Buildings, Reported BINs
NYC OpenData	150428_2014_nyc_1184_disclosure.xlsx	Street Name, Borough, Zip Code, DOF Benchmarking Submission Status, "Site EUI (kBtu/ft2)", "Weather Normalized Site EUI (kBtu/ft2)", "Source EUI (kBtu/ft2)", "Weather Normalized Source EUI (kBtu/ft2)", Municipally Supplied Potable Water - Indoor Intensity (gal/ft ²), Automatic Water Benchmarking Eligible, Reported Water Method, ENERGY STAR Score, "Total GHG Emissions (MtCO2e)", "Direct GHG Emissions (MtCO2e)", "Indirect GHG Emissions (MtCO2e)", Reported Property Floor Area (Building(s)) (ft ²), "DOF Property Floor Area (Buildngs and Parking) (ft2)", Primary Property Type - Self Selected, DOF Number of Buildings
NYC OpenData	2015_nyc_cy2014_1184_disclosure_data.xlsx	Record Number, NYC Borough, Block, and Lot (BBL), Co-reported BBL Status, BBLs Co-reported, Reported NYC Building Identification Numbers (BINs), Street Number, Street Name, Borough, Zip Code, BBL on the Covered Buildings List, DOF Benchmarking Submission Status, "Site EUI (kBtu/ft2)", "Weather Normalized Site EUI (kBtu/ft2)", "Source EUI (kBtu/ft2)", "Weather Normalized Source EUI (kBtu/ft2)", Municipally Supplied Potable Water - Indoor Intensity (gal/ft ²), Automatic Water Benchmarking Eligible, Reported Water Method, ENERGY STAR Score, "Total GHG Emissions

		(MtCO2e)", "Direct GHG Emissions (MtCO2e)", "Indirect GHG Emissions (MtCO2e)", Reported Property Floor Area (Building(s)) (ft ²), "DOF Property Floor Area (Buildngs and Parking) (ft2)", Primary Property Type - Self Selected, DOF Number of Buildings
NYC OpenData	nyc_benchmarking_disclosure_data_reported_in_2016.xlsx	Record Number, NYC Borough, Block, and Lot (BBL), Co-reported BBL Status, BBLs Co-reported, Reported NYC Building Identification Numbers (BINs), Property Name, Parent Property Id, Parent Property Name, Street Number, Street Name, Zip Code, Borough, DOF Benchmarking, Submission Status, Primary Property Type - Self Selected, List of All Property Use, Types at Property, Largest Property Use Type, Largest Property Use Type - Gross Floor Area (ft ²), 2nd Largest Property Use Type, 2nd Largest Property Use - Gross Floor Area (ft ²), 3rd Largest Property Use Type, 3rd Largest Property Use Type - Gross Floor Area (ft ²), Year Built, Number of Buildings - Self-reported, Occupancy, Metered Areas (Energy), Metered Areas (Water), ENERGY STAR Score, Site EUI (kBtu/ft ²), Weather Normalized Site EUI (kBtu/ft ²), Weather Normalized Site Electricity Intensity (kWh/ft ²), Weather Normalized Site Natural Gas Intensity (therms/ft ²), Source EUI (kBtu/ft ²), Weather Normalized Source EUI (kBtu/ft ²), Fuel Oil #1 Use (kBtu), Fuel Oil #2 Use (kBtu), Fuel Oil #4 Use (kBtu), Fuel Oil #5 & 6 Use (kBtu), Diesel #2 Use (kBtu), District Steam Use (kBtu), District Hot Water Use (kBtu), District Chilled Water Use (kBtu), Natural Gas Use (kBtu), Weather Normalized Site Natural Gas Use (therms), Electricity Use - Grid Purchase (kBtu), Weather Normalized Site Electricity (kWh), Total GHG Emissions (Metric Tons CO2e), Direct GHG Emissions (Metric Tons CO2e), Indirect GHG Emissions (Metric Tons CO2e), DOF Property Floor Area (ft ²), Property GFA - Self-reported (ft ²), Water Use (All Water Sources) (kgal), Municipally Supplied Potable Water - Indoor Intensity (gal/ft ²), Release Date, DEP Provided Water Use (kgal), Automatic Water Benchmarking Eligible, Reported Water Method

3. Data Preparation

Due to missing values and wrong data types, the data was cleaned up and prepared. The table below describes the data transformation that took place before analysis.

BlocPower_T.csv Transformation to DataFrame			
Column Name	Info	Rename	Summary of Data Transformation
UTSUM Electricity Usage	98 non-null object	Energy Usage	Remove unwanted characters, change data type to float and filled NaN with mean values.
INFO Year of Construction	100 non-null object	Year of Construction	Convert to float type.
INFO Number of Stories	103 non-null int64	Number of Stories	Leave as is.
INFO Total Square Feet	103 non-null object	Square Feet	Remove unwanted characters, change data type to float.
PLEI 1 Quantity	95 non-null float64	PLEI 1	NaN values, interpreted as 0 plugged in electrical equipment, so fill NaNs with 0.
PLEI 3 Quantity	88 non-null object	PLEI 3	Convert column to float type and fill NaNs with 0.

clusterEnergyLocation.csv Transformation to DataFrame			
Column Name	Info	Rename	Summary of Data Transformation
AddressID	103 non-null object	Address ID	Leave as is.
property_name	103 non-null object	Property Name	Leave as is.
Address	103 non-null object	Address	Leave as is.
Zipcode	103 non-null int64	Zipcode	Leave as is.
Long	103 non-null float64	Longitude	Leave as is.
Lat	103 non-null float64	Latitude	Leave as is.
Annual Energy Bill (USD)	103 non-null object	Annual Energy Bill (USD)	Remove unwanted characters and change data type to float.

