

HACKATHON

Pioneering Strategies for
Reducing Carbon Emission

Jeonseo Lee

TABLE OF CONTENTS

- 01 Background (CO2 Emission Trend)
- 02 Description of Dataset
- 03 Data Preprocessing
- 04 Traditional OLS
- 05 Machine Learning Algorithm & Predictions

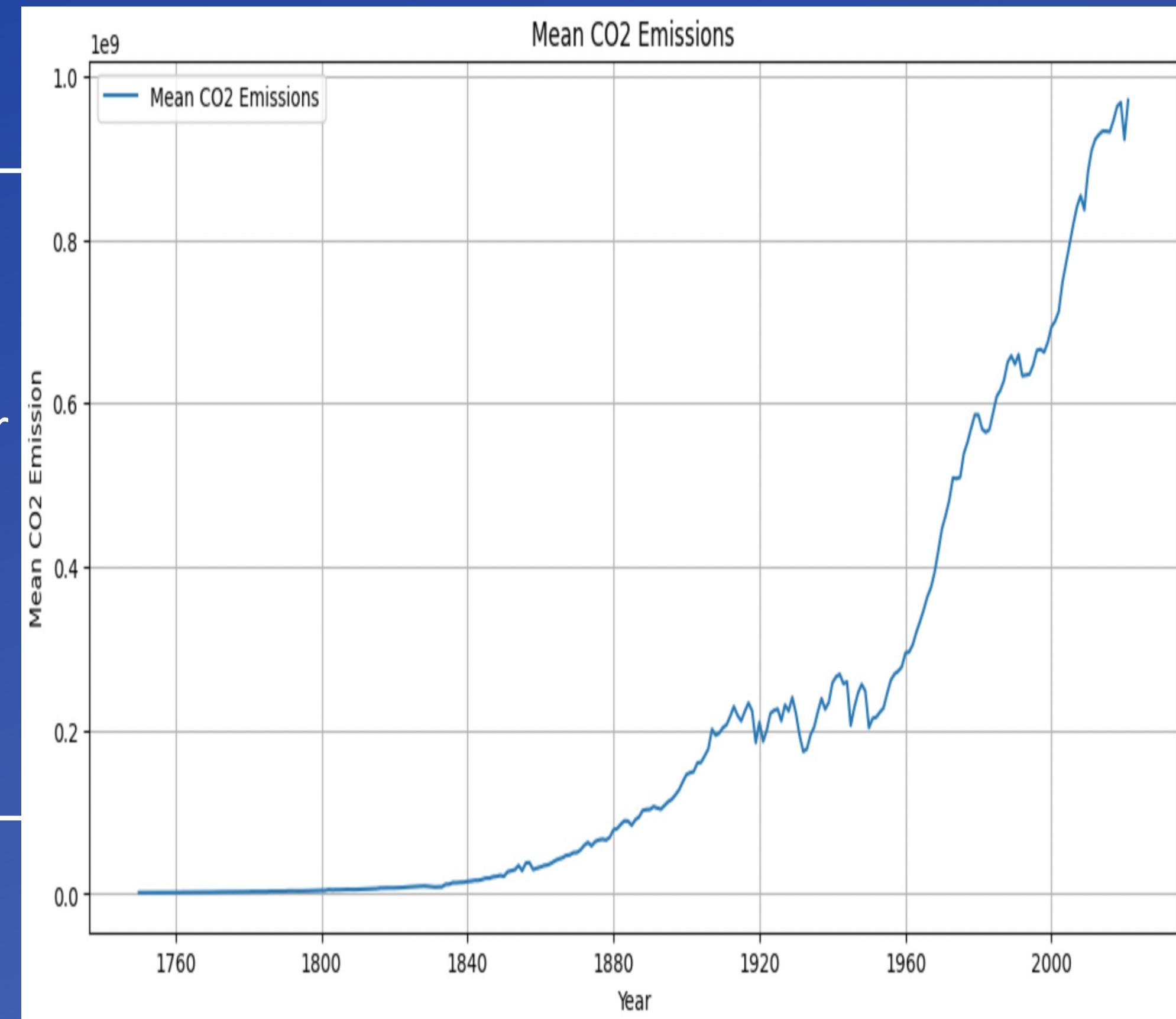
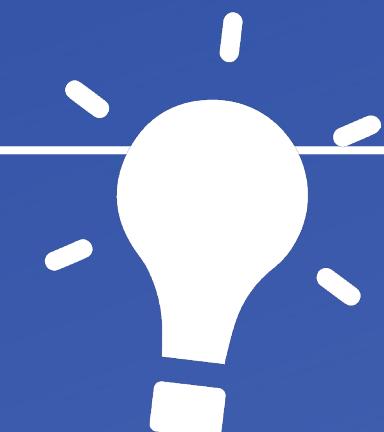
BACKGROUND

Over the years, there has been a consistent upward trend in global CO₂ emissions.



