

Projet 3 : Anticipez les besoins en consommation électrique de bâtiments

Claire Gayral
Juillet 2021

Introduction

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression

- Données bâtiments Seattle
- Deux prédictions à faire :
 - Consommation d'énergie
 - Émission de CO2
 - + *Energy Star score VS émission de CO2*



from <https://www.freepng.fr>

Introduction - Plan

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression

1 La mission

2 Analyse exploratoire

- Nettoyage des données
- Analyse univariée
- Analyse multivariée

3 Régression

1 La mission

2 Analyse exploratoire

3 Régression

Les données

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression

Les données d'entrée :

- Deux années : 2015 et 2016
- Bâtiments non résidentiels
- Test uniquement sur le permis de construction

Les données de sortie :

- Différentes variables énergétiques (gaz, fuel, électricité ...)
- Deux variables d'émission de CO₂

Conseils et avertissements de Douglas

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression

- Fuite de données
- Pertinence des variables
- Transformation d'échelle
- Paramétrisation par validation croisée

1 La mission

2 Analyse exploratoire

- Nettoyage des données
- Analyse univariée
- Analyse multivariée

3 Régression

Fusion des tables

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression

Avant fusion :

- Homogénéisation des noms de variables
- Choix des types des variables

Méthode de fusion :

- Moyenne sur les variables numériques
- Remplissage des valeurs manquantes dans 2016 à partir de 2015 pour les variables catégorielles

Prétraitements divers

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression

3432 bâtiments et **44** variables

- **Suppression variables et doublons**
- Inférence valeurs manquantes
- Fusion variables corrélées
- Suppression batiments résidentiels

—9 ni numérique ni catégorielle

—2 variables : 95% valeurs manquantes

0 doublons

20533 valeurs manquantes (13.6%)

Prétraitements divers

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

Analyse
exploratoire

Nettoyage des
données

Analyse univariée

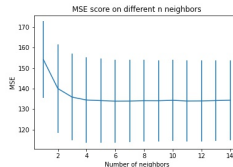
Analyse multivariée

Régression

3432 bâtiments et 33 variables

- Suppression variable et doublons
- **Inférence valeurs manquantes**
- Fusion variables corrélées
- Suppression batiments résidentiels

Par KNN avec 4 voisins



9 variables catégorielles et 24 numériques
1339 valeurs manquantes (1.2%)

Prétraitements divers

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

3432 bâtiments et **33** variables

- Suppression variable et doublons
- Inférence valeurs manquantes
- **Fusion variables corrélées**
- Suppression batiments résidentiels

fusion de 4 blocs :

- électricité
- surface au sol
- gaz naturel
- énergie

9 variables catégorielles et **24** numériques
122 valeurs manquantes (0.1%)

La mission

Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression

Matrice des corrélations

Projet 3 : Anticipez les besoins en consommation électrique de bâtiments

Claire Gavral

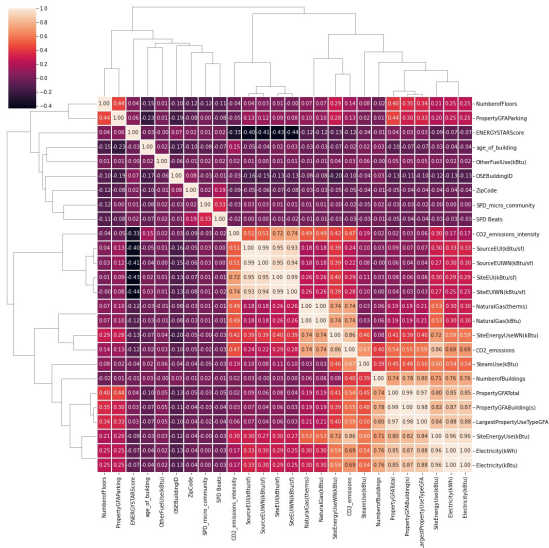
La mission

Analyse
exploratoire

Nettoyage des données

Analyse univariée

Analyse multivariée



Prétraitements divers

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

3432 bâtiments et 26 variables

- Suppression variable et doublons
- Inférence valeurs manquantes
- Fusion variables corrélées
- **Suppression batiments résidentiels**

