



Seattle

**Anticipez les
besoins en
consommation
de bâtiments**

Sommaire



Contexte

Analyse exploratoire



Modélisation

Conclusion



Objectif ville neutre en émissions de carbone en 2050



Seattle



Prédictions Consommation en énergie et émissions

- Cible : bâtiments non destinés à l'habitation
- Relevés de consommations et émissions coûteux : premier relevé de référence la première année
- A partir des données structurelles des bâtiments

Intérêt de l'Energy Star score

- Indicateur de l'efficacité énergétique relative d'un bâtiment
- Calcul de l'Energy Star score fastidieux
- Evaluer son intérêt pour la prédiction des variables cibles

Données disponibles, relevés de 2016



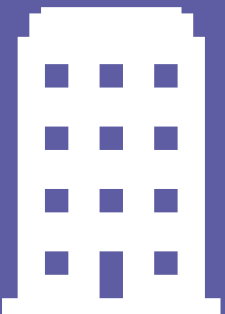
Données géographiques :

- Adresse, code postal
- Latitude et longitude
- Numéro de parcelle
- Quartier, code de district



Données d'usage des bâtiments :

- Type de bâtiments
- Différents usages du bâtiments et surfaces associées



Données de construction :

- Nombre d'étages
- Nombre de bâtiments
- Surfaces
- Année de construction / rénovation complète



Données consommation / émissions carbone :

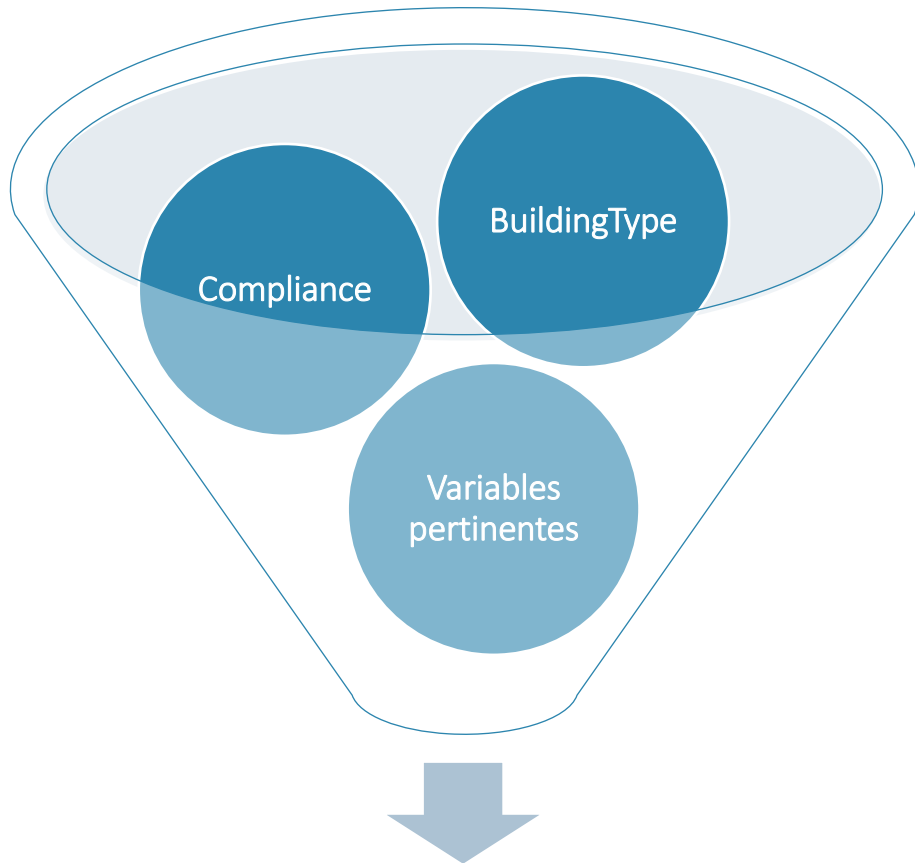
- Consommation énergétique totale et en électricité, gaz, vapeur
- Emissions carbone
- Fiabilité des données
- Energy Star Score



Analyse exploratoire

Nettoyage des données

Lignes : 3376, colonnes : 42



Lignes : 1324, colonnes : 14

- **Filtre et suppression de variables :**
 - BuildingType = Non residential
 - ComplianceStatus = 'Compliant'
 - Données géographiques non pertinentes
 - Données utilisation et structure du bâtiment
 - Variables avec beaucoup de valeurs manquantes
- **Valeurs aberrantes**
- **Ajout de nouvelles variables :**
 - Mix énergétique
 - % surface
 - Age du bâtiment
- **Targets :**
 - SiteEnergyUse(kBtu)
 - TotalGHGEmissions

Variables sélectionnées



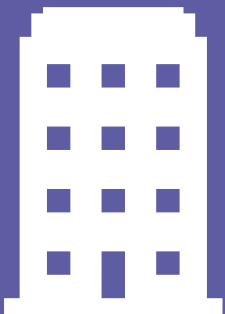
Données géographiques :

- Neighborhood



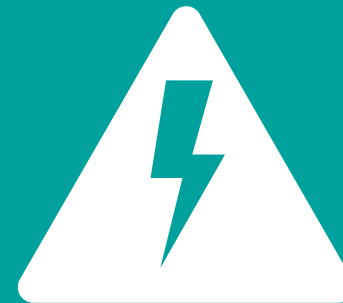
Données d'usage des bâtiments :

- PrimaryPropertyType
- PropertyGFAParking(%)
- PropertyGFABuilding(%)



Données de construction :

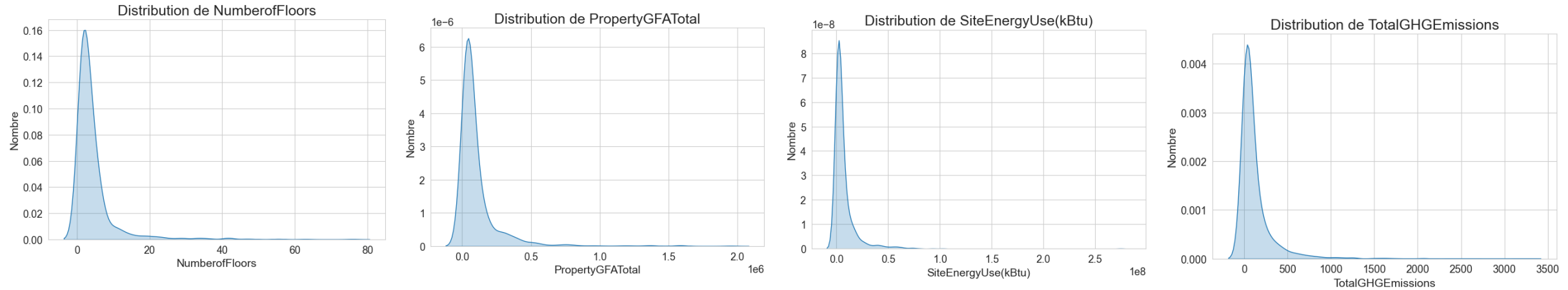
- NumberofFloors
- NumberofBuilding
- PropertyGFTotal
- BuildingAge



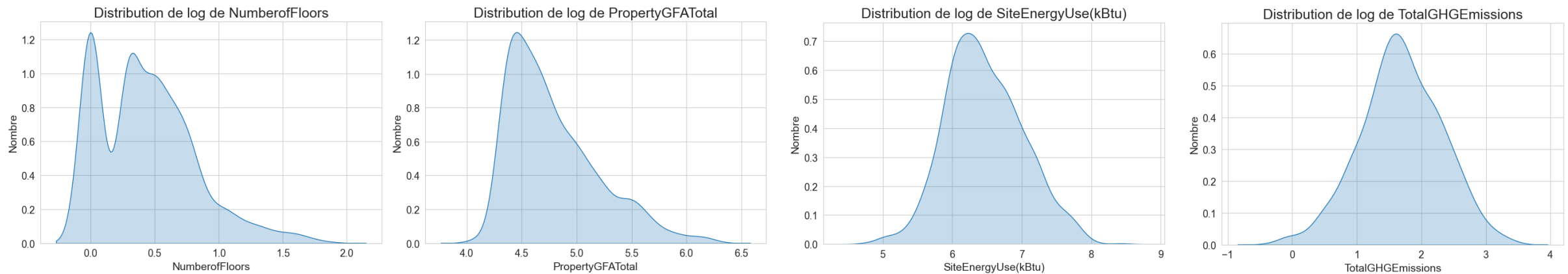
Données consommation / émissions carbone :

- SteamUse(%)
- NaturalGazUse(%)
- ElectricityUse(%)
- SiteEnergyUse(kBtu)
- TotalGHGEmissions
- Energy Star Score

Des distributions désaxées à droite

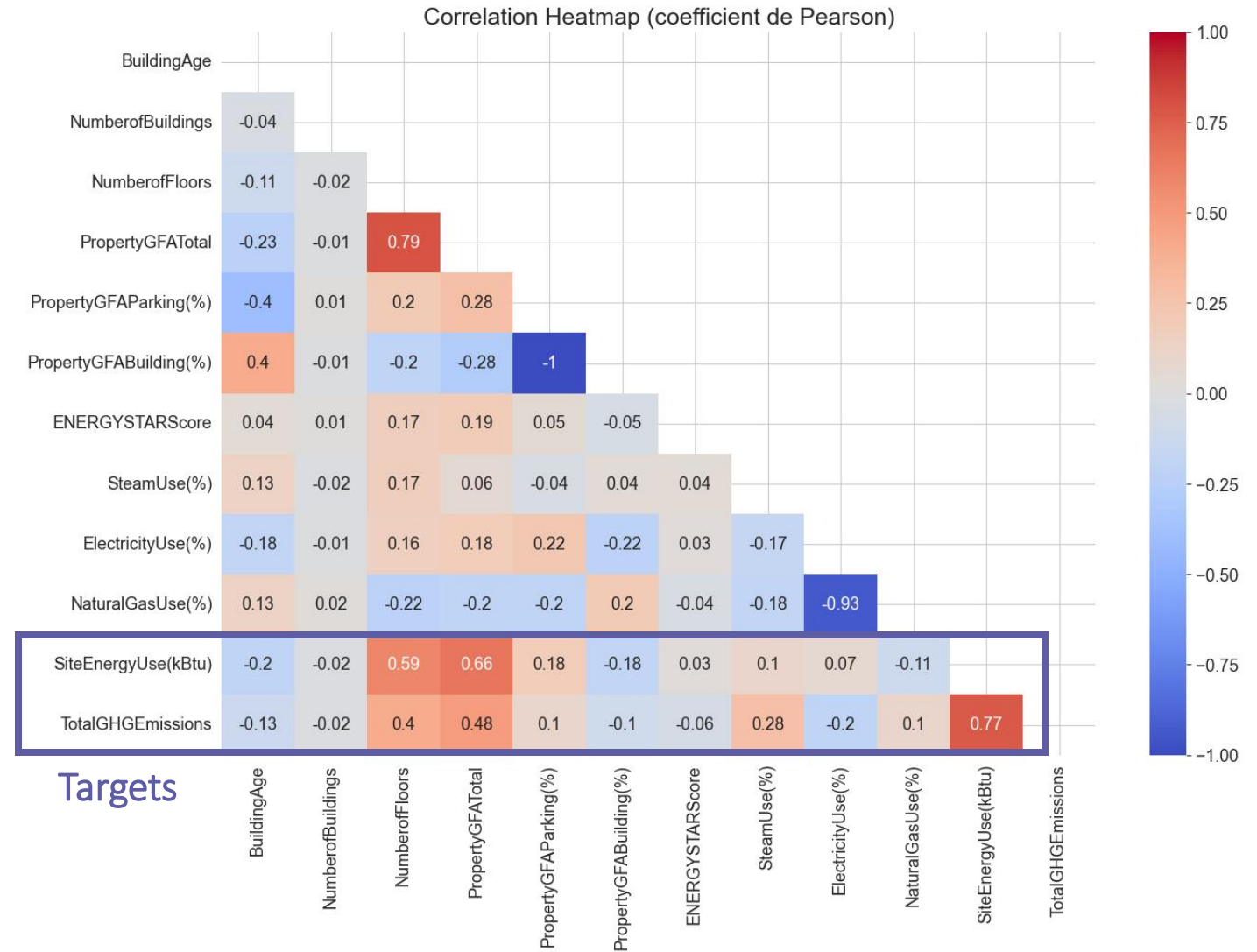


Transformation logarithmique



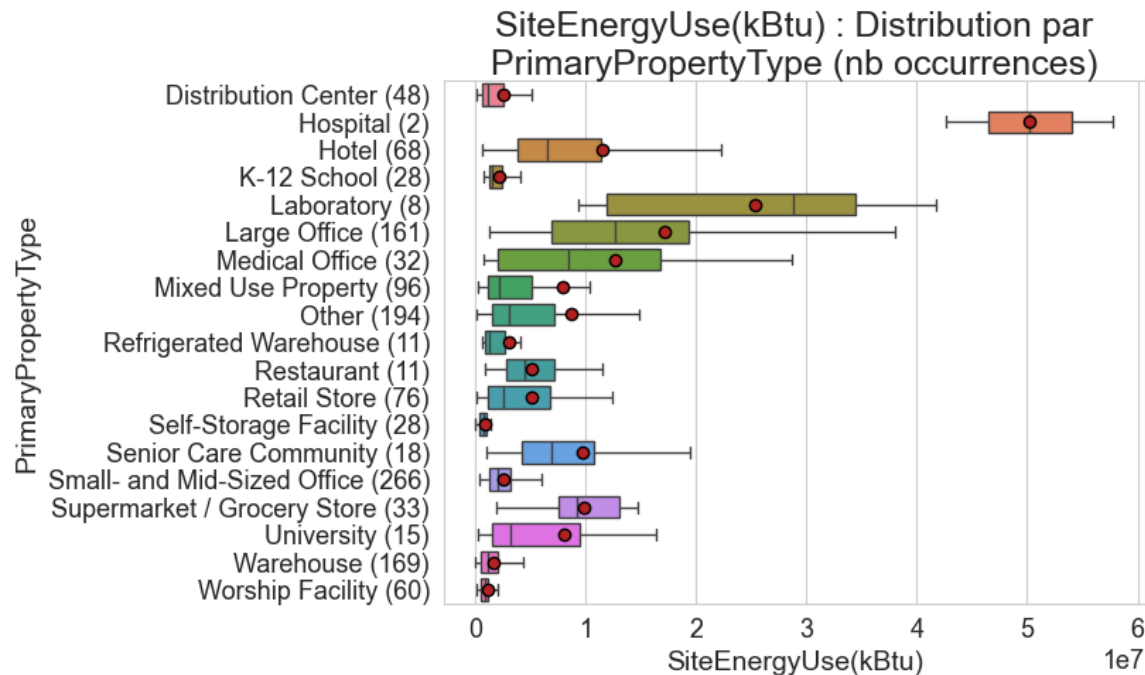
Corrélation entre les variables

- **Data leakage** : pas de corrélation entre les variables cibles (targets) et les variables de mix énergétique.
- Corrélation entre SiteEnergyUse et TotalGHGEmissions
- Corrélation entre SiteEnergyUse et PropertyGFATotal

