

# PROJET 3 : ANTICIPEZ LES BESOINS en consommation de BÂTIMENTS

MATTHIEU MARQUER  
23 OCTOBRE 2023



Seattle

## sommaire

1

### Le Contexte

Objectif neutralité  
carbone

2

### La Démarche

Nettoyage Graphique  
Modèles

3

### LES RÉSULTATS

## Le contexte

L'objectif est de proposer le model le plus adapté pour prédire les émissions de CO<sub>2</sub> ainsi que la consommation totale d'énergie des bâtiments non destinés à l'habitation.

---



## Le contexte



L'intégralité des données provient de Seattle Open Data:  
<https://data.seattle.gov/dataset/2016-Building-Energy-Benchmarking/2bpz-gwpy>

Données initiales :  
2016\_Building\_Energy\_Benchmarking.csv





2

## La Démarche

Nettoyage

Visualisation

Modèles

## Vérification:

- des types par variable
- des valeurs manquantes par variable
- des valeurs uniques par variable
- calcul du taux de remplissage des variables

## Suppression de variables:

"Comments", "Outlier", "DataYear", "City", "State", "YearsENERGYSTARCertified",  
"ThirdLargestPropertyUseType", "ThirdLargestPropertyUseTypeGFA"...

## Suppression des doublons:

TaxParcelIdentificationNumber et PropertyName

Ajout de la surface total pour chaque batiment et du ratio surface batiment/parking

