Anticipez les besoins en consommation de bâtiments

Objectifs

Seattle ville neutre en 2050, pour cela il faut anticiper la consommation d'énergie totale et les émissions en CO2 des bâtiments non destinés à l'habitation

Evaluer l'intérêt de l'ENERGY STAR Score sur la prédiction

Jeu de données

- Le jeu de données présente 3376 lignes et 46 colonnes

Soit 3376 bâtiments et 46 variables qui contiennent des informations sur les bâtiments:

- Variables d’identification et de localisation
- Variables de dimensions et d’utilisation
- Variables de consommations et d’émissions

Nettoyage des données

Sélection des variables d'intérêts

- Les variables d'identification et localisation
- Les types d'utilisation
- Surface totale et pour différente utilisation
- Les variables cibles, la consommation totale et les émissions de CO2

Filtrage des types de bâtiments

- Suppressions des bâtiments destinées à l'habitation tels que les logements.

Traitement des valeurs manquantes présentes dans les targets

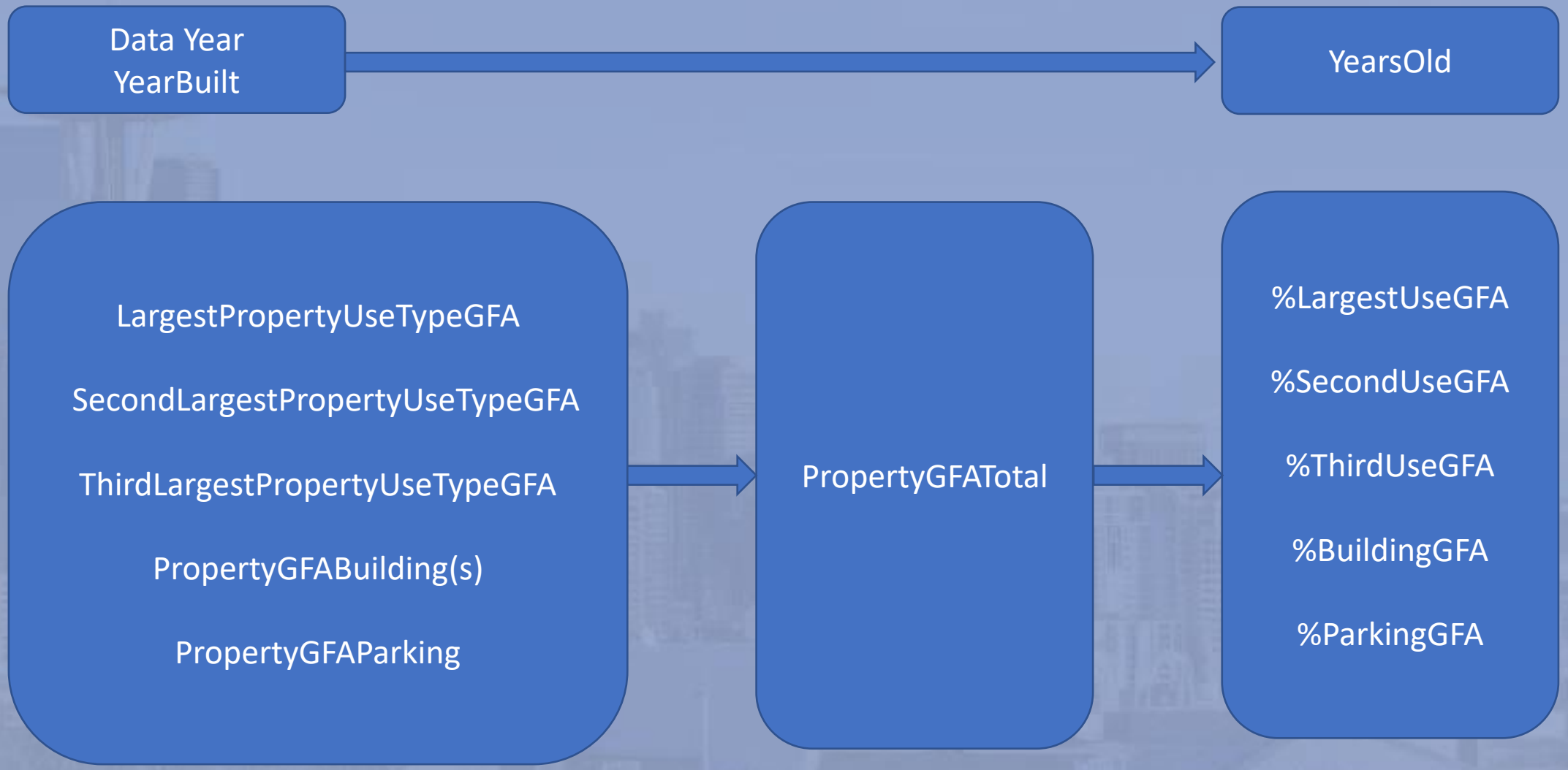
- Suppressions des valeurs manquantes pour les targets et imputation par une variable 'NotOtherUse' pour les variables catégorielles

