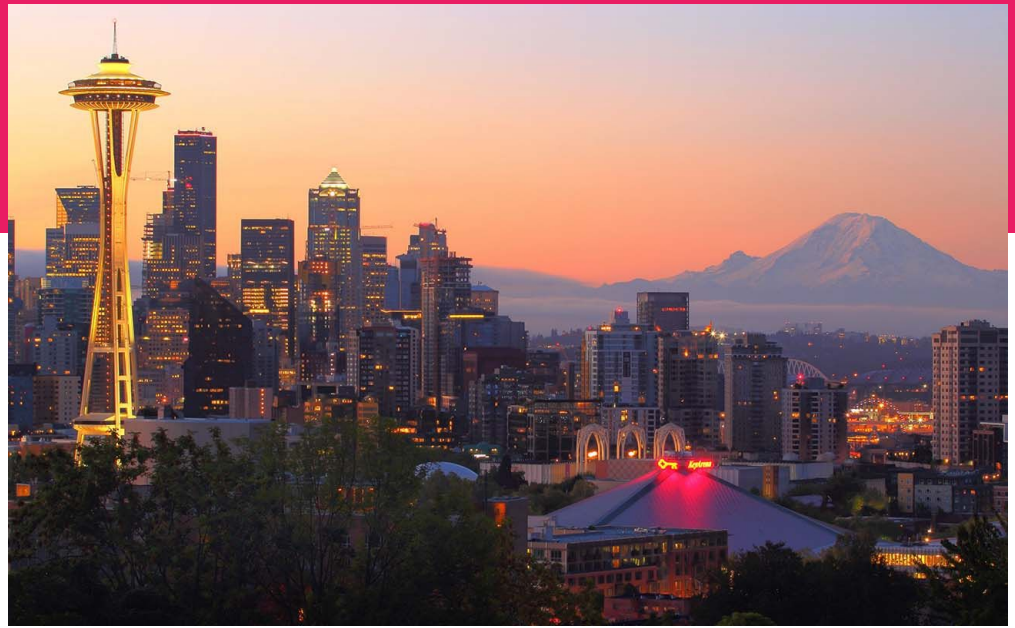
Anticipez les besoins en consommation électrique de bâtiments de la ville de Seattle



Sommaire

Problématique

Problématique

La ville de Seattle souhaite atteindre la neutralité carbone en 2050, dans ce but il nous a été demandé de mettre au point un modèle de prédiction des émissions de CO2 et de la consommation totale d'énergie de bâtiments non-résidentiels en se passant des relevés énergétiques, très coûteux à établir.

Il nous a aussi été demandé d'étudier l'intérêt de l'ENERGYSTAR Score qui est aussi très difficile à calculer.

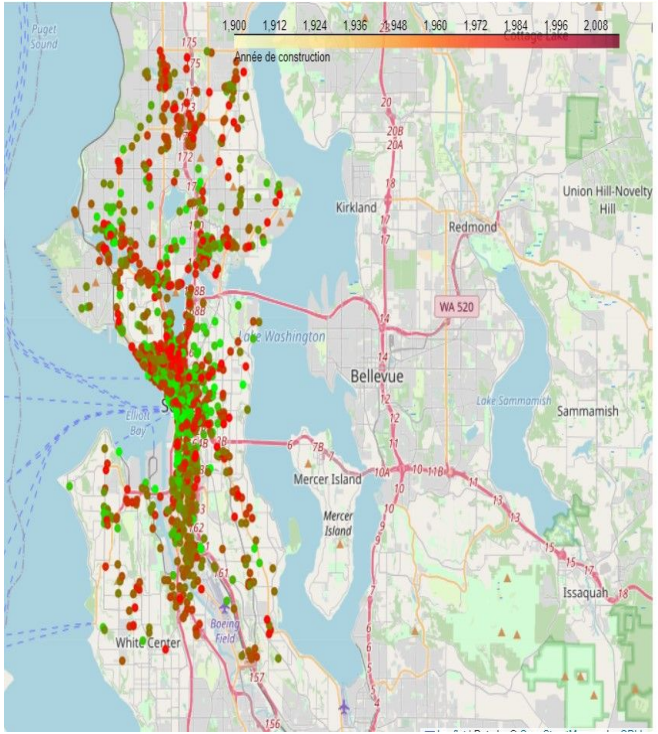
Les Données

Les Données

Le jeu de données nous donnes accès à 46 caractéristiques pour plus de 3000 bâtiments sur l’années 2015 ,

- Surface des bâtiments
- Le type de propriété
- L’années de construction
- La consommation d'énergie
- L'émission de CO2

…..



Géolocalisation des bâtiments par années de construction

Nettoyage des Données

1. Supprimer les colonnes qui ne contiennent que des valeurs manquantes.
2. Supprimer les propriétés avec la valeurs high ou low outlier on se basant sur les valeurs de la variable Outlier.
3. Sélectionnez les propriétés autres que résidentielles.
4. Sélectionnez que les données conformes (complication) correspondent aux données conformes.
5. Ne conservez que des propriétés avec un seul bâtiment.
6. Conserver uniquement les propriétés avec un seul bâtiment.
7. Supprimer les propriétés non énergétiques et les valeurs négatives.
8.