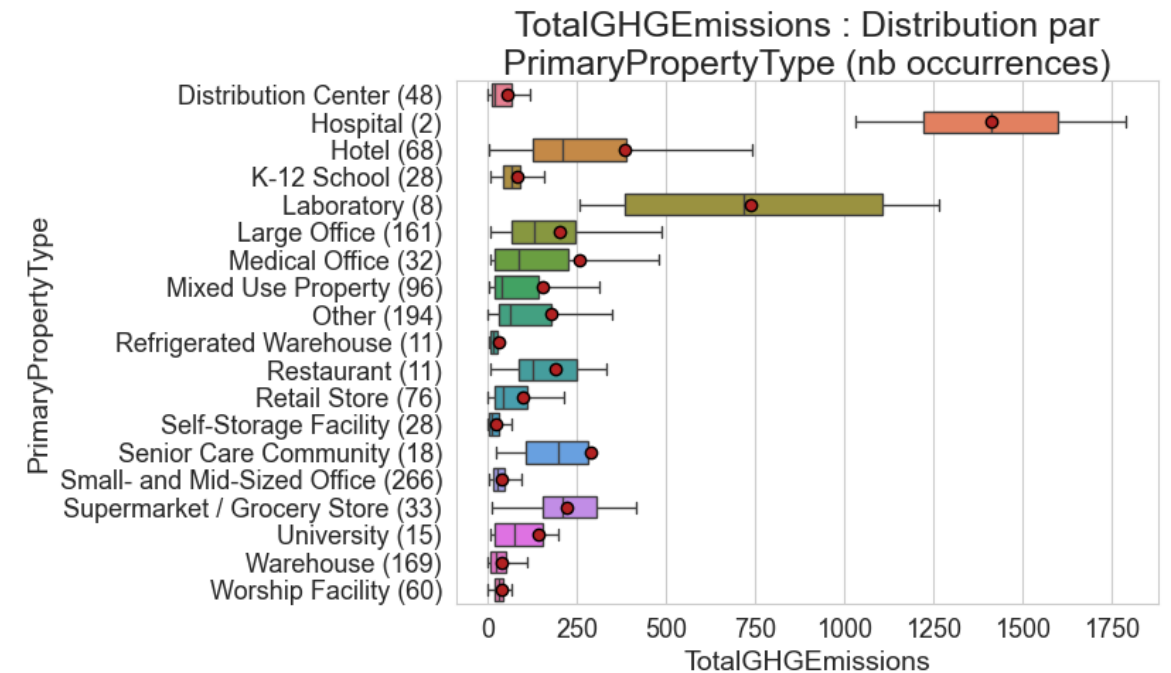


Hôpitaux et laboratoires : les plus grands consommateurs d'énergie et émetteurs carbone

SiteEnergyUse



TotalGHGEmissions

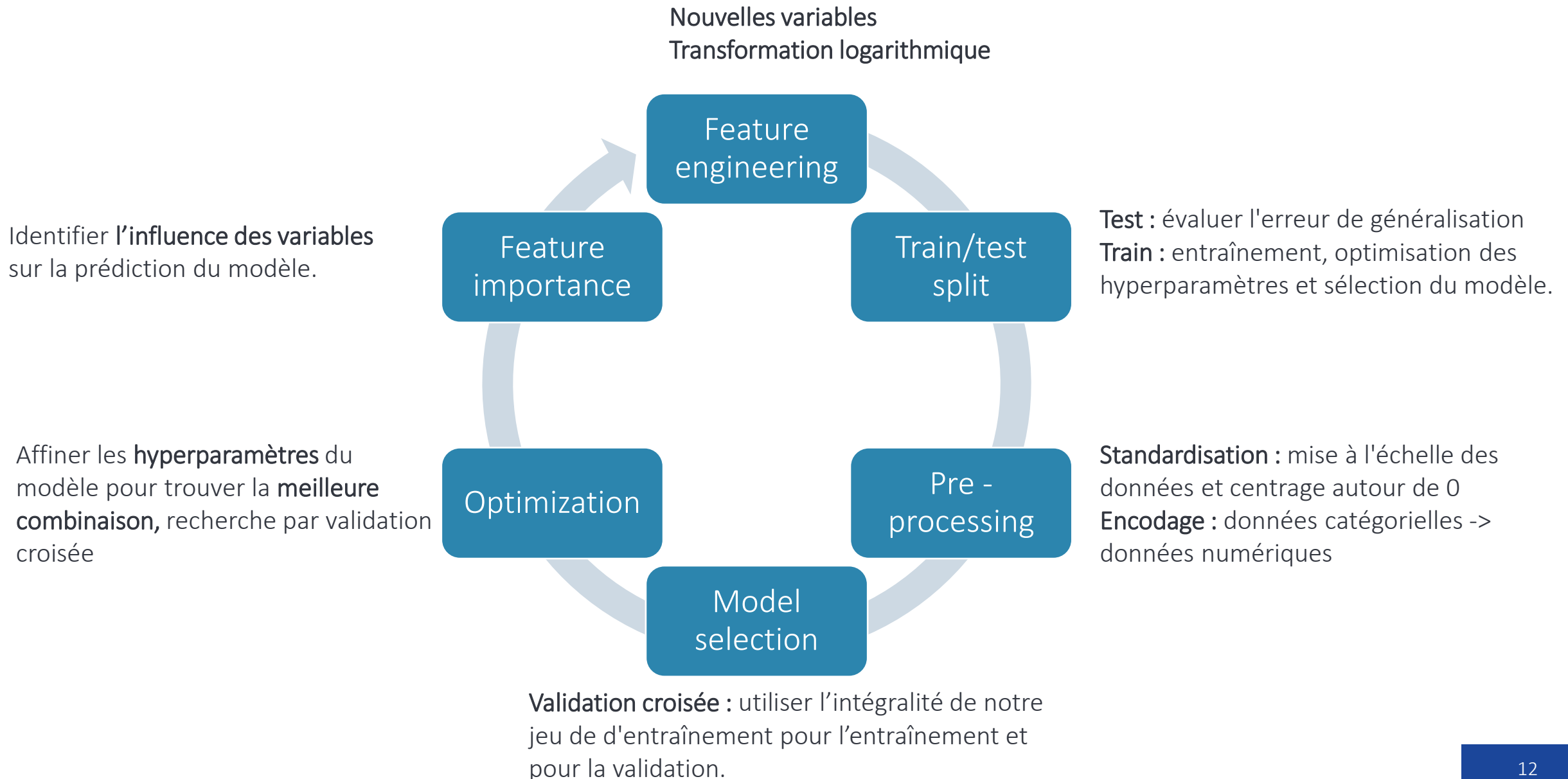


Des classes très hétérogènes en taille et une majorité de bureaux dans les bâtiments non résidentiels



Modélisation

Une démarche itérative



Les modèles testés



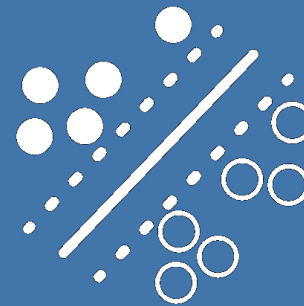
Baseline

Dummy Regressor
(approche naïve :
mean)

LinearRegression

Ridge
Lasso
ElasticNet

Méthodes linéaires
régularisées

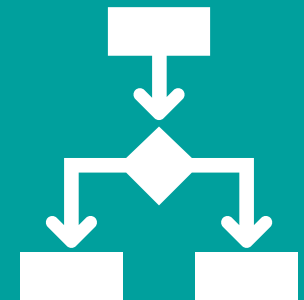


Méthodes non linéaires

**Support Vector
Regressor (SVR) :**
Support Vector
Machine à noyau

**Random Forest
Regressor (parallèle)**
**Gradient Boosting
Regressor**
(séquentielle)

Méthodes ensemblistes



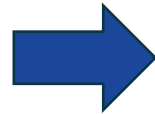
SiteEnergyUse : Sélection du modèle (1)

Scalers testés :

- StandardScaler
- RobustScaler

Encoders testés :

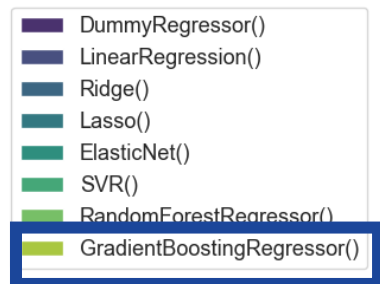
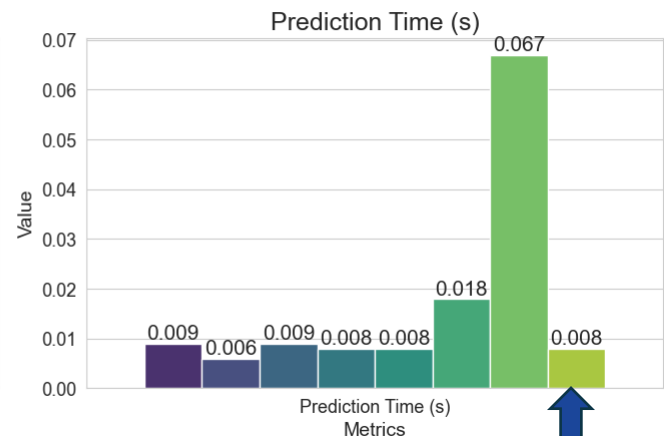
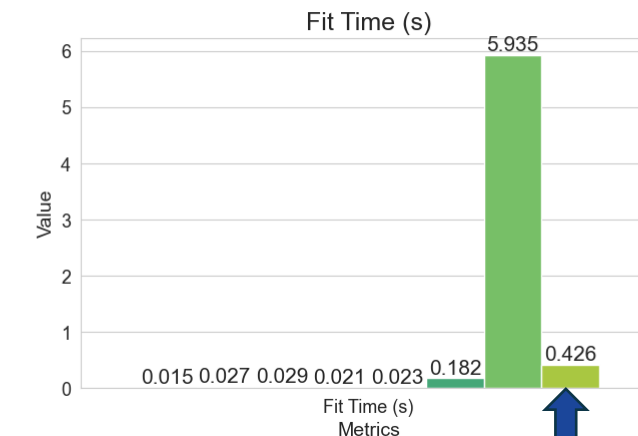
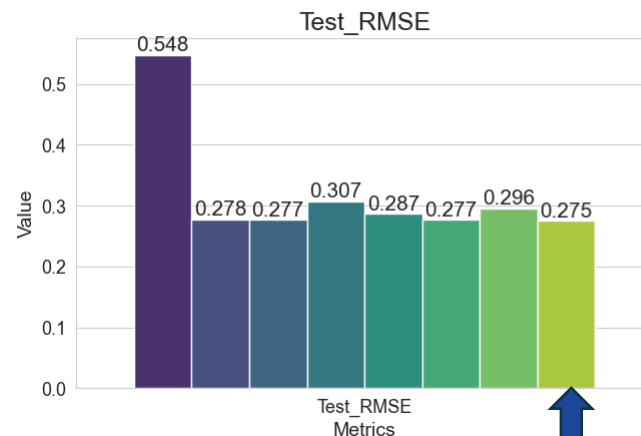
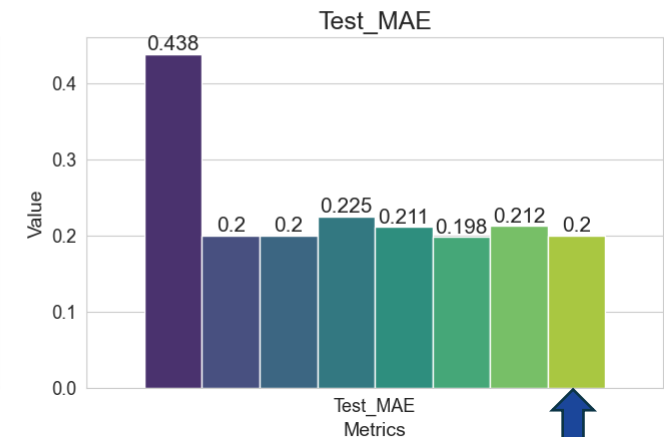
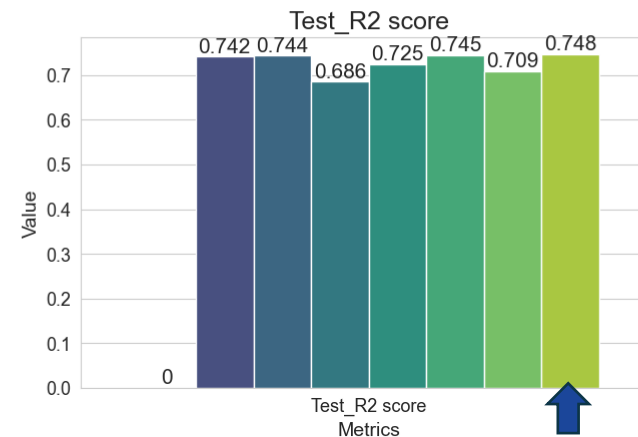
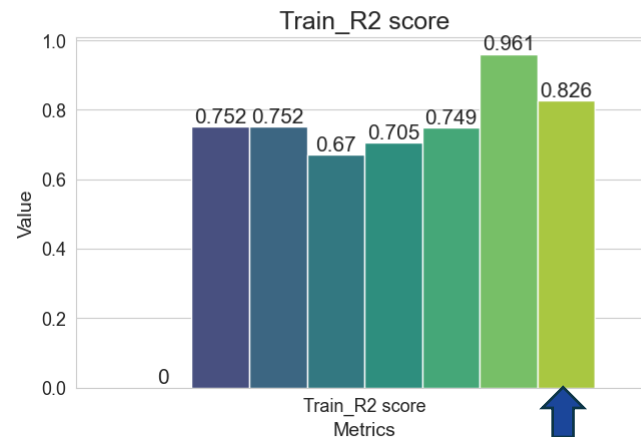
- TargetEncoder
- OneHotEncoder



	Value
Encoder	OneHotEncoder()
Scaler	RobustScaler()
Model	GradientBoostingRegressor()
Parameters	{'gradientboostingregressor__n_estimators': 100}
CV_R2 score	0.730 (+/- 0.013)
Train_R2 score	0.826
Test_R2 score	0.748
Test_MAE	0.2
Test_RMSE	0.275
Fit Time (s)	0.426
Prediction Time (s)	0.008

