

# **Anticipez les besoins en consommation électrique de bâtiments**



Objectif : Ville neutre en émissions de carbone

# Problématiques

- **Relevés coûteux**



- **Prédiction à partir des données déclaratives**
- **Évaluer l'intérêt de l'Energy Star Score**



# Interprétation

- **Valeur continue → Régression**
- **Impact de l'Energy Star Score**

# Base de donnée

- **Diffusée par ville de Seattle**
- **Contient pour ~3000 propriétés :**
  - Les informations déclaratives
  - La consommation énergétique
  - Les émissions de gazs à effets de serre
  - Le Score Energy Star



# Nettoyages préliminaires

- **Clé primaire : OSEBuilding**
- **Formatage des noms de colonnes**
- **Différences entre les colonnes**



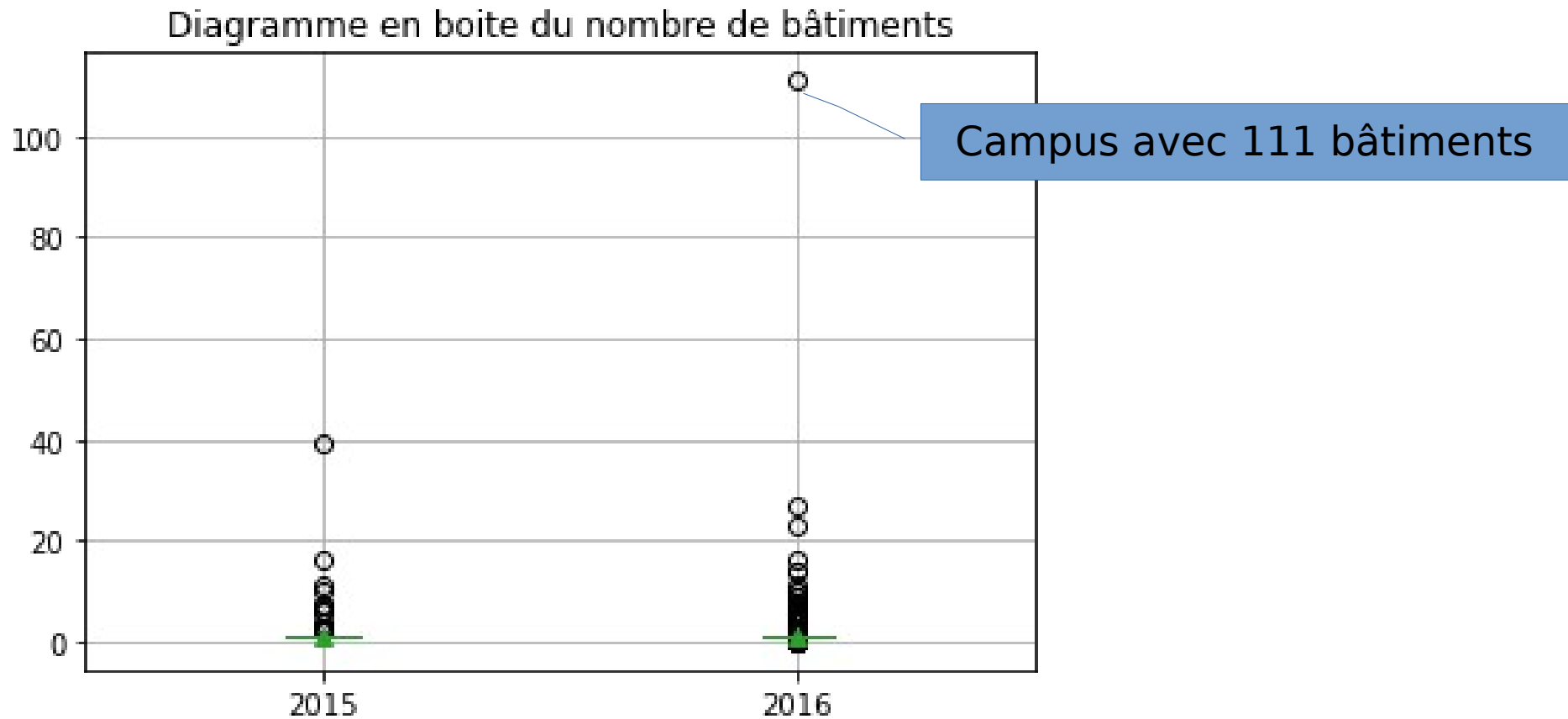
# Sélection des colonnes

- **YearBuilt**
- **NumberofBuildings**
- **NumberofFloors**
- **LargestPropertyUseType**
- **LargestPropertyUseTypeGFA**
- **PropertyGFABuilding(s)**
- **Latitude**
- **Longitude**