Avant le nettoyage

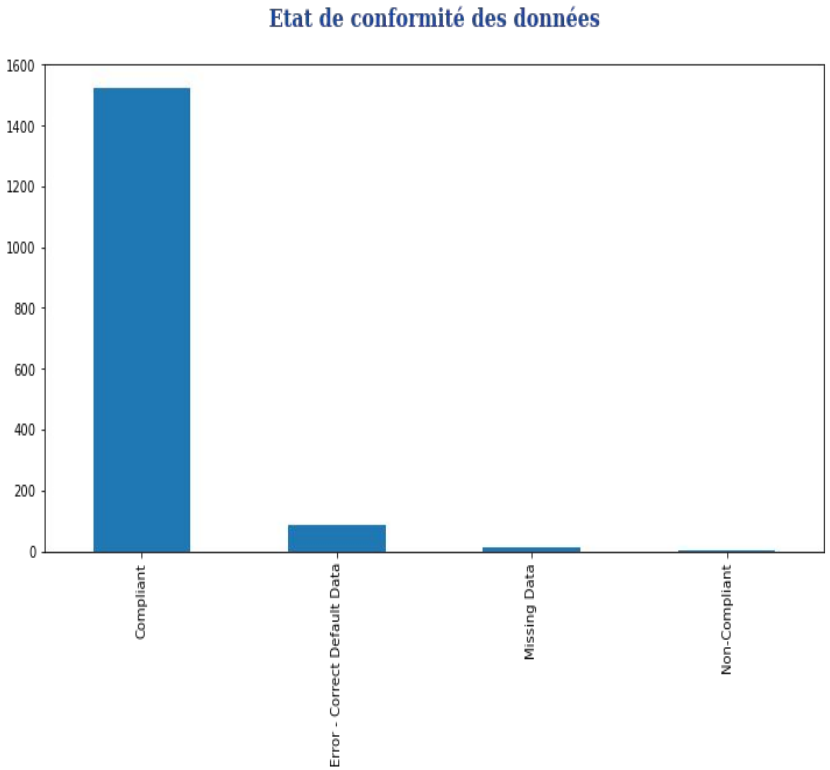
3376 et 46 colonnes

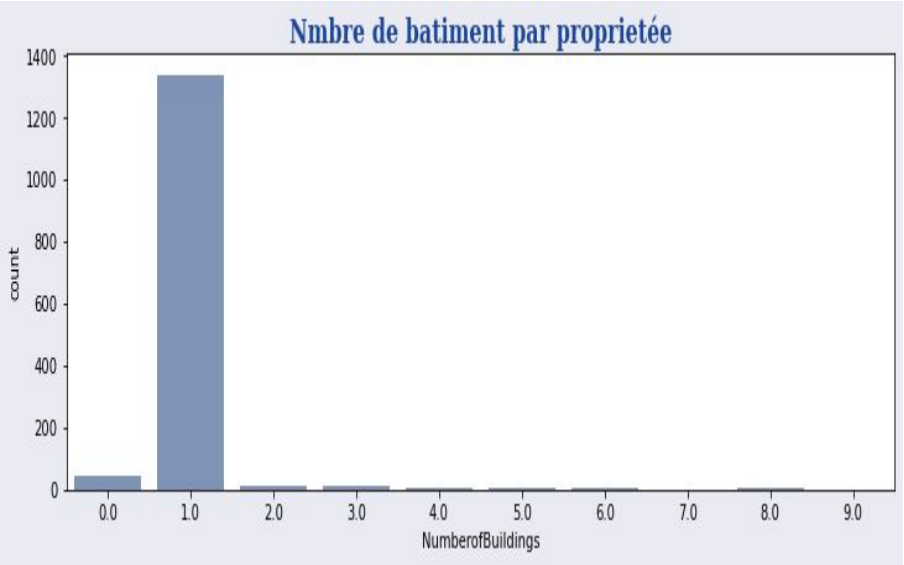
Après le nettoyage

1409 lignes et 24 colonnes

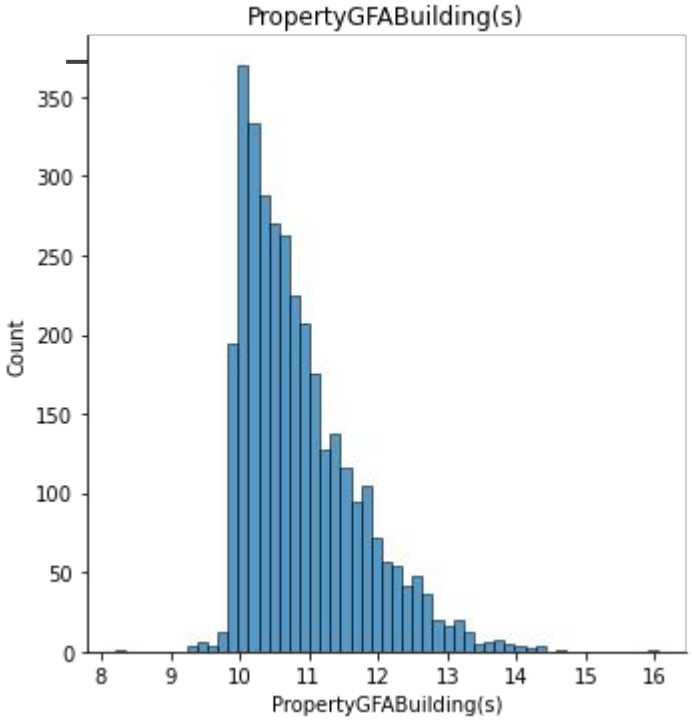