3376 lignes
46 Colonnes

1618 lignes
32 colonnes

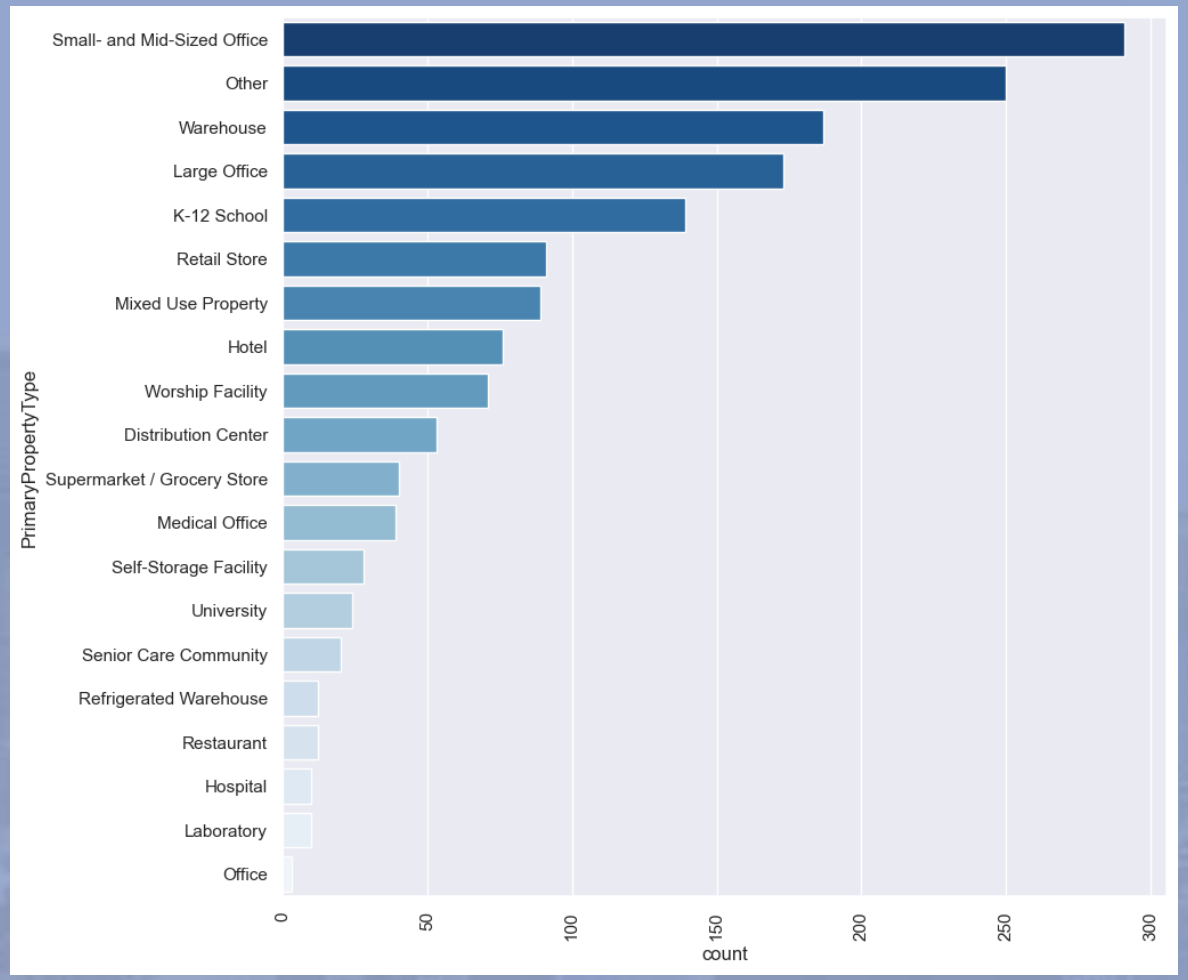
Seattle

Création de variables



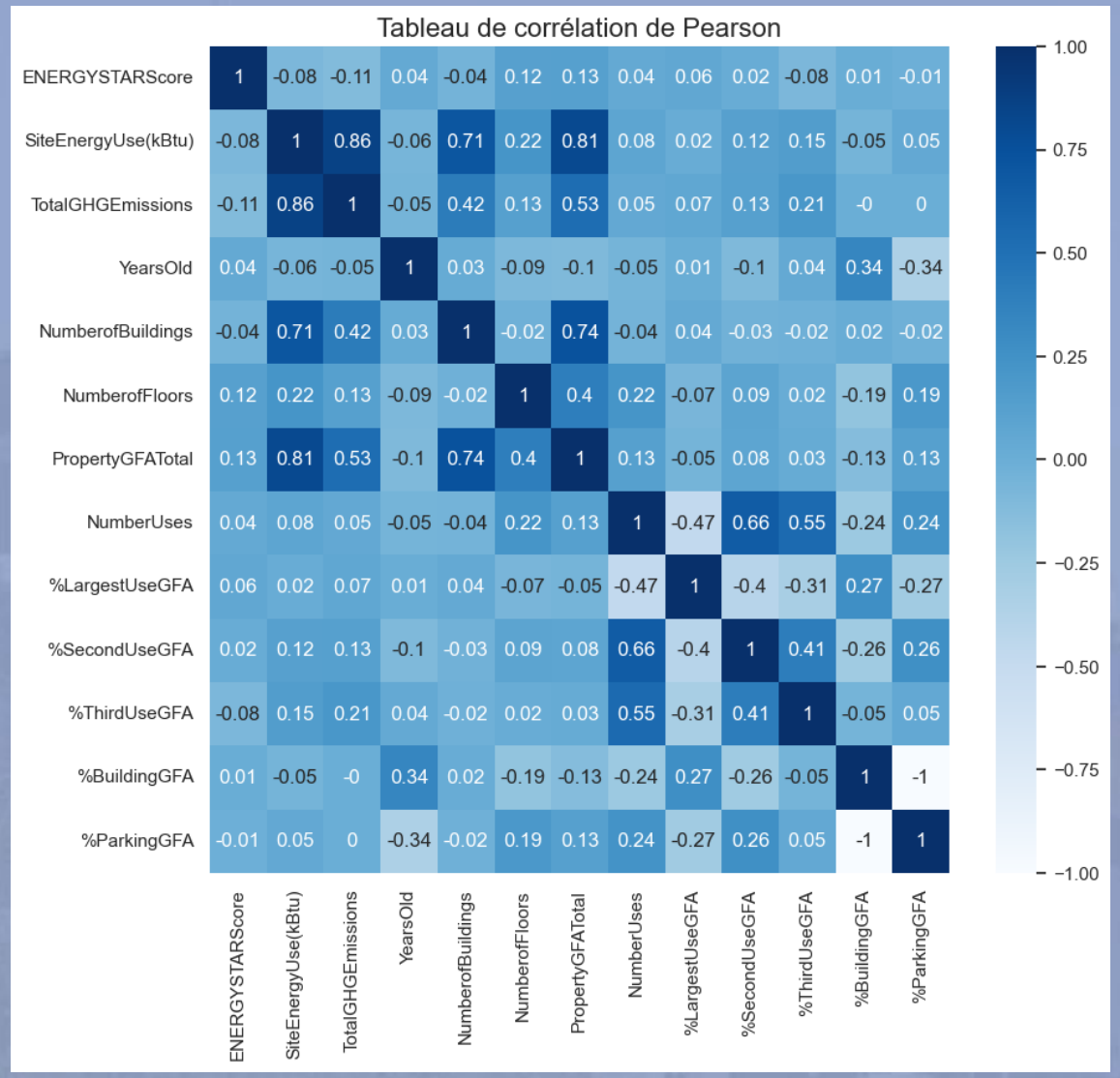
Analyse

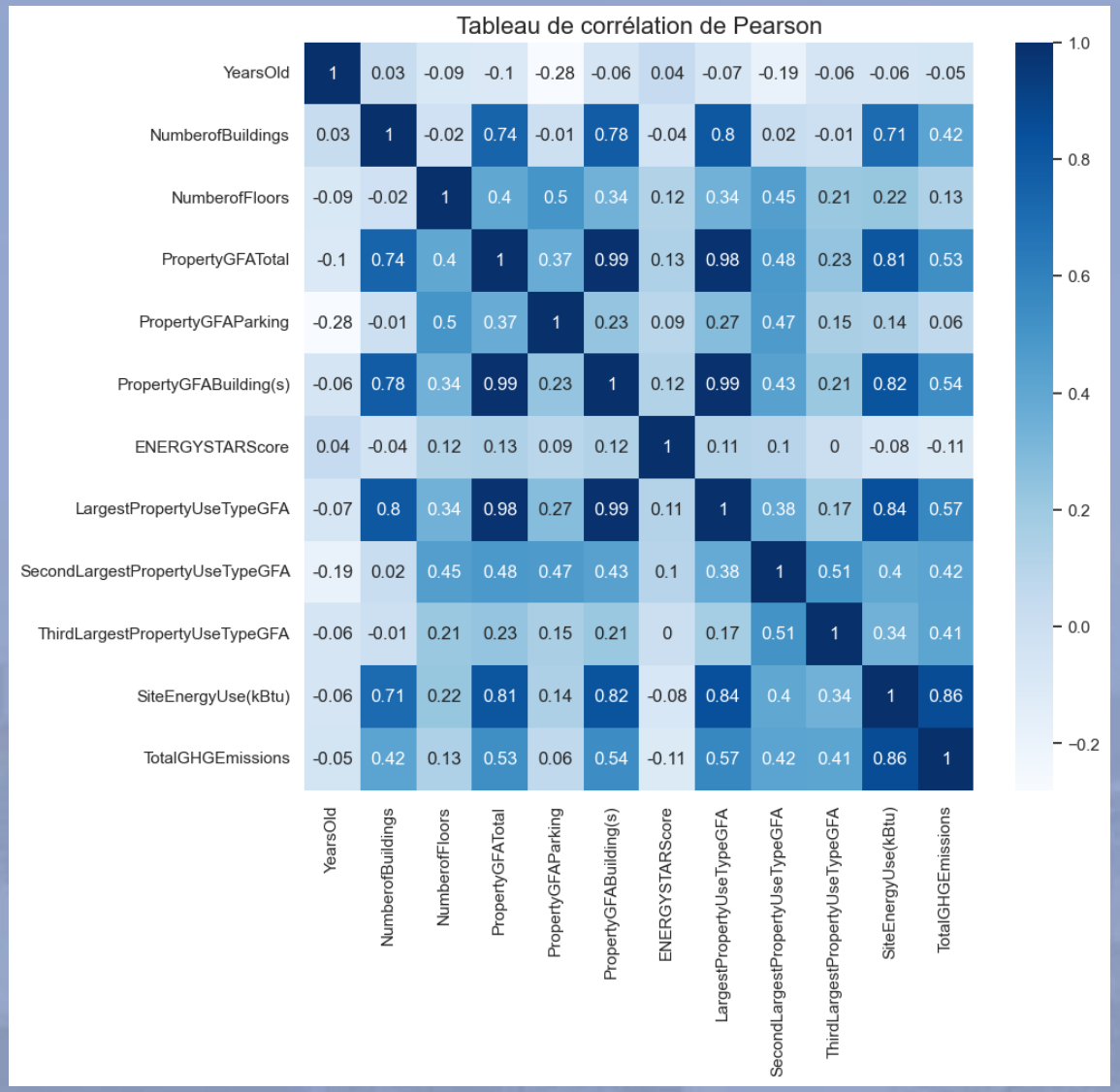
Répartition de la catégorie ‘PrimaryPropertyType’