# Valeurs aberrantes

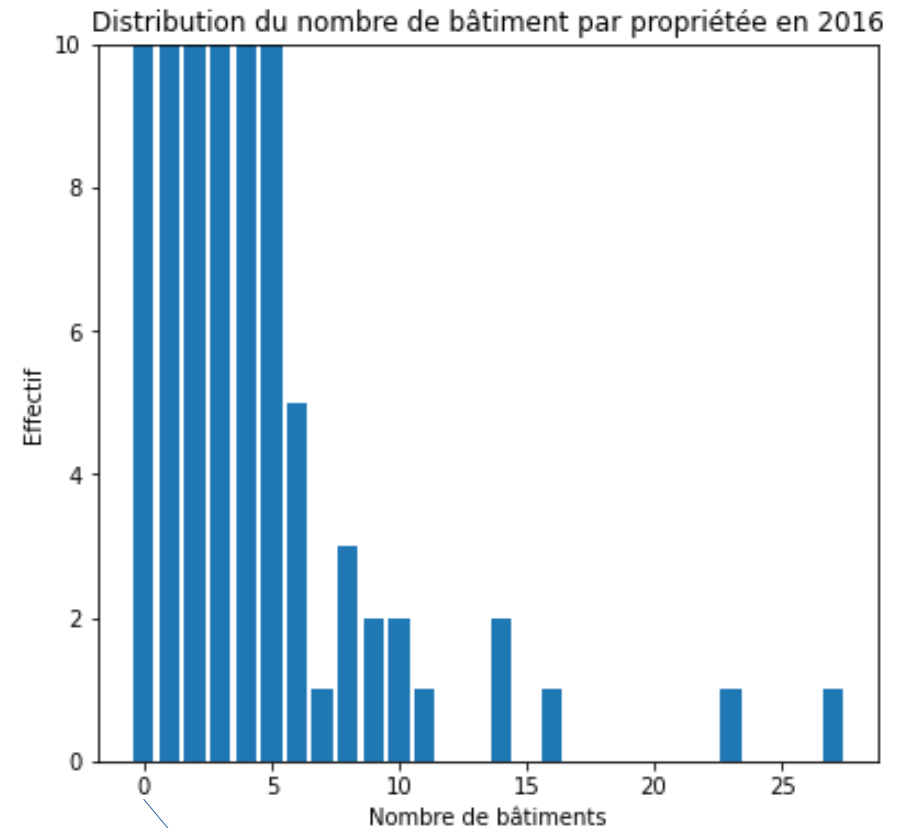
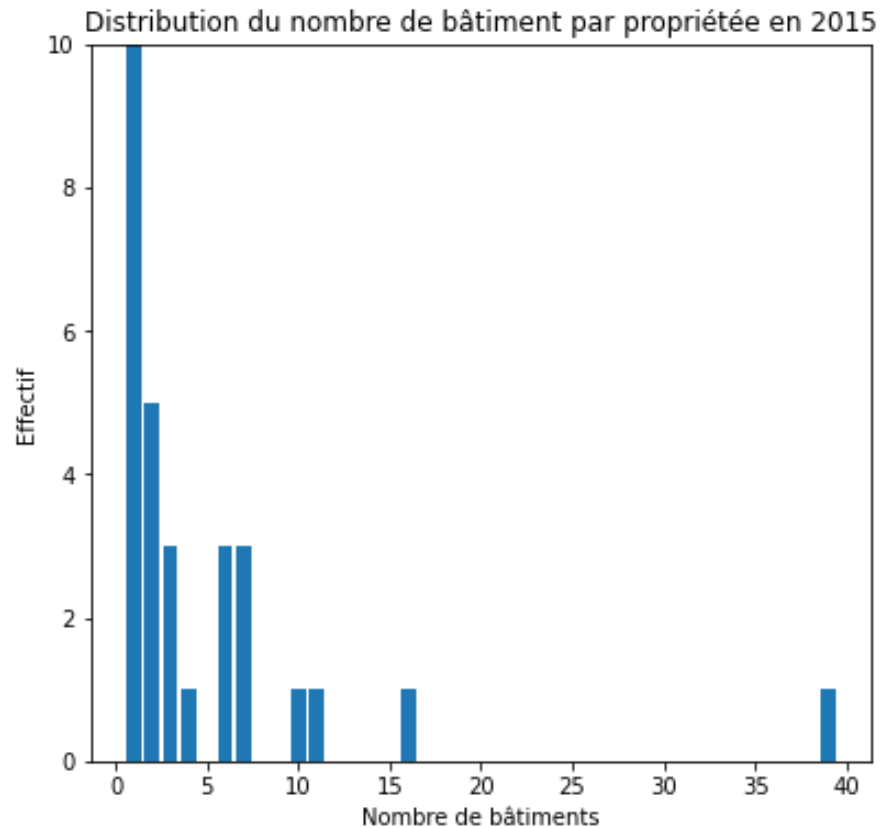
- **Pas d'expertise dans le domaine**
- **Boxplot**
  - Valeurs extrêmes
- **Affichage des distributions**
  - Détection des mouvements minoritaires