Rising CO₂ levels are a major driver of global climate change and its associated effects.

International cooperation is critical to achieving meaningful **reductions** in emissions.



DATASET

NYC Open Data on Energy Usage

<https://opendata.cityofnewyork.us/data/>



(2000, 249)

	property_id	property_name	parent_property_id	parent_property_name	year_ending	nyc_borough_block_and_lot	nyc_building_identification	ac
0	21205224	Astoria Blvd Property	Not Applicable: Standalone Property	Not Applicable: Standalone Property	2021-12-31T00:00:00.000	4006520042		4538607
1	2665352	Stellar - 70 West 93rd Street	20599688	Stellar - Campus West 93rd Street	2021-12-31T00:00:00.000	1-01206-0001		1031627
2	2665400	Stellar - 28-50 West 97th Street	Not Applicable: Standalone Property	Not Applicable: Standalone Property	2021-12-31T00:00:00.000	1-01832-0043		1055265 W
3	2665405	Stellar - 310 Greenwich St	Not Applicable: Standalone Property	Not Applicable: Standalone Property	2021-12-31T00:00:00.000	1-00142-0025	1083157;1083235;1083237;1083240;1083241;108324...	G St
4	2665415	Stellar - 87 Hamilton Place	Not Applicable: Standalone Property	Not Applicable: Standalone Property	2021-12-31T00:00:00.000	1-02072-0039		1061772

5 rows x 249 columns

DATA PROCESSING

01

Convert into Numeric Values

```
17
18 df_cleaned.isna().sum()
19
20 ✓ 0.0s
...
[254] ... /var/folders/2c/cd06n1c95mv7lsfybtv2zt1w0000
A value is trying to be set on a copy of a s
See the caveats in the documentation: https://
df_cleaned.fillna(df_cleaned.median(), inp
property_id          0
postal_code           0
largest_property_use_type_1 0
year_built            0
number_of_buildings   0
occupancy              0
national_median_energy_star 0
site_eui_kbtu_ft      0
weather_normalized_site_eui 0
national_median_site_eui 0
site_energy_use_kbtu   0
weather_normalized_site_energy 0
weather_normalized_site 0
source_eui_kbtu_ft     0
weather_normalized_source 0
national_median_source_eui 0
source_energy_use_kbtu 0
weather_normalized_source_1 0
natural_gas_use_kbtu   0
natural_gas_use_therms 0
electricity_use_grid_purchase 0
electricity_use_grid_purchase_1 0
weather_normalized_site_3 0
electricity_use_grid_purchase_2 0
electricity_use_grid_purchase_3 0
...
longitude             0
community_board        0
council_district       0
census_tract            0
dtype: int64
```

02

Set Threshold for NaN Values

03

Fill NA with Median to Proceed

The # of NAs are set to
be 0 across all columns

TRADITIONAL OLS

JANUARY
Sorted out independent variables
that appear to have strong
relationship with CO2 emission
amounts