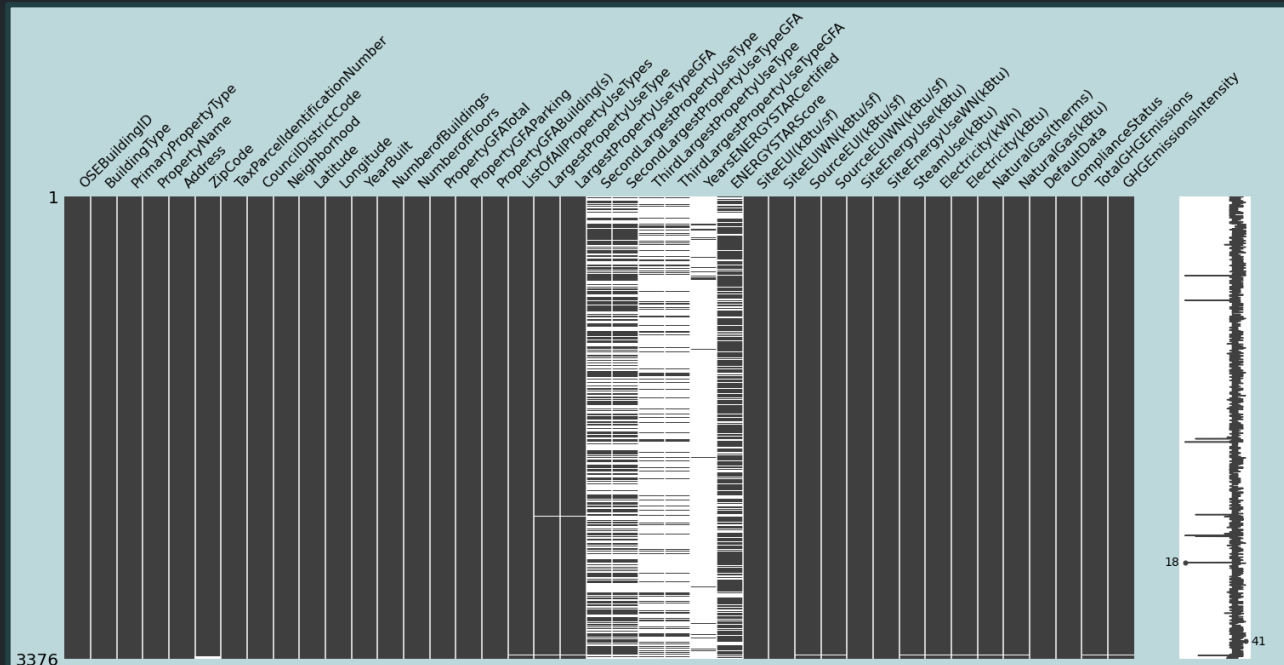
Suppression des batiments résidentiels

Binarisation de Electricity(kBtu) et GHGEmissionsIntensity

Utilisation de OneHotEncoder sur les variables catégorielles

# GRAPHIQUE

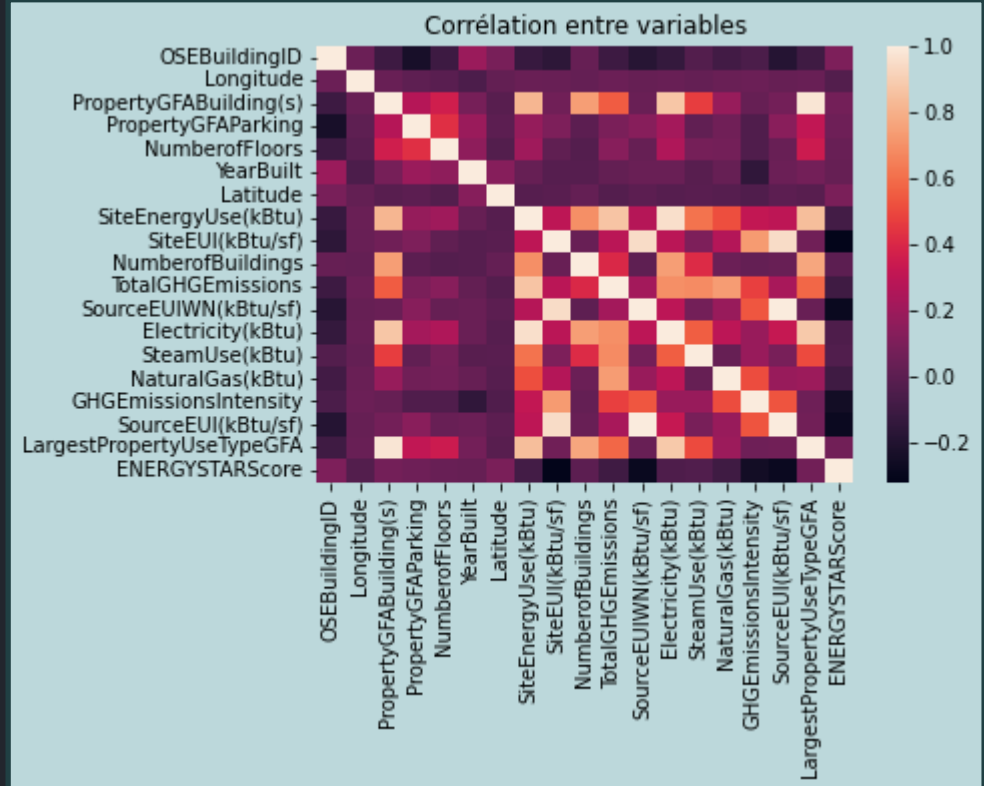
Utilisation de  
missingno





# Heatmap

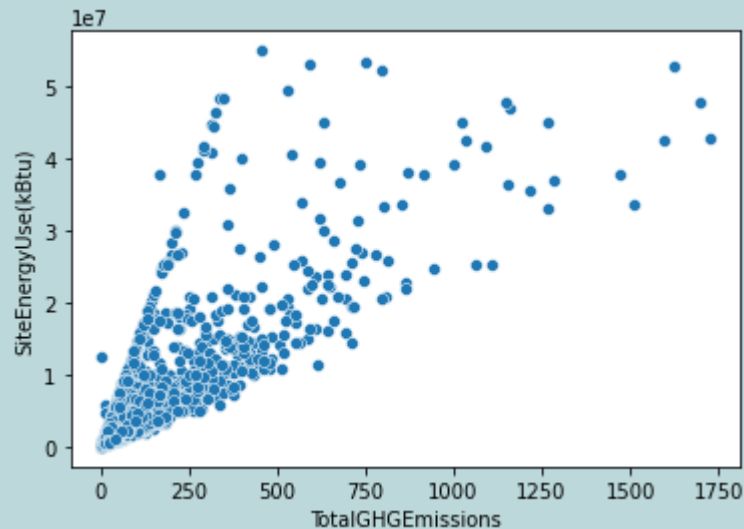
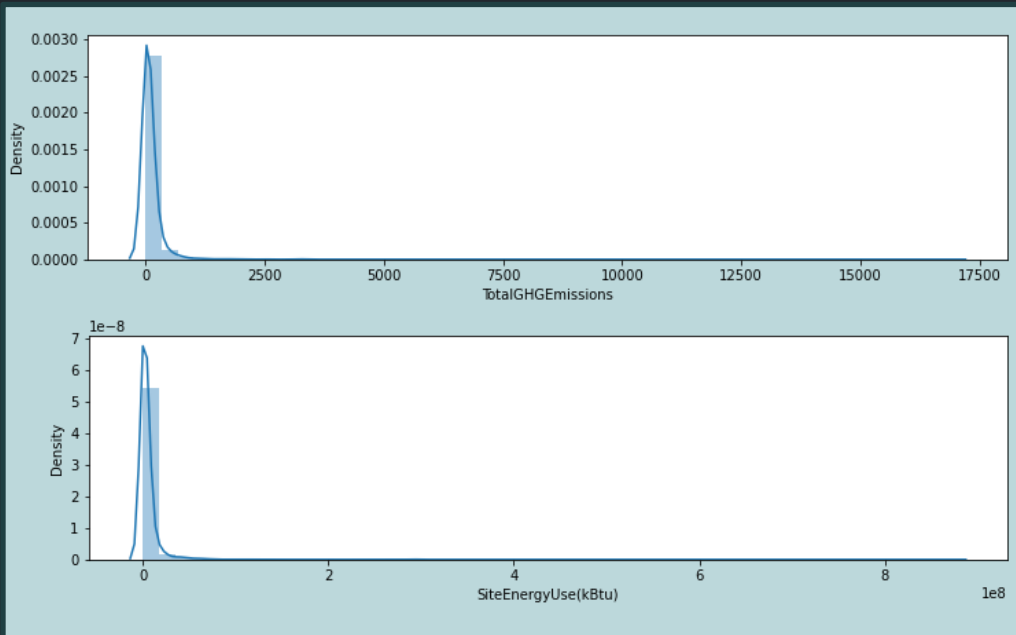
## Corrélation entre les différentes variables