# NumberofBuildings





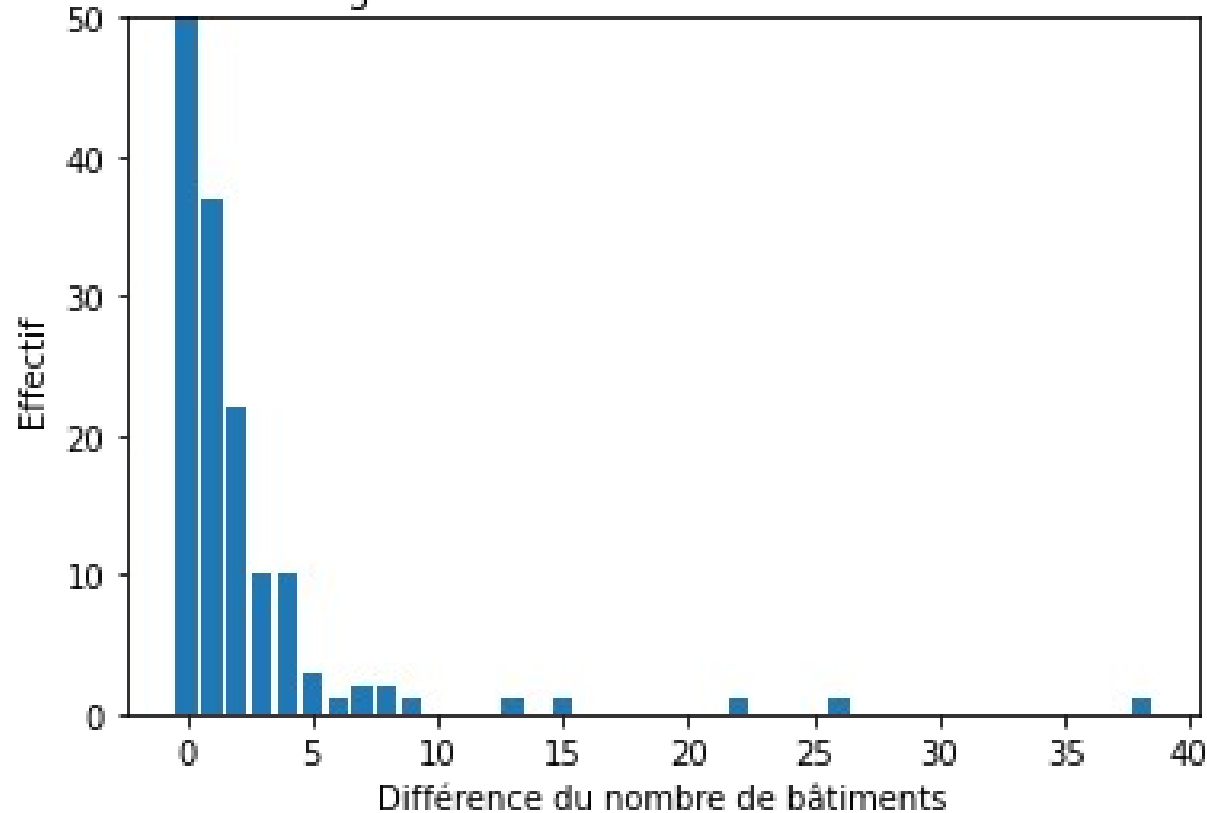
# NumberofBuildings



Aucun bâtiment

# NumberofBuildings

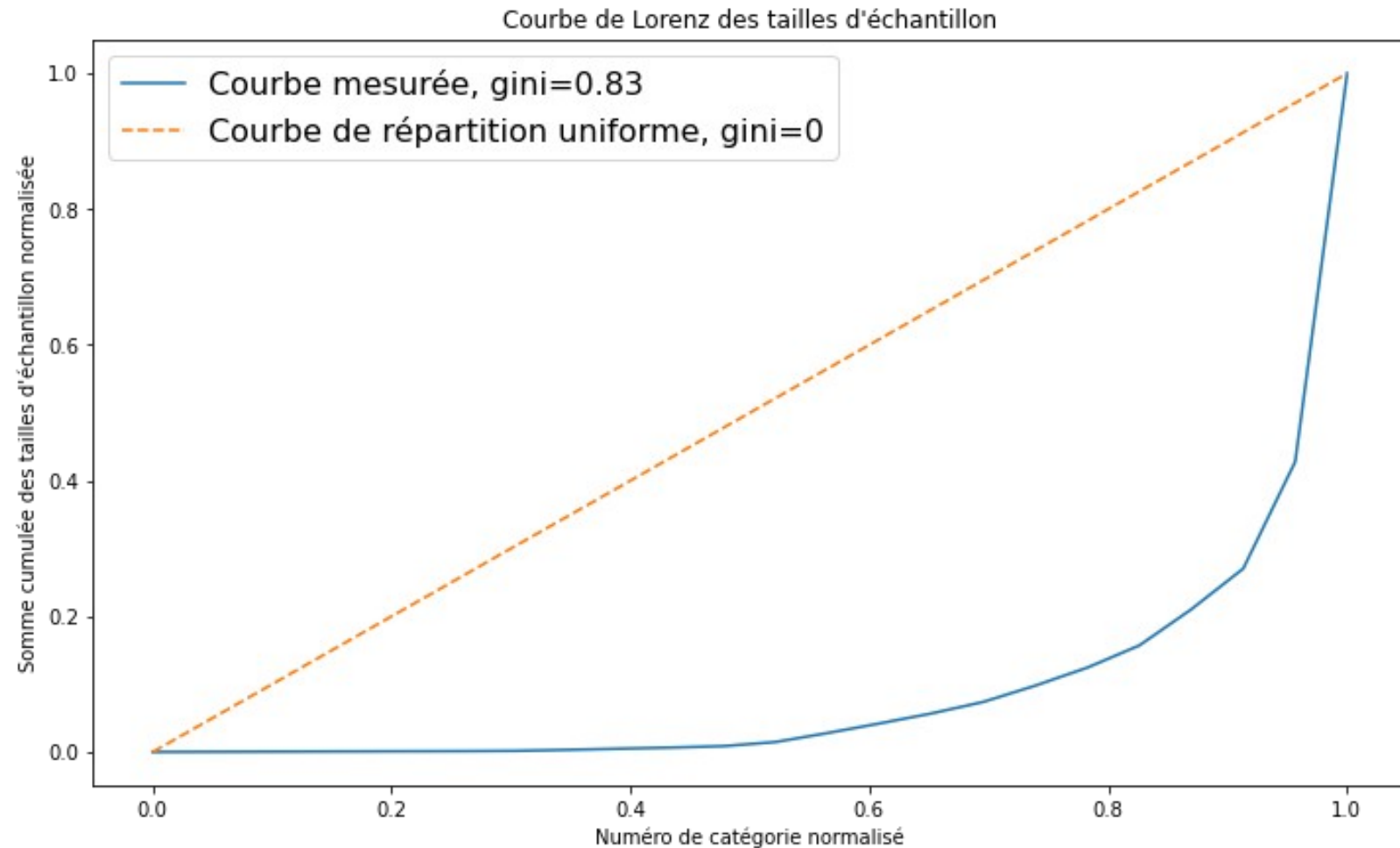
Distribution du changement du nombre de bâtiments entre 2015 et 2016