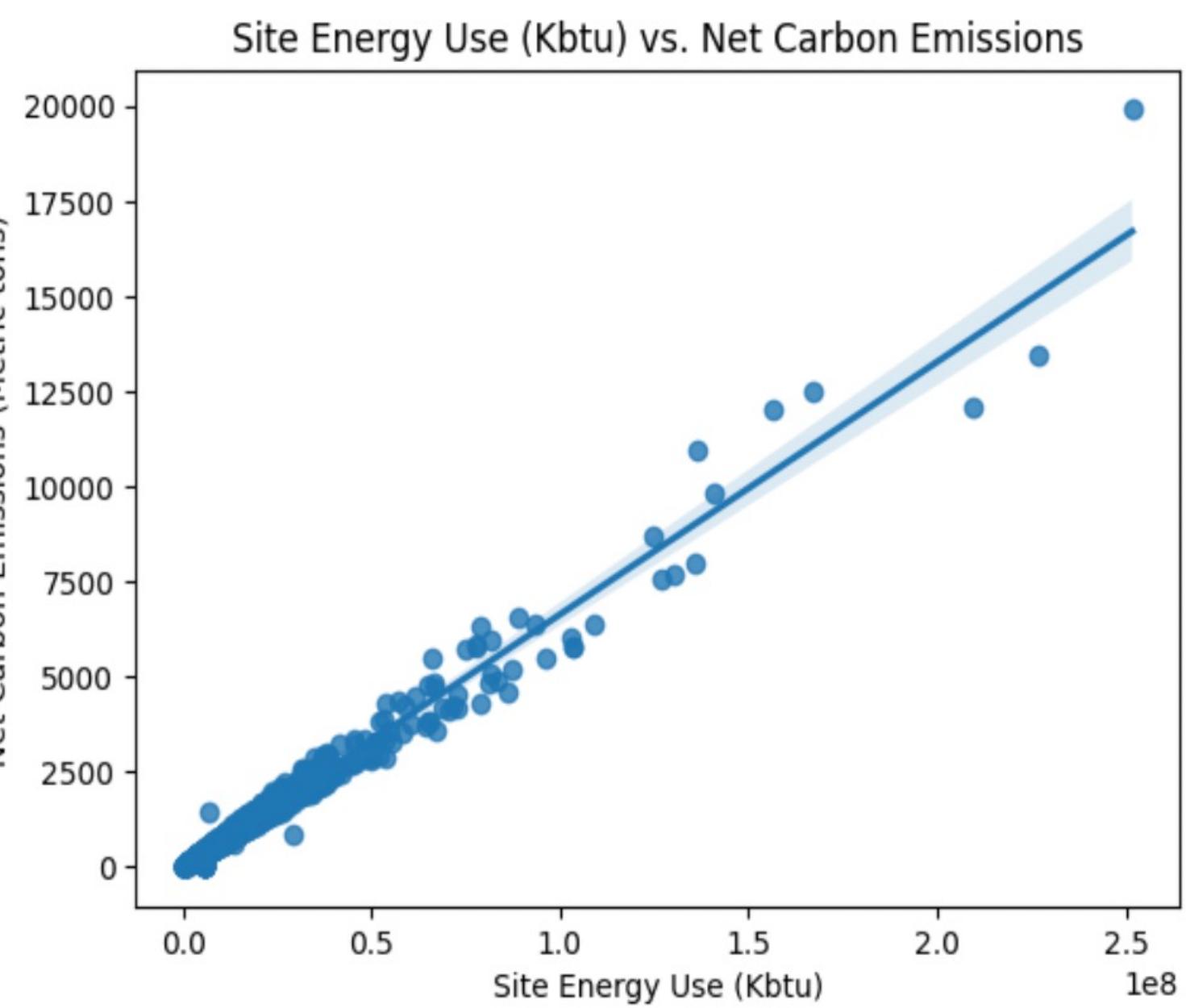
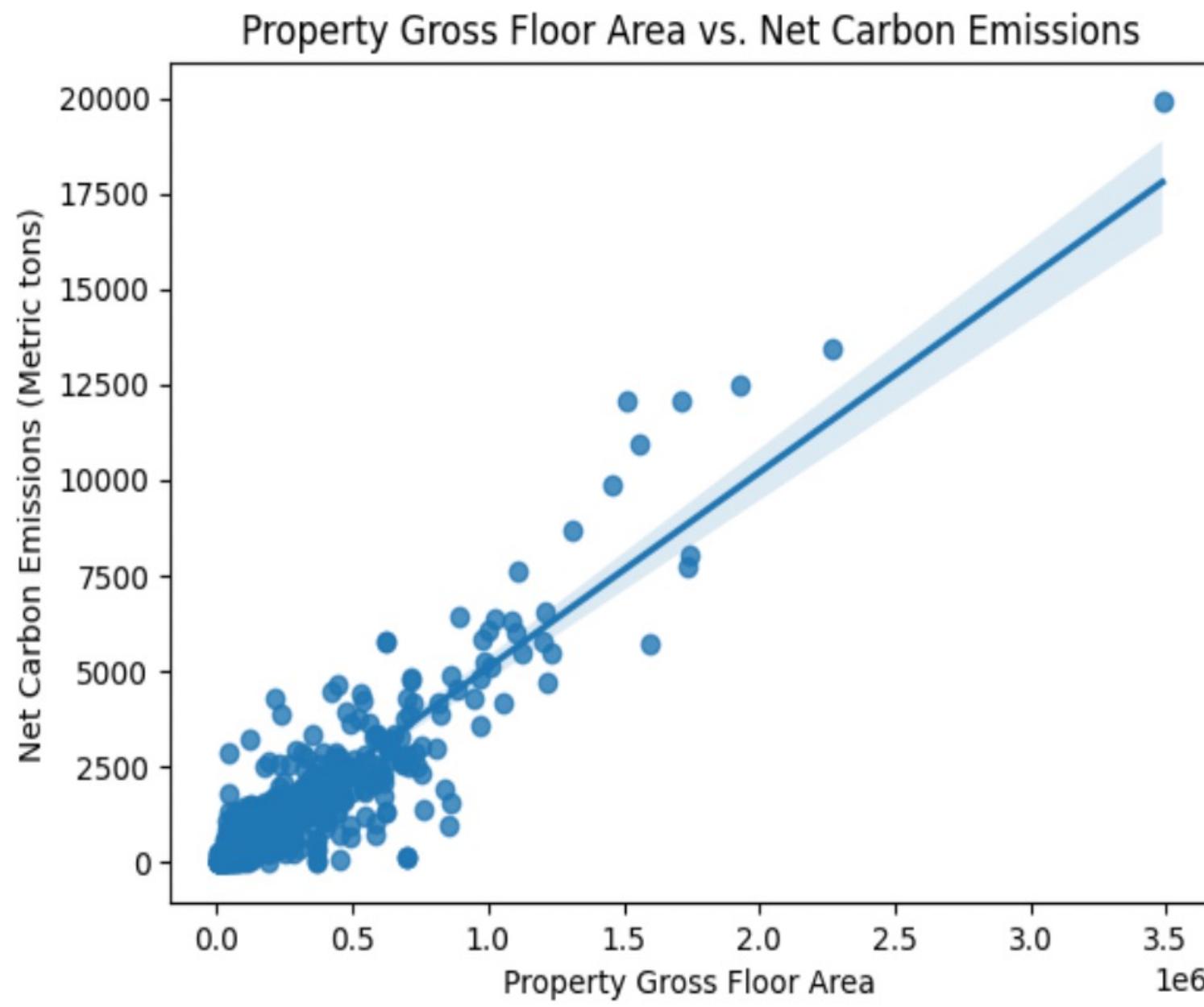
- MARCH
- 1) Number of Buildings
 - 2) Year Built
 - 3) Occupancy
 - 4) Property Gross Floor Area
 - 5) Weather Normalized Site (EUI)
 - 6) Site Energy Use