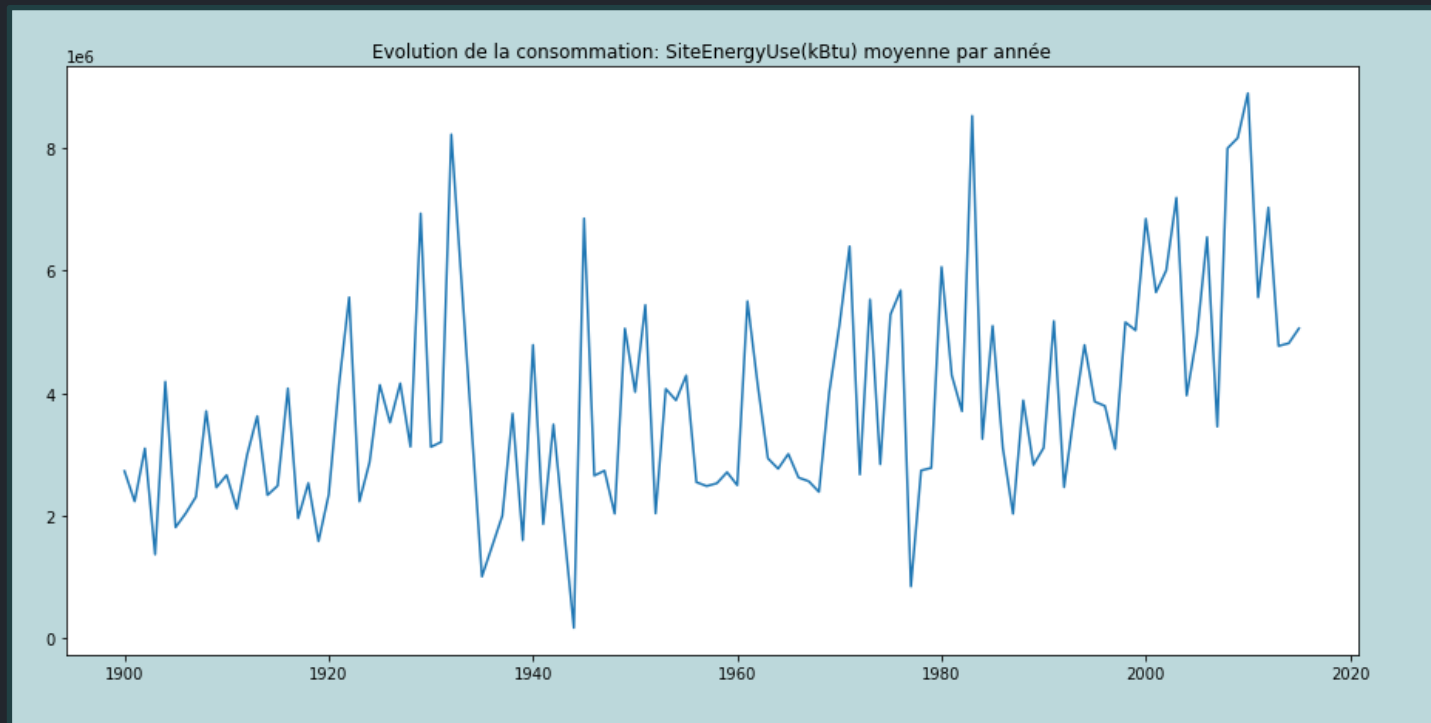
## densité et corrélation

Vérification de la densité sur les variables cibles:



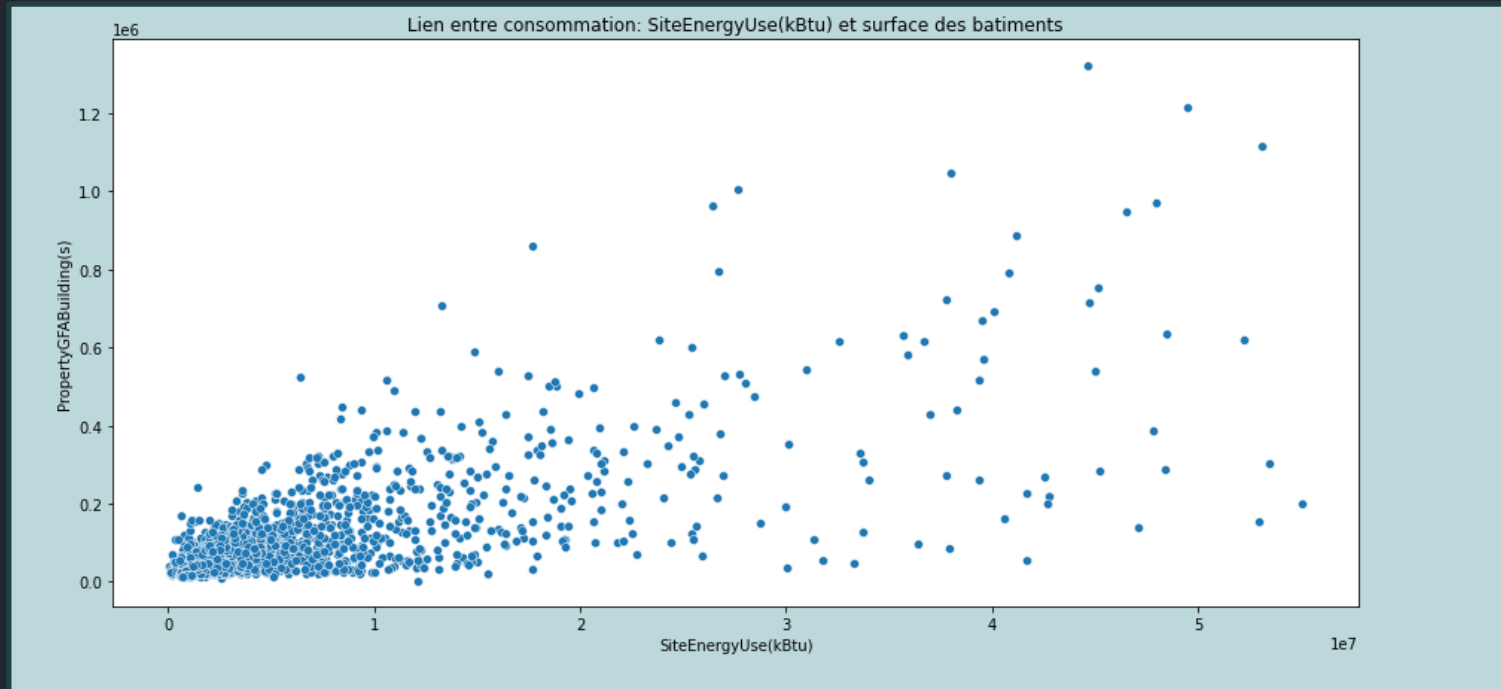
Ainsi que la corrélation entre ces deux variables:

Evolution de la consommation moyenne par  
année pour SiteEnergyUse(kBtu)