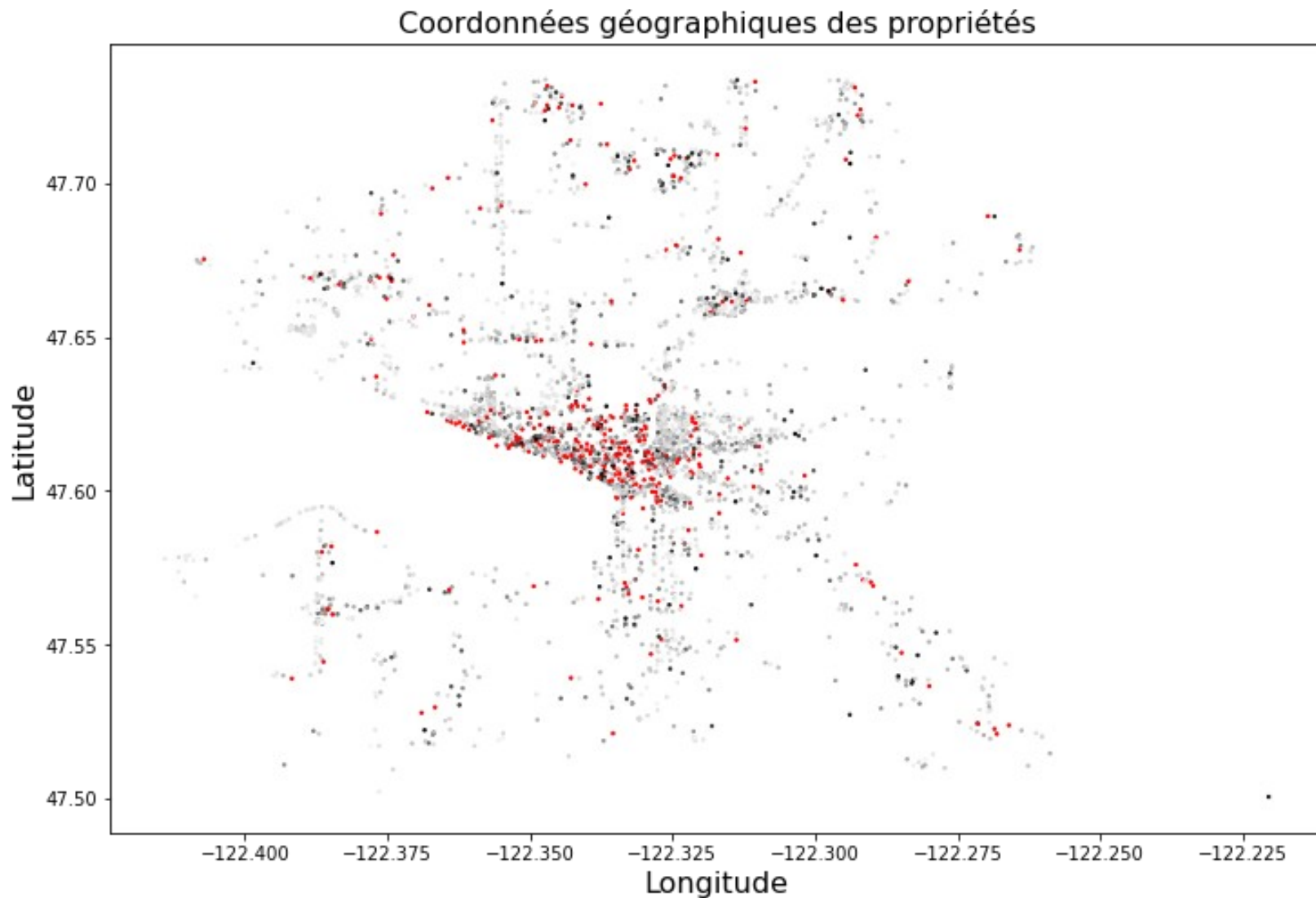
# Valeurs manquantes

```
Taux de remplissage des variables de 2015:  
Comments                0.218341  
Outlier                  2.432938  
YearsENERGYSTARCertified 2.838428  
ThirdLargestPropertyUseTypeGFA 16.094822  
ThirdLargestPropertyUseType 16.094822  
SecondLargestPropertyUseTypeGFA 46.163444  
SecondLargestPropertyUseType 46.163444  
ENERGYSTARScore          77.386151  
LargestPropertyUseTypeGFA 96.163444  
LargestPropertyUseType    96.163444  
dtype: float64
```