CDD-HDD-Features.csv Transformation to DataFrame			
Column Name	Info	Rename	Summary of Data Transformation
Property Name	103 non-null object	Property Name	Leave as is.
plug_load_consumption	103 non-null float64	Plug Load Consumption	Leave as is.
ac_consumption	103 non-null float64	AC Consumption	Leave as is.
domestic_gas	103 non-null float64	Domestic Gas	Leave as is.
heating_gas	103 non-null float64	Heating Gas	Leave as is.

2012_nyc_cy2011_l184_disclosure_data.xlsx Transformation to DataFrame			
Column Name	Info	Rename	Summary of Data Transformation
BBL	3875 non-null int64	BBL	Leave as is.
Street Number	3828 non-null object	Street Number	Leave as is.
Street Name	3875 non-null object	Street Name	Leave as is.
Borough	3875 non-null object	Borough	Leave as is.
Zip	3875 non-null int64	Zip	Leave as is.
Site EUI	2126 non-null float64	Site EUI 2012	Leave as is.

2013_nyc_l184_disclosure.xlsx Transformation to DataFrame			
Column Name	Info	Rename	Summary of Data Transformation
BBL	13196 non-null int64	BBL	Leave as is.
Street Number	13145 non-null object	Street Number	Leave as is.
Street Name	3875 non-null object	Street Name	Leave as is.
Borough	13190 non-null object	Borough	Leave as is.
Zip	13196 non-null int64	Zip	Leave as is.
Site EUI (kBtu/ft2)	10249 non-null float64	Site EUI 2013	Leave as is.

150428_2014_nyc_l184_disclosure.xlsx Transformation to DataFrame			
Column Name	Info	Rename	Summary of Data Transformation
NYC Borough, Block, and Lot (BBL)	14907 non-null float64	BBL	Leave as is.
Street Number	13435 non-null object	Street Number	Leave as is.
Street Name	13469 non-null object	Street Name	Leave as is.
Borough	14908 non-null object	Borough	Leave as is.
Zip	13445 non-null float64	Zip	Leave as is.
Site EUI (kBtu/ft2)	11237 non-null object	Site EUI 2014	Convert to float type.

2015_nyc_cy2014_l184_disclosure_data.xlsx Transformation to DataFrame			
Column Name	Info	Rename	Summary of Data Transformation
NYC Borough, Block, and Lot (BBL)	13933 non-null int64	BBL	Leave as is.
Street Number	13671 non-null object	Street Number	Leave as is.
Street Name	13671 non-null object	Street Name	Leave as is.
Borough	13933 non-null object	Borough	Leave as is.
Zip Code	11611 non-null float64	Zip	Leave as is.
Site EUI (kBtu/ft2)	11879 non-null object	Site EUI 2015	Convert to float type.

nyc_benchmarking_disclosure_data_reported_in_2016.xlsx Transformation to DataFrame			
Column Name	Info	Rename	Summary of Data Transformation
NYC Borough, Block, and Lot (BBL)	14907 non-null float64	BBL	Leave as is.
Street Number	13435 non-null object	Street Number	Leave as is.
Street Name	13469 non-null object	Street Name	Leave as is.
Borough	14908 non-null object	Borough	Leave as is.
Zip Code	13445 non-null float64	Zip	Leave as is.
Site EUI (kBtu/ft2)	11237 non-null object	Site EUI 2016	Convert to float type.

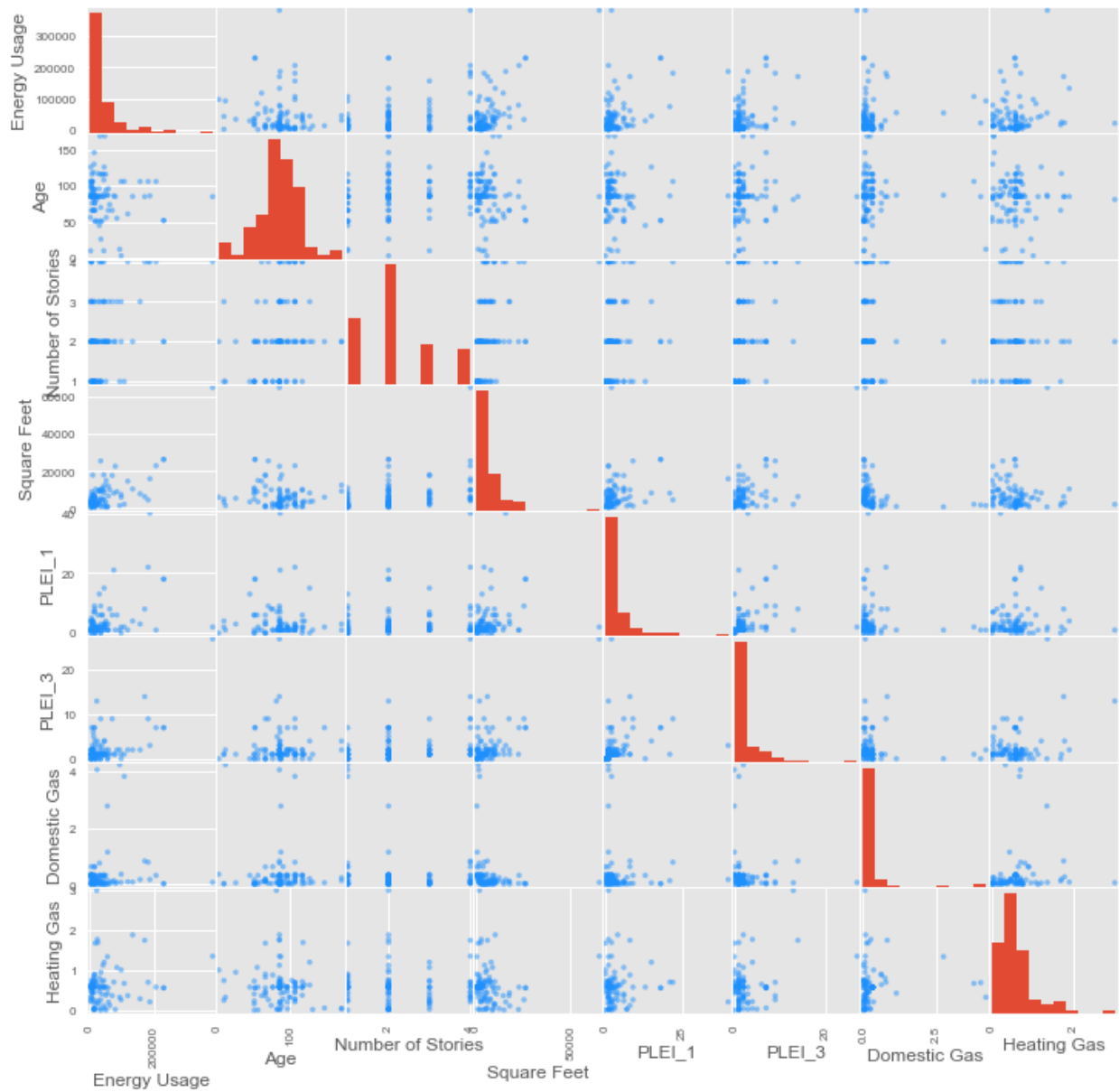
Final Data Sets:

bloc_df		
Column	Info	Sample
Property Name	103 non-null object	ChurchofStCeciliaReport
Energy Usage	103 non-null float64	117870
Age	103 non-null float64	61
Number of Stories	103 non-null int64	4
Square Feet	103 non-null float64	14600
PLEI_1	103 non-null float64	1
PLEI_3	103 non-null float64	2
Domestic Gas	103 non-null float64	0.096226
Heating Gas	103 non-null float64	0.366193
Plug Load Consumption	103 non-null float64	11.651406
AC Consumption	103 non-null float64	0.983531
Annual Energy Bill (USD)	103 non-null float64	21216.6
Year of Construction	96 non-null float64	1955
Address ID	103 non-null object	125 East 105th Street10029
Address	103 non-null object	125 East 105th Street
Zipcode	103 non-null int64	10029
Longitude	103 non-null float64	-73.947326
Latitude	103 non-null float64	40.791919
kwh/sq ft	103 non-null float64	8.07329
Inefficient	03 non-null bool	False
GHG Emission (mt)	103 non-null float64	82.8626
Emissions Intensity	103 non-null float64	0.00567552

buildings		
Column	Info	Sample
BBL	2710 non-null object	1000090001
Street Number	2702 non-null object	34
Street Name	2710 non-null object	WHITEHALL STREET
Borough	2710 non-null object	MANHATTAN
Zip	2710 non-null int64	10004
Site EUI 2012	2503 non-null float64	157.9
Site EUI 2013	2503 non-null float64	99.5
Site EUI 2014	2503 non-null float64	78.5
Site EUI 2015	2503 non-null float64	81.3
Site EUI 2016	2503 non-null float64	71.4
sq ft.	2705 non-null float64	845018.0
Address	2710 non-null object	21216 34 WHITEHALL STREET MANHATTAN
Lon	2710 non-null float64	-73.989353
Lat	2710 non-null float64	40.729733

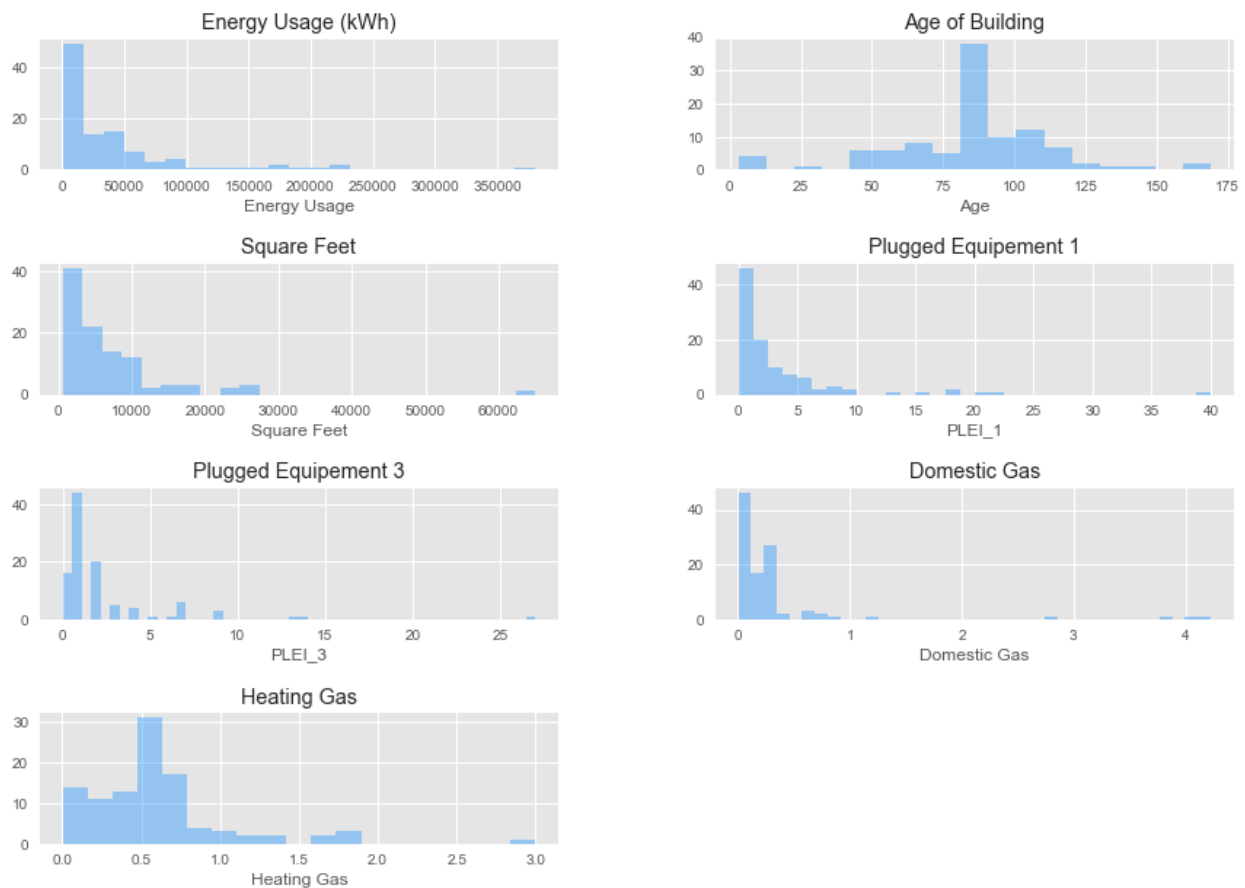
4. Exploratory Analysis: 103 NY buildings