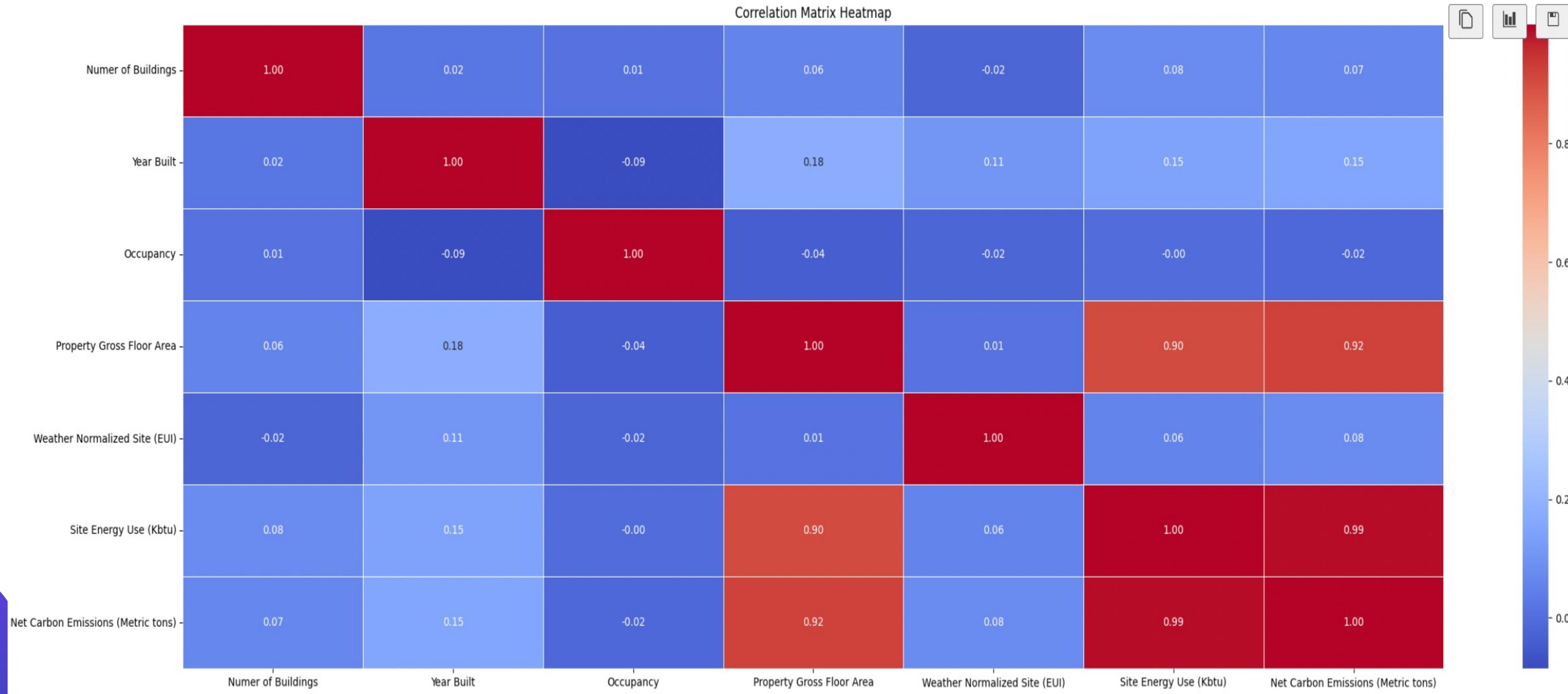
OLS Regression Results						
Dep. Variable:	Net Carbon Emissions (Metric tons)	R-squared:	0.981			
Model:	OLS	Adj. R-squared:	0.981			
Method:	Least Squares	F-statistic:	1.391e+04			
Date:	Sat, 07 Oct 2023	Prob (F-statistic):	0.00			
Time:	21:52:43	Log-Likelihood:	-10359.			
No. Observations:	1600	AIC:	2.073e+04			
Df Residuals:	1593	BIC:	2.077e+04			
Df Model:	6					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	120.6097	243.937	0.494	0.621	-357.862	599.081
Numer of Buildings	-1.5140	0.901	-1.680	0.093	-3.282	0.254
Year Built	-0.0378	0.122	-0.309	0.757	-0.278	0.202
Occupancy	-0.9278	0.400	-2.317	0.021	-1.713	-0.143
Property Gross Floor Area	0.0009	4.47e-05	20.094	0.000	0.001	0.001
Weather Normalized Site (EUI)	2.9986	0.405	7.400	0.000	2.204	3.793
Site Energy Use (Kbtu)	5.6e-05	5.44e-07	103.025	0.000	5.49e-05	5.71e-05
Omnibus:	1334.690	Durbin-Watson:	1.940			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	392306.513			
Skew:	2.943	Prob(JB):	0.00			

