Capstone 2

Modelling Building Energy Efficiency in New York

Guillermo Guerra

Summary

The purpose of the analysis is to classify 103 buildings based on the efficient consumption of energy and to create a model to identify inefficient energy buildings. Predictive variables from this model will be used to generalize recommendations for buildings required to benchmark their energy consumption in accordance to a city law placed in 2012, designed to reduce emissions.

103 Buildings with predictive features

12,000+ buildings required to benchmark Energy Intensity Use from 2012-2016

5 New York City Boroughs

Data Sources:

Data fields and column names in appendix.

Data collected from multiple source:

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.

Proprietary Data (BlocPower): BlocPower_T.csv, clusterEnergyLocation.csv

Public Heating & Cooling: CDD-HDDFeatures.csv

Public City Building Data (NYC OpenData): NYC_Municipal_Building_Energy_ Benchmarking_Re sults__2014_.csv, 2012_nyc_cy2011 _ll84_disclosur e_data.xlsx, 2013_nyc_ll84_disclosure.xlsx, 150428_2014_nyc _ll84_disclosur e.xlsx, 2015_nyc_cy2014 __ll84_disclosu re_data.xls, nyc_benchmarking_disclosure_da ta_reported_in_ 2016.xlsx

Data Preparation

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

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.

Data Preparation

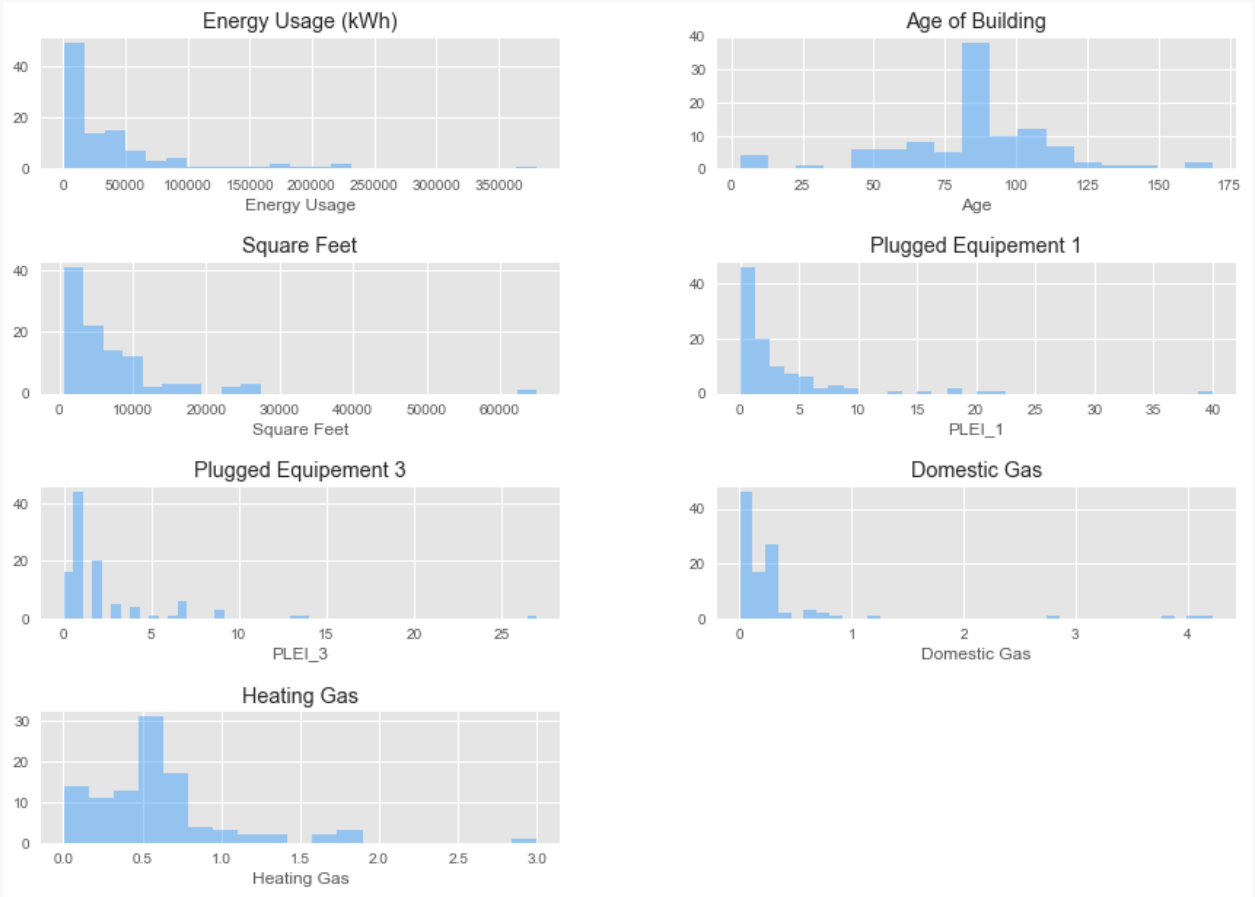
Final datasets after cleanup and transformation.