Observation des Données

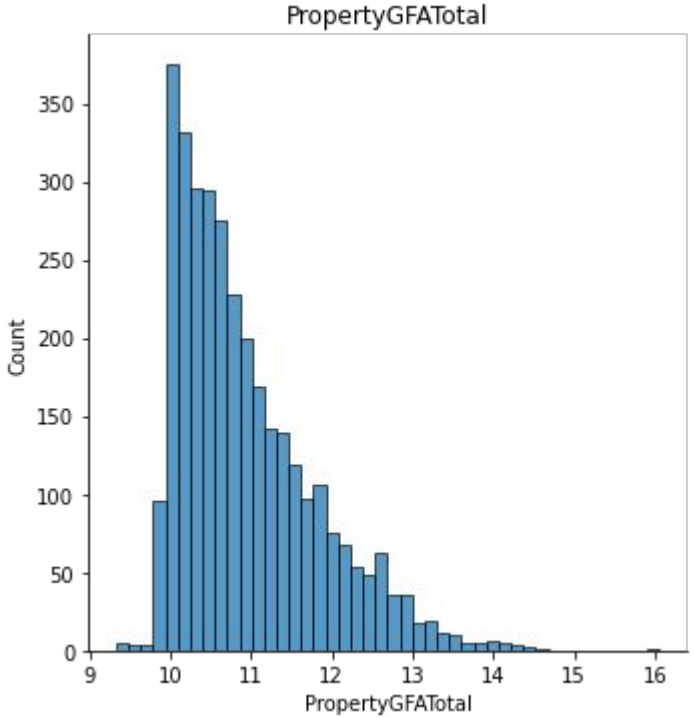
Observation des des variables catégorielles:



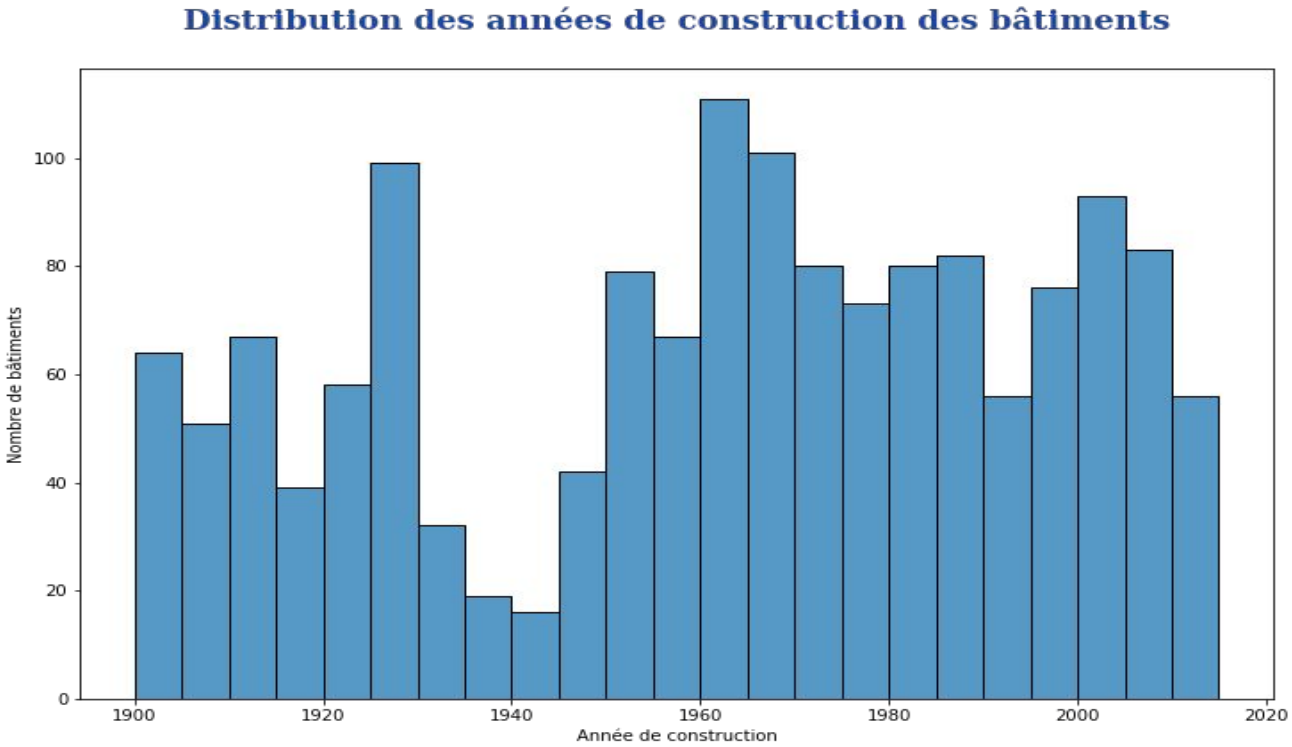


Observation des des variables numériques:

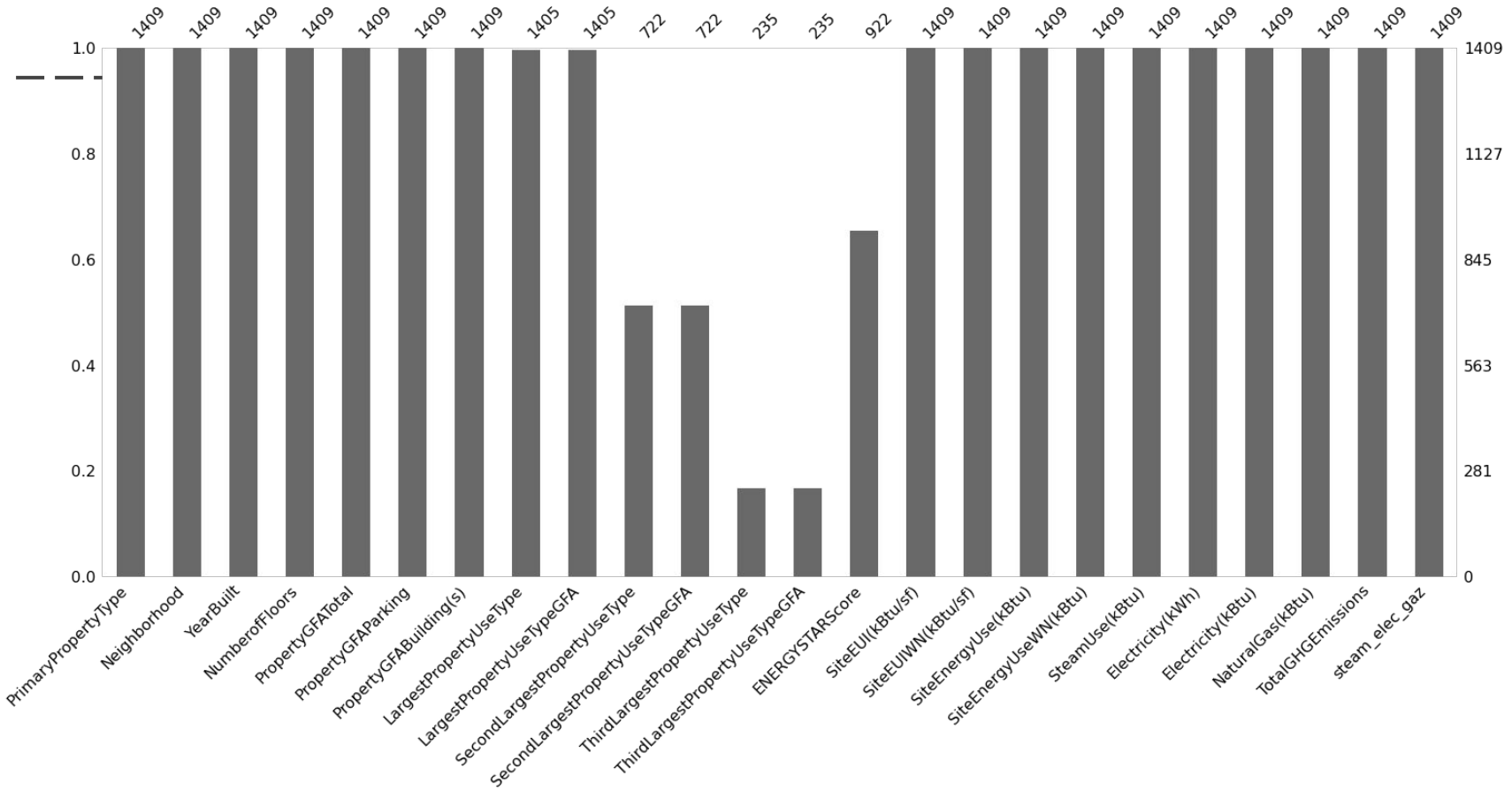




Observation des des variables numériques:



Les valeurs manquantes



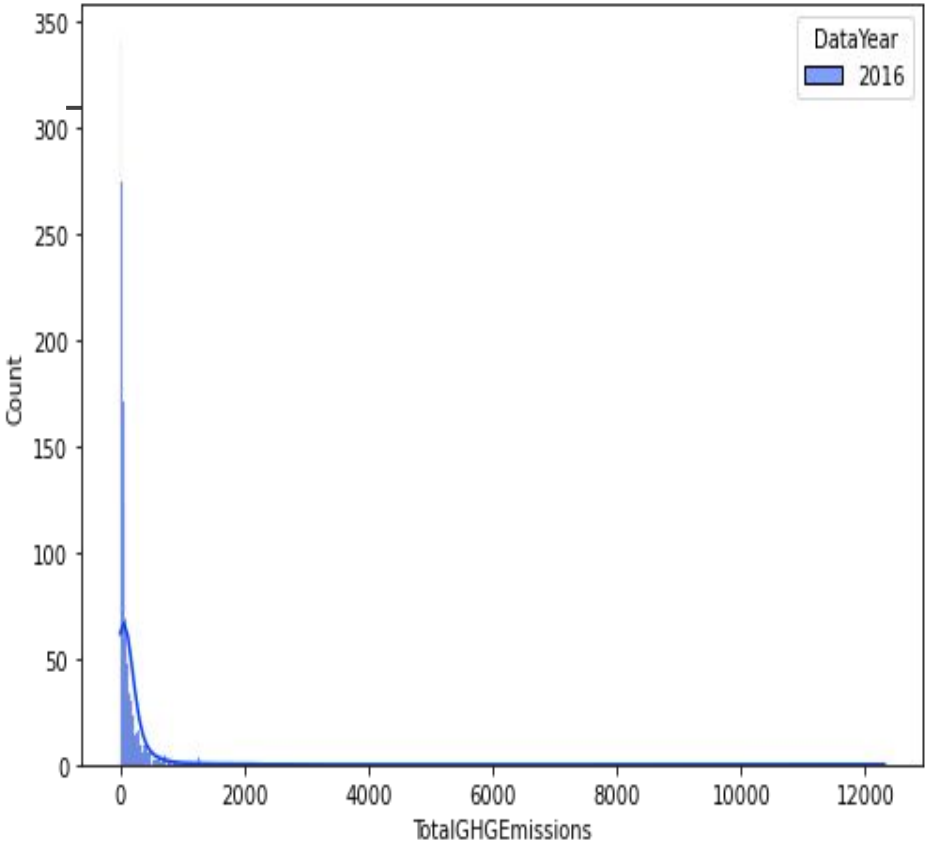
Target : SiteEUIWN(kBtu/sf)