SiteEnergyUse : Sélection du modèle (2)

Comparaison des performances des différents modèles pour la meilleure combinaison Encoder : OneHotEncoder(), Scaler : RobustScaler()



SiteEnergyUse : Optimisation du modèle

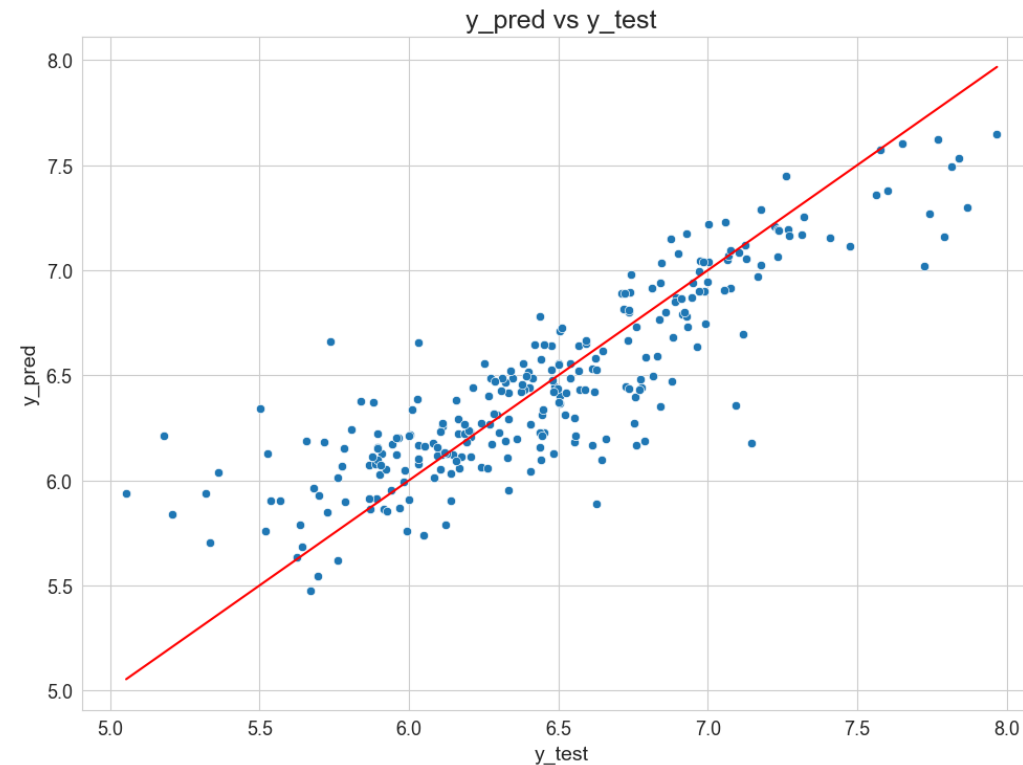
Méthodes :

- GridSearchCV : systématique
- RandomizedSearchCV : aléatoire

Hyperparamètres :

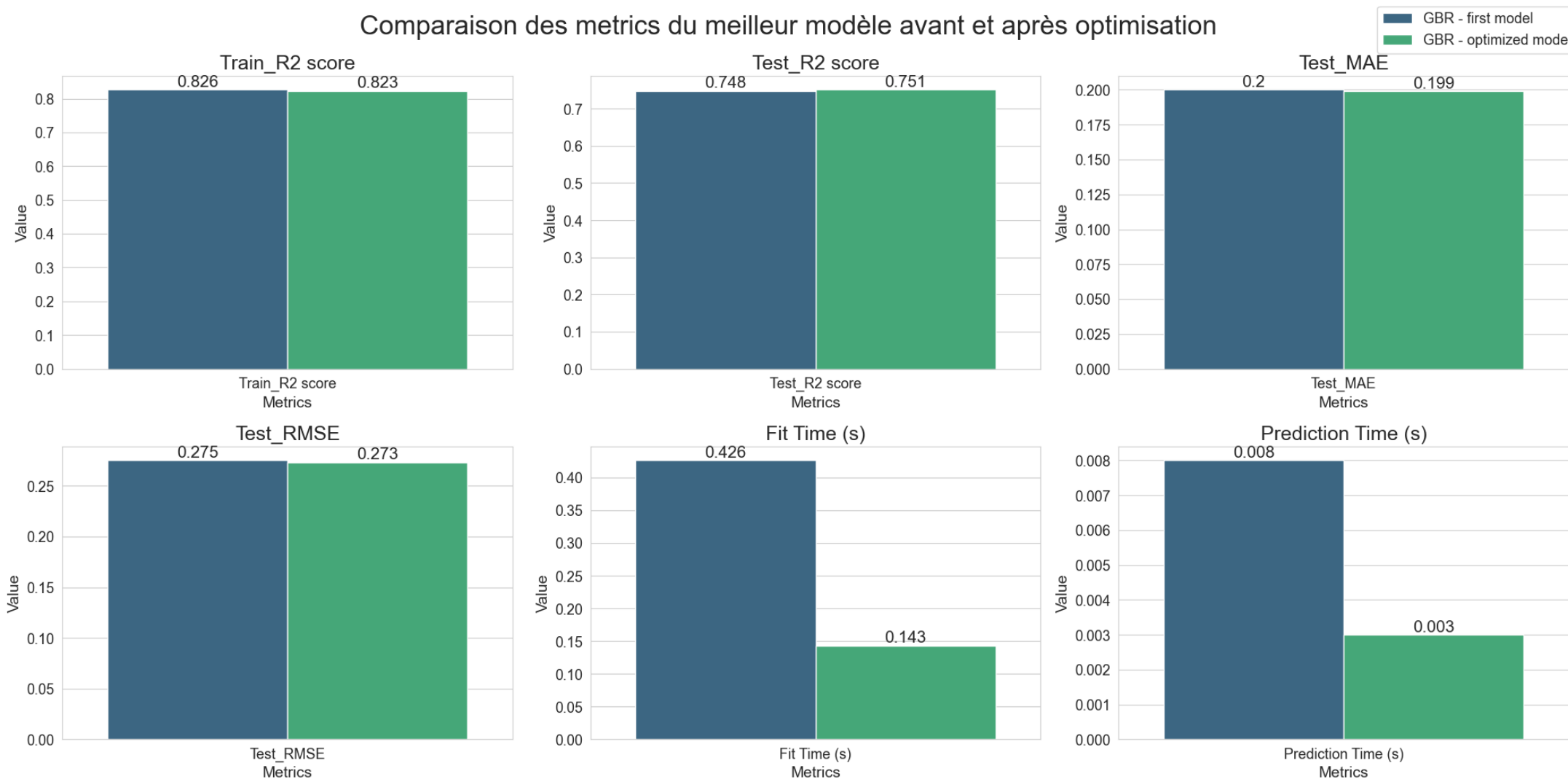
- n_estimators
- learning_rate
- max_depth
- min_samples_split

Résultat du modèle après optimisation



SiteEnergyUse : Optimisation du modèle

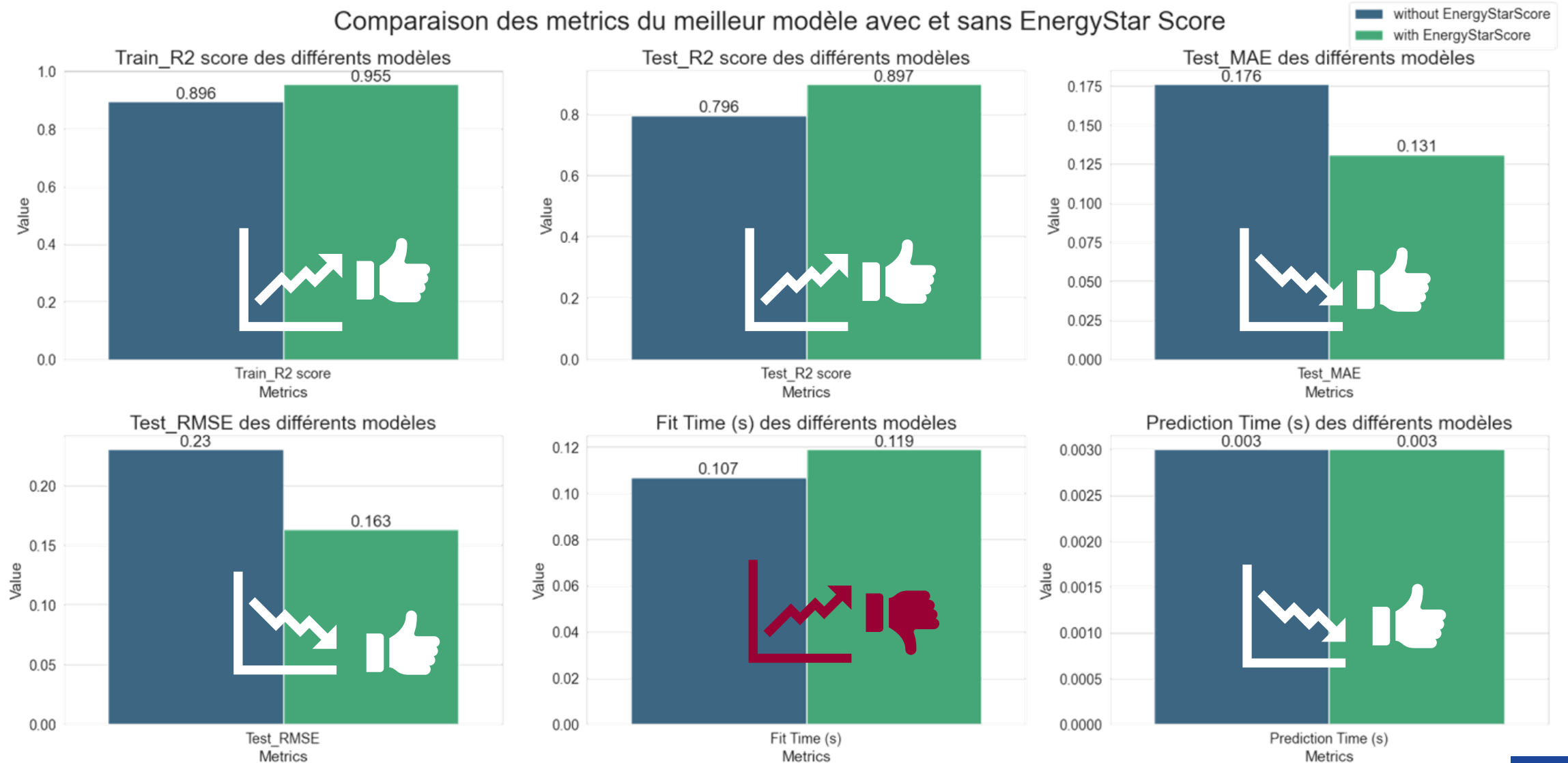
Comparaison des metrics du meilleur modèle avant et après optimisation



Peu d'améliorations des performances du modèle après optimisation.

L'Energy Star Score améliore les performances du modèle pour la prédiction de SiteEnergyUse

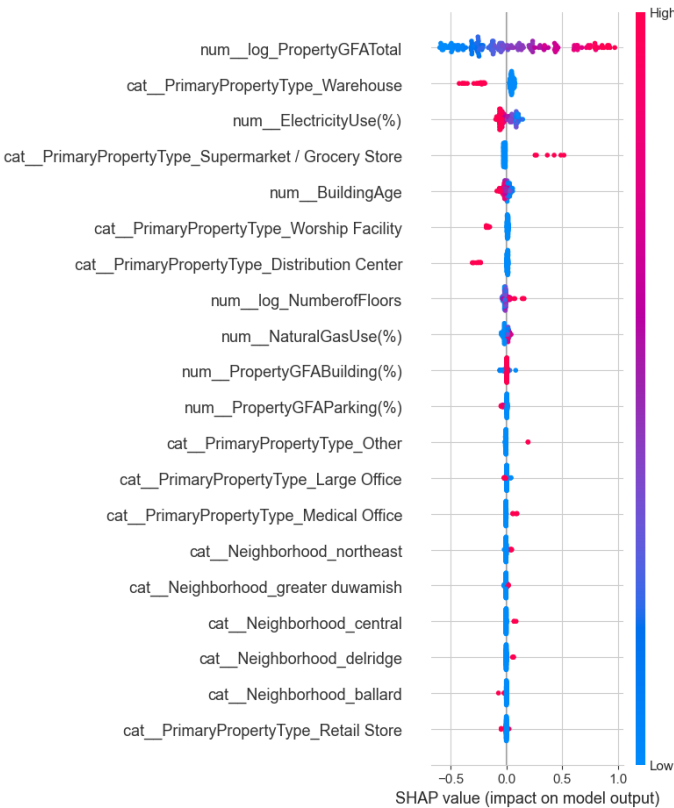
Comparaison des metrics du meilleur modèle avec et sans EnergyStar Score



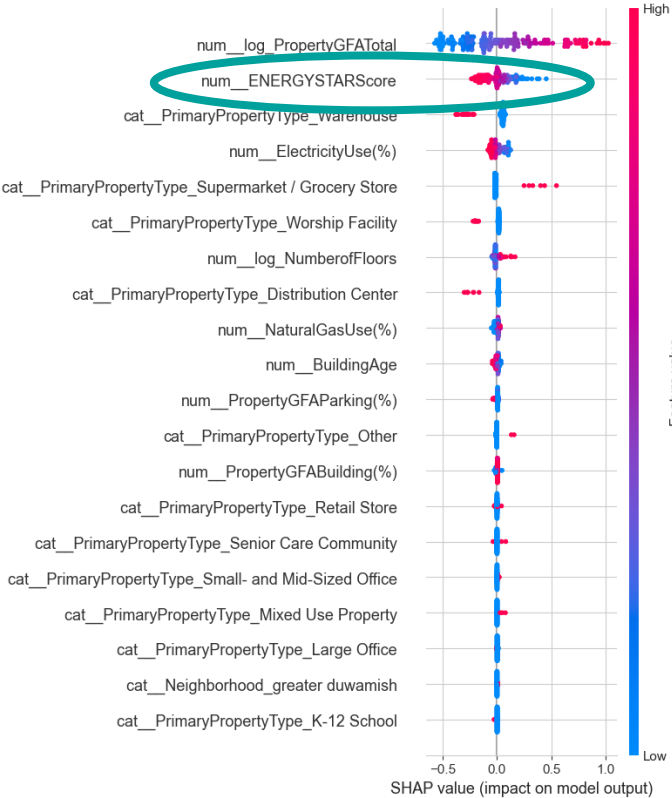
L'Energy Star Score a une contribution importante