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

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

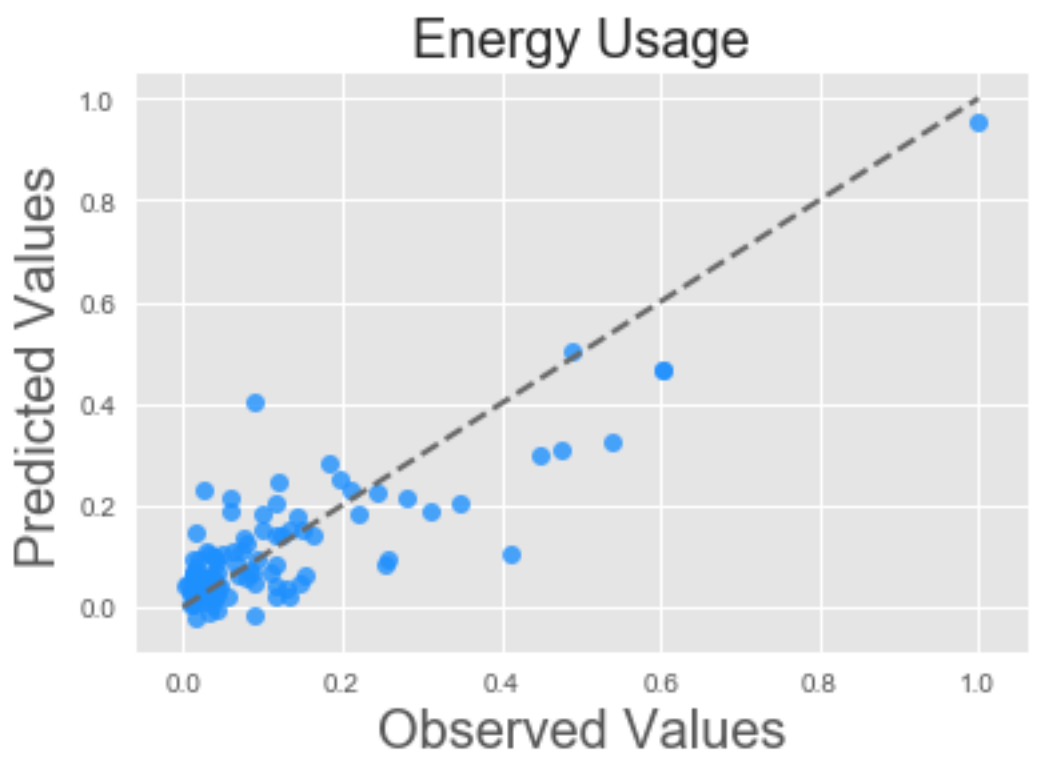
Exploratory Analysis: 103 NY buildings

Age of Building shows somewhat of a normal distribution with an age building median close to 85 years. The other variables display a skewed right distribution.



Regression Analysis

Square Feet is the best predictor of energy usage for a building.