A correlation matrix using relevant parameters shows relationships between energy usage and building characteristics.



Square feet appears to have the highest correlation with energy usage, and it does not seem there is a strong correlation between any of the parameters.

To look at a graphical distribution of parameters, histograms are plotted.

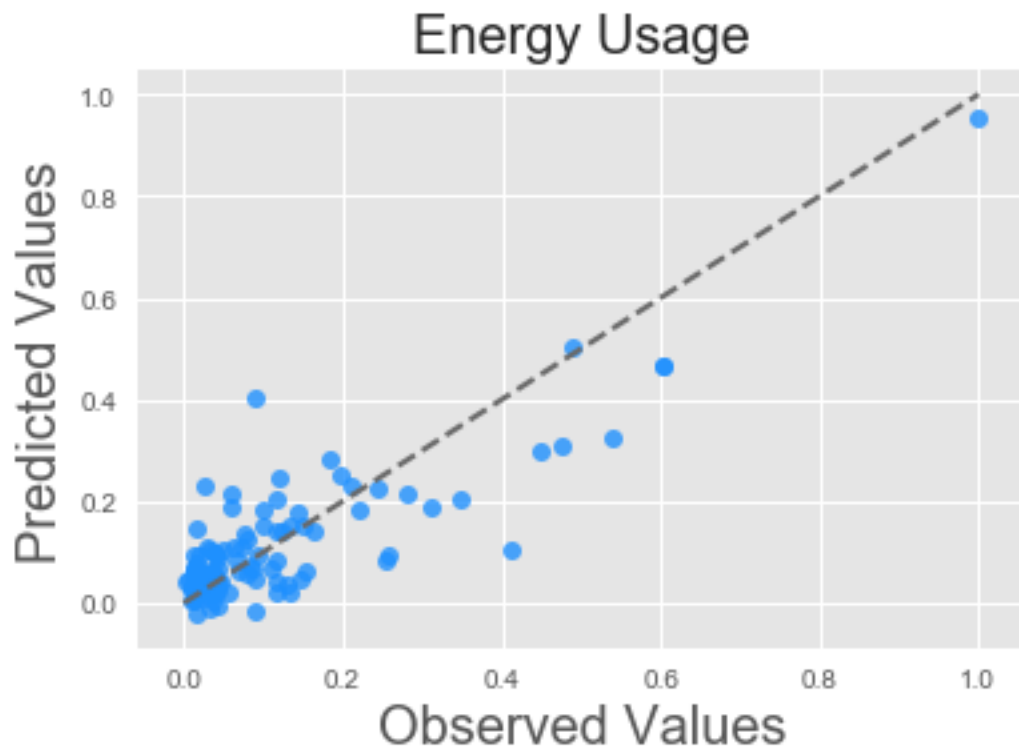


Age of Buildings shows somewhat of a normal distribution with an age building median close to 85 years. The other variables display a skewed right distribution. A regression analysis will help build a predictive model and show what variables help predict energy usage. To perform this analysis, first each parameters is scaled using sklearn's preprocessing.MaxAbsScaler(), so that the maximum value for each variable is 1. Next, a linear regression is run using linear_model.LinearRegression. The coefficients are:

Parameter	Coefficient
Intercept	-0.06714
Age	-0.0235
Number of Stories	0.048774
Square Feet	0.777122
PLEI 1	0.312308
PLEI 3	0.122954
Domestic Gas	0.229171
Heating Gas	0.143661