SiteEnergyUse

Sans EnergyStar score

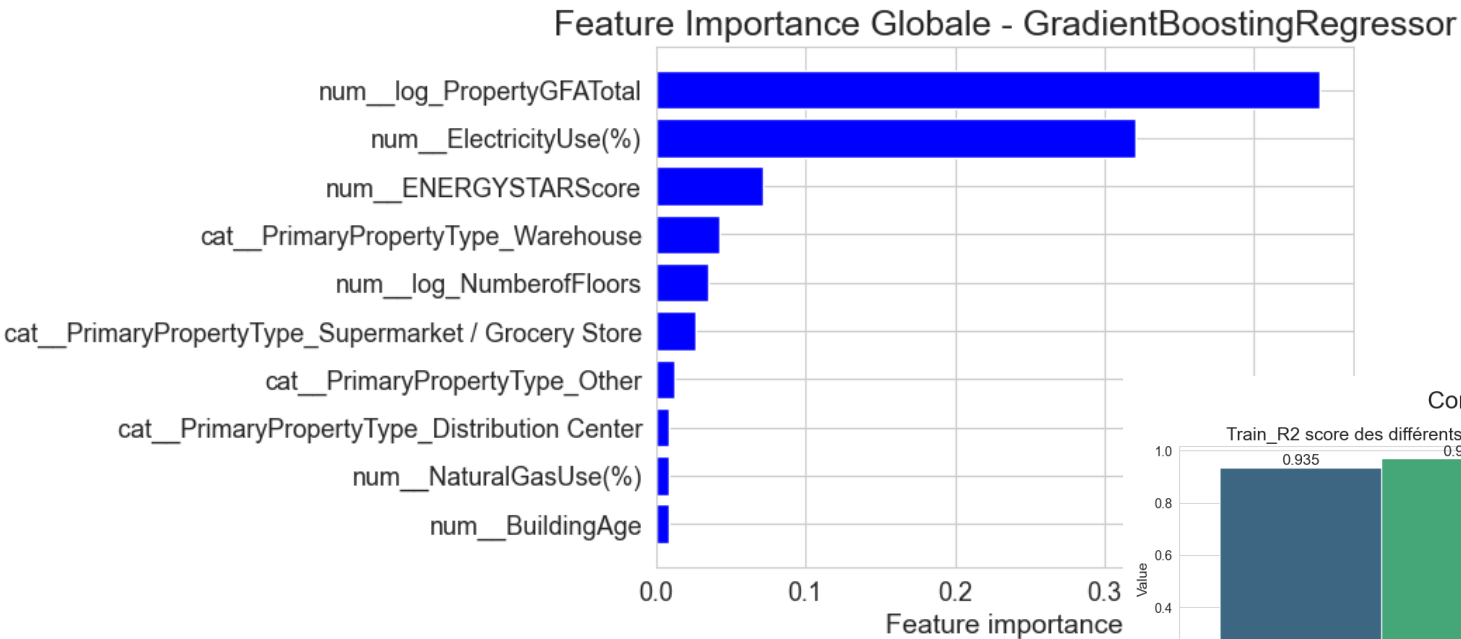


Avec EnergyStar score

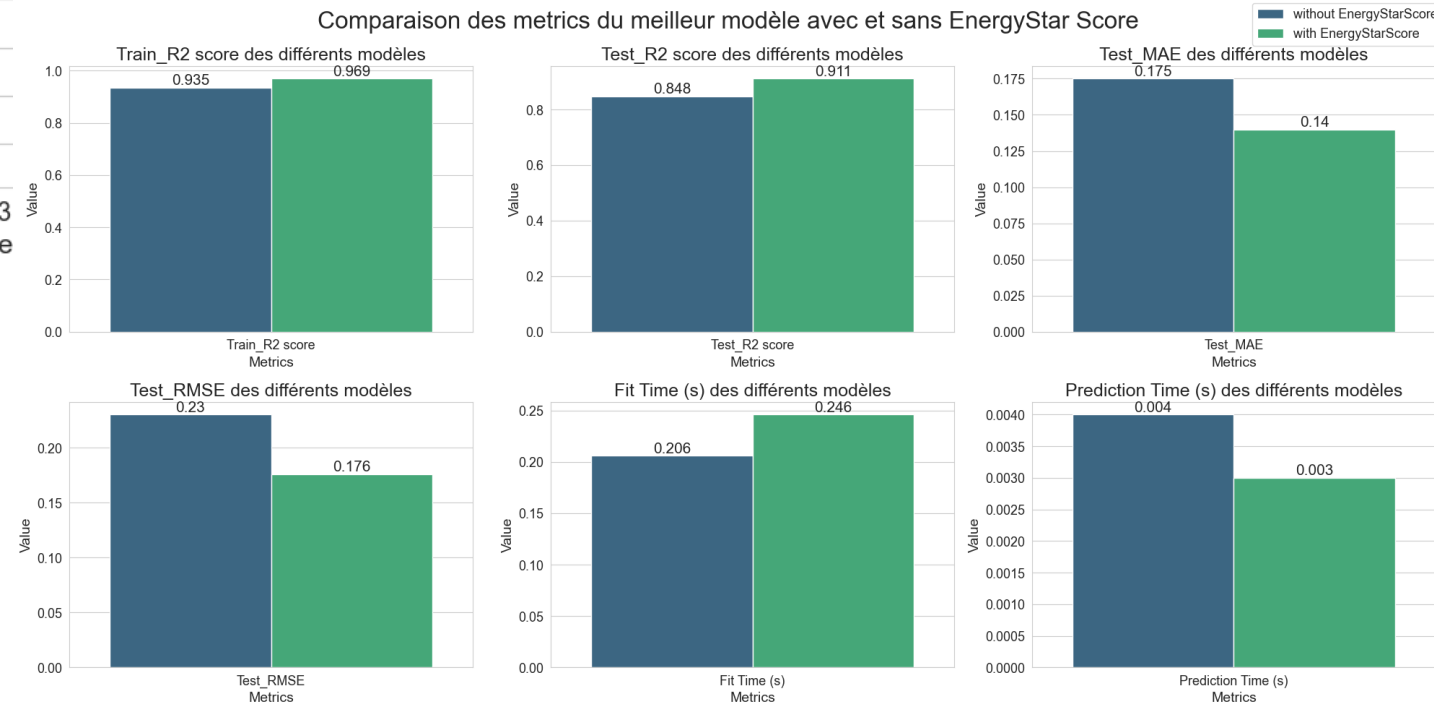


L'EnergyStar score est une variable influente pour la prédiction finale.

Total GHG Emissions – synthèse des résultats



L'EnergyStar score a une influence moindre pour la prédiction de TotalGHGEmissions.



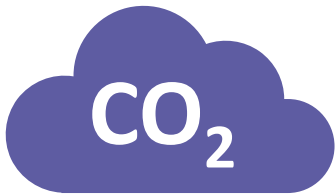


Conclusion & Perspectives

Conclusion

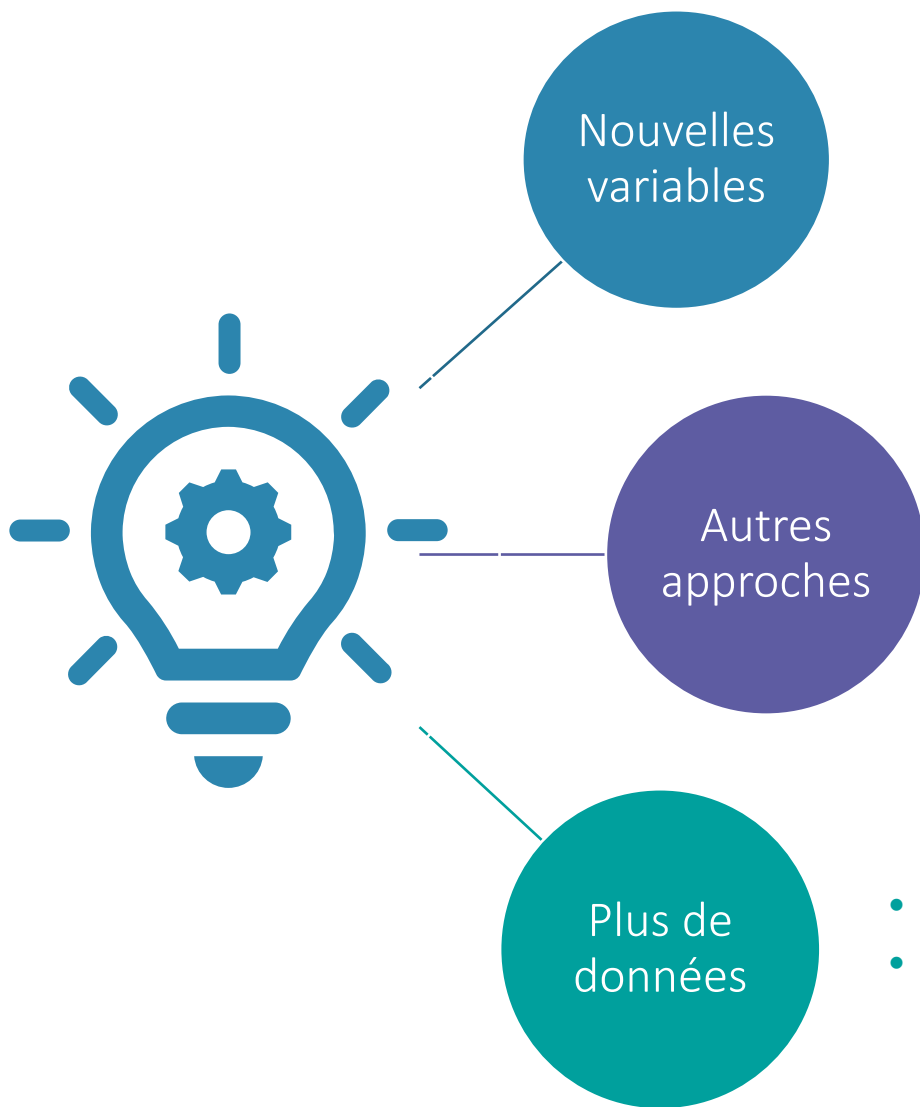


- La prédiction de **Site Energy Use (kBtu)** à partir des données structurelles des bâtiments est possible.
- Le meilleur modèle est le **Gradient Boosting Regressor**.
- L'**Energy Star Score** est une variable qui a une **forte contribution** sur la prédiction.



- La prédiction de **Total GHG Emissions** à partir des données structurelles des bâtiments est possible.
- Le meilleur modèle est le **Gradient Boosting Regressor**.
- L'**Energy Star Score** est **moins pertinent** pour la prédiction de **Total GHG Emissions**.

Perspectives



- **PropertyUseType** avec le % de surface totale
- **Latitude** et **longitude** au lieu de Neighborhood

- Estimer de l'**Energy Star Score**
- Prédire les **2 targets** en même temps
- **1 modèle** / type de bâtiment

- Vérifier les **hypothèses de mix énergétique** (données 2015)
- **Augmenter le nombre de données** des catégories les moins représentées