Seattle

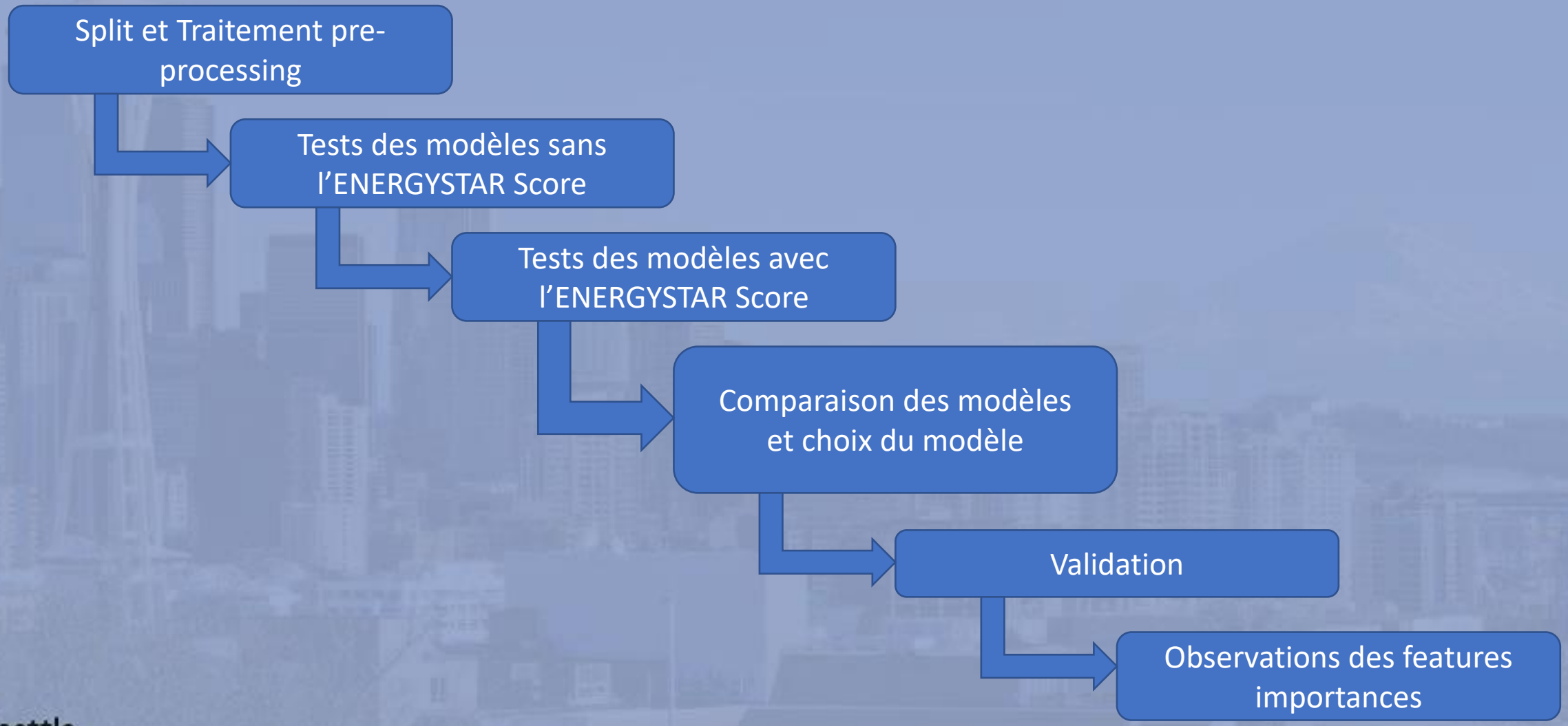
• Corrélation de Pearson entre les variables quantitatives