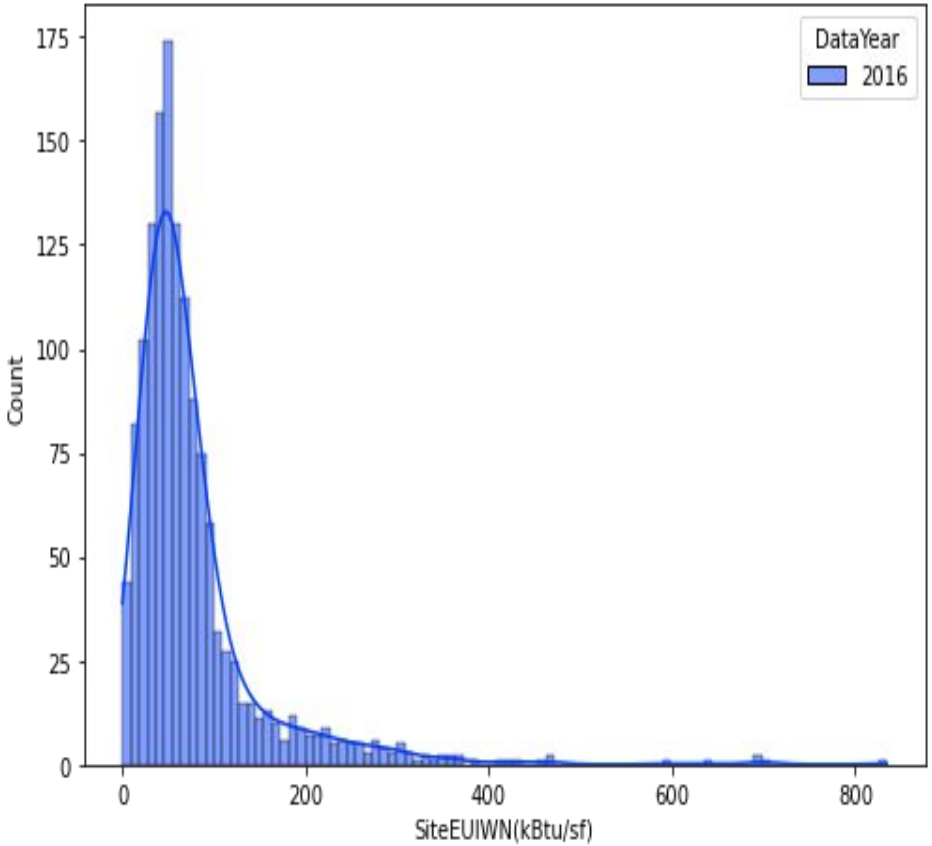
L'intensité de consommation d'énergie du site (EUI) normalisée selon les conditions météorologiques (WN) et WN normalisé selon la superficie (en pieds carrés).

Target : TotalGHGEmissions

La quantité totale d'émissions de gaz à effet de serre, y compris les gaz de dioxyde de carbone, de méthane et d'oxyde d'azote, rejetées dans l'atmosphère en raison de la consommation d'énergie de la propriété

Observation des Target:





Les features:

_ _ _

- PropertyGFAParking
- PropertyGFABuilding(s)
- PrimaryPropertyType
- NumberofFloors
- YearBuilt
- Neighborhood
- EberguStarScore

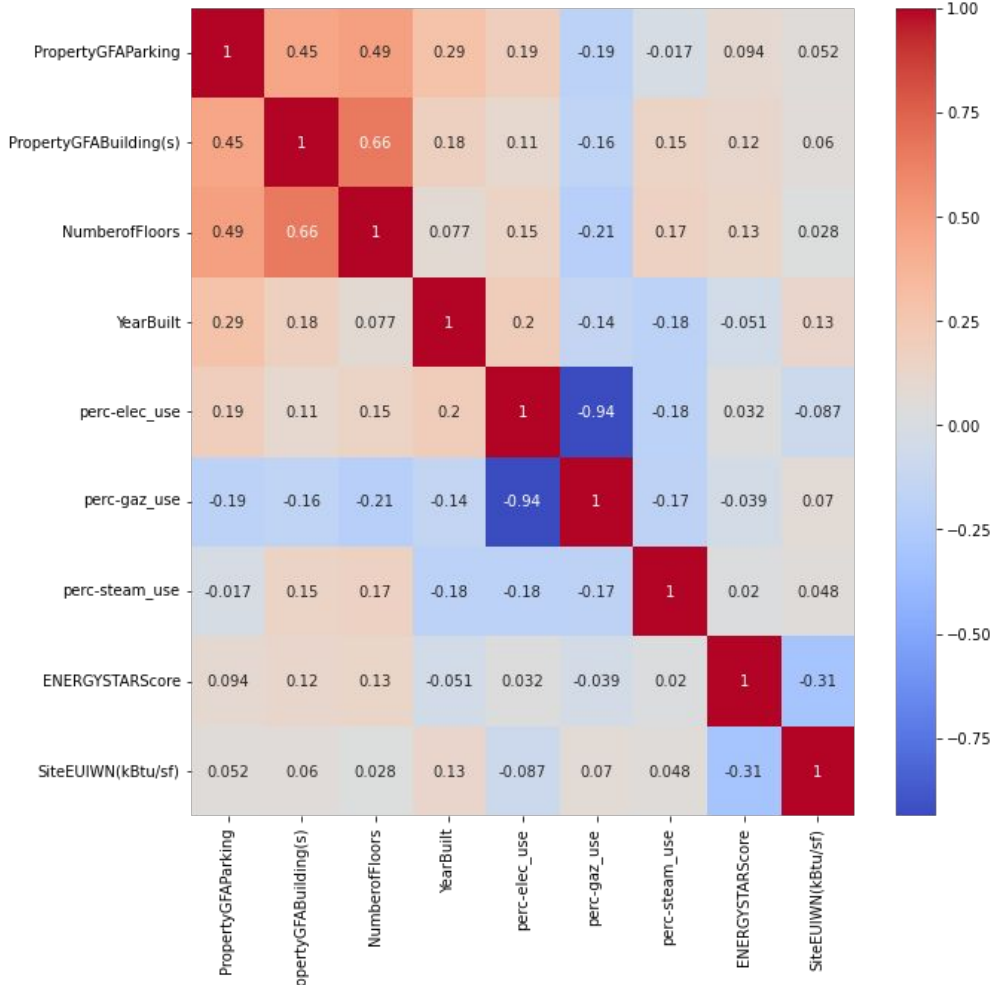
Features engineering :