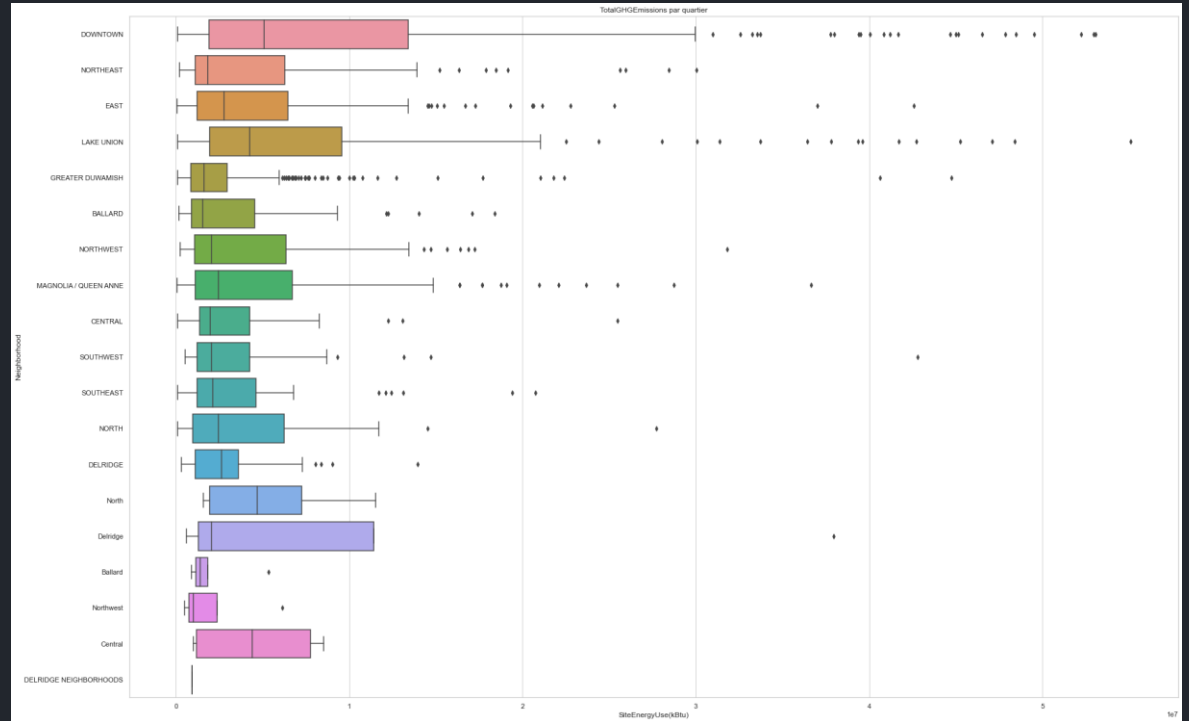
consommation /  
surface

Lien entre surface des bâtiments et  
SiteEnergyUse(kBtu)



