

**OpenClassrooms**

**Projet P4:**

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

N. Alves

# Anticipez les besoins en consommation électrique de bâtiments

## Sommaire

- Contexte
- Présentation des informations disponibles
- Étude du jeu de données
- Analyse
- Modélisation
- Conclusion et perspectives

# Anticipez les besoins en consommation électrique de bâtiments

- Contexte
- Présentation des informations disponibles
- Étude du jeu de données
- Analyse
- Modélisation
- Conclusion et perspectives

# Anticipez les besoins en consommation électrique de bâtiments

## Contexte

Dans le cadre d'un plan action climat, Seattle a mis en place une base de donnée répertoriant les bâtiments de la ville afin d'en comparer leur performance énergétique.

⇒ On souhaite estimer les niveaux de gaz à effet de serre et de consommation d'énergie de nouveaux bâtiments. Afin de limiter les relevés coûteux, on créera des modèles numériques basés sur les données existantes.



Seattle

# Anticipez les besoins en consommation électrique de bâtiments

- Contexte
- **Présentation des informations disponibles**
- Étude du jeu de données
- Analyse
- Modélisation
- Conclusion et perspectives

# Anticipez les besoins en consommation électrique de bâtiments

## Présentation des informations disponibles

Deux fichiers au format ‘csv’, correspondant aux années 2015 et 2016, sont disponibles.

- Le fichier de 2015 : 3340 lignes x 46 colonnes,
- Le fichier de 2016 : 3376 lignes x 47 colonnes.

Chaque ligne représente un bâtiment et chaque colonne une caractéristique de ce bâtiment.

- Ces caractéristiques se déclinent sur différents thèmes et sur différents niveaux de précision:  
Surface, type d’usage, localisation, consommation d’énergie, émissions de gaz à effets de serre etc.

# Anticipez les besoins en consommation électrique de bâtiments

- Contexte
- Présentation des informations disponibles
- **Étude du jeu de données**
- Analyse
- Modélisation
- Conclusion et perspectives

# Anticipez les besoins en consommation électrique de bâtiments

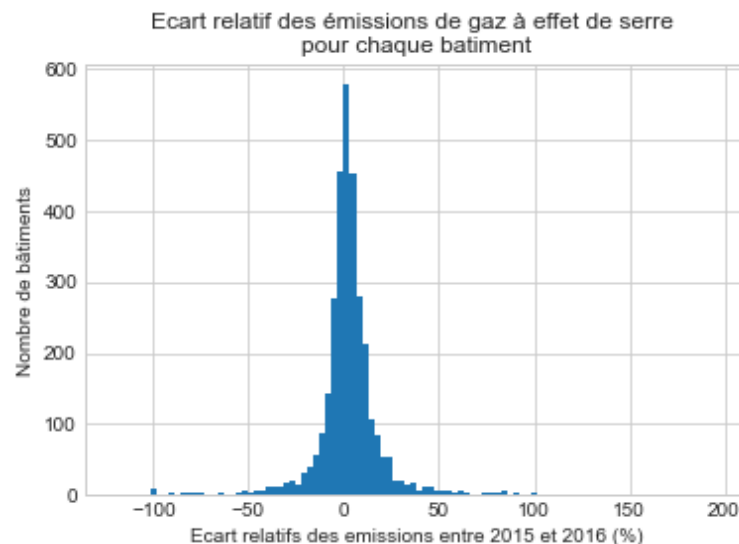
## Étude du jeu de données

### Création du jeu de donnée :

On fusionne les deux fichiers en un seul jeu de donnée pour enrichir les observations qui vont alimenter les modèles.

- On conserve tous les bâtiments de 2016 et les bâtiments uniquement présents dans le fichier de 2015.
- Si un bâtiment est présent sur les deux années, on étudie les évolutions de ses émissions de gaz à effet de serre. Si elles ont fortement varié (de 20 % à 100 %), on conserve alors aussi les observations de 2015. On le considérera comme deux bâtiments distincts dans les modèles.

Ces fortes variations peuvent correspondre par exemple à une exploitation différente du bâtiment ou à des travaux de rénovation.