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

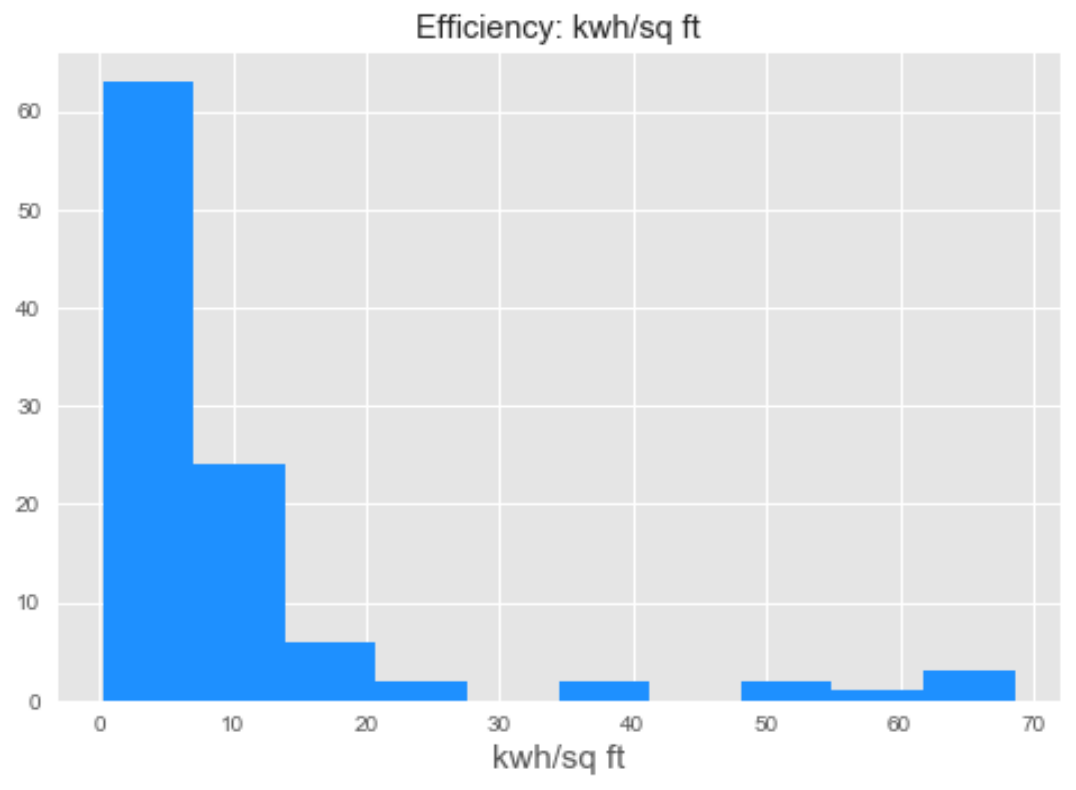
R-Squared: 0.72

F-statistic: 34.47

Labelling Inefficient Buildings

Total energy usage /square feet ratio as inefficiency metric

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 building that consume energy inefficiently.



Modelling: Unsupervised Learning

Kmeans clustering reveals two groups.