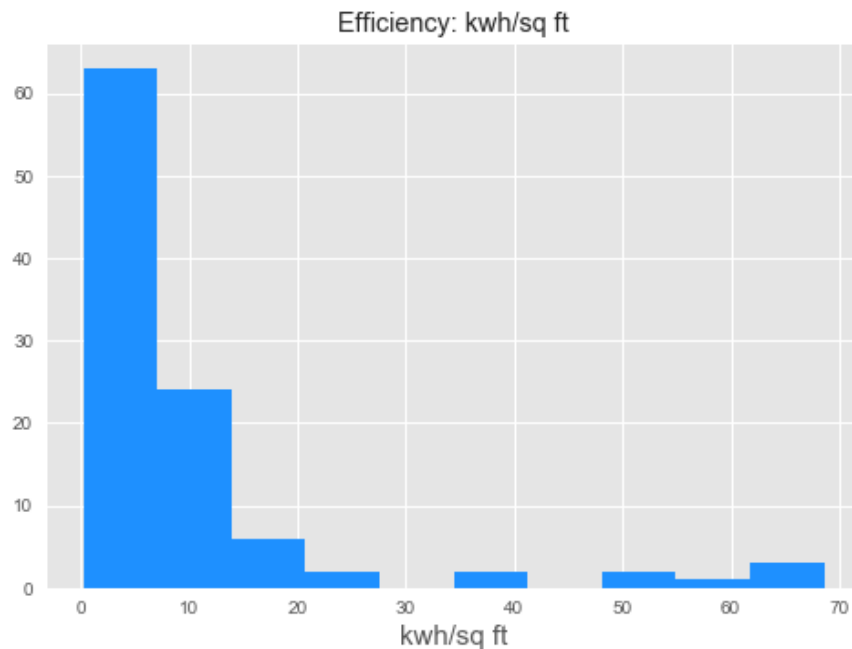
R-Squared: 0.71750454564

Square Feet is the best predictor of energy usage for a building, which is not all surprising. These parameters are then used to predict a building's energy usage. Predicted values are plotted against real energy usage value below to show accuracy of model. The dotted line represents a perfect model.



4. Labelling Inefficient Buildings

Since square footage appears to be the most relevant independent variable, energy usage /square feet is used to determine energy efficiency. The higher the ratio, the more inefficient a building is. To determine an efficiency threshold, a histogram of all ratios is plotted first.

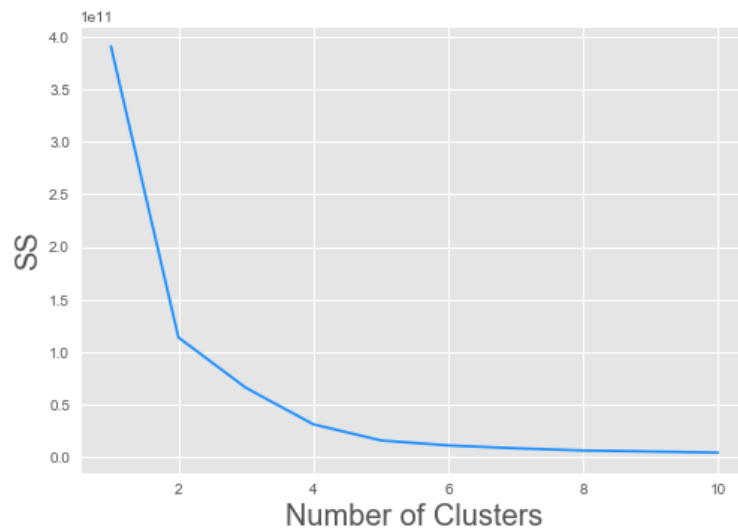


The skewed distribution shows there are a number of inefficient buildings, so buildings with an efficiency ratio > 20 are labeled “True”. After labelling each building, there are 10 out of 103 buildings that consume energy inefficiently.

5. Modelling: Unsupervised Learning.

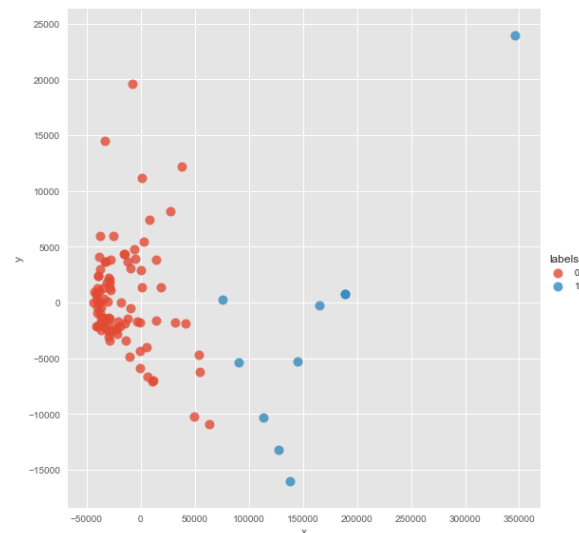
Even though the purpose of this analysis is to classify buildings into two groups, a support vector machine using sklearn's cluster KMeans classifies observation into k cluster. A graph with the sum of squares from each observation to the nearest cluster and the different k clusters is plotted. Using the elbow method, k is chosen where adding another cluster does not provide with better modelling.

Using all variables (Energy Usage, Age, Number of Stories, Square Feet, PLEI_1, PLEI_3, Domestic Gas, Heating Gas, Plug Load Consumption, AC Consumption, Annual Energy Bill (USD)):

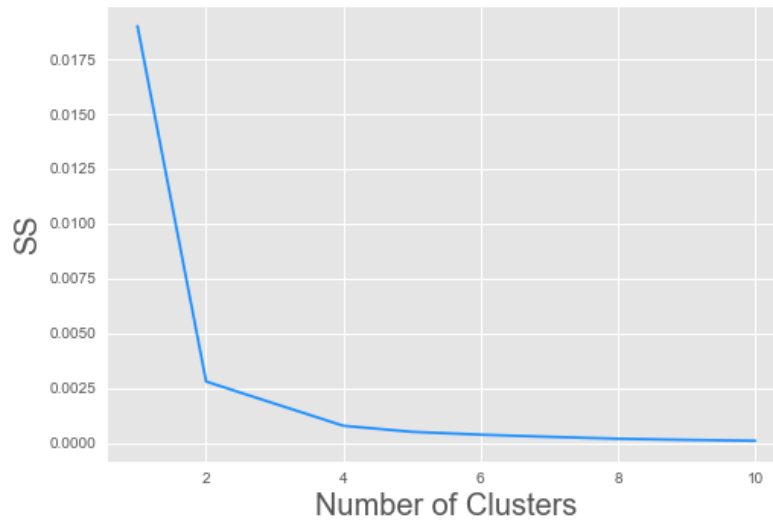


K=2.