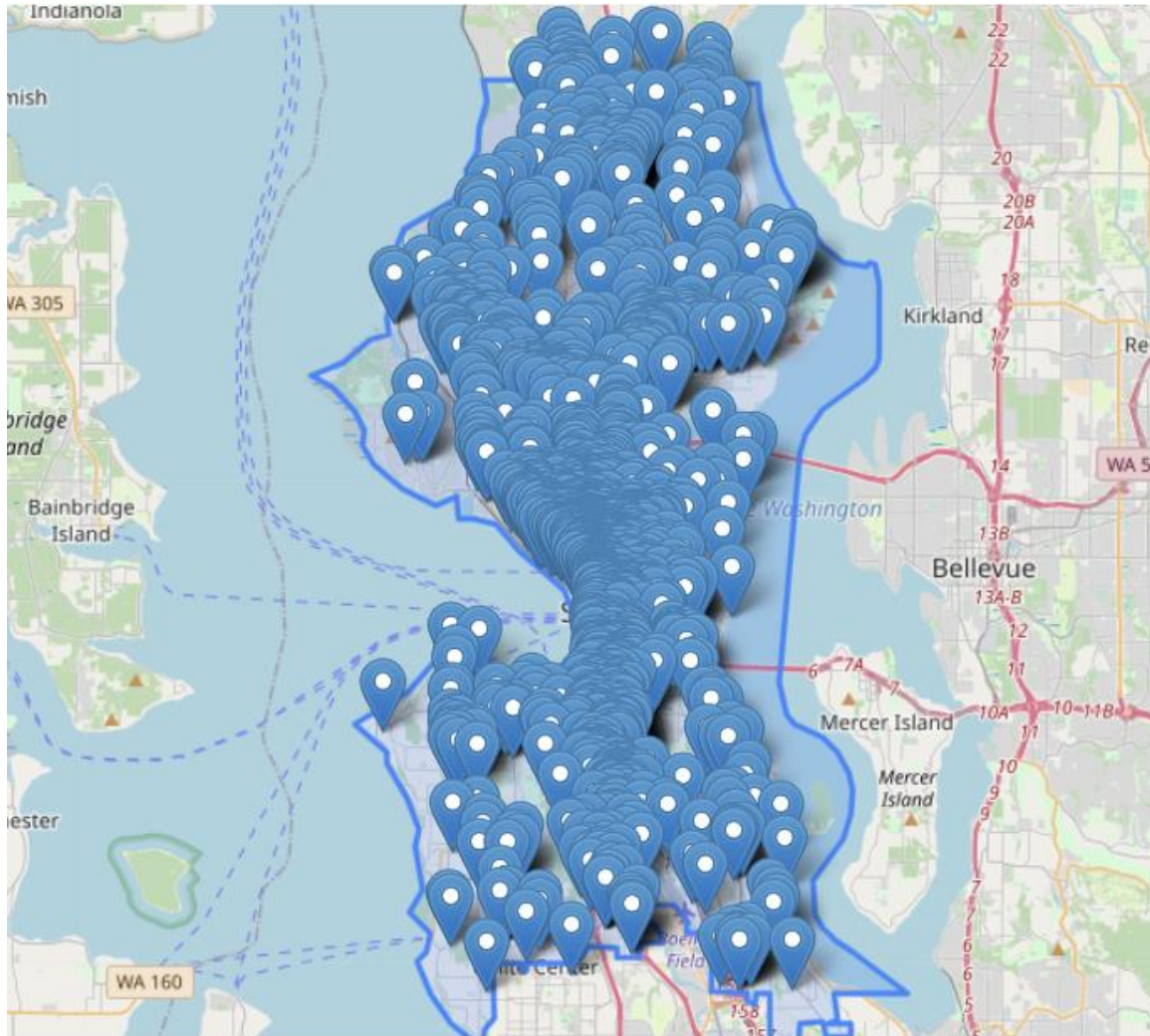
Merci de votre attention



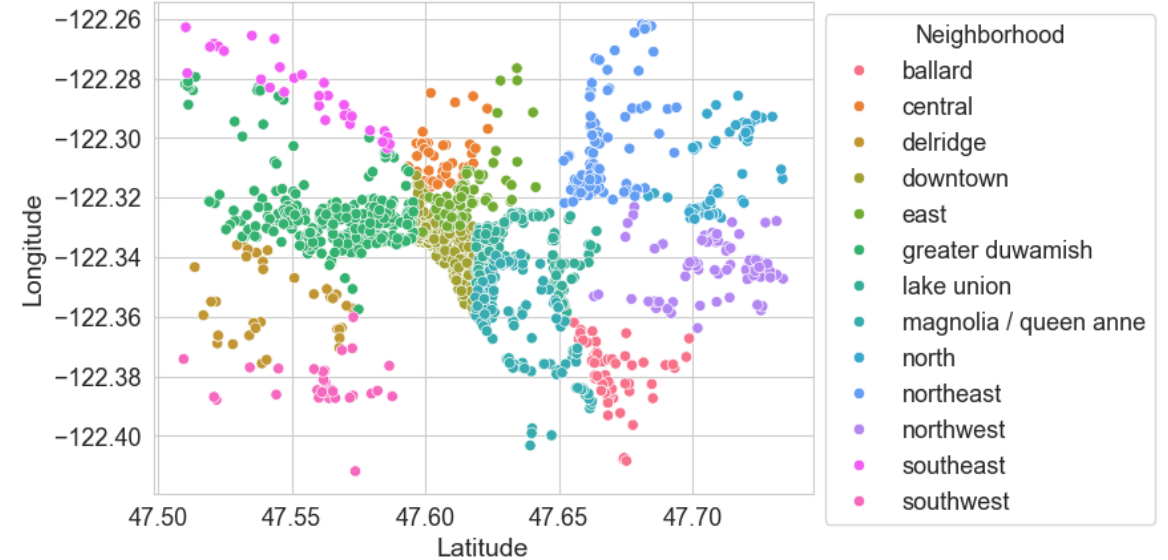
Back-up slides



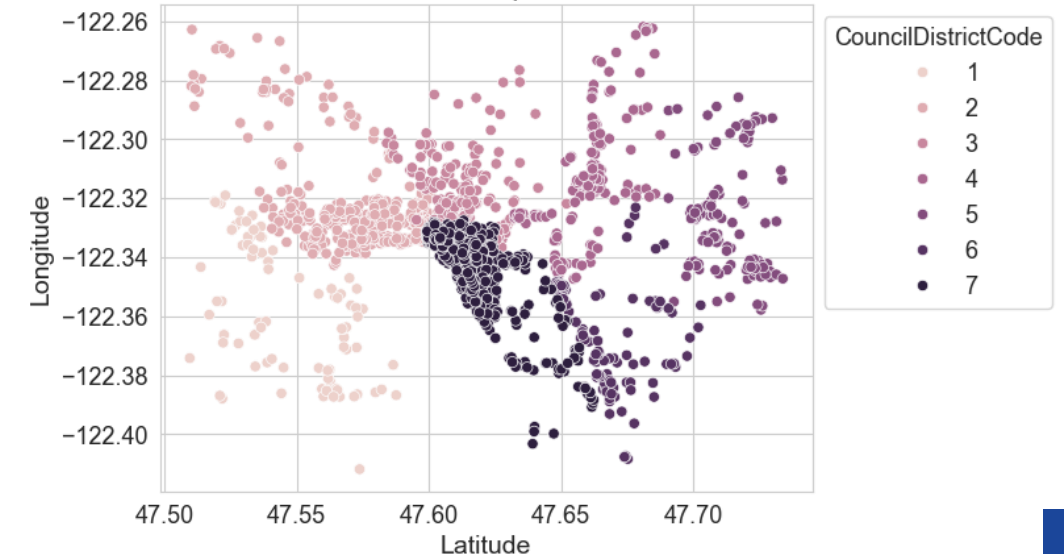
Répartition géographique des bâtiments non résidentiels



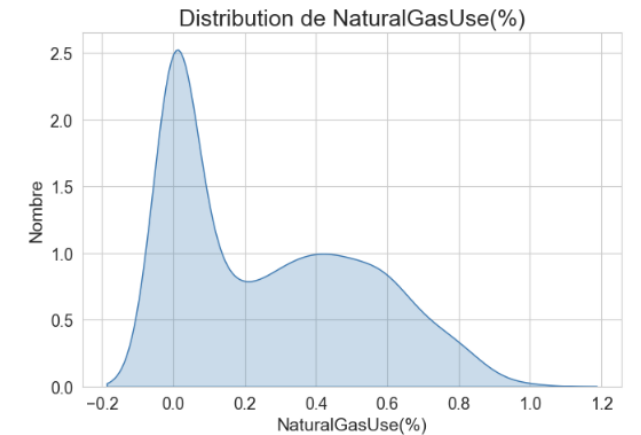
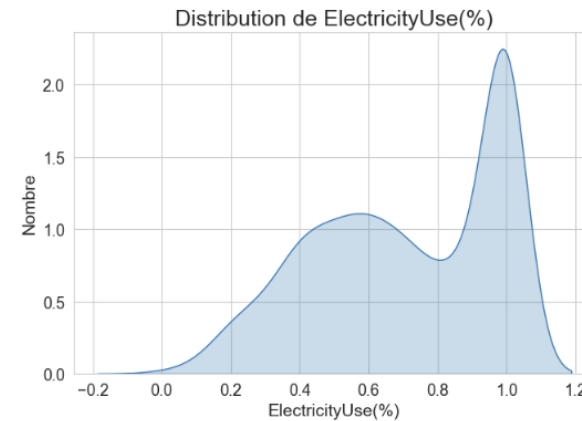
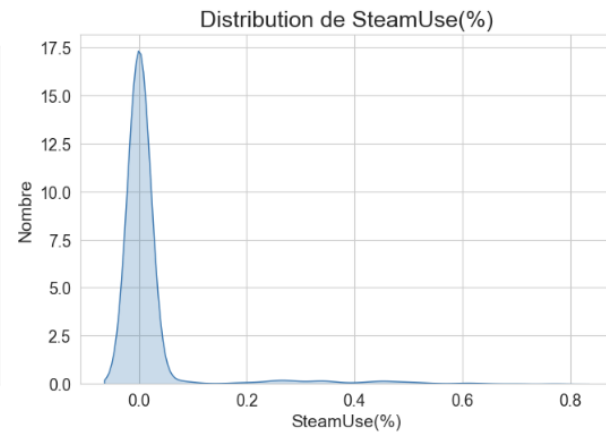
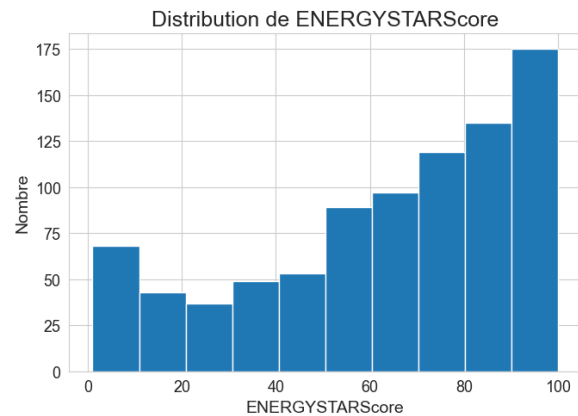
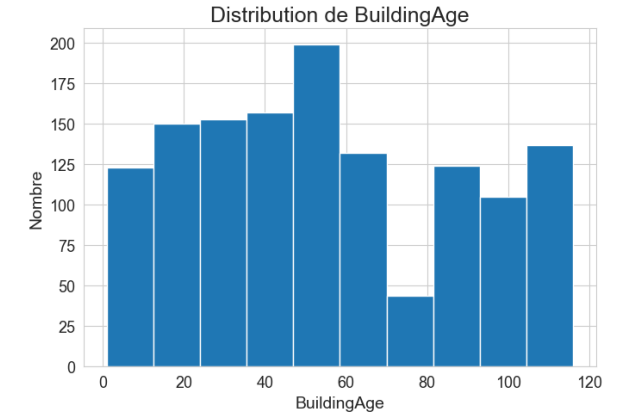
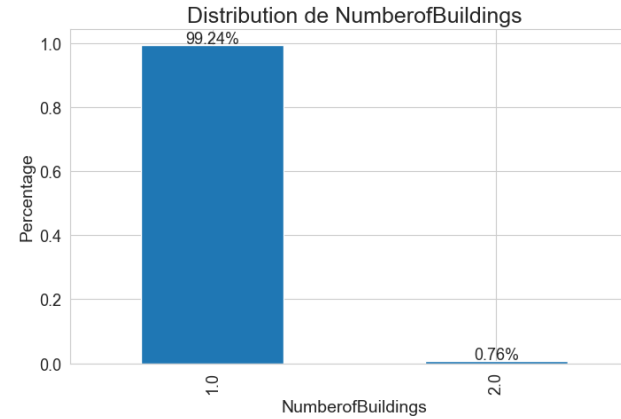
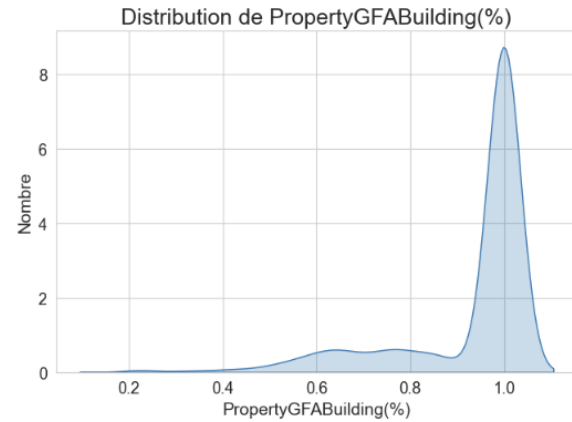
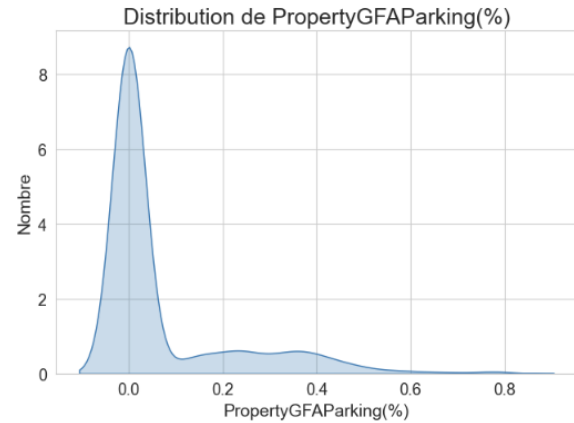
Localisation des bâtiments par Neighborhood



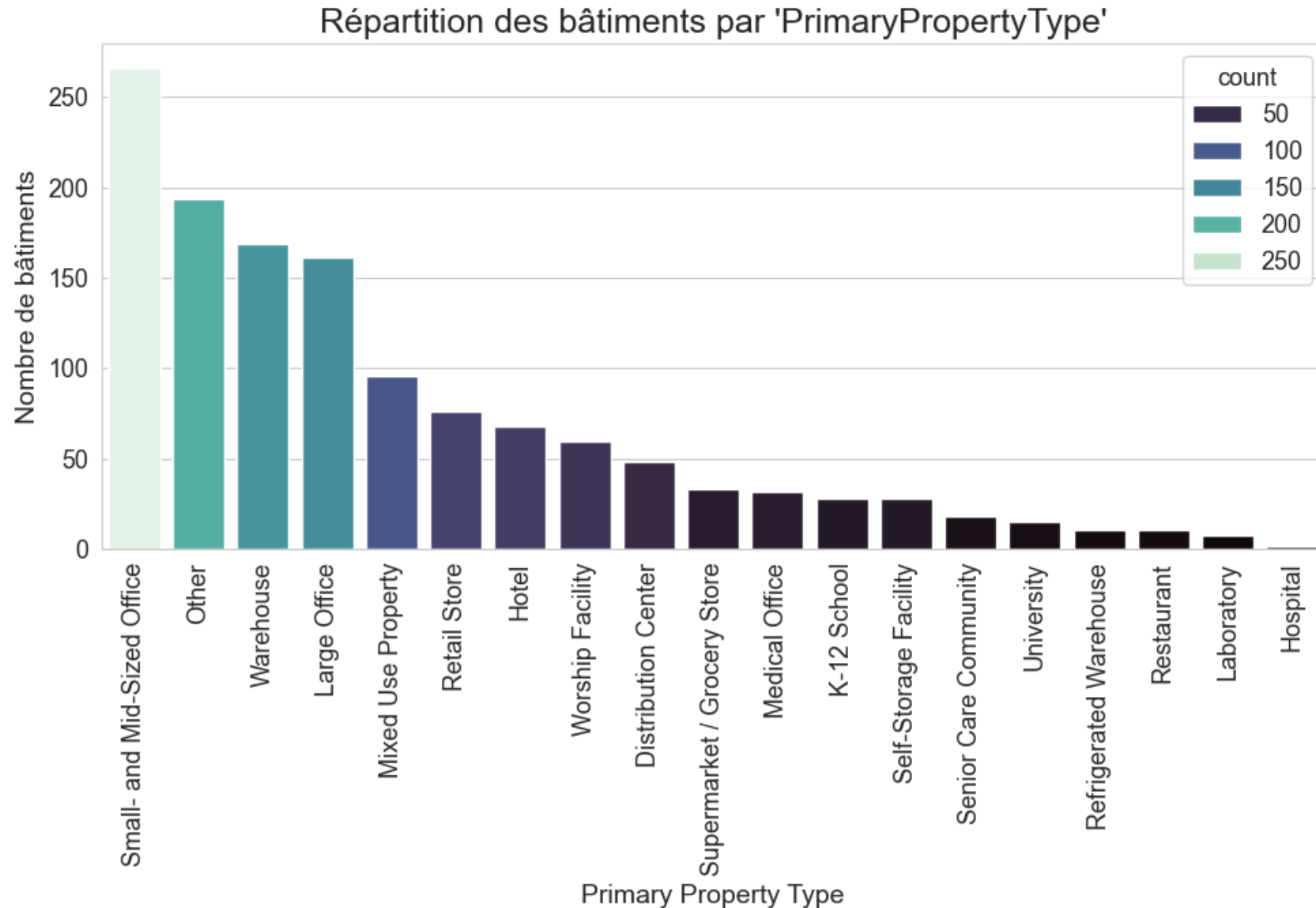
Localisation des batiments par CouncilDistrictCode