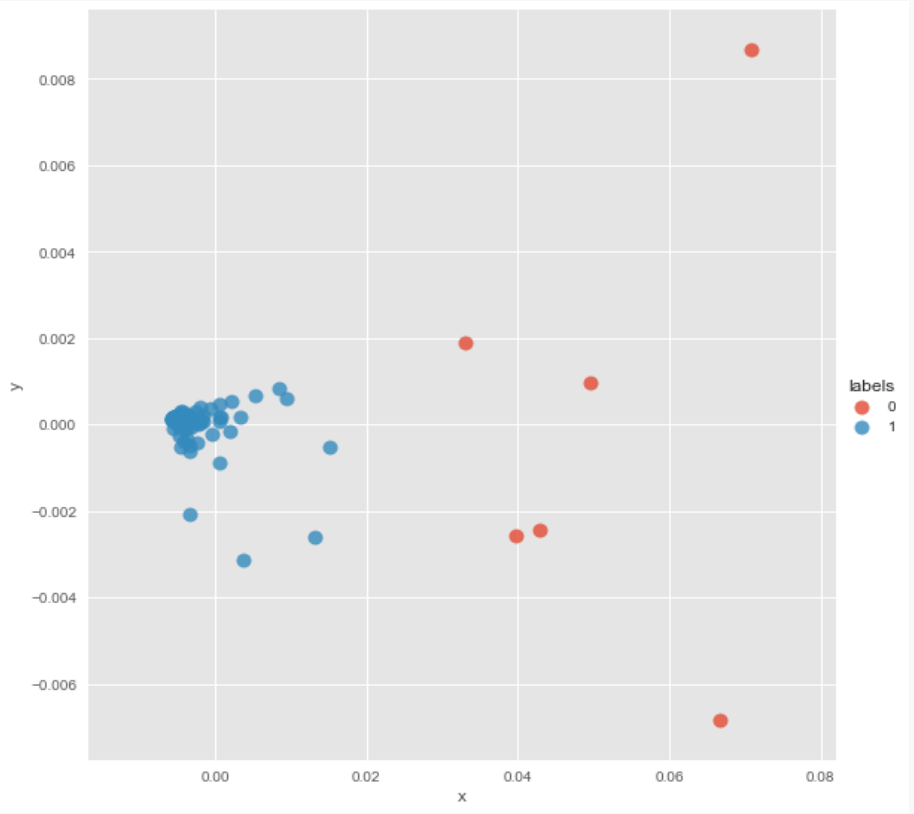
Highly Predictive:

Running Kmeans using only energy independent variables (Domestic Gas, Heating Gas, Plug Load Consumption, AC Consumption), each divided by the square footage yields a model that predicts 96% of the testing data.

Clusters:

Efficient buildings are closely clustered shown in blue. It appears there could be clusters of inefficient buildings.

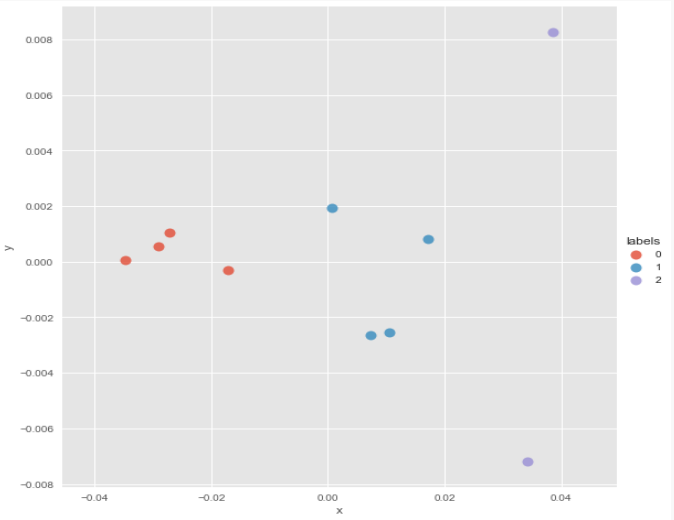
Clustering Using Energy Parameters



Modelling: Unsupervised Learning

Clusters among inefficient buildings reveal types of inefficiencies.

Inefficient Buildings Clustering



Labels 0 show inefficiency caused by PLEI_1 and PLEI_3. One of the buildings is a restaurant and another one is a church by looking at the property names. Labels 1 show inefficiency by the total Plug Load Consumption. From looking at the property names, these seem like restaurants or a food establishment. Buildings labeled 2 are the two outliers. Both of them have large Plug Load consumption and one of them has a high AC consumptions. It is worth noting that both of them are coffee restaurants.