8 variables catégorielles et 18 numériques
122 valeurs manquantes (0.1%)

La mission

Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression

Prétraitements divers - bilan

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression

1698 bâtiments et 26 variables

8 variables catégorielles et 18 numériques
59 valeurs manquantes (0.1%)

Variables d'apprentissage :

- 6 numériques
- 5 catégorielles

Variables cibles

- "SiteEnergyUse(kBtu)"
comme proxy
- "CO2_emissions"

Variables d'apprentissage numériques

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

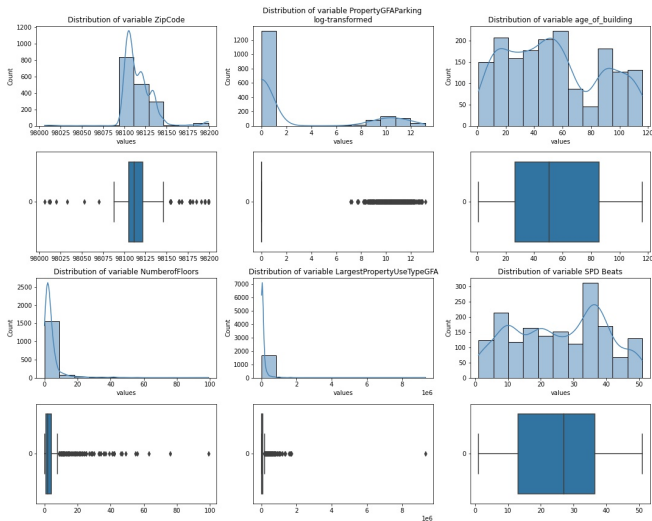
Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression



Variables d'apprentissage catégorielles

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

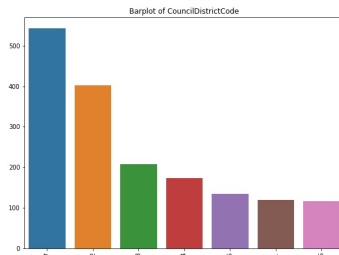
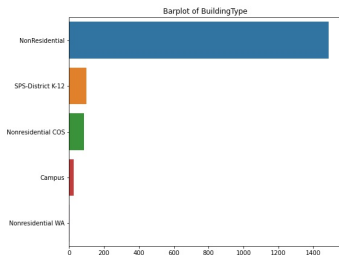
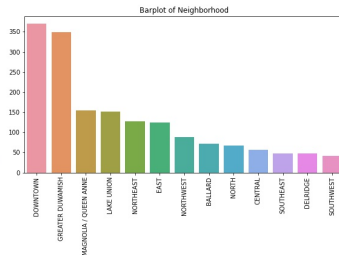
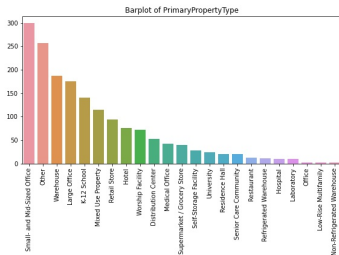
Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression



Variables d'apprentissage numériques

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

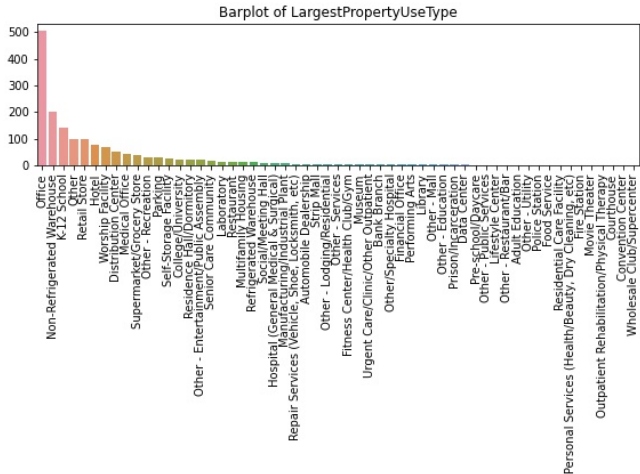
Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression



Variables de sortie

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

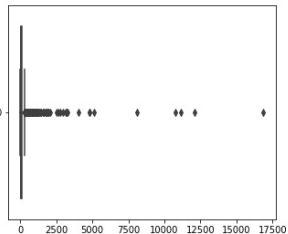
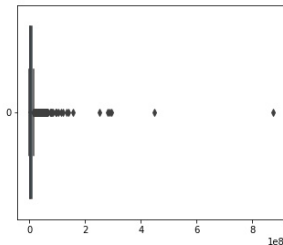
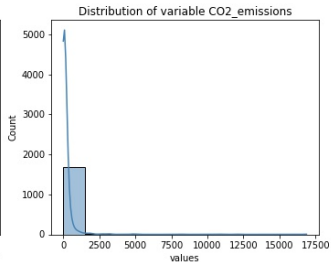
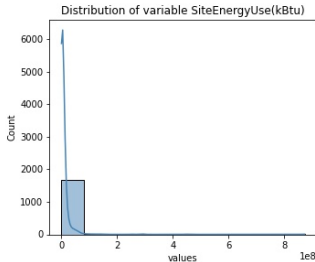
Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression



Variables de sortie

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

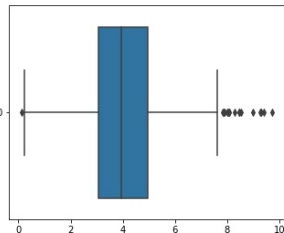
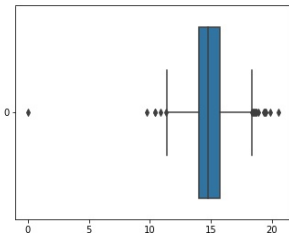
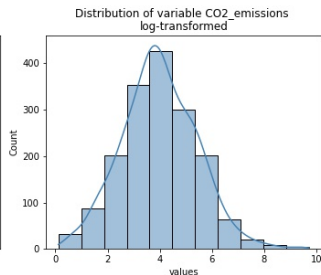
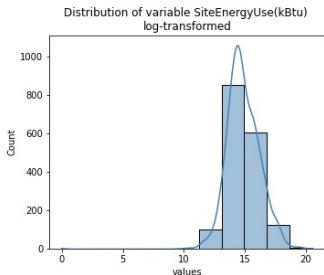
Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

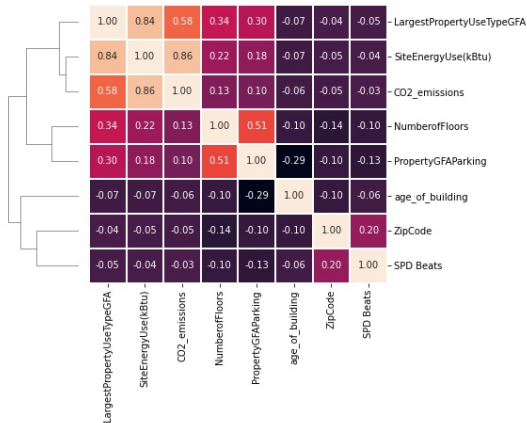
Régression



Matrice des corrélations

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral



La mission

Analyse
exploratoire

Nettoyage des
données

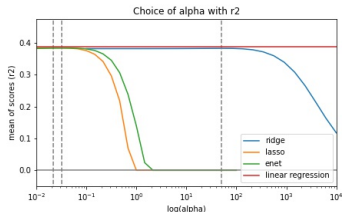
Analyse univariée

Analyse multivariée

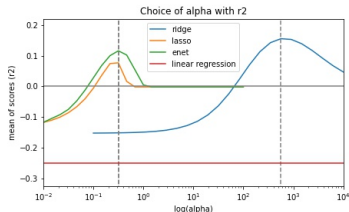
Régression

Régression linéaire - pénalisation

Consommation d'énergie :