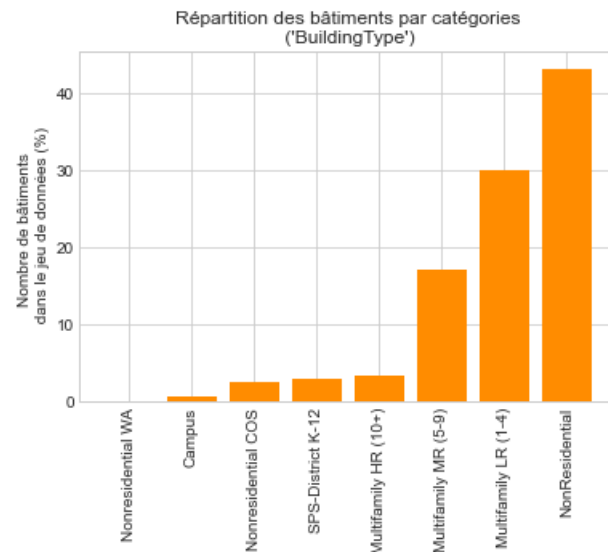
# Anticipez les besoins en consommation électrique de bâtiments

## Étude du jeu de données

### Traitement des lignes :

- Selection de lignes: on effectue un filtrage par type de bâtiments. L'étude concerne uniquement les bâtiments à usage non résidentiel.
- Les hôpitaux ne sont pas pris en compte dans cette étude.
- Valeurs aberrantes :

Si la somme des consommations par type d'énergie ne couvrent pas 100 % de la consommation totale d'énergie, on enlève le bâtiment correspondant du jeu de donnée.



# Anticipez les besoins en consommation électrique de bâtiments

## Étude du jeu de données

### Traitement des colonnes

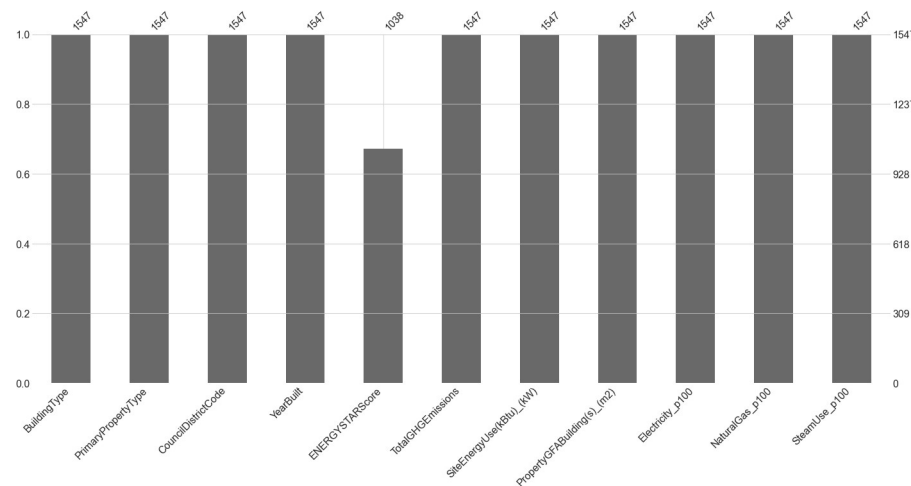
- Selection de colonnes : on effectue un filtrage par thème retenu :

Lieu, surface, type d'utilisation, date de construction.

- Création de colonnes :

On ajoute trois nouvelles variables représentant la nature de l'énergie et sa proportion dans l'énergie totale utilisée : proportion d'énergie sous forme de gaz, de vapeur et d'électricité.

Taux de remplissage par colonne du jeu de donnée final



⇒ Après traitement, le jeu de donnée comporte 1500 lignes et 11 colonnes.

# Anticipez les besoins en consommation électrique de bâtiments

- Contexte
- Présentation des informations disponibles
- Étude du jeu de données
- **Analyse**
- Modélisation
- Conclusion et perspectives

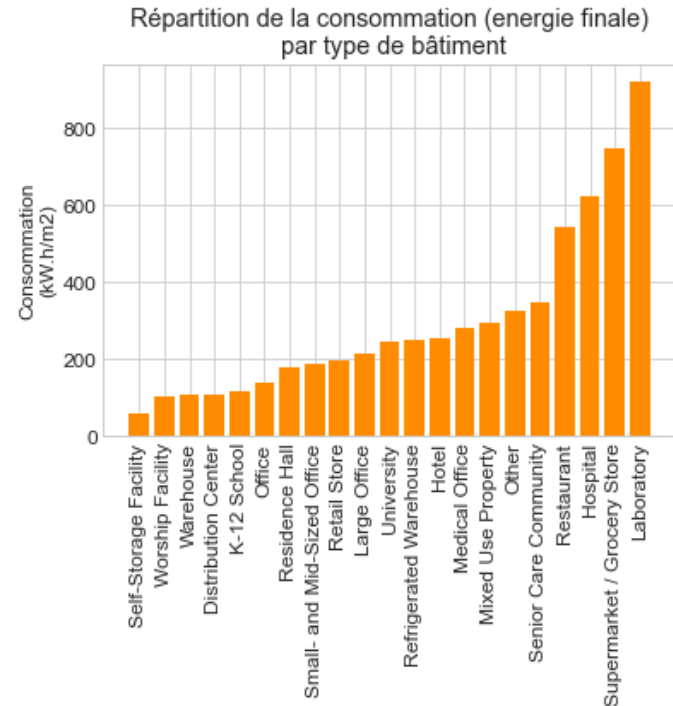
# Anticipez les besoins en consommation électrique de bâtiments

## Analyse

### Consommation d'énergie finale<sup>(\*)</sup>

- Les consommations varient selon le type de bâtiment de 0 à 900 kW.h/m<sup>2</sup> sur une année.