- perc-elec_use
- perc-gaz_use
- perc-steam_use
-

Analyse de corrélation:

— — —

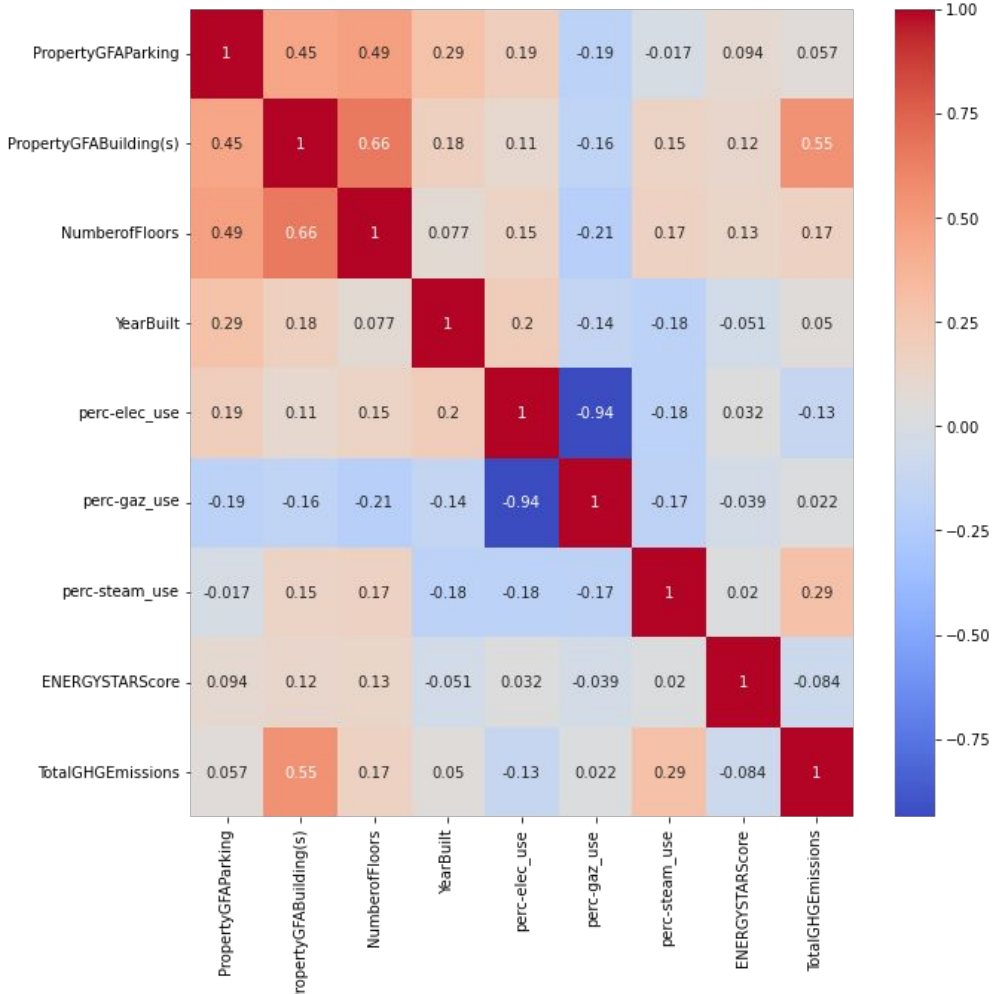
Target :
SiteEUIWN(kBtu/sf)
est le plus corrélé avec ENERGY STAR Score



Analyse de corrélation:

Target :
TotalGHGEmissions : est le plus corrélé avec

PropertyGFABuilding(s)



Modélisation

Features engineering :

Calculer le pourcentage d'énergie consommée par bâtiment, nous utiliserons le total sans le WN. Ceci est basé sur la somme des trois énergies (électricité, vapeur et gaz) :
perc-elec_use , perc-gaz_use, perc-steam_use

Utiliser les colonnes avec les surfaces et les valeurs des variables proprety use type pour calculer le pourcentage par rapport à la surface total

Apprentissage supervisé
Problème de régression

Les modèles testés :

– – –

Modèle Naïf(Dummy regressor)