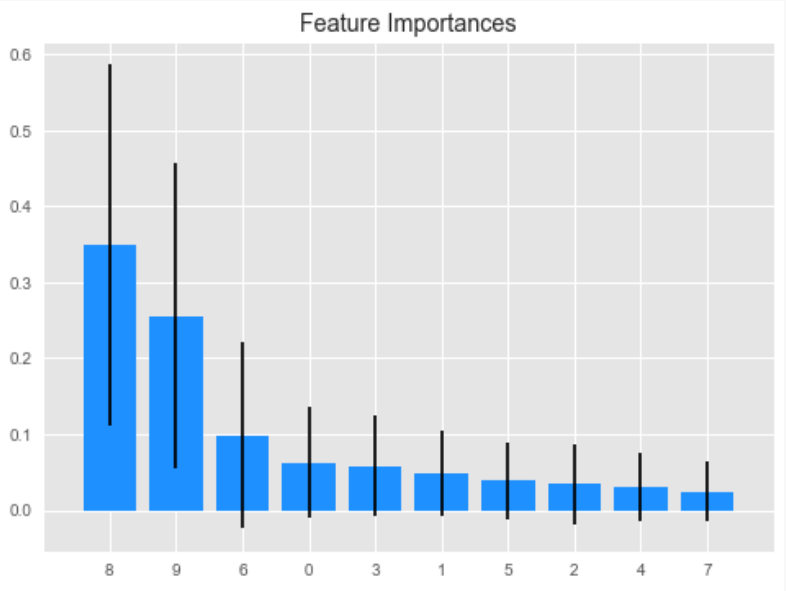
Property Name	Inefficiency	Labels
MoonbluIncdbaJoyBurgerBar	PLEI_1 and/or PLEI_3	0
AnchorHouse-ParkPl	PLEI_1 and/or PLEI_3	0
AnchorHouse-BergenSt	PLEI_1 and/or PLEI_3	0
NewTestamentChristianChurch	PLEI_1 and/or PLEI_3	0
69thLaneStudio	Plug Load Consumption	1
NYSERDA_Energy_.Tavern_Corp_-…	Plug Load Consumption	1
CAAABagels	Plug Load Consumption	1
Curran'sSuperiorMeatsReport	Plug Load Consumption	1
PaninicoCafeReport	Plug Load & AC Consumption	2
LunaNYCafeCorpReport	Plug Load & AC Consumption	2

Modelling: Supervised Learning

Random forrest classifier reveal two important features: Plug Load & AC Consumption

First the data is split into training (80%) and testing (20%) testing data. Using the RandomForrestClassifier 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. Among the predictive features, load consumption from electrical equipment, followed by air conditioning consumption, and gas used for domestic purposes are the most important leaves in the model.

#	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



Measuring Environmental Performance

Greenhouse gas emissions as metric for environmental performance

- Greenhouse gas emissions are calculated using the Environmental Protection Agency's Emission Factor: 7.03 × 10-4 metric tons CO2 / kWh.
- 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.
- 23 buildings (listed on the right) including all inefficient buildings do not meet the benchmark.