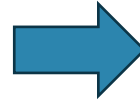
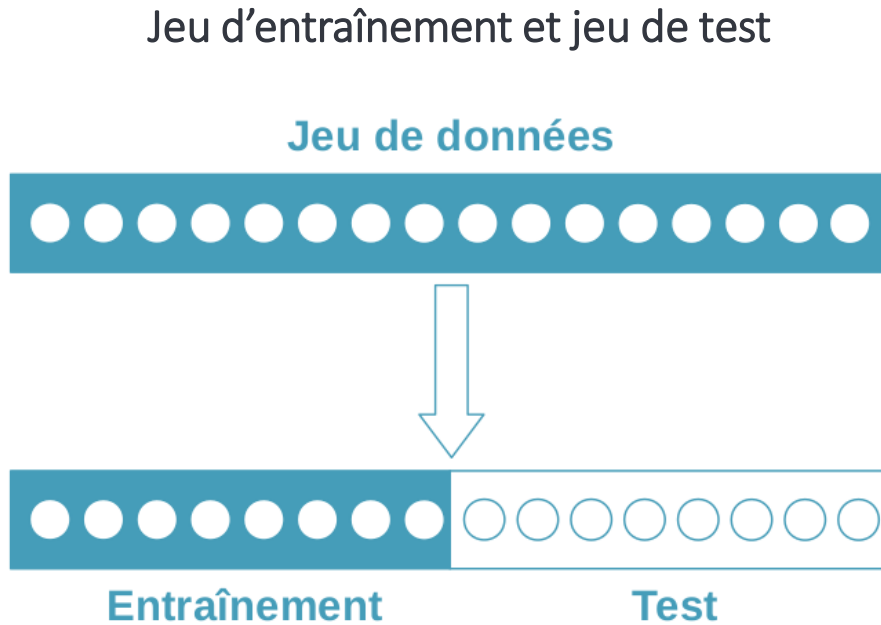
Distribution des autres variables



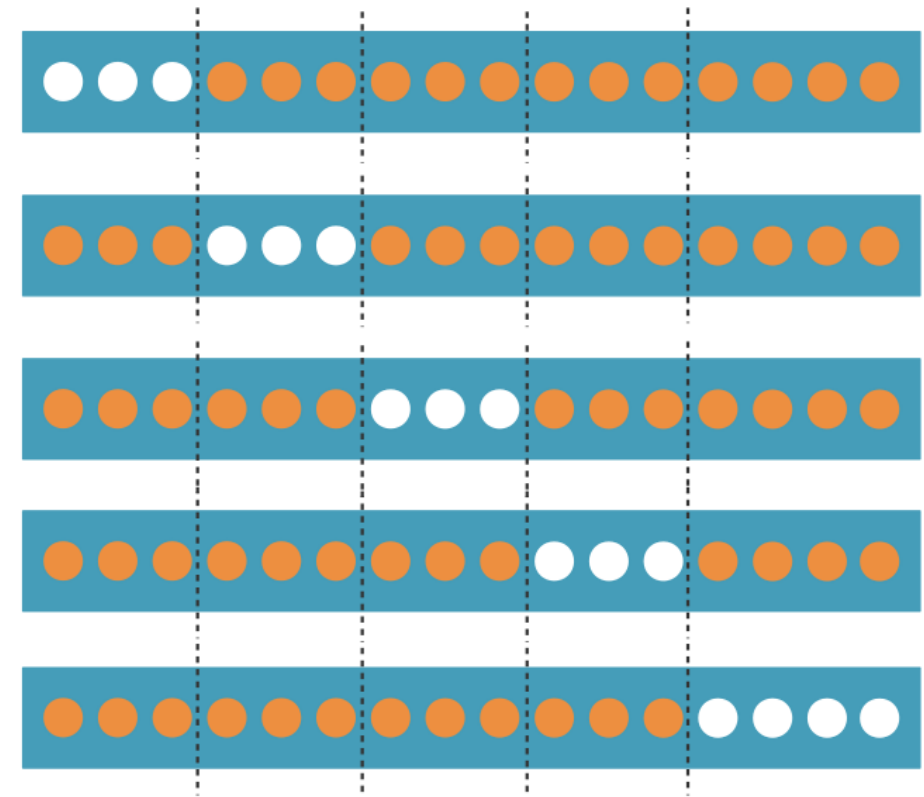
Majorité de bureaux dans les bâtiments non résidentiels



Validation croisée

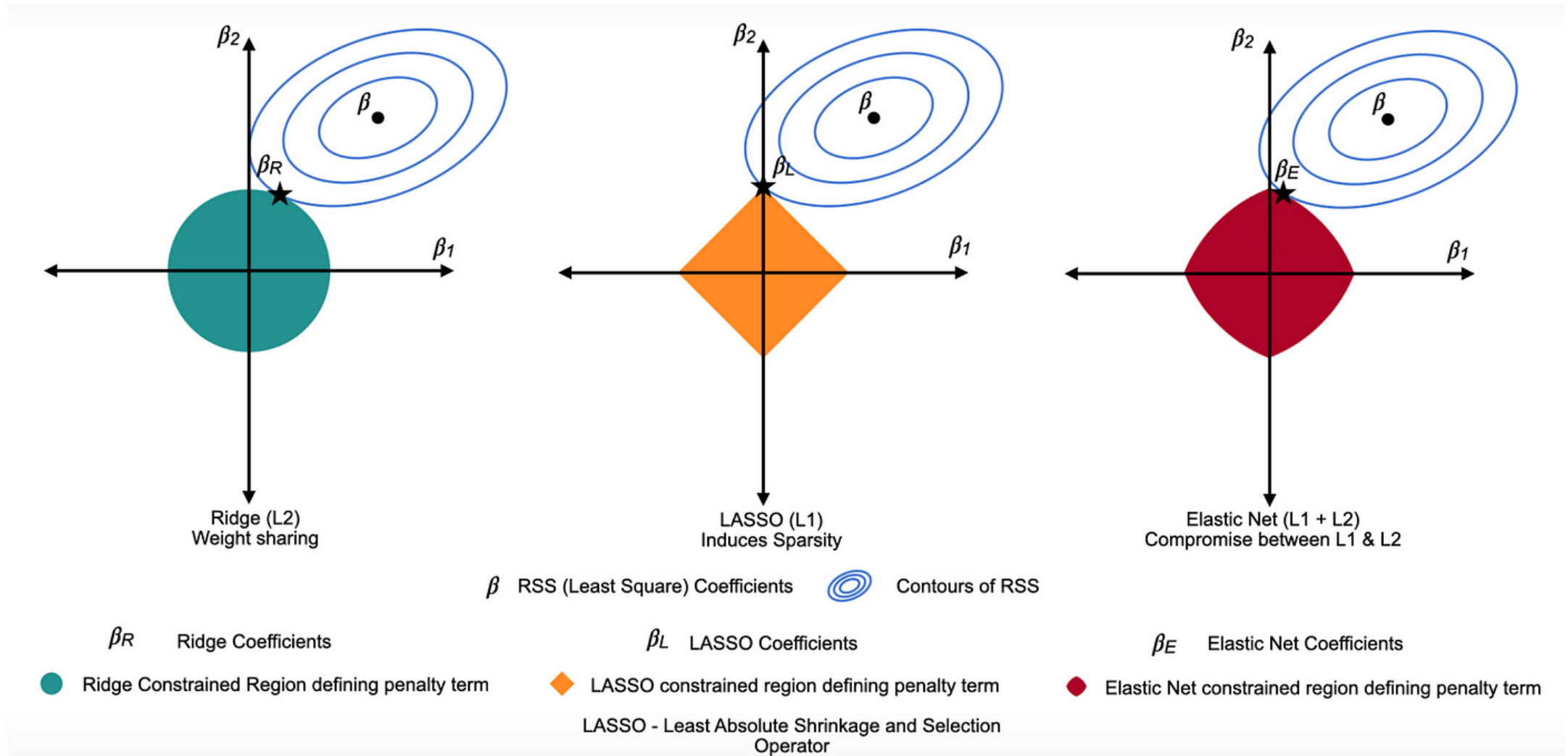


Validation croisée sur le jeu d'entraînement



- Entraînement
- Validation

Méthodes linéaires avec régularisation



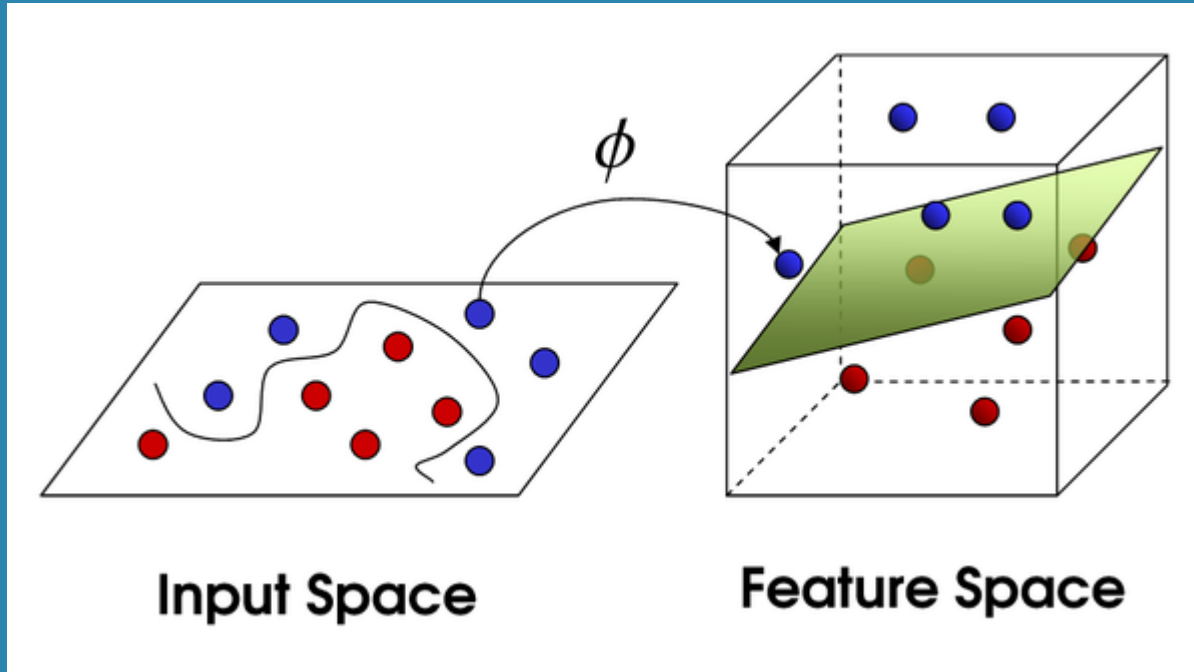
Eviter le surapprentissage en restreignant l'amplitude des poids

Obtenir un modèle parcimonieux, sélection de variables et de réduction de dimension supervisée

Combine les régularisations Ridge et Lasso

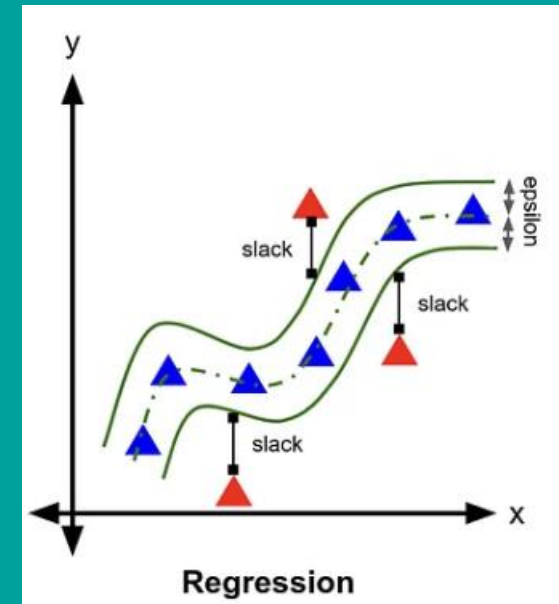
Support Vector Machine Regression (SVR)

Support Vector Machine



Trouve l'hyperplan qui sépare au mieux les données en classes distinctes, en maximisant la marge entre les points les plus proches de chaque classe