Régression linéaire

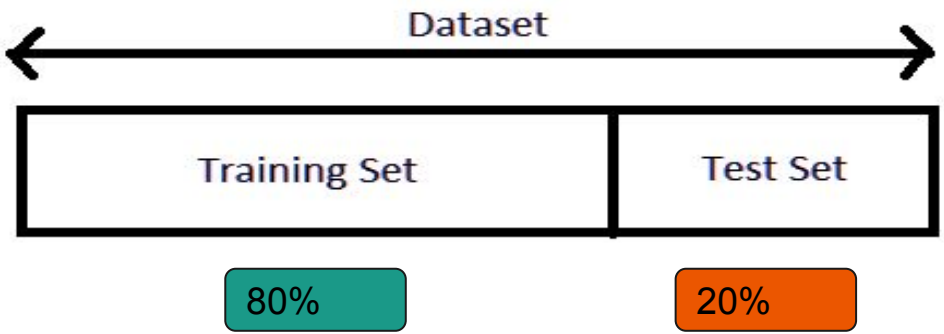
Ridge

Lasso

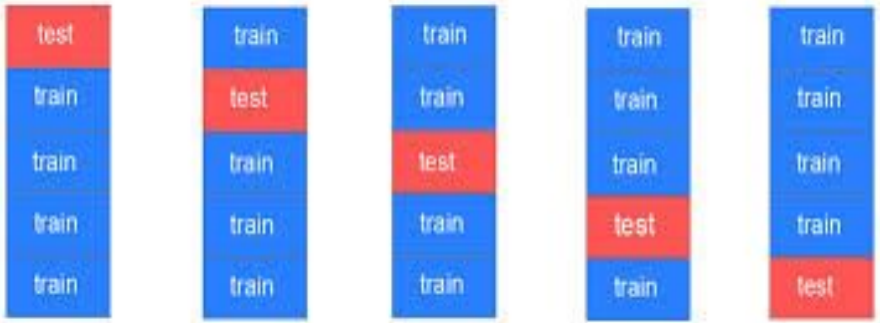
Arbre de décision

Forêt aléatoire

Partitionnement des données:



Cross validation et Grid search



Métriques d’ évaluation:

RMSE : racine carré de MSE (erreur quadratique moyenne)

R2 : coefficient de détermination

Optimisation des modèles

Grid search

Mettre la target a l’échelle logarithmique

Features engineering

Résultats des modèles :
Prédiction de prédiction des émissions de CO2

Modele	R2 avant features engineering	R2 après(Energy Star Score)
Ridge	Test :0.39 sans le log Train: 0.47	Test :0.66 sans le log Train: 0.78
Lasso	Test :0.36 sans le log Train: 0.34	Test :0.65 sans le log Train: 0.72
Arbre de décision	Test :0.63 sans le log Train: 0.93	Test :0.67 avec le log Train: 0.76
Forêt aléatoire	Test :0.36 avec le log Train: 0.76	Test :0.68 avec le log Train: 0.81

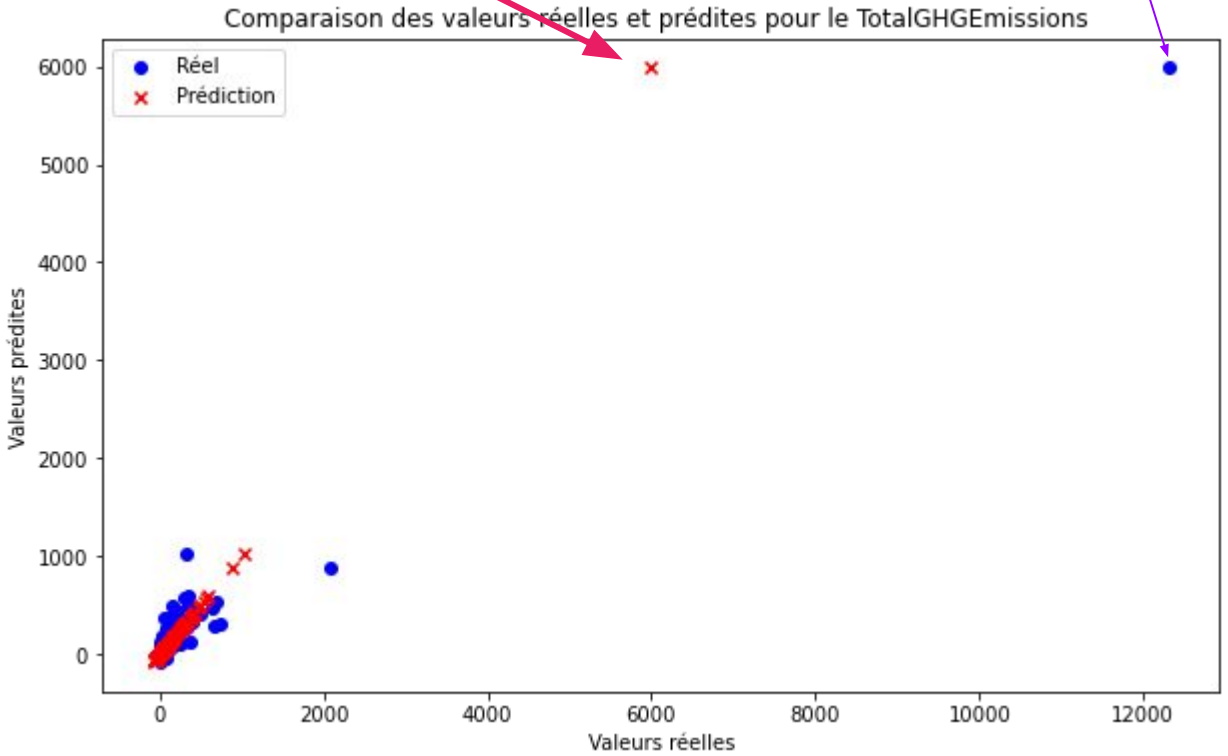
Modèle final	R2
Ridge	Train 0.74 Test 0. 71 sans le log