(\*) L'énergie finale est l'énergie utilisée par le bâtiment directement.

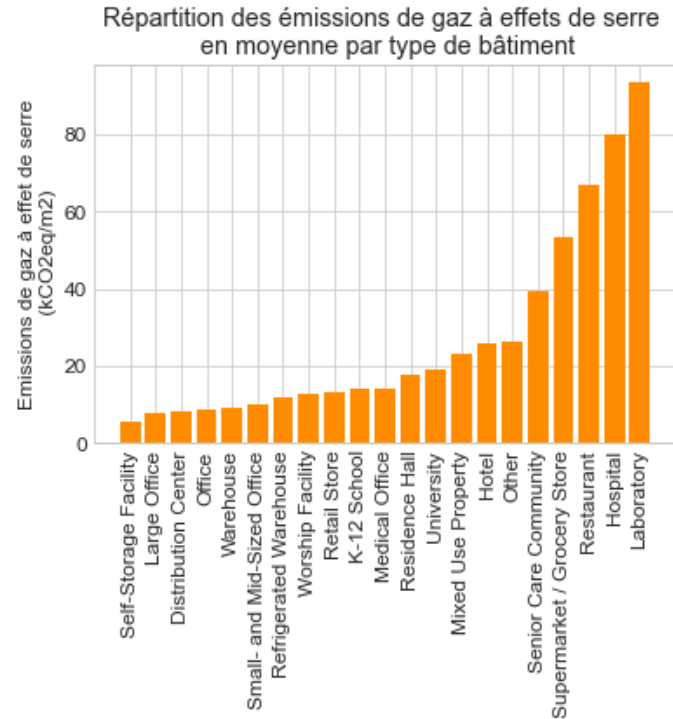


# Anticipez les besoins en consommation électrique de bâtiments

## Analyse

### Emissions de gaz à effet de serre

- Les émissions varient de 0 à 90 kg CO<sub>2</sub> eq/m<sup>2</sup> sur une année.
- Les laboratoires et hôpitaux sont les bâtiments les plus émissifs.



# Anticipez les besoins en consommation électrique de bâtiments

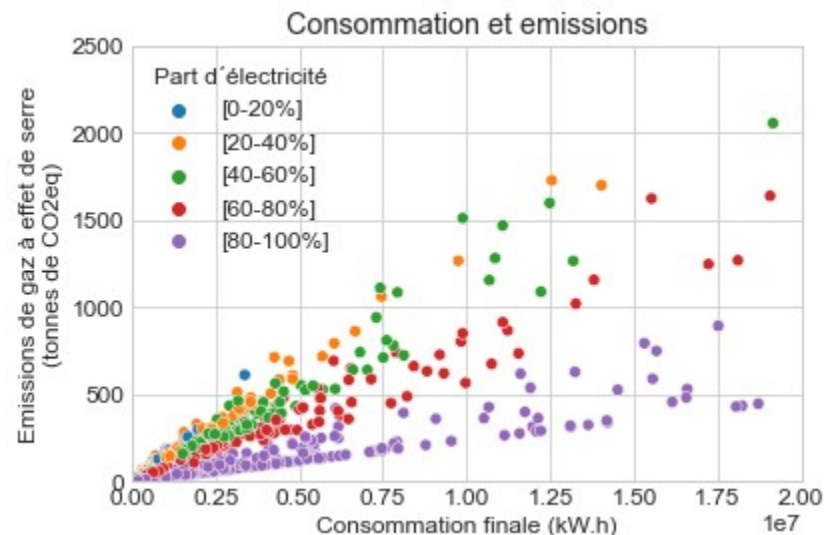
## Analyse

### Consommation et émissions:

- La consommation d'énergie et les émissions de gaz à effet de serre sont liés.
- Elles dépendent également des types et proportions d'énergie utilisées.