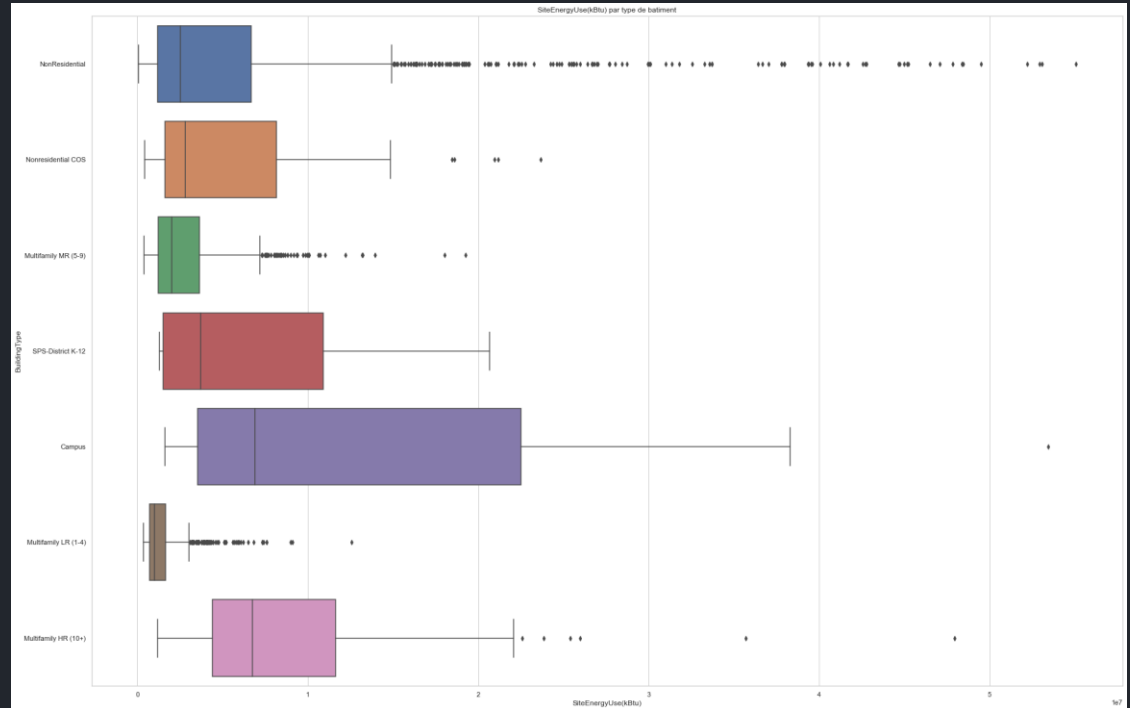
# consommation par quartier

Représentation des consommations de TotalGHGEmissions par quartier

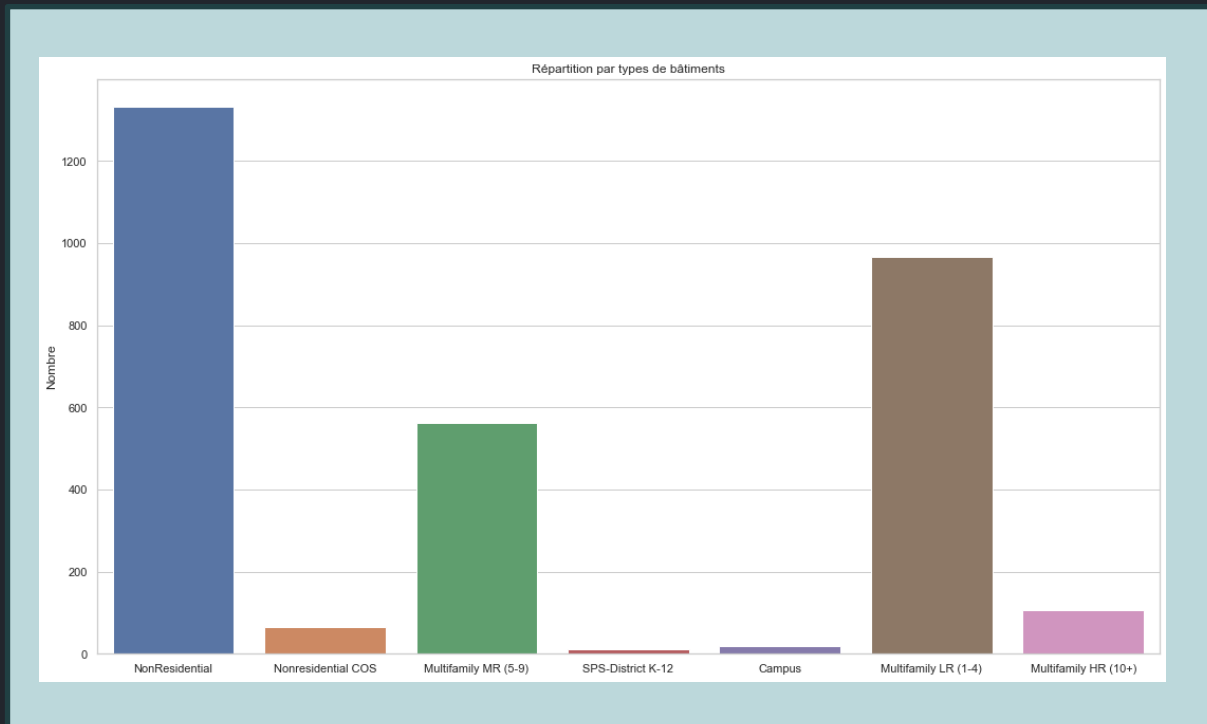


## consommation par type de bâtiment

Représentation des  
consommations de  
SiteEnergyUse(kBtu)  
par type de bâtiment



## RÉPARTITION PAR TYPES DE BÂTIMENTS



Utilisation de `train_test_split` pour découpe du jeu en 70% d'entraînement et 30% de test.

Liste des modèles utilisés:

DummyRegressor, LinearRegression

Lasso, Ridge, ElasticNet

DecisionTreeRegressor, RandomForestRegressor, XGBRegressor

Métriques utilisées:

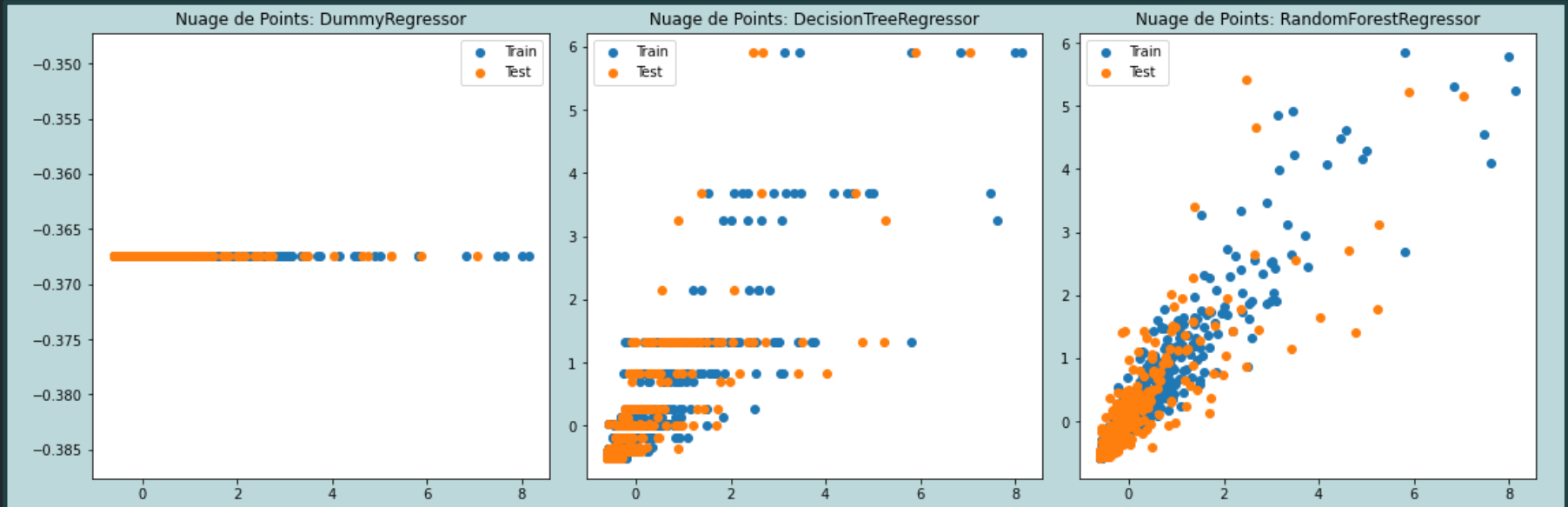
RMSE, MAE,  $R^2$

Passage d'une partie des hyperparamètres au logarithme



Graphiques des nuage d'erreur  
pour comparaison des  
différents modèles utilisé