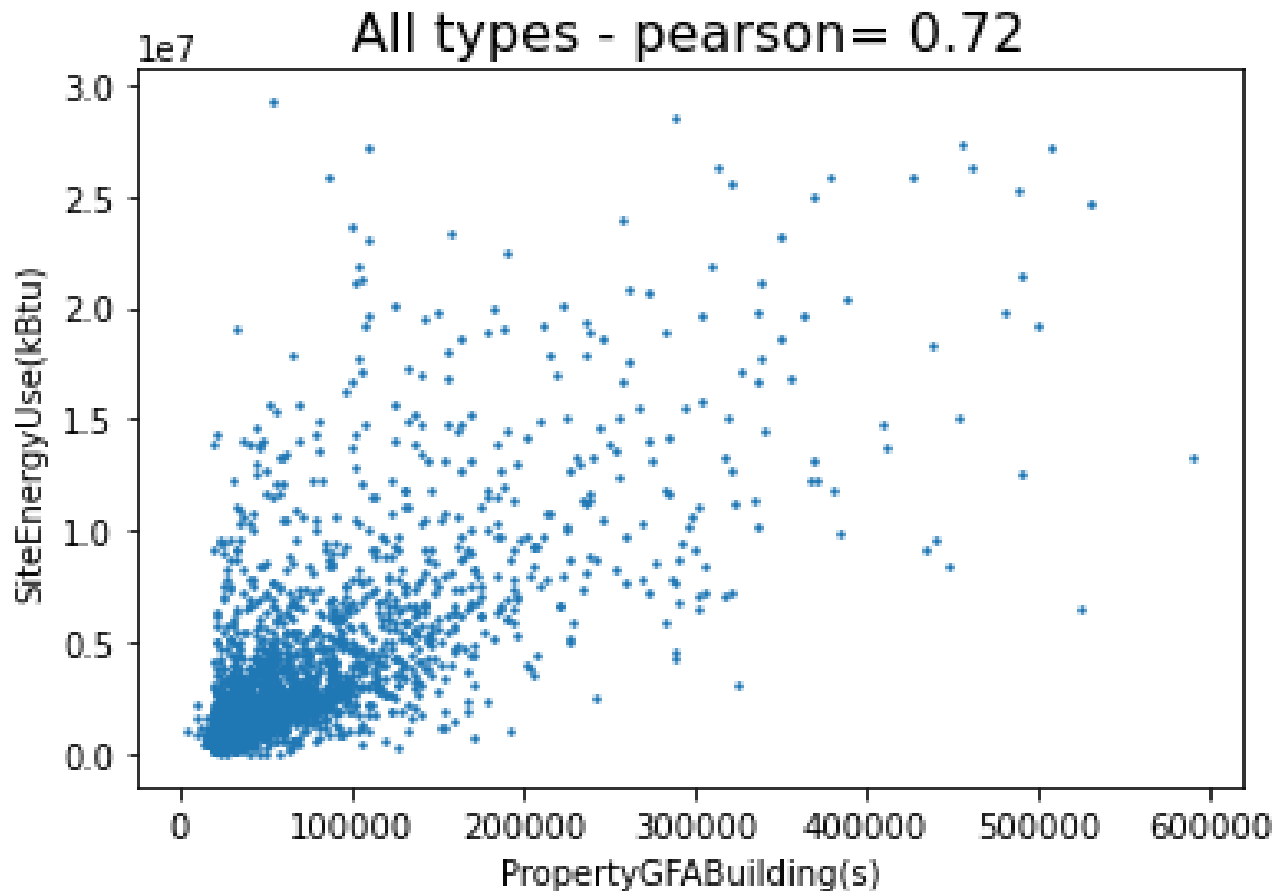
# Tailles des échantillons par activité



# Répartition géographique de la consommation

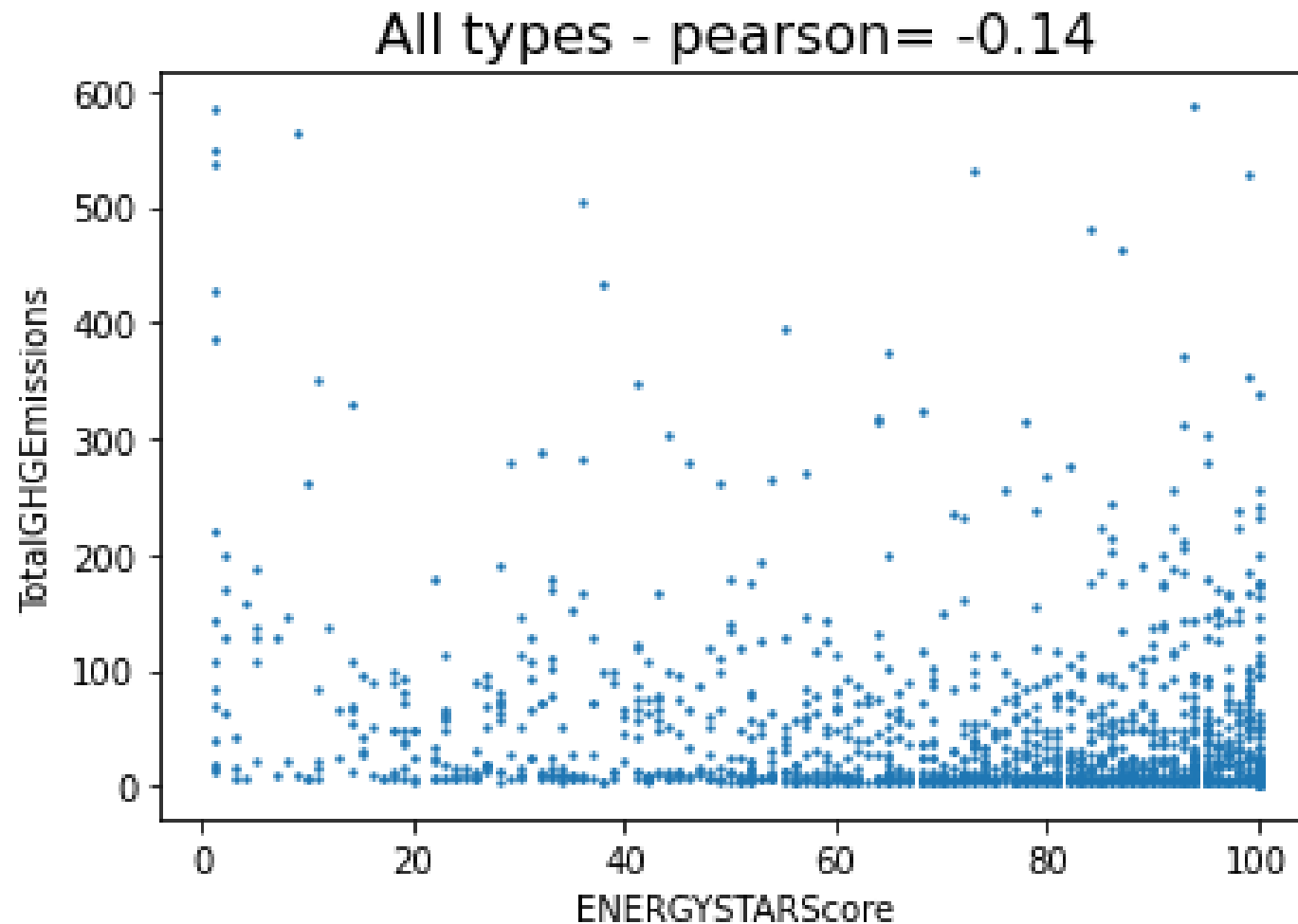


# Corrélation : GFA et consommation

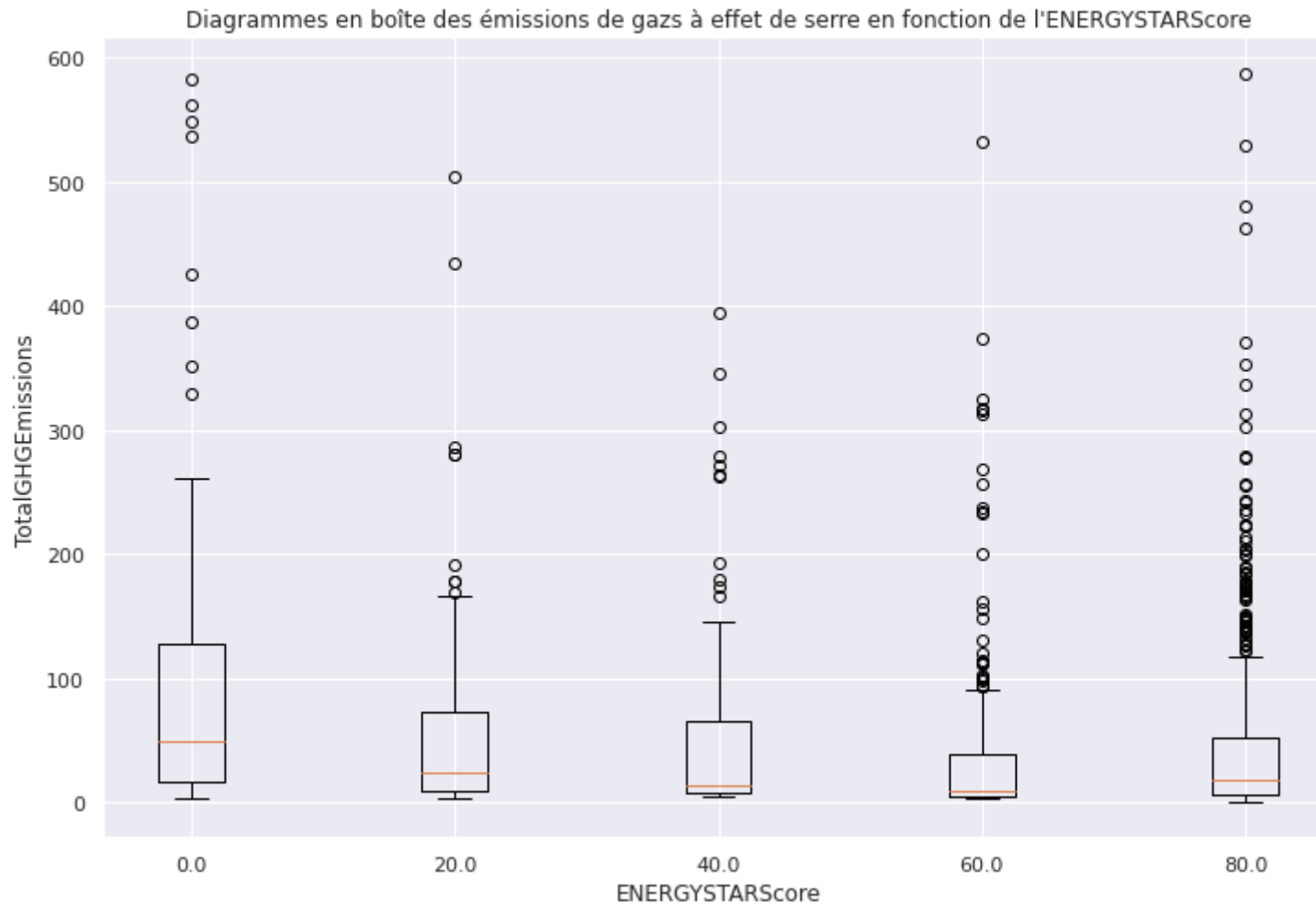


Par type de bâtiment : 0.7 – 0.94

# Energy Star Score - Émissions

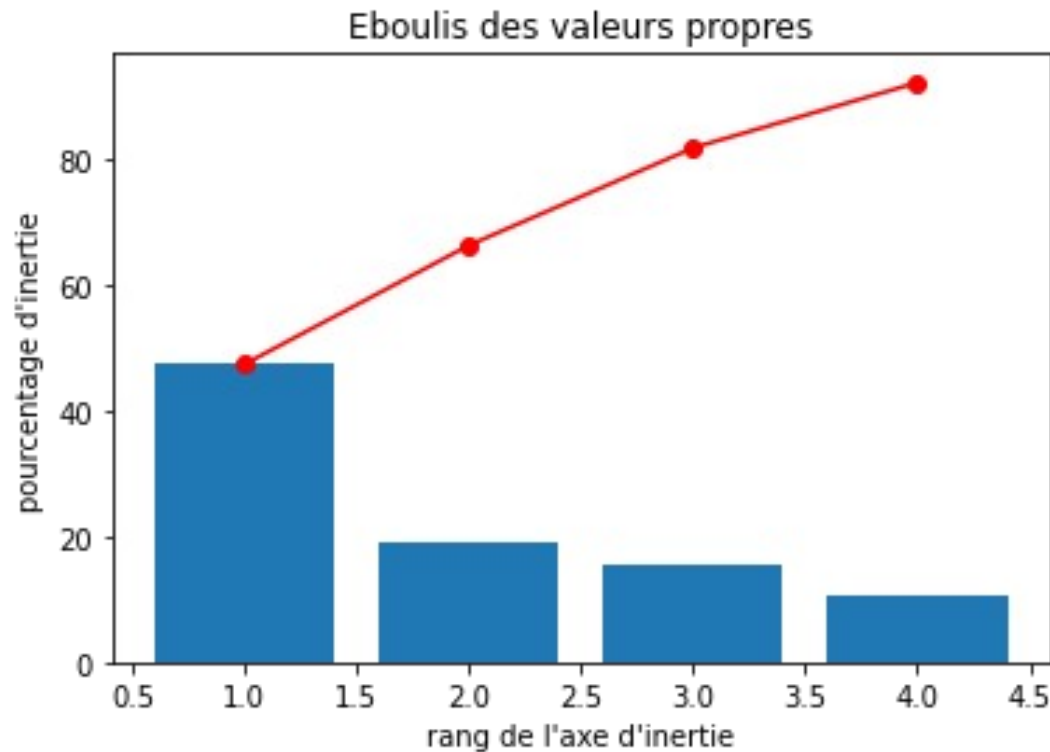


# Energy Star Score - Émissions

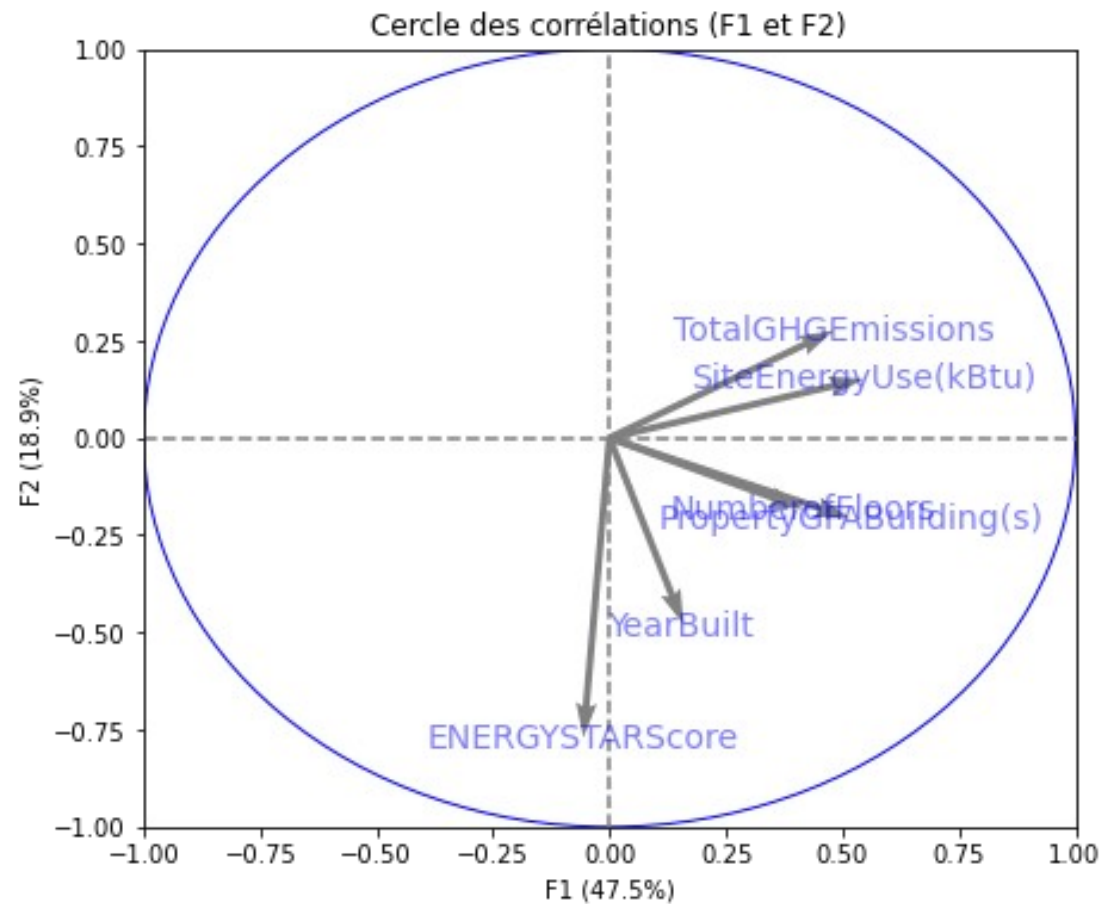