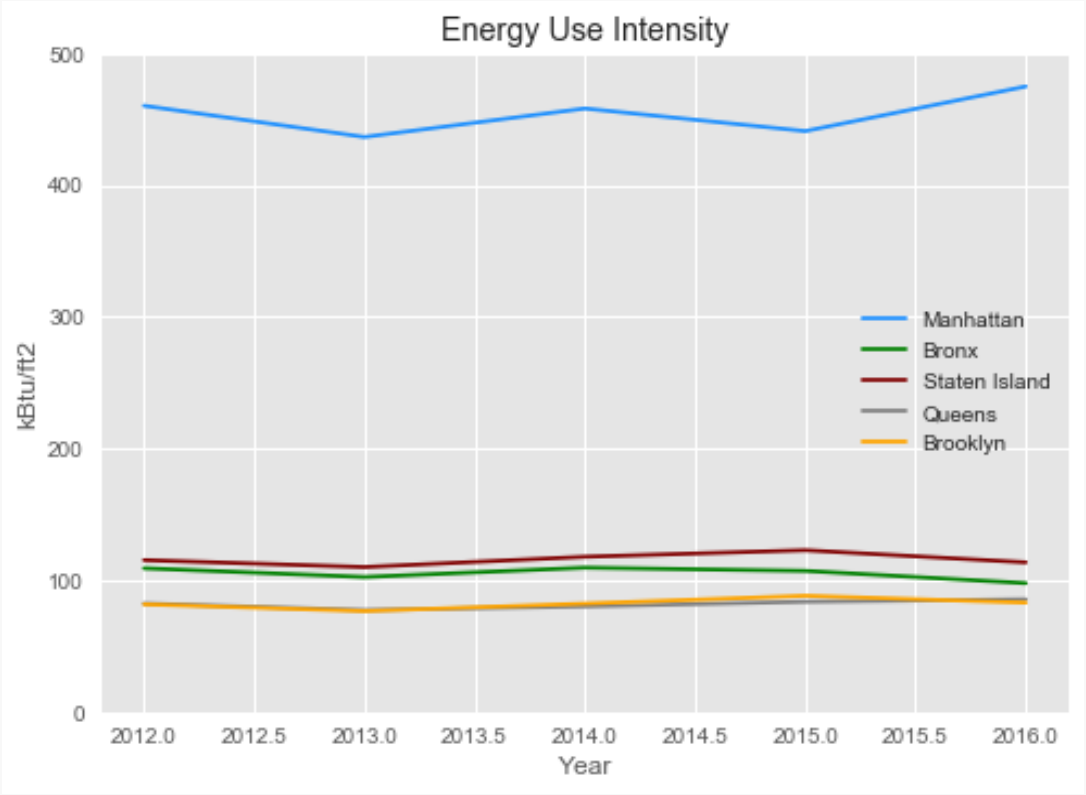
Property Name	Emissions Intensity (GHG Emission (mt)/Sq ft.)	Inefficient
PaninicoCafeReport	0.048208	TRUE
LunaNYCafeCorpReport	0.044468	TRUE
CAAABagels	0.043554	TRUE
Curran'sSuperiorMeatsReport	0.039017	TRUE
NewTestamentChristianChurch	0.034172	TRUE
MoonbluIncdbaJoyBurgerBar	0.034172	TRUE
AnchorHouse-ParkPl	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

NYC's Greener Greater Buildings Plan

Energy Use Intensity (EUI) : Energy consumed/total square feet

US: Local Law 84 of 2009 requires buildings to disclose energy consumption of the city's largest buildings. Staten Island shows higher EUIs when taking population density numbers into consideration.

New York City's five boroughs V·T·E						
Jurisdiction		Population	Land area		Density	
Borough	*County*	*Estimate (2016)*[177]	*square miles*	*square km*	*persons / sq. mi*	*persons / sq. km*
Manhattan	New York	1,643,734	22.83	59.1	72,033	27,826
The Bronx	Bronx	1,455,720	42	110	34,653	13,231
Brooklyn	Kings	2,629,150	71	180	37,137	14,649
Queens	Queens	2,333,054	109	280	21,460	8,354
Staten Island	Richmond	476,015	58.5	152	8,112	3,132
City of New York		**8,537,673**	**303.33**	**781.1**	**28,188**	**10,947**