GRAPHIQUES DES  
RÉSULTATS

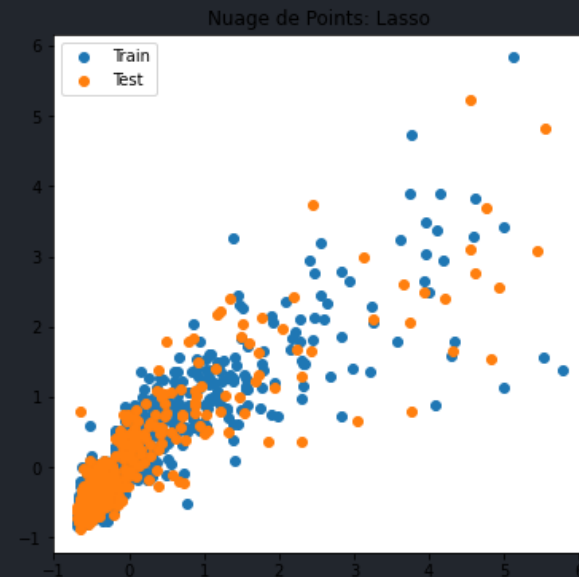




# LES RÉSULTATS

Pour SiteEnergyUse(kBtu):

	Model	RMSE	MAE	R <sup>2</sup>
0	DummyRegressor	1.16	0.57	-0.15
1	LinearRegression	54446427422.71	3104053900.17	-2558863623205183029248.00
2	Lasso	0.51	0.30	0.77
3	Ridge	0.52	0.30	0.76
4	ElasticNet	0.52	0.30	0.77
5	DecisionTreeRegressor	0.55	0.28	0.74
6	RandomForestRegressor	0.54	0.27	0.75
7	XGBRegressor	0.55	0.25	0.74



Pour SiteEnergyUse(kBtu):

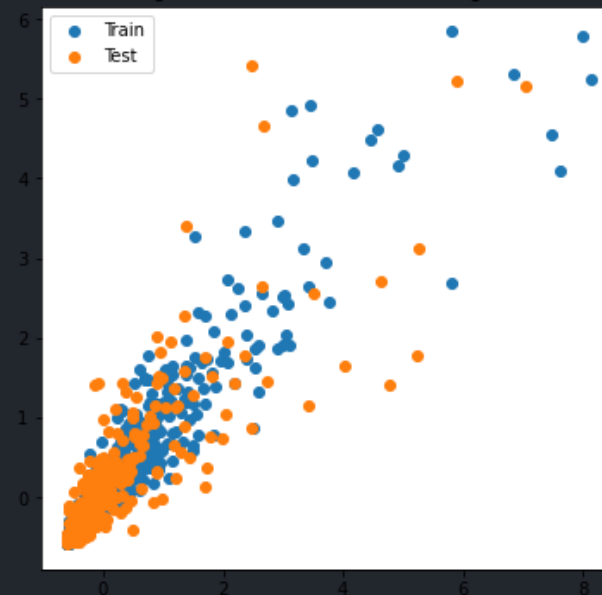
	EnergyStarScore	Model	RMSE	MAE	R <sup>2</sup>
0	Avec	Lasso	0.49	0.25	0.77
1	Sans	Lasso	0.49	0.25	0.77

# LES RÉSULTATS

Pour TotalGHGEmissions:

	Model	RMSE	MAE	R <sup>2</sup>
0	DummyRegressor	1.04	0.50	-0.14
1	LinearRegression	11295499.16	553078.52	-134886047540599.64
2	Lasso	0.62	0.38	0.59
3	Ridge	0.61	0.38	0.60
4	ElasticNet	0.61	0.38	0.60
5	DecisionTreeRegressor	0.59	0.29	0.64
6	RandomForestRegressor	0.51	0.26	0.72
7	XGBRegressor	0.53	0.26	0.70

Nuage de Points: RandomForestRegressor



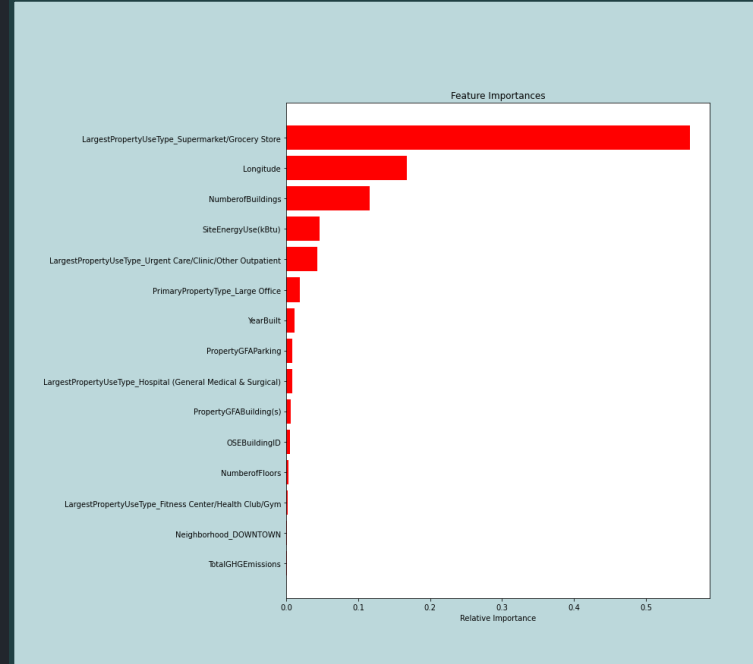
Pour TotalGHGEmissions:

EnergyStarScore			Model	RMSE	MAE	R <sup>2</sup>
0	Avec	RandomForestRegressor		0.51	0.25	0.76
1	Sans	RandomForestRegressor		0.51	0.25	0.75

# Interprétation GLOBAL

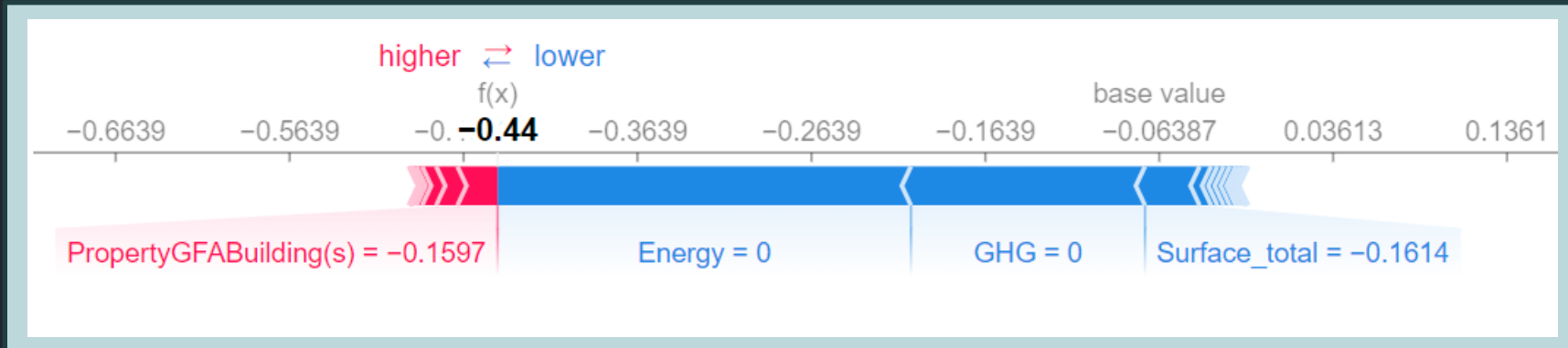
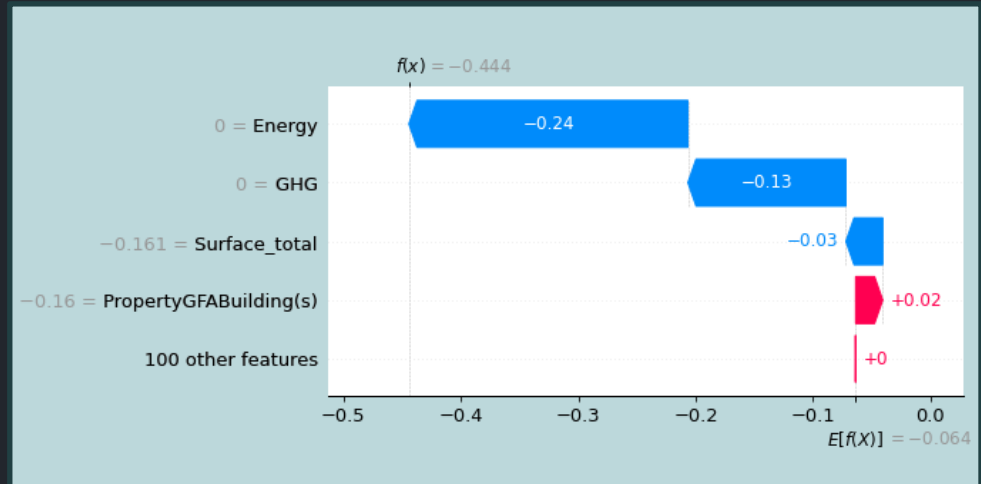
Utilisation de GridSearchCV et de  
la cross-validation

Visuel de Feature Importances



## Interprétation local

Utilisation de SHAP pour une interprétation local



# CONCLUSION

Il est plus facile de prédire TotalEnergieUse que TotalGHGEmissions,

Nous pourrions envisager de prendre des données sur plusieurs années puis le comparer entre elles

Ajout de variables liées au temps moyen sur l'année, ainsi que les pics haut et bas des températures

Concernant EnergieStarScore sont utilisation ne change pas les résultats sur SiteEnergyUse et l'améliore que très peu pour TotalGHGEmissions