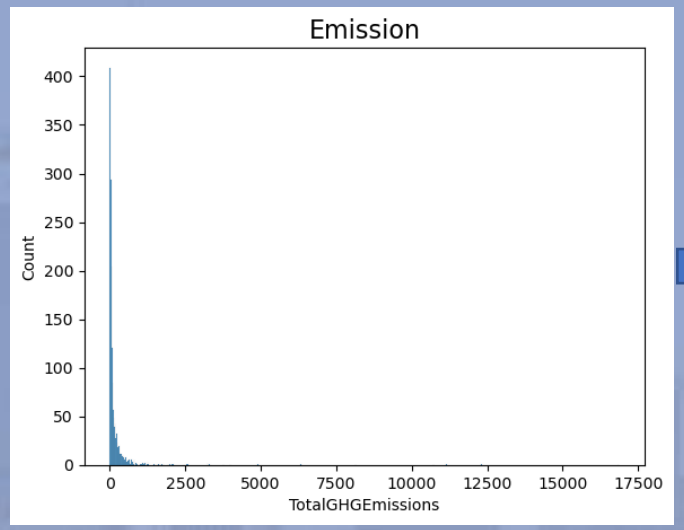
Seattle

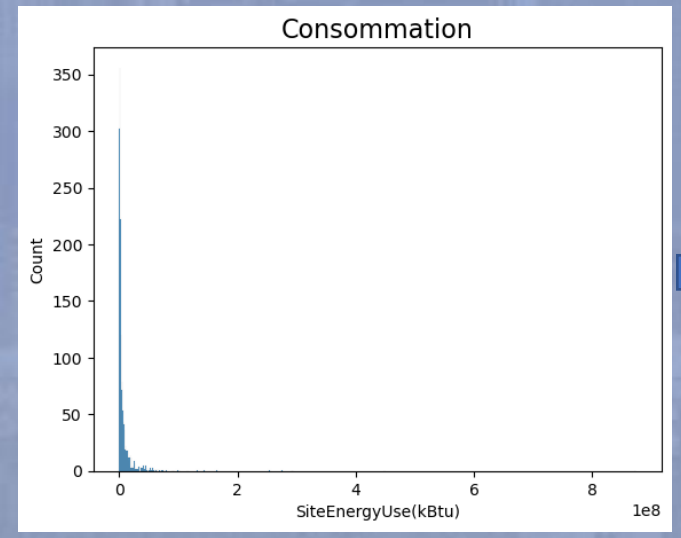
Méthodologie



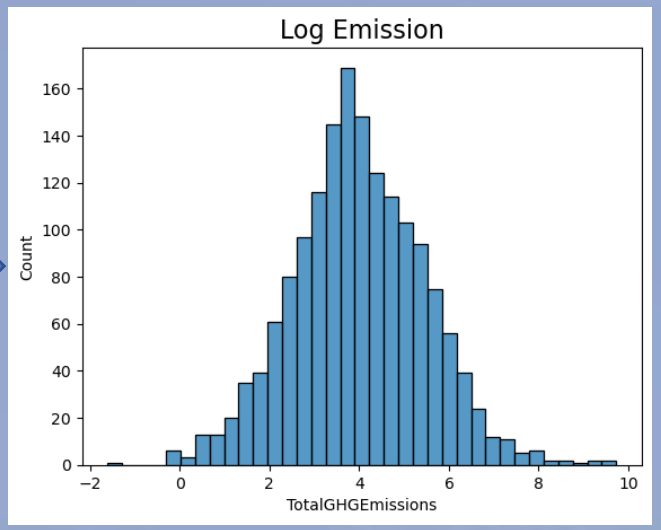
Pré-traitement des variables

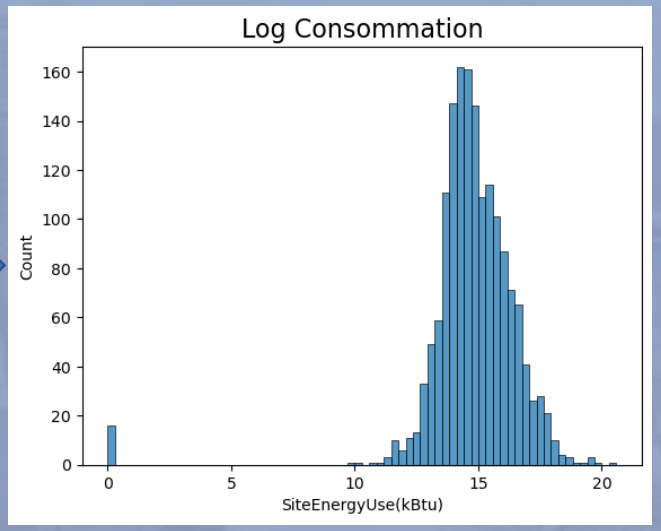
Transformation log pour les valeurs cibles





Transformation
Logarithme



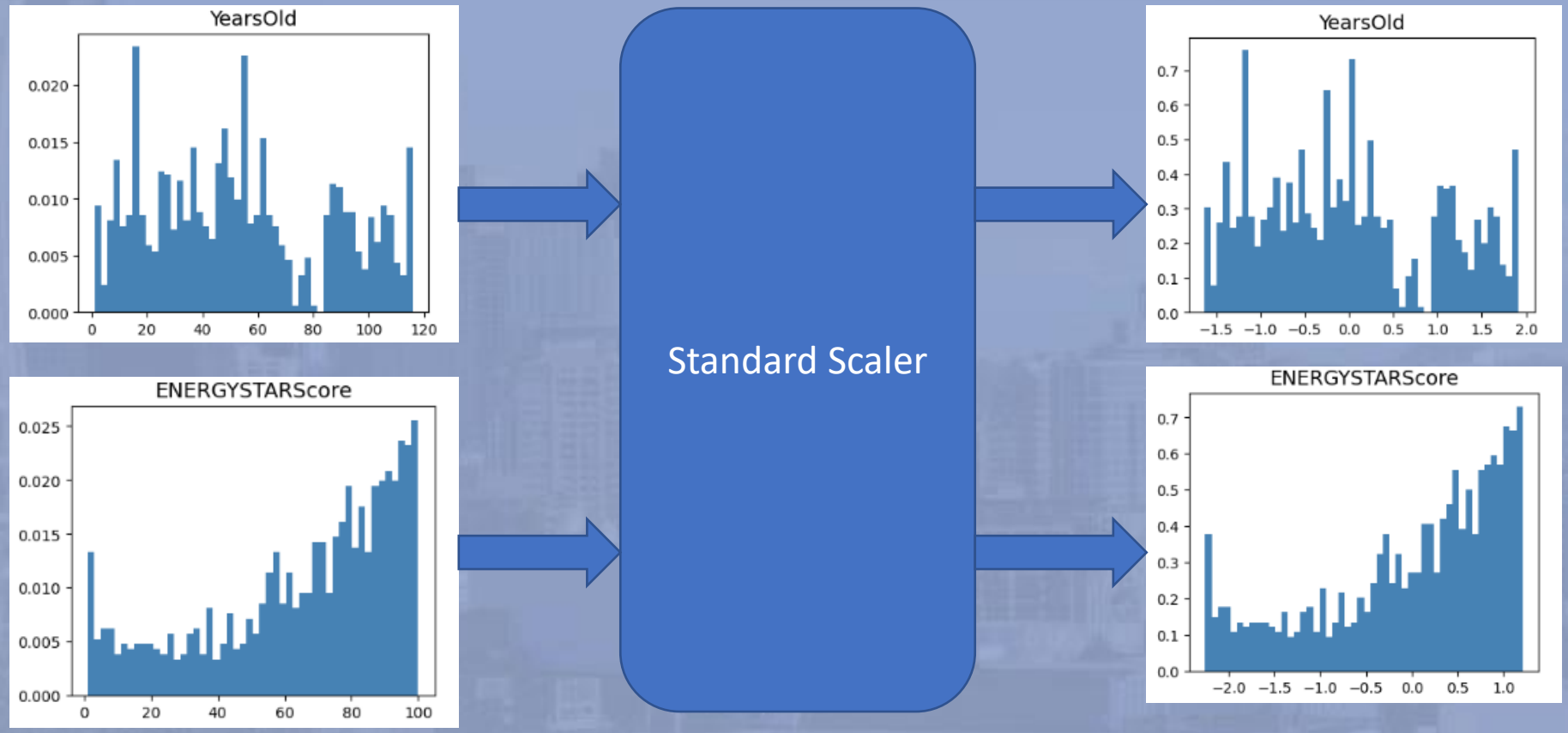


Seattle

Normalisation et encodage

- Normalisation avec Standard Scaler

  Permet de réaliser une mise à l'échelle afin que les données soient centrées à 0 avec un écart type de 1



- Encodage avec One Hot Encoder

PrimaryPropertyType
Warehouse
Other
Hotel
Large Office
…..

Warehouse	Other	Hotel	Large Office	….
1	0	0	0	
0	1	0	0	
0	0	1	0	
0	0	0	1	

MSE , RMSE et R^2

- MSE :

Moyenne des distances euclidiennes entre les valeurs réelles et les valeurs prédites.

- RMSE :

Racine carrée de la moyenne des distances euclidiennes entre les valeurs réelles et les valeurs prédites.

L'erreur moyenne sur le log de la valeur cible

- R^2 (Coefficient de Corrélation):

somme des distances euclidiennes entre valeurs prédites et valeurs réelles divisé par la somme des distances euclidiennes entre valeurs réelles et moyenne.

La capacité de la prédiction à expliquer la variance réelle (sur le log).

Test des différents modèles

- Les modèles qui vont être tester:

  - Régression Ridge
  - Régression Lasso
  - Régression Elastic Net
  - kNN Regressor
  - Suport Vector Regressor
  - XGBoost Regressor
  - LightGBM Regressor
  - Random Forest Regressor