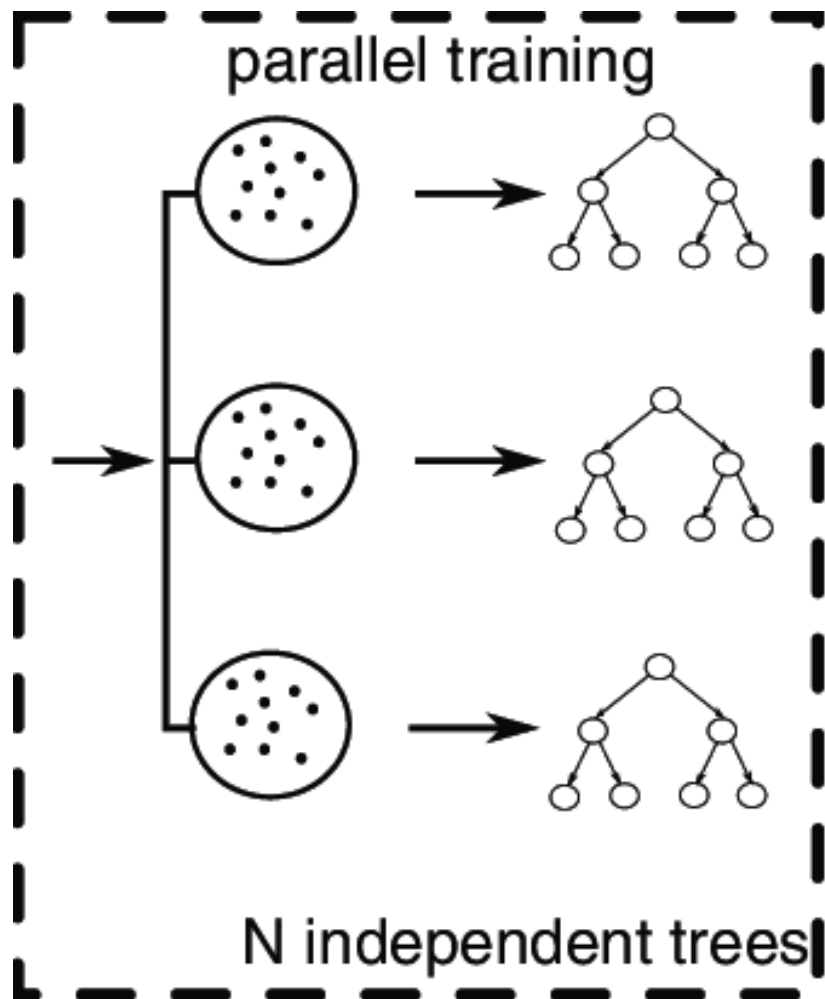
Support Vector Machine Regression



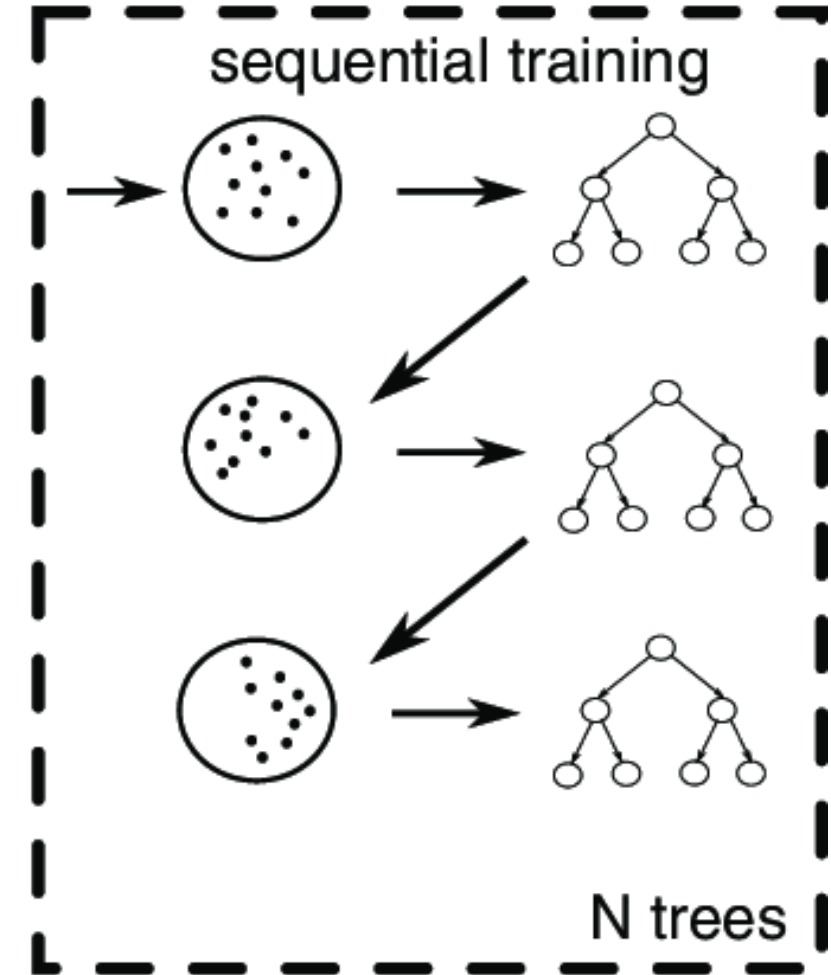
Utilise une marge d'erreur et des vecteurs de support pour ajuster une fonction qui minimise la différence entre les prédictions et les valeurs réelles

Méthodes Ensemblistes

Random Forest



Gradient Boosting



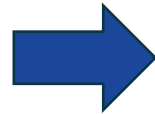
TotalGHGEmissions : Sélection du modèle (1)

Scalers testés :

- StandardScaler
- RobustScaler

Encoders testés :

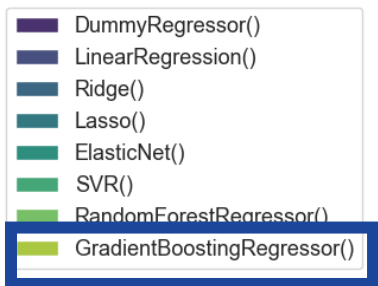
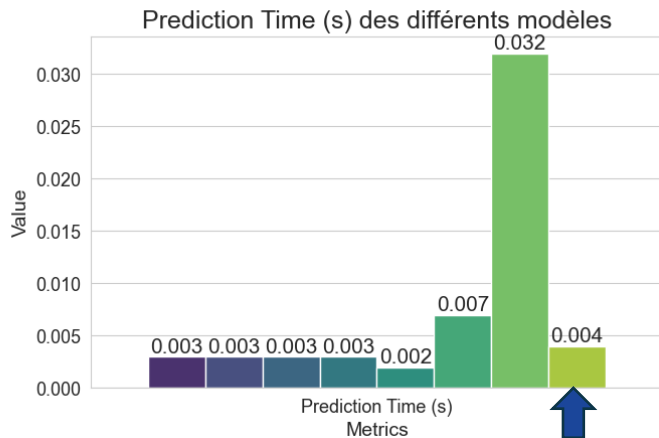
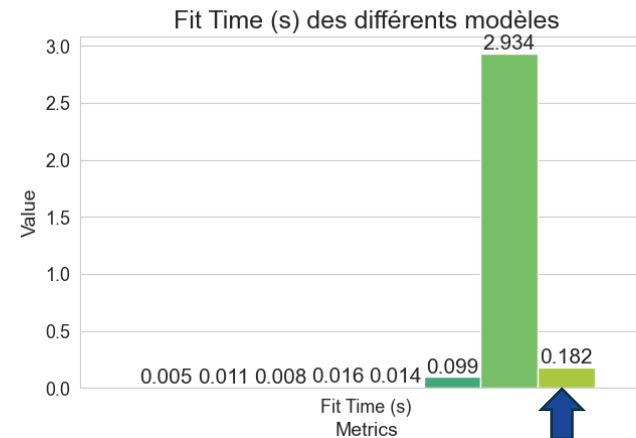
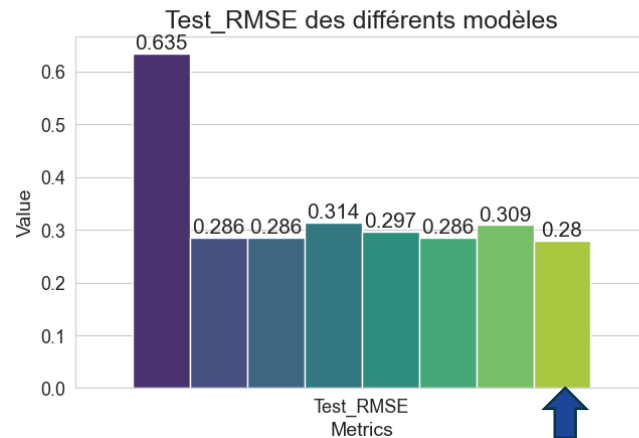
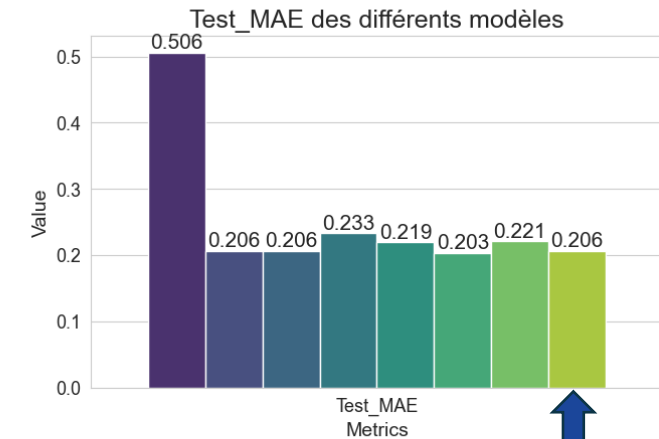
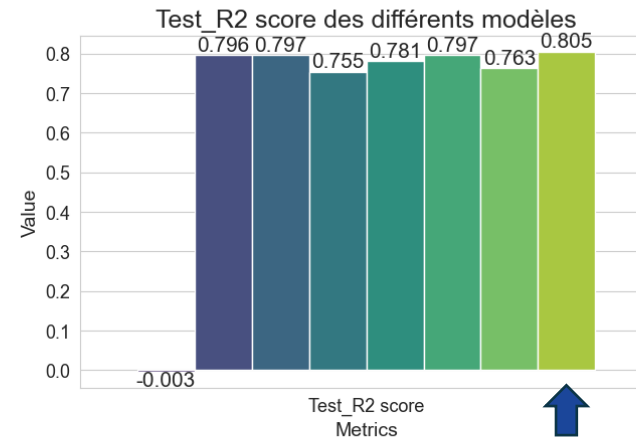
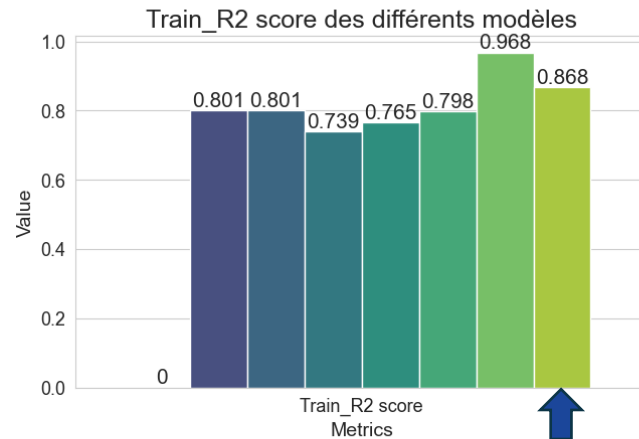
- TargetEncoder
- OneHotEncoder



	Value
Encoder	OneHotEncoder()
Scaler	StandardScaler()
Model	GradientBoostingRegressor()
Parameters	{'gradientboostingregressor__n_estimators': 100}
CV_R2 score	0.781 (+/- 0.021)
Train_R2 score	0.868
Test_R2 score	0.805
Test_MAE	0.206
Test_RMSE	0.28
Fit Time (s)	0.182
Prediction Time (s)	0.004

TotalGHGEmissions : Sélection du modèle (2)

Comparaison des performances des différents modèles pour la meilleure combinaison Encoder : OneHotEncoder(), Scaler : StandardScaler()



TotalGHGEmissions : Optimisation du modèle

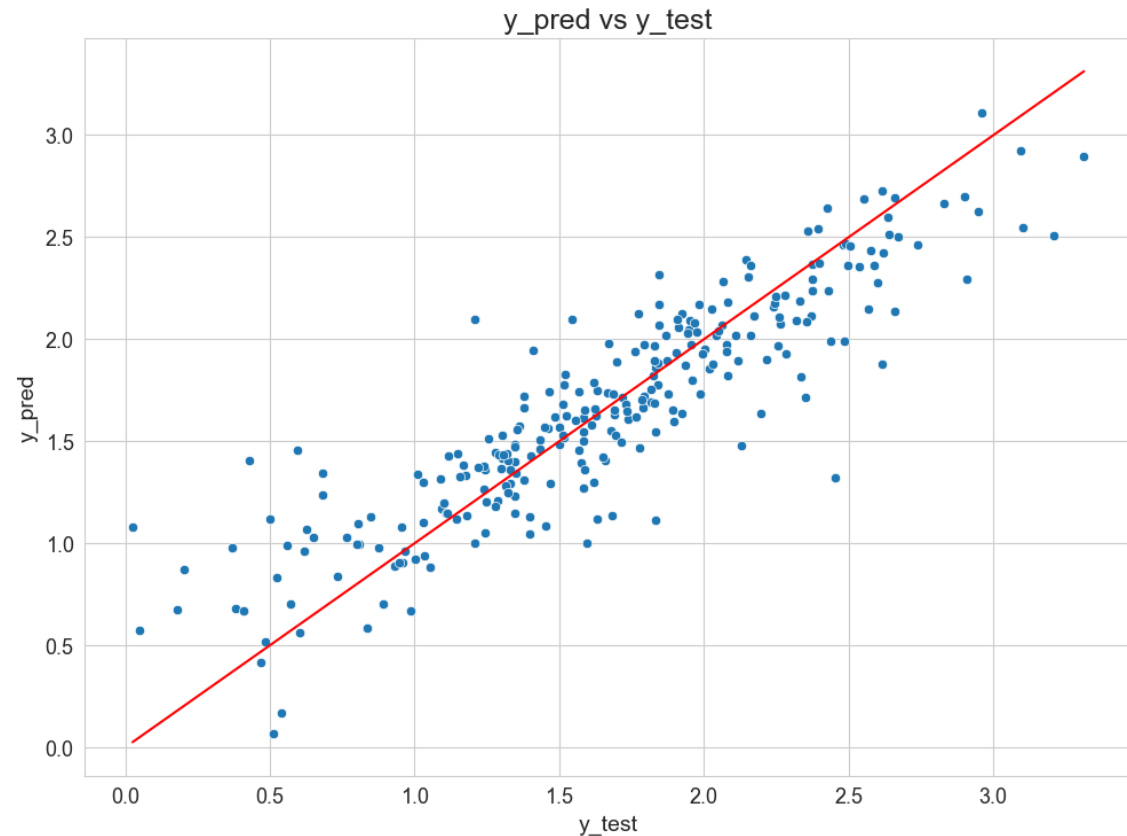
Méthodes :

- GridSearchCV : systématique
- RandomizedSearchCV : aléatoire

Hyperparamètres :