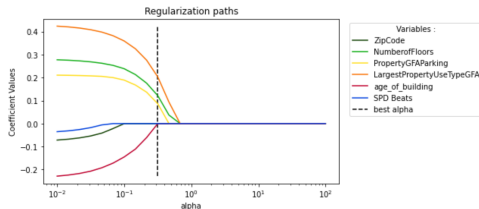


Émission de CO2 :

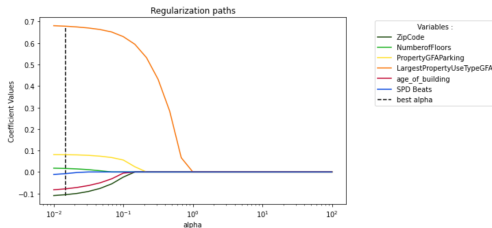


Régression Lasso - Chemins de régularisation

Consommation d'énergie :

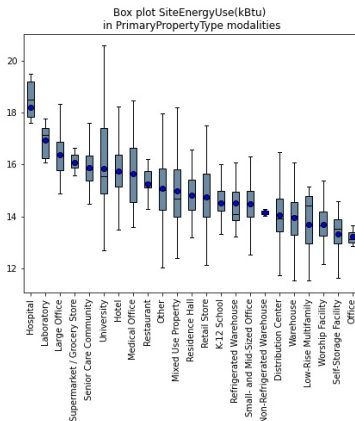


Émission de CO2 :



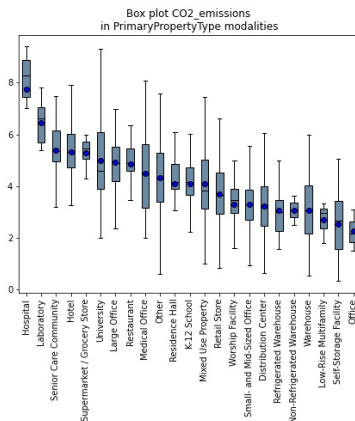
ANOVA - Destination du bâtiment

Consommation d'énergie :



$$\eta^2 = 0.372$$

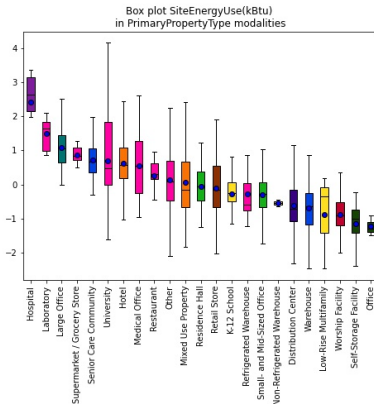
Émission de CO2 :



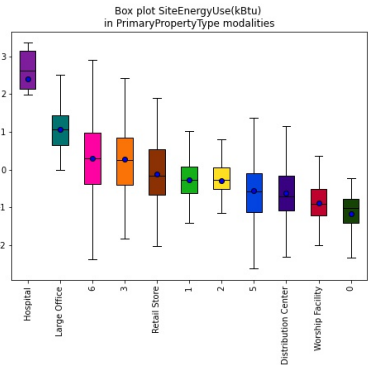
$$\eta^2 = 0.321$$

Regroupement des modalités - clustering semi-supervisé

Consommation d'énergie :



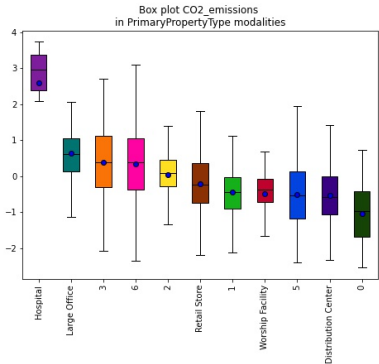
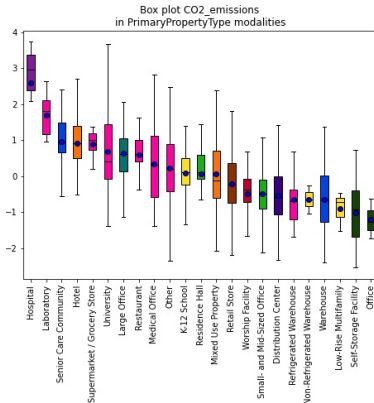
$$\eta^2 = 0.372$$



$$\eta^2 = 0.316$$

Regroupement des modalités - clustering semi-supervisé

Émission de CO2 :