Cross Validation

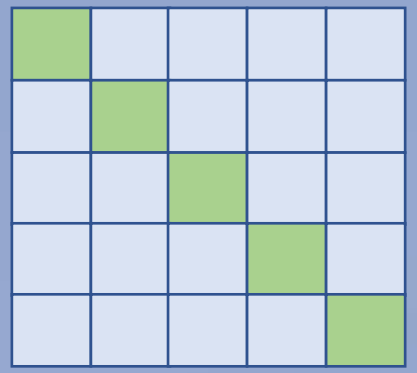
Split

Training set

Test set

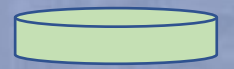
Repeated K-Fold

Grid Search
Cross Validation



Train

Validation
Set



Modèles

Meilleurs paramètres

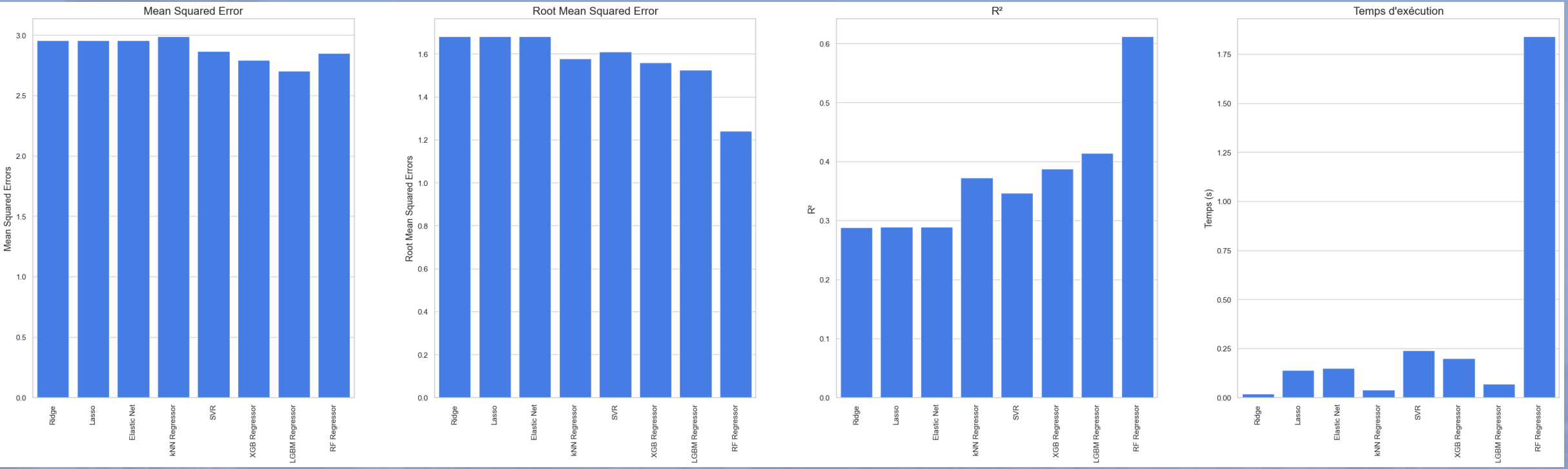
Modèles

Test

Comparaison

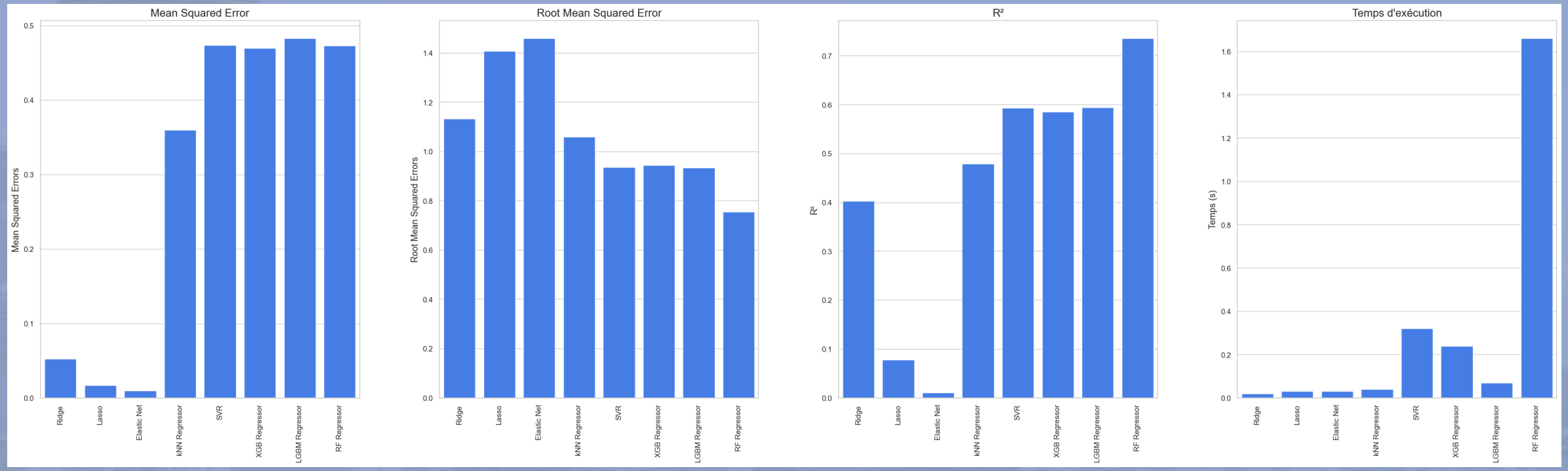
Evaluation et
sélections du
modèle

Comparaison des modèles pour la consommation d’énergie



Seattle

Comparaison des modèles pour l’émission de CO2



Seattle

Consommation d'énergie

	R^2 Train	R^2 Test
Ridge	0.303815	-0.728460
Lasso	0.304169	-0.700649
Elastic Net	0.303641	-0.735701
kNN Regressor	0.422469	0.362411
SVR	0.360211	0.322745
XGB Regressor	0.445751	0.373037
LGBM Regressor	0.405397	0.354933
RF Regressor	0.592719	0.297099

- Problème d'overfitting pour certains modèles
- Comportement sur le test set

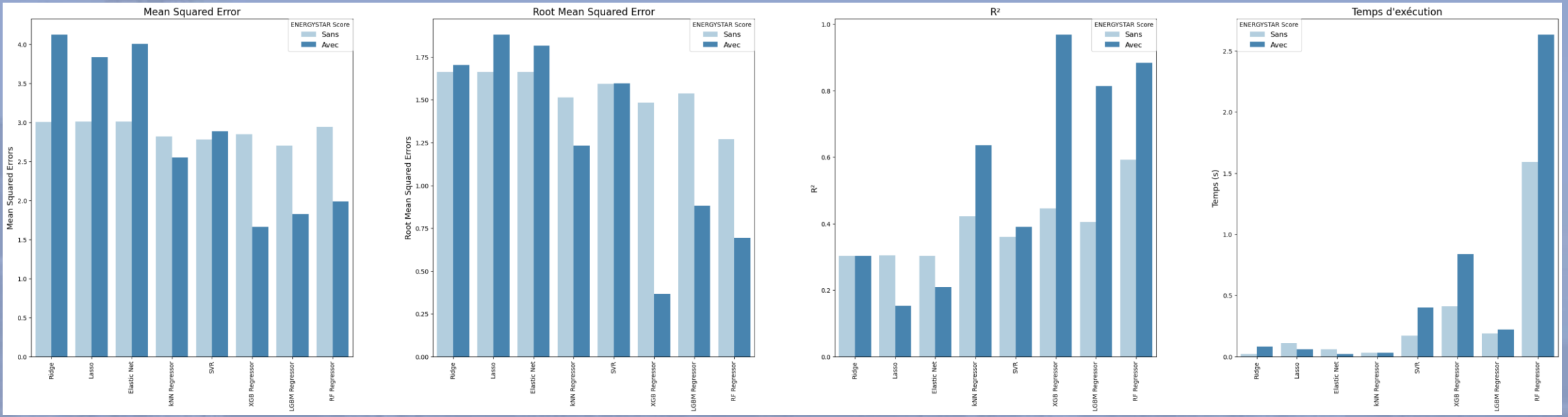
Emission C02

	R^2 Train	R^2 Test
Ridge	0.402866	0.361437
Lasso	0.078094	0.061417
Elastic Net	0.010933	0.007445
kNN Regressor	0.479059	0.412488
SVR	0.593057	0.524511
XGB Regressor	0.585509	0.538961
LGBM Regressor	0.594466	0.508146
RF Regressor	0.734652	0.476599