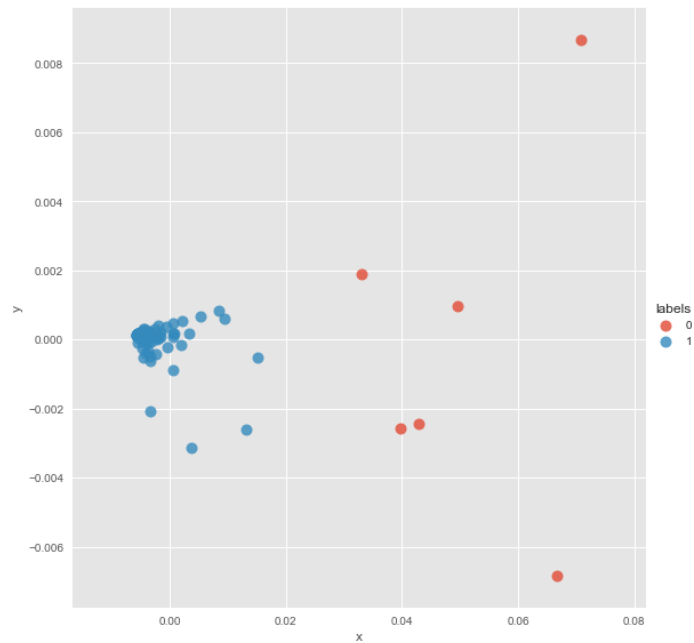
To display the clusters in a two dimensional graph, a principal components analysis is run on the variable and are plotted below.



The model accurately predicts 84.5% of the observations. Using only energy independent variables (Domestic Gas, Heating Gas, Plug Load Consumption, AC Consumption), each divided by the square footage should yield better results.



K =2

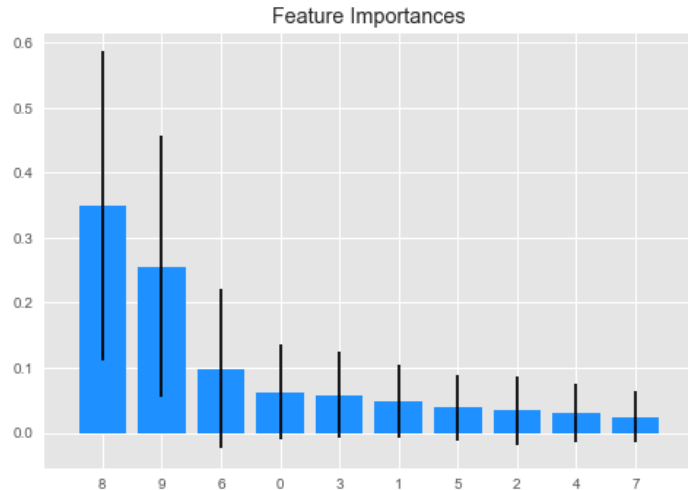


As expected, this model is more accurate, predicting 96% of the data.

6. Supervised Learning

Since each building is labeled, an ensemble method, may obtain better predictive performance. First the data is split into training (80%) and testing (20%) testing data. Using the RandomForestClassifier on the training data, predictions are then made with the parameters from the testing data. As it turns out, the model is able to accurately predict 100% of the testing data labels. To see which features helped the classifier best, feature importance is plotted below:

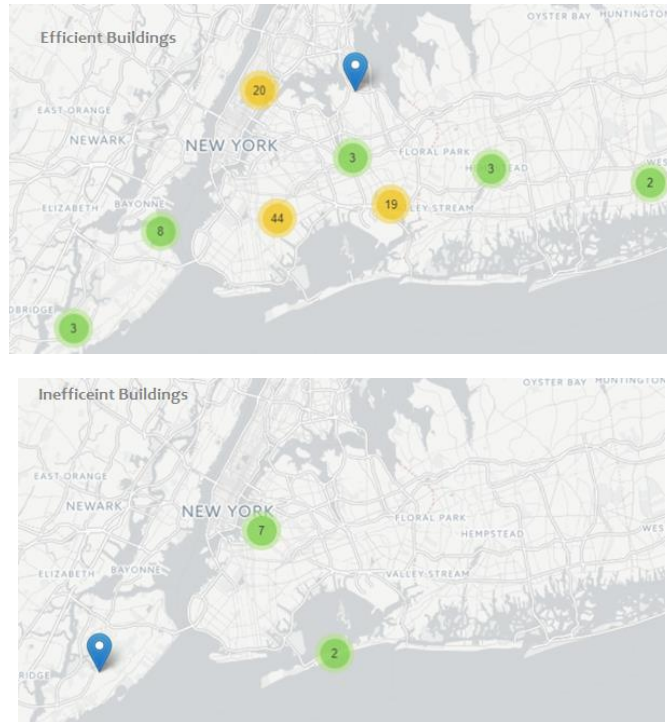
#	Feature
0	Energy Usage
1	Age
2	Number of Stories
3	Square Feet
4	PLEI 1
5	PLEI 3
6	Domestic Gas
7	Heating Gas
8	Plug Load Consumption
9	AC Consumption