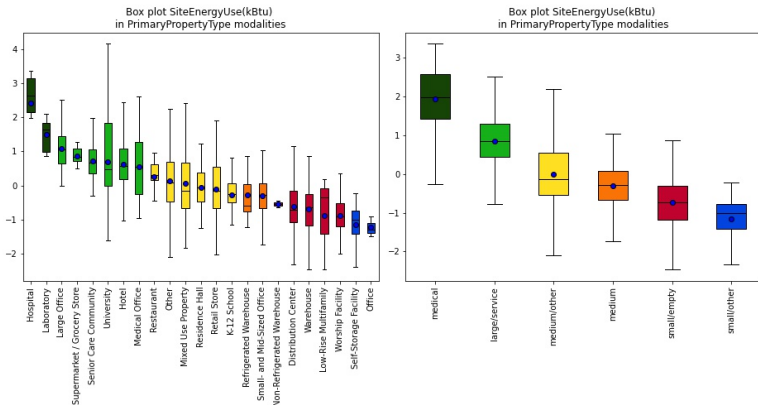


$$\eta^2 = 0.321$$

$$\eta^2 = 0.238$$

Regroupement des modalités - à la main

Consommation d'énergie :

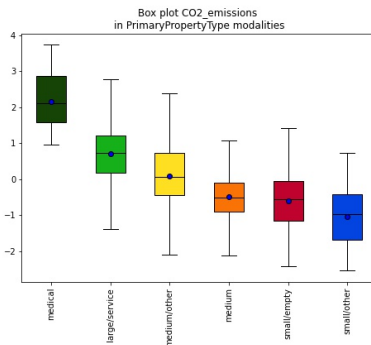
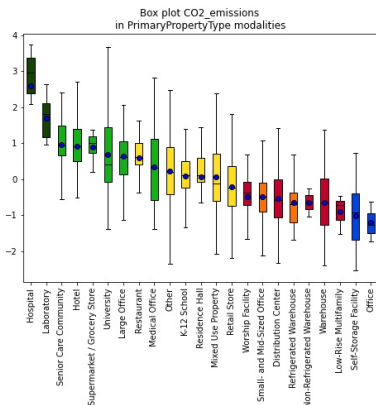


$$\eta^2 = 0.372$$

$$\eta^2 = 0.348$$

Regroupement des modalités - à la main

Émission de CO2 :



$$\eta^2 = 0.321$$

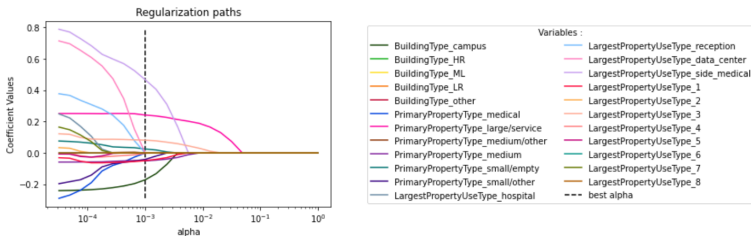
$$\eta^2 = 0.300$$

ANOVA lasso - sélection de modalités

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

6 variables numériques et 23 méta-modalités



→ *Retrait de 14 des 23 méta-modalités*

6 variables numériques et 9 variables binaires (modales)

1 La mission

2 Analyse exploratoire

3 Régression

Modèles de régression

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression

80% : Apprentissage par validation croisée (5 folds)

20% : Test, score prédiction R^2

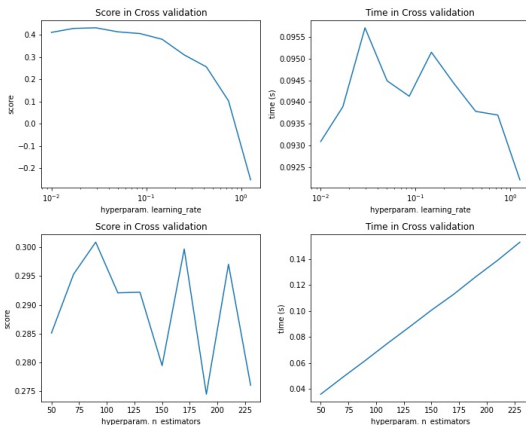
Modèles testés :

- Régression à noyaux
- K plus proches voisins
- Machine à vecteur de support
- Forêts aléatoires
- Gradient Boosting

Hyper-paramètres : nature du noyau, régularisation, nombre de proches voisins, taille de la marges, nombre d'arbres, ...