Le Modèle sélectionné Ridge:
Alpha = 29.5

R2:
Train 0.74

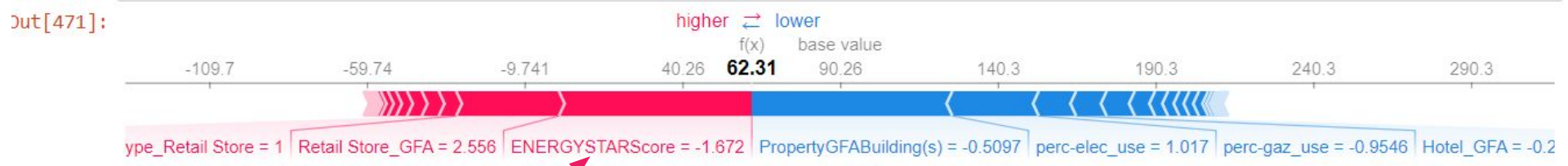
valeurs mal prédites par le modele

valeurs réel atypique

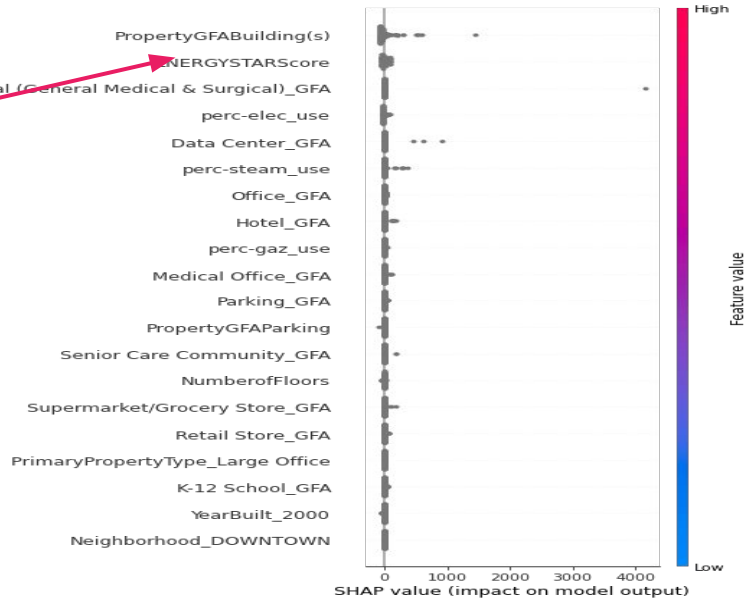


Test 0.71

La variable Energy star score



Features importance Globale et locale



Résultats des modèles :

Prédiction de la consommation totale d'énergie

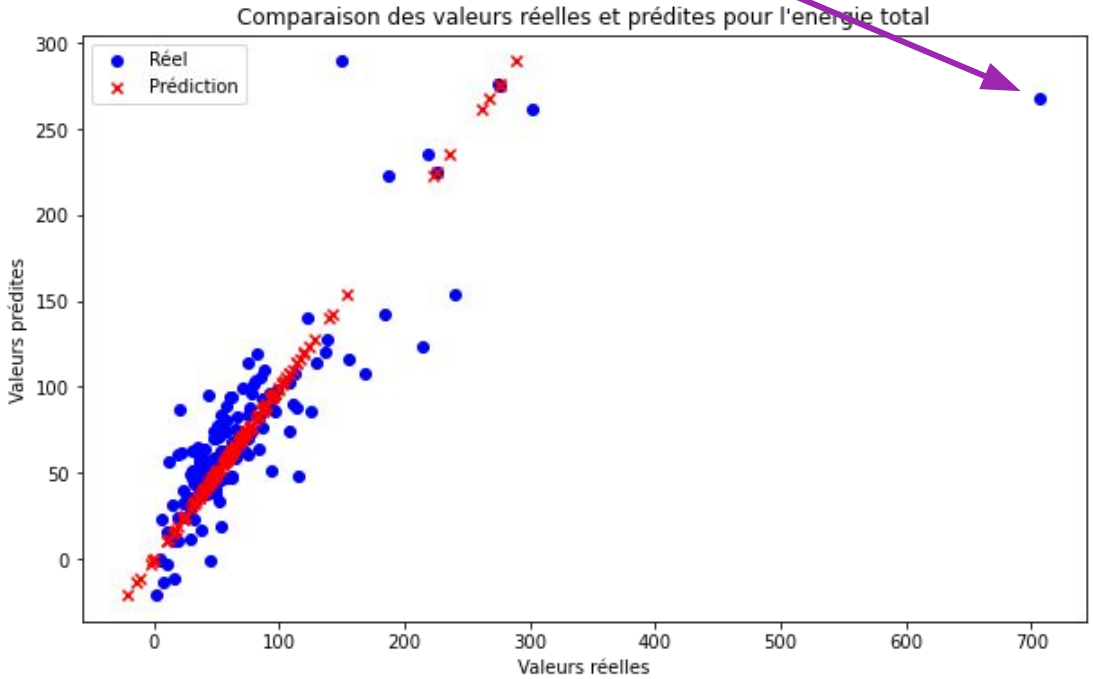
Model	R2 avant features engineering	R2 après features engineering
Ridge	Train 0.40 avec le log Test 0.62	Train 0.43 Test 0.65
Lasso	Train 0.54 avec le log Test 0.48	Train 0.34 Test 0.62 avec le log
Arbre de décision	Train 0.84 Test 0.63 avec le log	Train 0.95 Test 0.64 sans le log
Forêt aléatoire	Train 0.85 Test 0.71 sans le log	Train 0.85 avec le log Test 0.75

Model	R2
Ridge	Train 0.83 sans le log Test 0.65

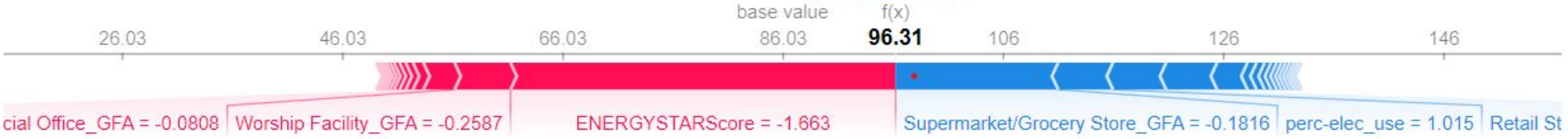
Valeurs atypiques

Le Modèle sélectionné
Ridge:
Alpha = 29

R2:
Train 0.83



Test 0.65



Features importance Globale et locale

Axe d’amélioration :

– – –

Data augmentations

Merci pour votre écoute.