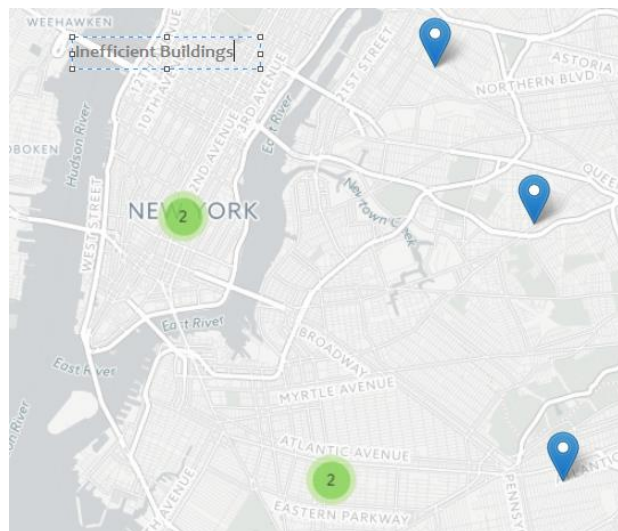
Load consumption from electrical equipment, followed by air conditioning consumption, and gas used for domestic purposes are the most important leaves in the model. This seems intuitive, since new appliances and air conditioning units have become increasingly more energy efficient in the last decade. It makes sense that inefficient energy buildings probably have old electrical equipment that consume more energy.

6. Geographic Visualization

Using MarkerCluster from Folium, efficient and inefficient buildings are mapped to see whether location plays a role in a building's energy efficiency.



It seems that most inefficient buildings (7) are clusters in a location.



When, zooming in, location does not appear to be a factor in determining a building's efficiency.

7. Environmental Performance

To measure environmental performance, greenhouse gas emissions are calculated using the Environmental Protection Agency's Emission Factor: 7.03×10^{-4} metric tons CO₂ / kWh. The Emissions Intensity is then calculated by dividing the emission from the total square footage. Based on emissions only, is not enough to determine a building's environmental performance.

A benchmark is calculated by analyzing 2,477 buildings from New York's municipal benchmark 2014 results (latest benchmark result). The mean emission intensity (0.0077) will be the benchmark.

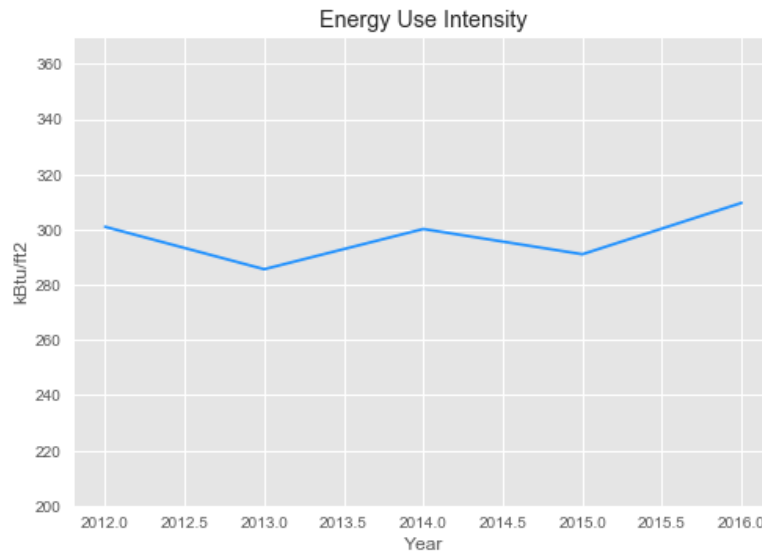
Using this benchmark, 23 buildings including all the inefficient buildings out of the 103 fail to meet this benchmark. The table below lists environmentally poor performing buildings sorted by their emissions intensity.

Property Name	Emissions Intensity (GHG Emission (mt)/Sq ft.)	Inefficient
PaninicoCafeReport	0.048208	TRUE
LunaNYCafeCorpReport	0.044468	TRUE
CAAABagels	0.043554	TRUE
Curran'sSuperiorMeatsReport	0.039017	TRUE
NewTestamentChristianChurch	0.034172	TRUE
MoonbluIncdbaJoyBurgerBar	0.034172	TRUE
AnchorHouse-ParkPl	0.02791	TRUE
NYSERDA_Energy_Assessment_-_1011_Tavern_Corp_- ...	0.026384	TRUE
69thLaneStudio	0.018989	TRUE
AnchorHouse-BergenSt	0.018985	TRUE
CongregationOhabZedekSynagogue	0.013983	FALSE
NYSERDA_Energy_Assessment_-_ _DAC_Unisex_Beauty_...	0.012874	FALSE
PentecostalHouseofPrayer	0.012815	FALSE
ThaiRock	0.011434	FALSE
NYSERDA_Energy_Assessment_-_Island_Bay_Grill_a...	0.010762	FALSE
NewMtZionBaptistChurch	0.01027	FALSE
MountLebanonBaptistChurch	0.009111	FALSE
WaysideBaptistChurch	0.008971	FALSE
TheStarPeople'sLaundromat	0.008893	FALSE
4-1_Vision_Education_Media,_LLC_-_NYSERDA_Ener...	0.008261	FALSE
BrooklynLegalServicesCorpA	0.00823	FALSE
NYSERDA_Energy_Assessment_-_Nelly's_Nails_and_...	0.008104	FALSE
HarrietTubmanFannieLouHamerCollectivedbaSistas...	0.008026	FALSE