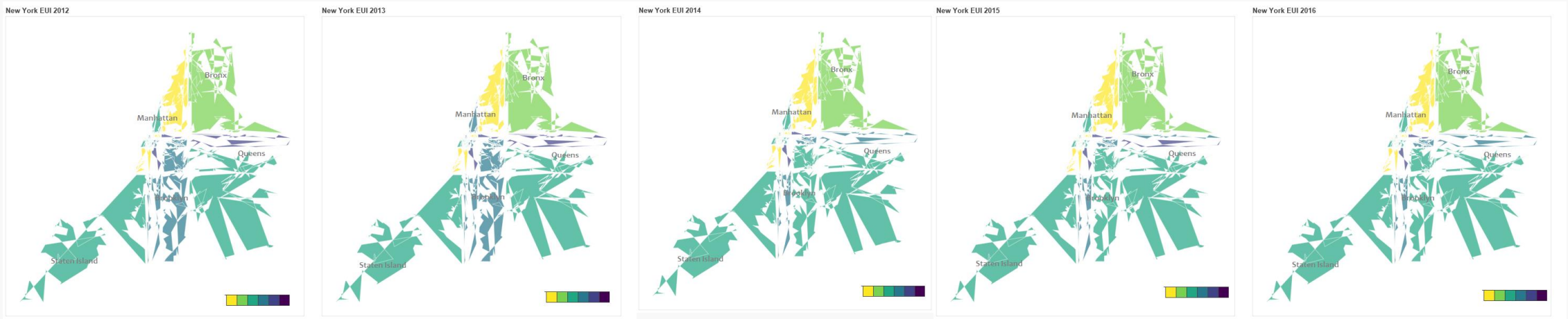


Source: https://en.wikipedia.org/wiki/New_York_City

NYC's Greener Greater Buildings Plan

Energy Use Intensity (EUI) : Energy consumed/total square feet

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.



Limitations

Analysis of 103 Buildings do not indicate a representative sample

Due to limited public information, insights revealed by analyzing the first dataset of 103 buildings were used to provide recommendations for the full dataset of 2710 buildings in New York. Although the 103 buildings are scattered across the five boroughs, there is no indication of it being a representative sample. Furthermore, the most inefficient buildings in the analysis are food establishments, which are more likely to consume more energy than residential properties. A proper analysis would be to separate by building type before identifying inefficiencies, but due to missing data this was not possible. With that said, recommendations are made from the analysis.

Recommendations

Analysis of 103 Buildings do not indicate a representative sample

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 Staten Island 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/

New York City. (2017, August 14). Retrieved August 14, 2017, from https://en.wikipedia.org/wiki/New_York_City

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/ll84_scores.shtml

Appendix

Source	CSV	Fields
Bloc Power	BlocPower_T.csv	UTSUM_Electricity_Usage, INFO_Year of Construction, INFO_Number of Stories, INFO_Total Square Feet, PLEI_1_Quantity, PLEI_3_Quantity
Bloc Power	clusterEnergyLocation.csv	AddressID, property name, Adress, 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 Open Data	NYC_Municipal_Building_Energy_Benchmarking_Results__2014_.csv	Borough, Block, Lot, BIN, Building, Agency, 2010 Score *, 2010Source EUI (kBtu/ftÂ²)*, 2010 GHG Emissions Intensity (kgCO2e/ftÂ²)*, 2014 Score *, 2014 Source EUI (kBtu/ftÂ²)*, 2014 GHG Emissions Intensity (kgCO2e/ftÂ²)*
NYC Open Data	2012_nyc_cy2011_ll84_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 Open Data	2013_nyc_ll84_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 Open Data	150428_2014_nyc_ll84_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/ft²), 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)) (ft²), "DOF Property Floor Area (Buildngs and Parking) (ft2)", Primary Property Type - Self Selected, DOF Number of Buildings
NYC Open Data	2015_nyc_cy2014__ll84_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/ft²), 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)) (ft²), "DOF Property Floor Area (Buildngs and Parking) (ft2)", Primary Property Type - Self Selected, DOF Number of Buildings
NYC Open Data	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 (ft²), 2nd Largest Property Use Type, 2nd Largest Property Use - Gross Floor Area (ft²), 3rd Largest Property Use Type, 3rd Largest Property Use Type - Gross Floor Area (ft²), Year Built, Number of Buildings - Self-reported, Occupancy, Metered Areas (Energy), Metered Areas (Water), ENERGY STAR Score, Site EUI (kBtu/ft²), Weather Normalized Site EUI (kBtu/ft²), Weather Normalized Site Electricity Intensity (kWh/ft²), Weather Normalized Site Natural Gas Intensity (therms/ft²), Source EUI (kBtu/ft²), Weather Normalized Source EUI (kBtu/ft²), 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 (ft²), Property GFA - Self-reported (ft²), Water Use (All Water Sources) (kgal), Municipally Supplied Potable Water - Indoor Intensity (gal/ft²), Release Date, DEP Provided Water Use (kgal), Automatic Water Benchmarking Eligible, Reported Water Method