Résultats validation croisée - exemple 1

- Consommation d'énergie en log
- Régression par renforcement du gradient (Gradient Boosting)

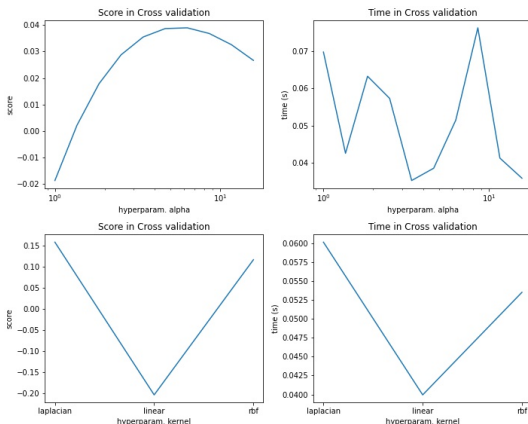


Résultats validation croisée - exemple 1

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

- Émission de CO2
- Régression à noyaux (Kernel Ridge)



Résultats sur l'ensemble test - exemple 2

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression

R^2 score :

	co2	co2_log	energy	energy_log
kernel_ridge	0.212894	0.415290	0.462811	0.386568
KNN	0.520360	0.369095	0.614207	0.510462
SVR	0.188895	0.365437	-0.060280	0.529694
random_forest	0.301204	0.377591	0.416001	0.392492
gradient_boosting	0.069889	0.406552	0.176548	0.463589

Temps d'exécution :

	co2	co2_log	energy	energy_log
kernel_ridge	0.041531	0.040600	0.040638	0.048025
KNN	0.007096	0.007656	0.007869	0.008411
SVR	0.096416	0.058987	0.057608	0.019776
random_forest	0.152543	0.610430	0.344307	0.395365
gradient_boosting	0.042099	0.109928	0.042785	0.108870

Conclusion - Choix de modèle

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression

Émission de CO2 :

- Plus proches voisins (*KNN*)
- $n_neighbors = 12$
- $R^2 = 0.52$, $t = 0.007s$

Consommation d'énergie :

- Plus proches voisins (*KNN*)
- $n_neighbors = 12$
- $R^2 = 0.61$, $t = 0.007s$

Pistes d'amélioration :

- *KNN* \Rightarrow pas d'interprétation de variable
- Corrélation entre variables modales ?
- Classification automatique des modalités

Merci pour votre écoute !

Et l'ENERGY STAR SCORE ?

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

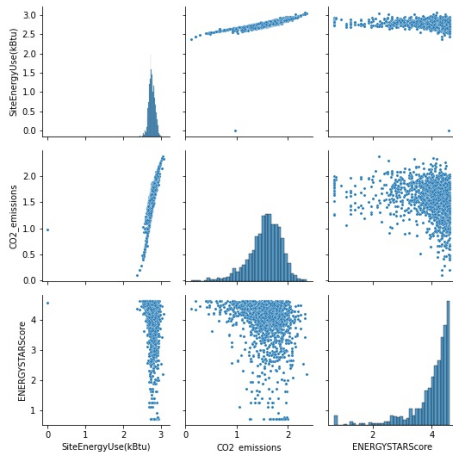
Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression



Et l'ENERGY STAR SCORE ?

Projet 3 :
Anticipez les
besoins en
consommation
électrique de
bâtiments

Claire Gayral

La mission

Analyse
exploratoire

Nettoyage des
données

Analyse univariée

Analyse multivariée

Régression

Elastic-net regression : $\text{EnergyStar score} = \text{CO2} + \text{Energy}$