# Analyse par Composantes Principales



# Analyse par Composantes Principales



# Modèles

- **Régression linéaire**
- **Régression Ridge, Lasso**
- **Algorithme des plus proches voisins**
- **Forêt aléatoire**
- **Gradient boosting**

# SiteEnergyUse

Régression linéaire	0.48
Ridge	0.48
Lasso	0.48
k-nn	0.72
Forêt aléatoire	0.80
Gradient boosting	0.85

alpha =25

alpha=1000

k=2

n\_estimators = 13  
max\_depth = 14

Learning\_rate=0.4  
n\_estimators=50  
max\_depth=10

# SiteEnergyUse : Optimisation

- Recherche par grille des meilleurs paramètres
- $R^2 = 0.87$
- + Energy Star Score :  $R^2 = 0.86$

# GHGEmissions

Régression linéaire	0.28
Ridge	0.28
Lasso	0.28
k-nn	0.57
Forêt aléatoire	0.71
Gradient boosting	0.79

alpha =25

alpha=1

k=2

n\_estimators = 13  
max\_depth = 14

Learning\_rate=0.4  
n\_estimators=50  
max\_depth=10

# **GHGEmissions: Optimisation**

- **Recherche par grille des meilleurs paramètres**
- **$R^2 = 0.81$**
- **+ Energy Star Score :  $R^2 = 0.81$**

# Conclusion

- **On fait des modèles applicables à tous les types de bâtiments / d'activité**
- **Avec le gradient boosting optimisé :**
  - SiteEnergyUse :  $R^2 = 0.87$
  - TotalGHGEmissions :  $R^2 = 0.81$



# Ouverture

**Des améliorations sont possibles**

- **Transformer les features de position**
- **Chercher une expertise métier pour améliorer la qualité du nettoyage**
- **Utiliser un réseau de neurones profond**