TRADITIONAL OLS

Two Most Relevant Variables that Can Contribute to CO2 Emission:
1) Property Gross Floor Area



TRADITIONAL OLS



MACHINE LEARNING

O1 KNN REGRESSOR

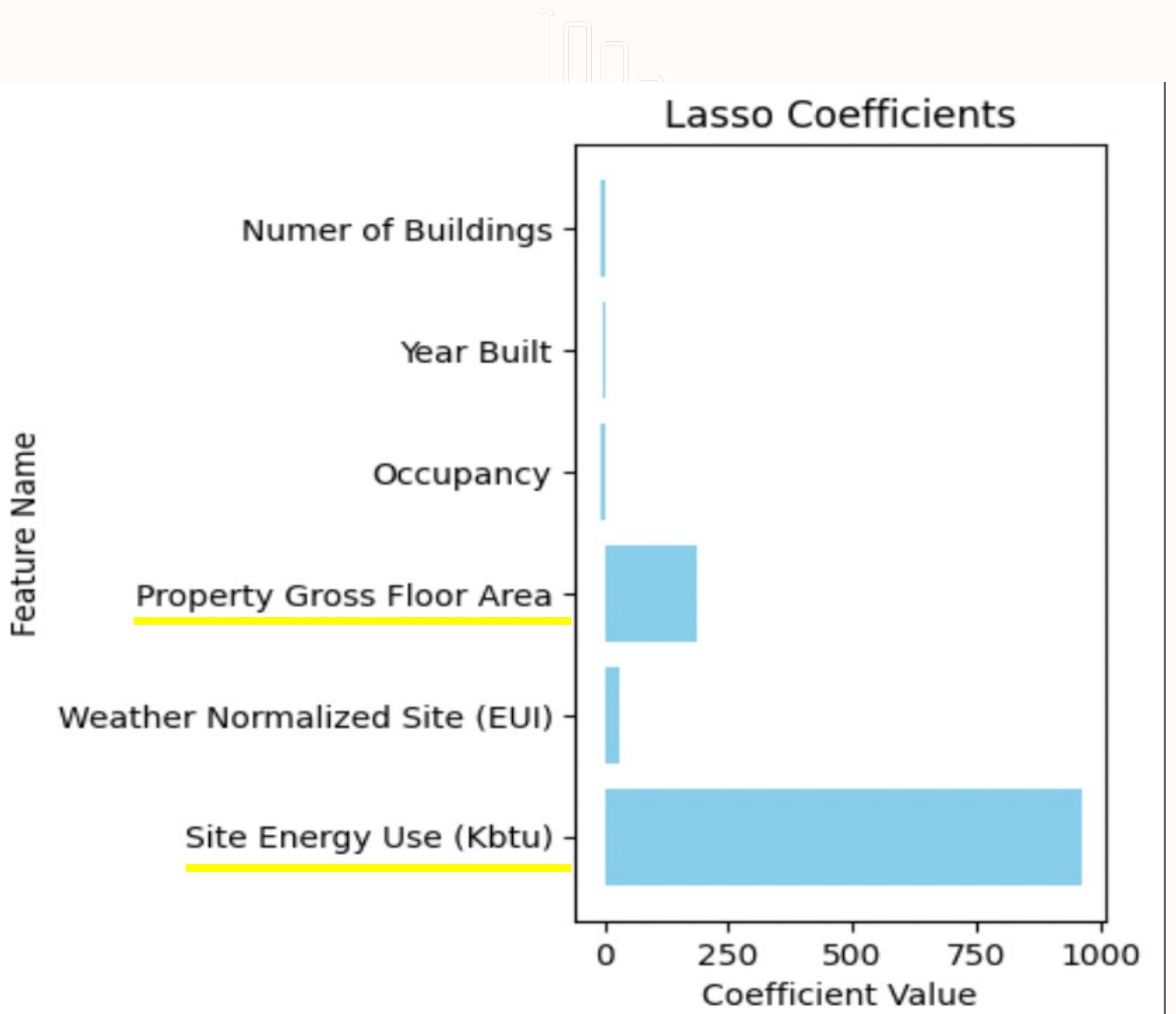
```
Training Set'S MSE: 103138.29061525  
Test Set'S MSE: 418225.091404  
Training-set score: 0.949  
Test-set score: 0.942
```

O2 LINEAR REGRESSOR

```
Training Set'S MSE: 32801.12761232202  
Test Set'S MSE: 45497.975063494756  
Training-set score: 0.981  
Test-set score: 0.978
```

O3 LASSO

```
Training Set'S MSE: 32798.52653356789  
Test Set'S MSE: 45485.813396932834  
Number of features used: 6  
Training-set score: 0.981  
Test-set score: 0.978
```



AND PREDICTION

- The LASSO regression model demonstrates superior performance, boasting an impressive accuracy score of **0.978** when incorporating all six explanatory variables.
- Notably, it's the two variables that particularly stood out in the traditional OLS model, **Property Gross Floor Area & Site Energy Use**, showcasing the most substantial contribution to enhancing its predictive accuracy.

**THANK'S
FOR
WATCHING**