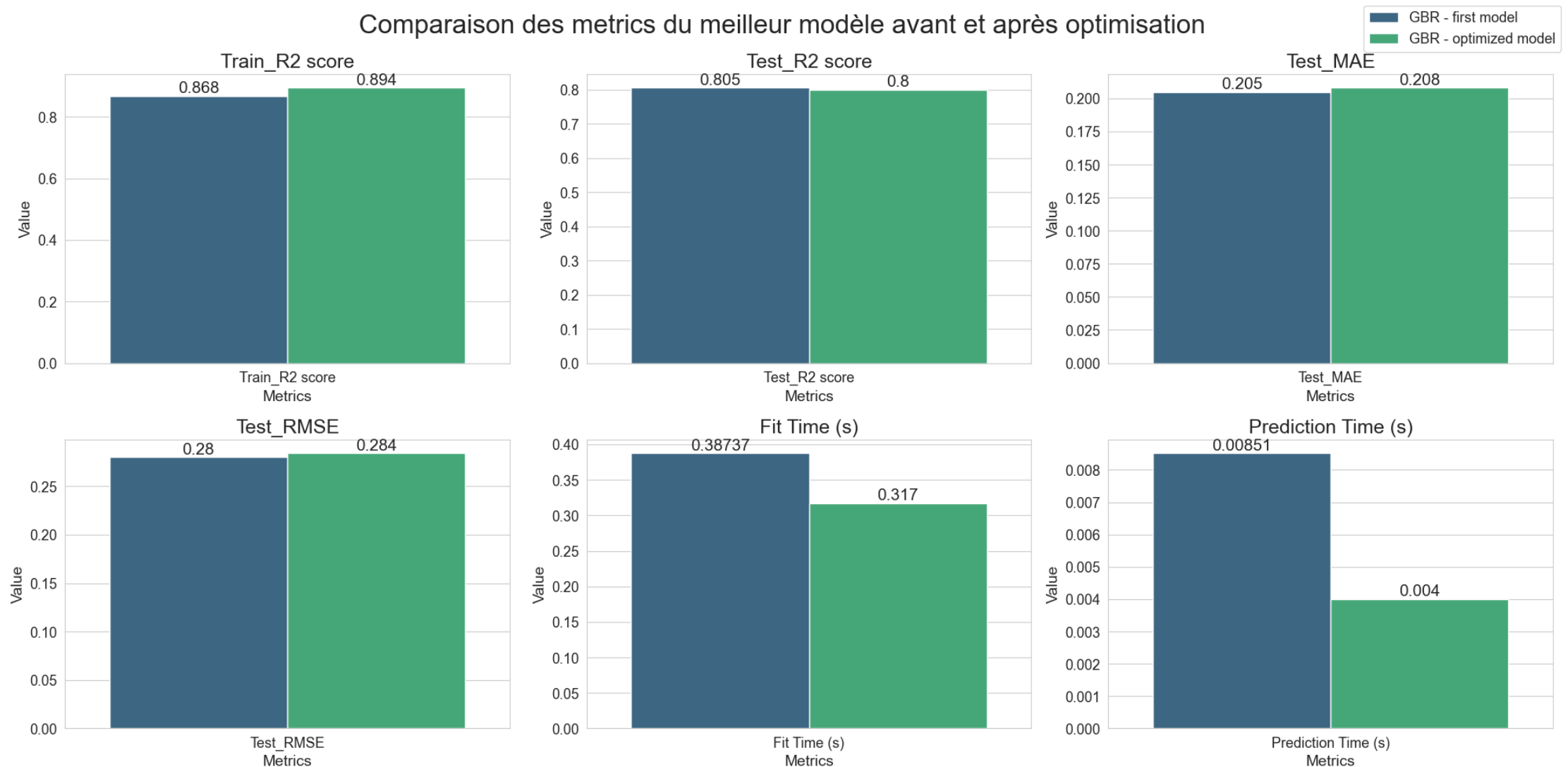
- n_estimators
- learning_rate
- max_depth
- min_samples_split

Résultat du modèle après optimisation



TotalGHGEmissions : Optimisation du modèle

Comparaison des metrics du meilleur modèle avant et après optimisation



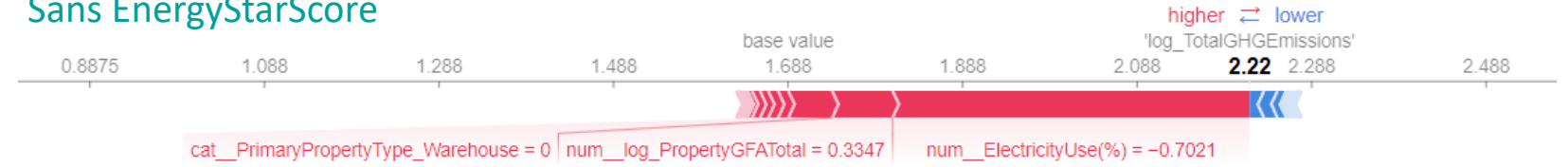
Pas d'améliorations des performances du modèle après optimisation.

Analyse d'un élément de la catégorie la plus fréquente 'Small- and Mid-Sized Office'

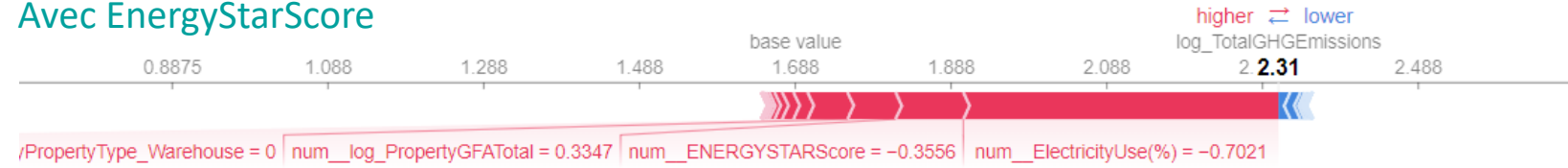
OSEBuildingID	575
PrimaryPropertyType	Small- and Mid-Sized Office
Neighborhood	northwest
NumberofBuildings	1.0
NumberofFloors	5
PropertyGFATotal	86400
ENERGYSTARScore	53.0
SiteEnergyUse(kBtu)	9080022.0
TotalGHGEmissions	293.55
BuildingAge	42
PropertyGFAParking(%)	0.0
PropertyGFABuilding(%)	1.0
SteamUse(%)	0.0
ElectricityUse(%)	0.45
NaturalGasUse(%)	0.55
log_NumberofFloors	0.69897
log_PropertyGFATotal	4.936514
log_TotalGHGEmissions	2.467682

TotalGHGEmissions

Sans EnergyStarScore

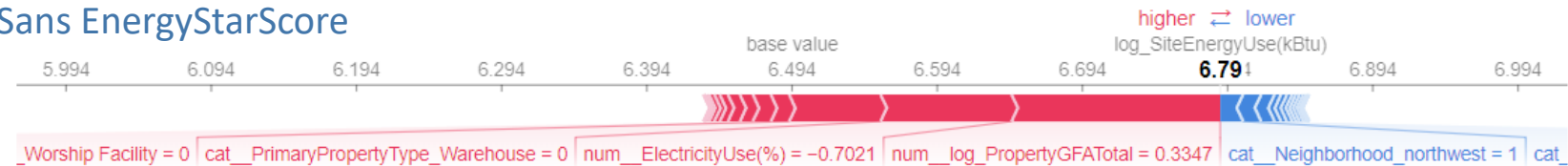


Avec EnergyStarScore

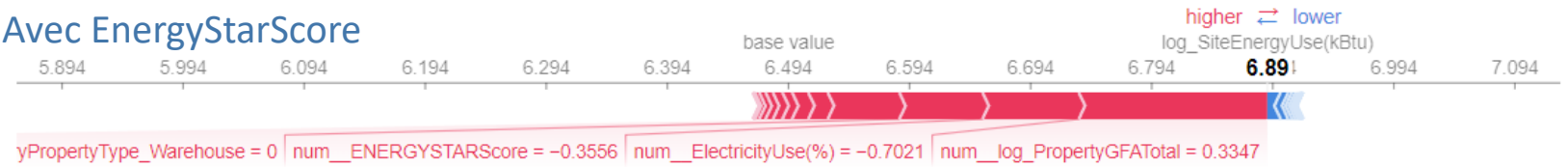


SiteEnergyUse (kBtu)

Sans EnergyStarScore



Avec EnergyStarScore

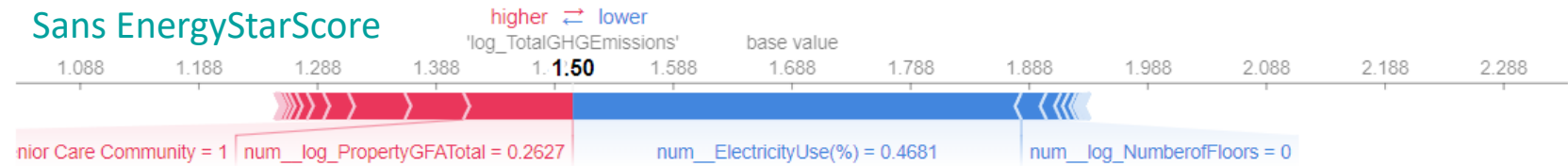


Analyse d'un élément d'une catégorie peu fréquente 'Senior Care Community'

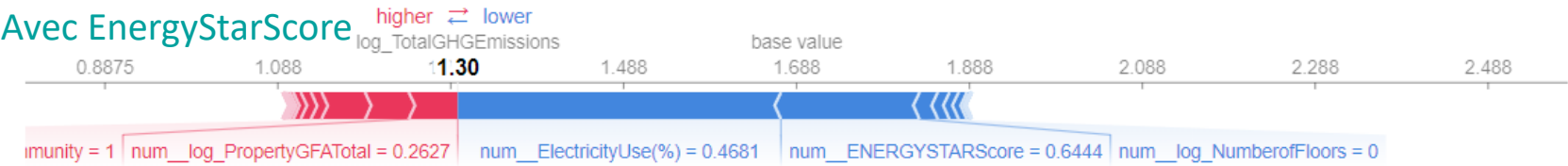
OSEBuildingID	20377
PrimaryPropertyType	Senior Care Community
Neighborhood	southeast
NumberOfBuildings	1.0
NumberOfFloors	3
PropertyGFATotal	78374
ENERGYSTARScore	98.0
SiteEnergyUse(kBtu)	3321035.75
TotalGHGEmissions	23.15
BuildingAge	35
PropertyGFAParking(%)	0.0
PropertyGFABuilding(%)	1.0
SteamUse(%)	0.0
ElectricityUse(%)	1.0
NaturalGasUse(%)	0.0
log_NumberofFloors	0.477121
log_PropertyGFATotal	4.894172
log_TotalGHGEmissions	1.364551

TotalGHGEmissions

Sans EnergyStarScore

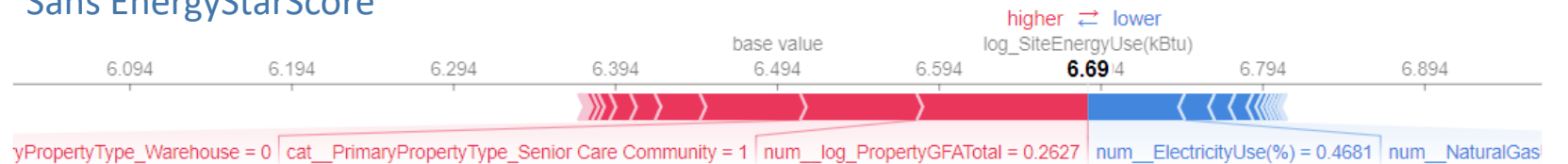


Avec EnergyStarScore

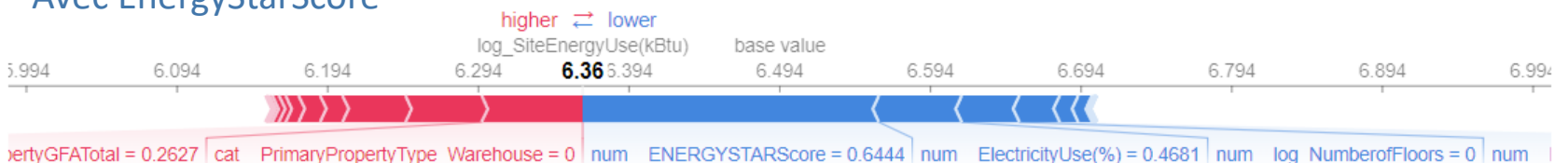


SiteEnergyUse (kBtu)

Sans EnergyStarScore



Avec EnergyStarScore

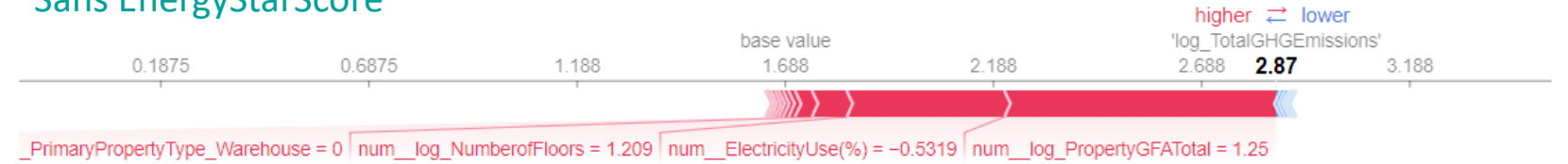


Analyse d'un outlier

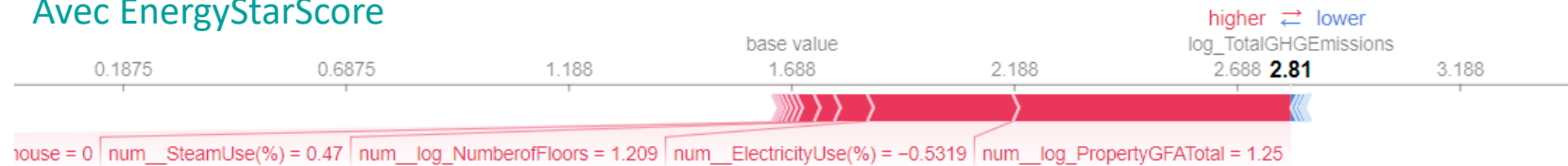
OSEBuildingID	387
PrimaryPropertyType	Large Office
Neighborhood	downtown
NumberOfBuildings	1.0
NumberOfFloors	21
PropertyGFATotal	298426
ENERGYSTARScore	79.0
SiteEnergyUse(kBtu)	27076922.0
TotalGHGEmissions	1084.79
BuildingAge	56
PropertyGFAParking(%)	0.0
PropertyGFABuilding(%)	1.0
SteamUse(%)	0.47
ElectricityUse(%)	0.53
NaturalGasUse(%)	0.0
log_NumberofFloors	1.322219
log_PropertyGFATotal	5.474837
log_TotalGHGEmissions	3.035346

OSEBuildingID	640
PrimaryPropertyType	Large Office
Neighborhood	greater duwamish
NumberOfBuildings	1.0
NumberOfFloors	6
PropertyGFATotal	1380959
ENERGYSTARScore	79.0
SiteEnergyUse(kBtu)	74130576.0
TotalGHGEmissions	810.7
BuildingAge	104
PropertyGFAParking(%)	0.0
PropertyGFABuilding(%)	1.0
SteamUse(%)	0.0
ElectricityUse(%)	0.91
NaturalGasUse(%)	0.09
log_NumberofFloors	0.778151
log_PropertyGFATotal	6.140181
log_SiteEnergyUse(kBtu)	7.869997

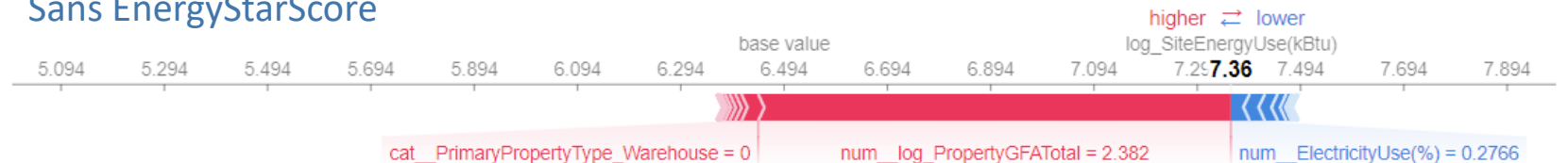
TotalGHGEmissions Sans EnergyStarScore



Avec EnergyStarScore



SiteEnergyUse (kBtu) Sans EnergyStarScore



Avec EnergyStarScore