- Pas d'overfitting
- Comportement sur le test set

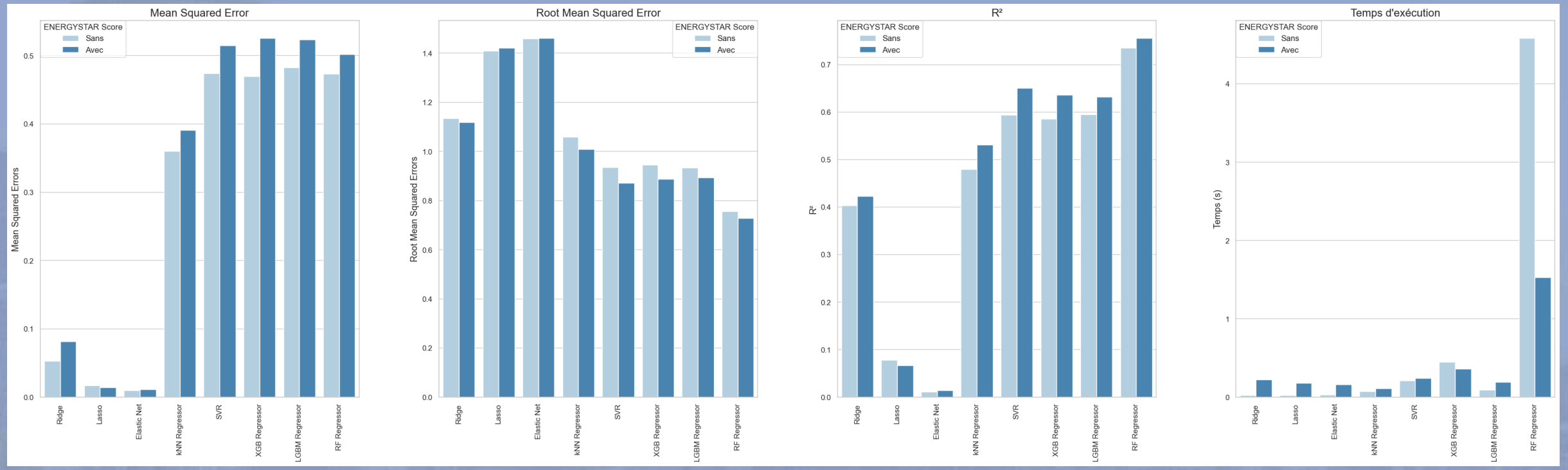
Nouveaux tests avec l’ENERGYSTAR Score

Consommation énergie



Seattle

Emission CO2



Seattle

Consommation d'énergie

	R^2 Train	R^2 Test
Ridge	0.303395	0.244833
Lasso	0.153382	0.141042
Elastic Net	0.209925	0.172316
kNN Regressor	0.635795	0.658304
SVR	0.390350	0.410527
XGB Regressor	0.967926	0.823031
LGBM Regressor	0.814276	0.761886
RF Regressor	0.884349	0.847370

- Pas d'overfitting
- Comportement sur le test set

Emission C02

	R^2 Train	R^2 Test
Ridge	0.423086	0.372486
Lasso	0.066071	0.046992
Elastic Net	0.014402	0.008995
kNN Regressor	0.530612	0.403665
SVR	0.649610	0.535243
XGB Regressor	0.635468	0.557803
LGBM Regressor	0.631510	0.533800
RF Regressor	0.754852	0.533090

- Pas d'overfitting
- Comportement sur le test set

Choix du modèle

Pour la prédiction de la consommation d'énergie:

Random Forest Regressor

Pour la prédiction de l'émission de C02:

XGB Regressor

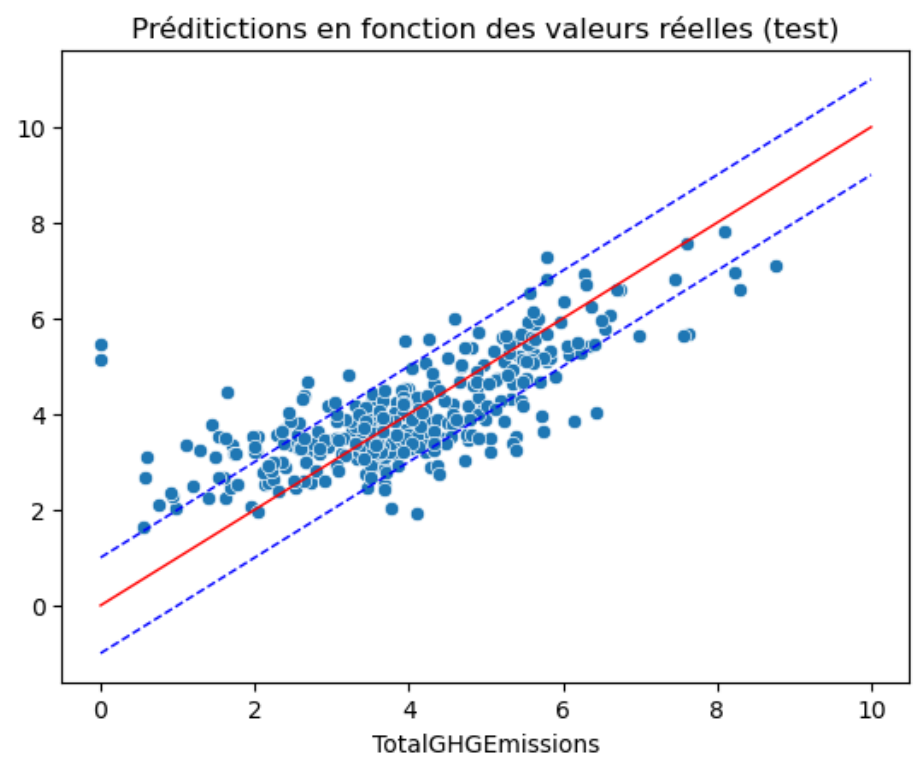
Validation et Prédictions

XGB Regressor

$R^2 = 0,56$

MSE = 1,01

RMSE = 1

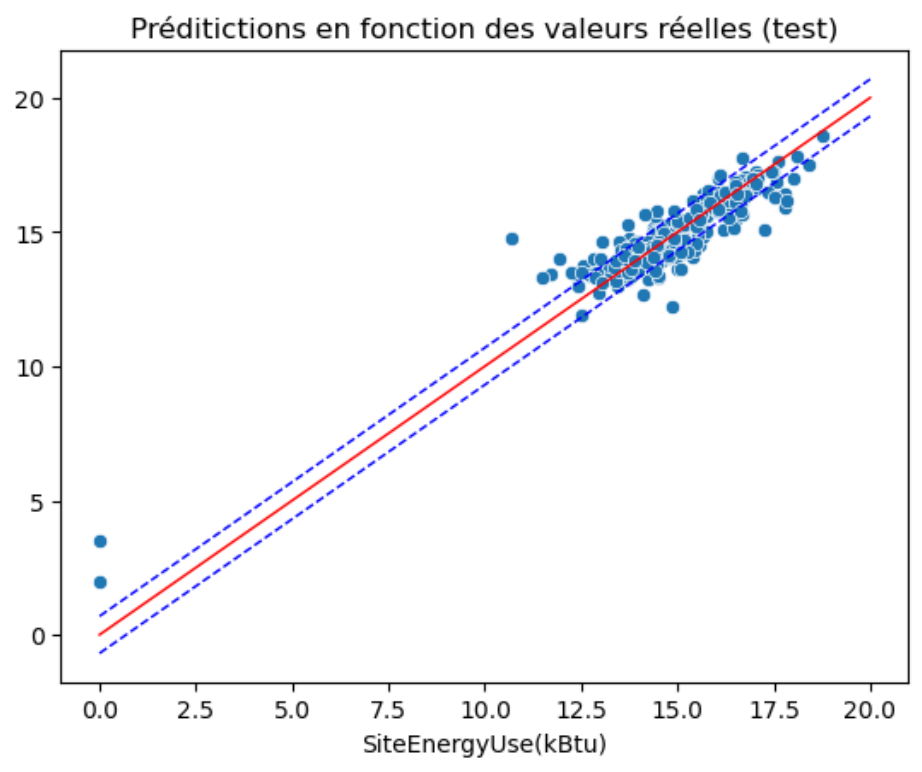


Random Forest Regressor

$R^2 = 0,84$

MSE = 0,48

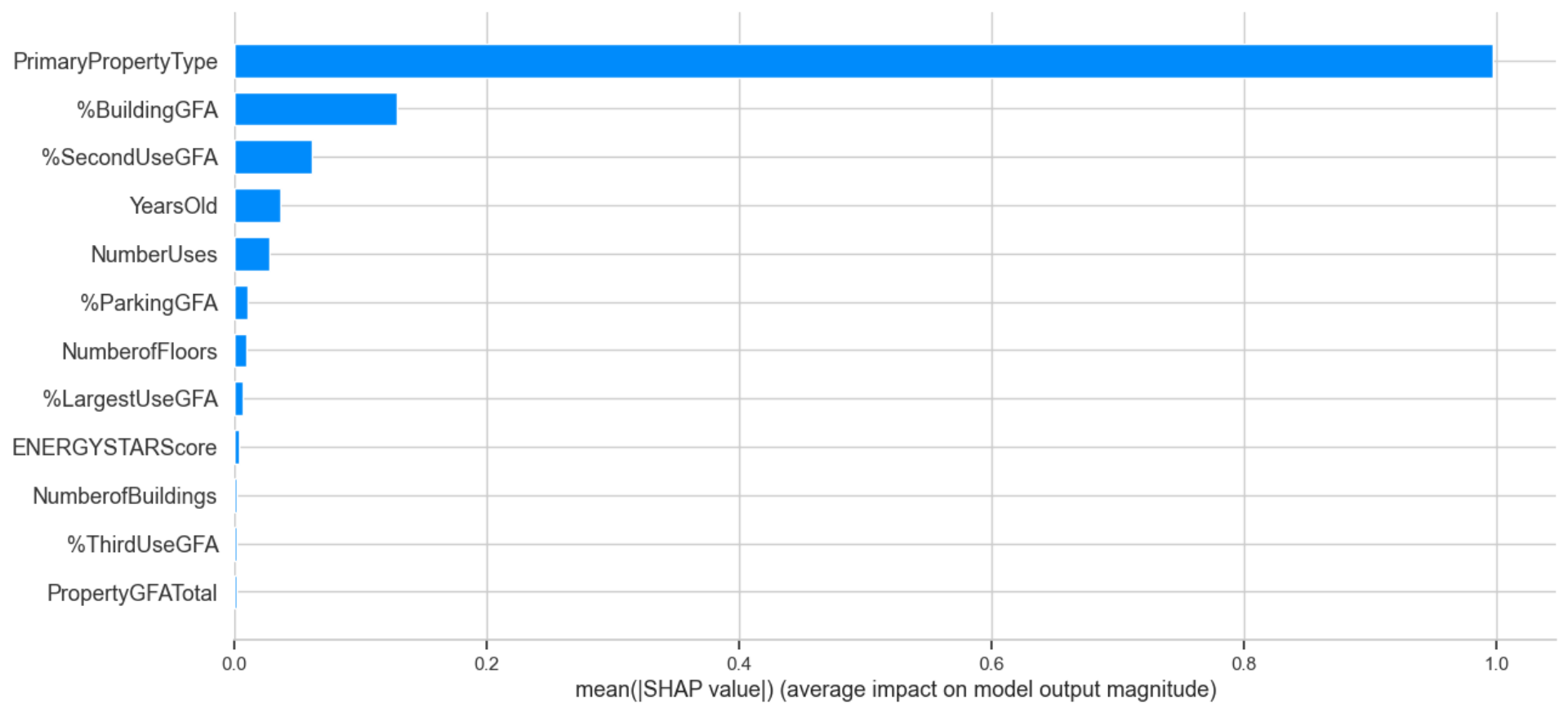
RMSE = 0,69



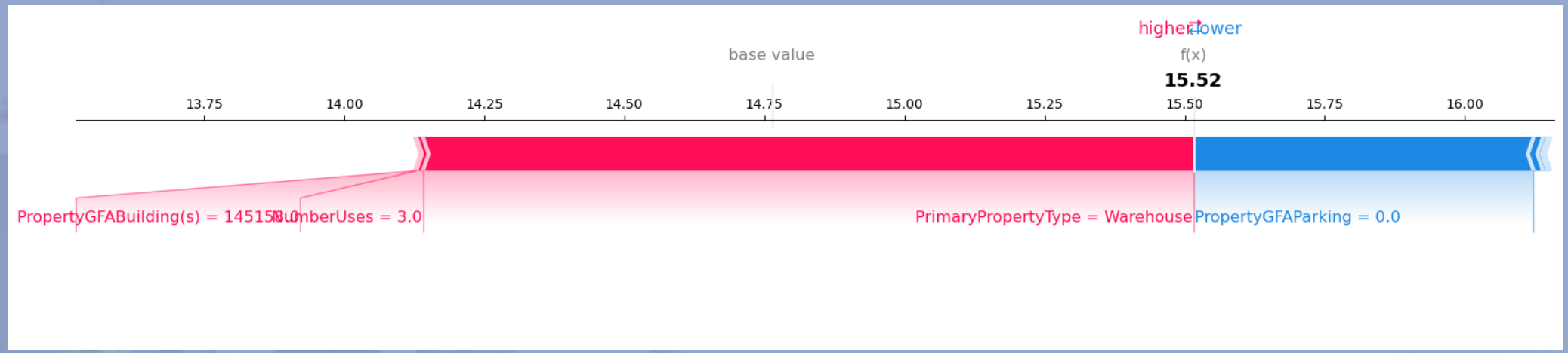
Seattle

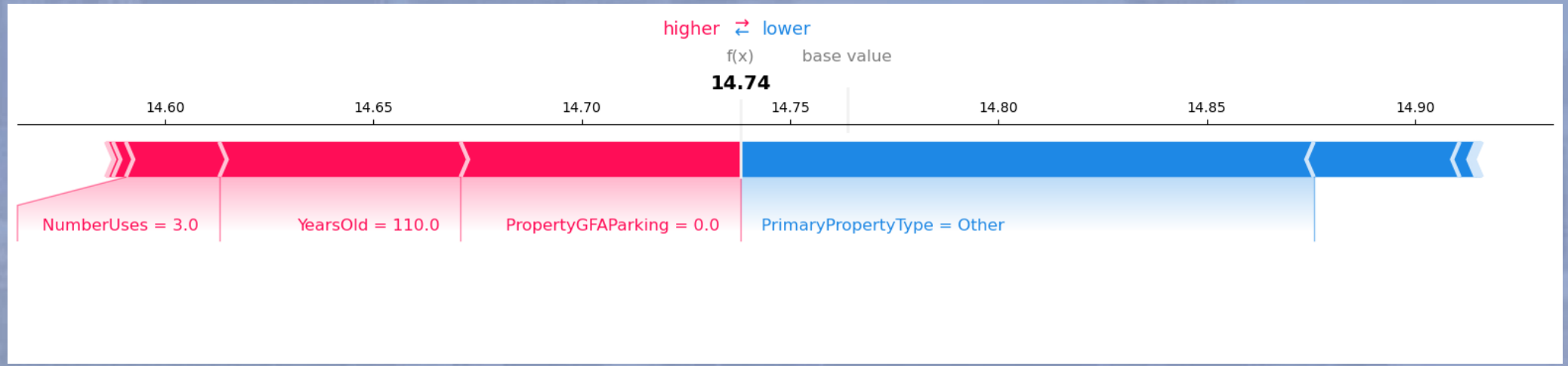
FEATURES IMPORTANCES

Consommation énergie

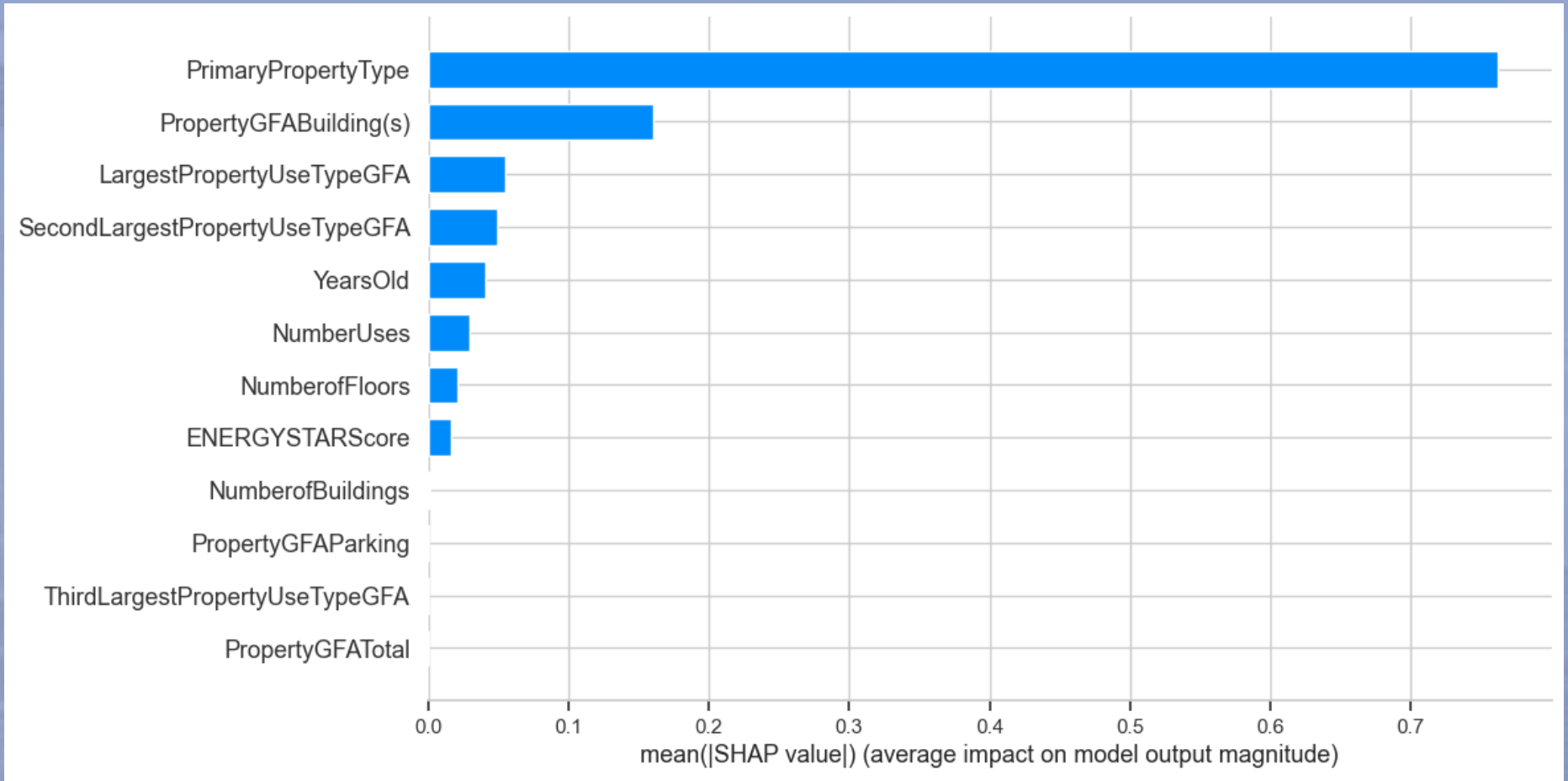


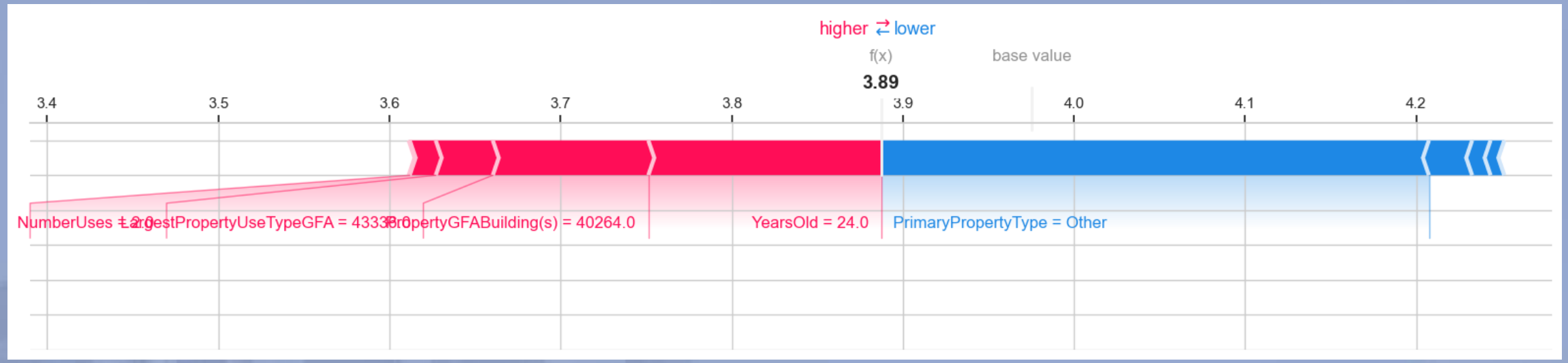
Seattle

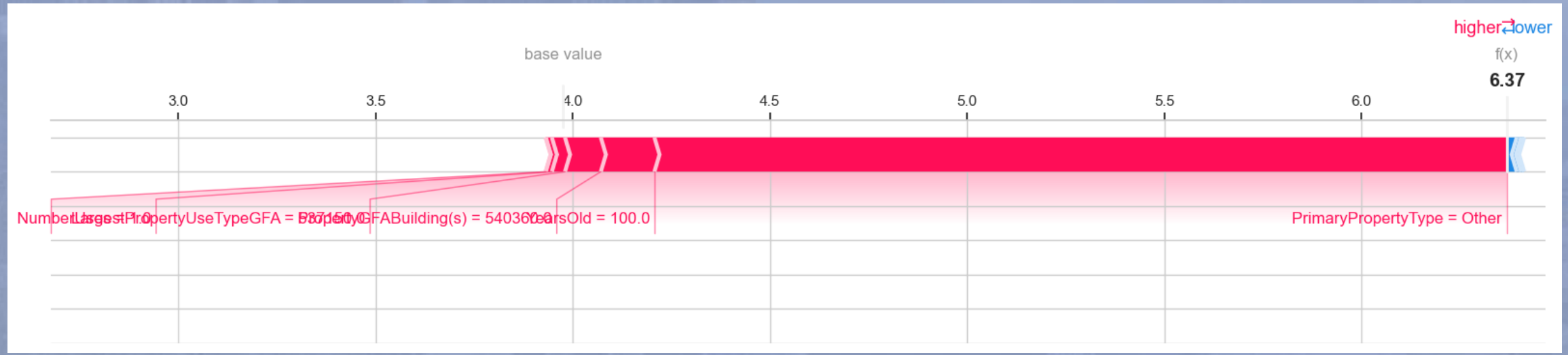

higher ⇄ lower
base value
f(x)
15.52
13.75
14.00
14.25
14.50
14.75
15.00
15.25
15.50
15.75
16.00
PropertyGFABuilding(s) = 145153.0
NumberUses = 3.0
PrimaryPropertyType = Warehouse
PropertyGFAParking = 0.0


higher ⇄ lower
f(x)
base value
14.74
14.60
14.65
14.70
14.75
14.80
14.85
14.90
NumberUses = 3.0
YearsOld = 110.0
PropertyGFAParking = 0.0
PrimaryPropertyType = Other

Emission CO2




higher ⇄ lower
f(x)
3.89
base value
3.4
3.5
3.6
3.7
3.8
3.9
4.0
4.1
4.2
NumberUses = 2.0
LargestPropertyUseTypeGFA = 433380.0
PropertyGFABuilding(s) = 40264.0
YearsOld = 24.0
PrimaryPropertyType = Other


higher ⇄ lower
base value
f(x)
6.37
3.0
3.5
4.0
4.5
5.0
5.5
6.0
NumberUses = 1.0
LargestPropertyUseTypeGFA = 837150.0
PropertyGFABuilding(s) = 540360.0
YearsOld = 100.0
PrimaryPropertyType = Other

Conclusion

- Prédiction satisfaisante mais lente pour la consommation d'énergie
- Prédiction décevante concernant l'émissions CO2
- L'ENERGYSTAR Score améliore les prédictions

MERCI DE VOTRE ATTENTION

Seattle