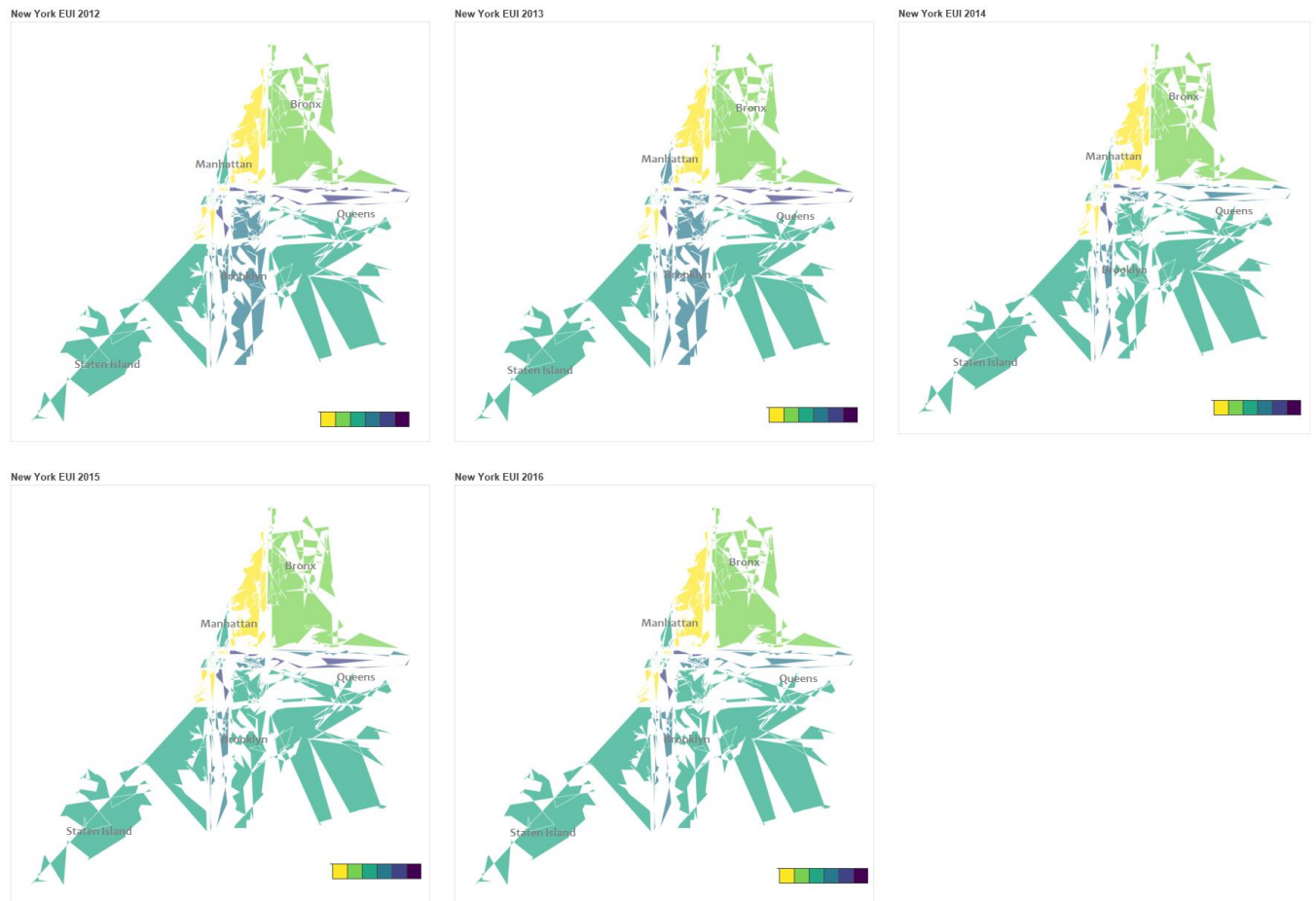
8. NYC's Greener Greater Buildings Plan

So why is this all relevant? Because US: Local Law 84 of 2009 requires buildings to disclose energy consumption of the city's largest buildings. Unfortunately, the data for these buildings do not list the feature analyzed for the 103 buildings analyzed previously. But we can translate some of the findings to provide appropriate recommendations.

NYC Mayor's Office of Sustainability has data from 2012 -2016. After merging all the data sets, mean Energy Use Intensity (EUI) are plotted for these years:



Average EUI appears to remain somewhat stable at around 300 kBtu/Sq. ft. Mapping all the EUI for each year is more insightful:



Brooklyn has improved the most over the years, but still has areas with high EUIs. EUIs for the rest of the borough seems to remain stable with little or no change. East of Manhattan and the Bronx are the areas with more efficient buildings. Queens and Staten Island followed closely Brooklyn regarding EUIs for 2016.

9. Recommendations

A. Replace all old electrical equipment including air conditioners at these 10 locations:

- 1) 69-71 Grand Avenue
- 2) 361 6th Ave
- 3) 976 Park Place
- 4) 1041 Bergen St.
- 5) 171 Avenue C
- 6) 758 Arthur Kill Road
- 7) 241 Beach 116th Street #1
- 8) 239 Beach 116th Street
- 9) 334 Ashford Street
- 10) 31-86 37th Street

B. For the city of New York: Test an energy efficiency program in Brooklyn focused on removing old inefficient electrical appliances and air conditioners by teaching tenants on topics of energy savings. Based on a pilot program provide incentives for replacing old inefficient electrical equipment.

C. For building managers: If building EUI above 0.0077, analyze electrical equipment usage and/or perform an energy audit of the site to see causes of inefficiencies.

Sources:

Batchcoordinates. (n.d.). Retrieved July 29, 2017, from <http://www.freegeocoding.com/>

City of New York, NYC Open Data. (n.d.). NYC Open Data. Retrieved July 29, 2017, from <https://opendata.cityofnewyork.us/>

Ruiz, A. (2016, June 24). Modeling energy usage in New York City. Retrieved July 29, 2017, from <https://datascience.ibm.com/blog/modeling-energy-usage-in-new-york-city/>

(n.d.). Retrieved July 29, 2017, from http://www.nyc.gov/html/gbee/html/plan/l184_scores.shtml