⇒ Pour une même consommation d'énergie :

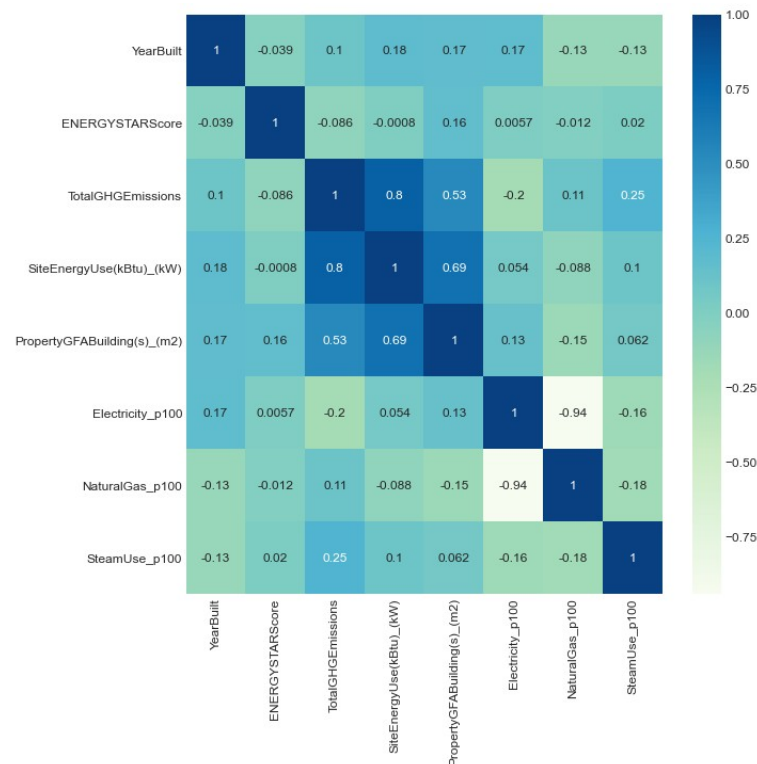
Plus la part d'électricité est importante, plus les émissions sont faibles.



# Anticipez les besoins en consommation électrique de bâtiments

## Analyse

- Consommation d'électricité et de gaz sont anti-corrélées
- Emissions et consommation sont liées
- Consommation et surface augmente dans la même direction
- Pas de corrélation saillante entre les autres variables.



Matrice de corrélation

# Anticipez les besoins en consommation électrique de bâtiments

- Contexte
- Présentation des informations disponibles
- Étude du jeu de données
- Analyse
- **Modélisation**
- Conclusion et perspectives



# Anticipez les besoins en consommation électrique de bâtiments

## Modélisation

- Méthode

- Séparation du jeu de donnée en un jeu d'entraînement et un jeu de test,
- Encodage des variables catégorielles,
- Standardisation des variables numériques,
- Passage au logarithme des cibles.
- Optimisation des hyperparamètres et validation croisée,
- Mesures récupérées:
  - $R^2$ ,
  - valeur absolue des écarts,
  - erreur relative (moyenne et écart type)

- Choix de modèles

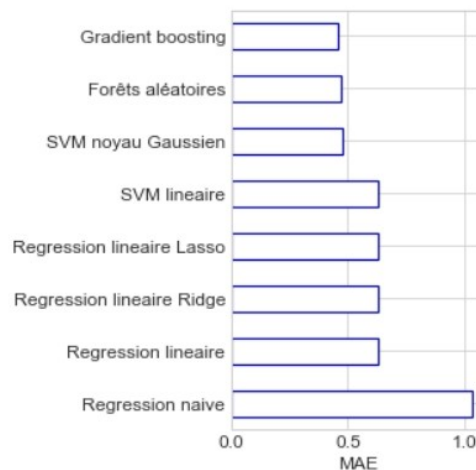
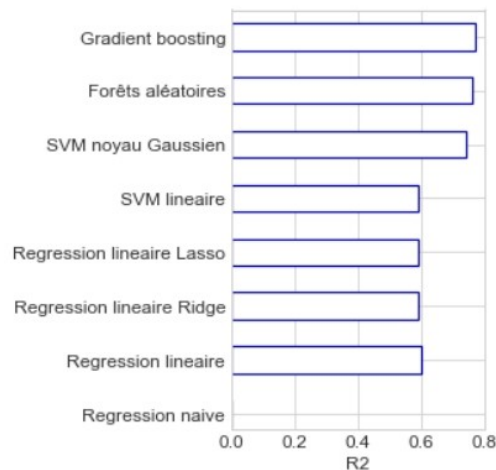
Les variables à simuler sont continues, le choix s'est porté sur des modèles de régression linéaires et non-linéaires

# Anticipez les besoins en consommation électrique de bâtiments

## Modélisation

### Résultat du modèle de consommation

Les algorithmes non linéaires et ensemblistes offrent les meilleurs résultats. Le choix se porte sur le modèle 'Boosting du gradient'. Il y a un écart entre les modèles linéaires et non-linéaires.



Calcul basé sur la cible  
 $\ln(\text{consommation})$

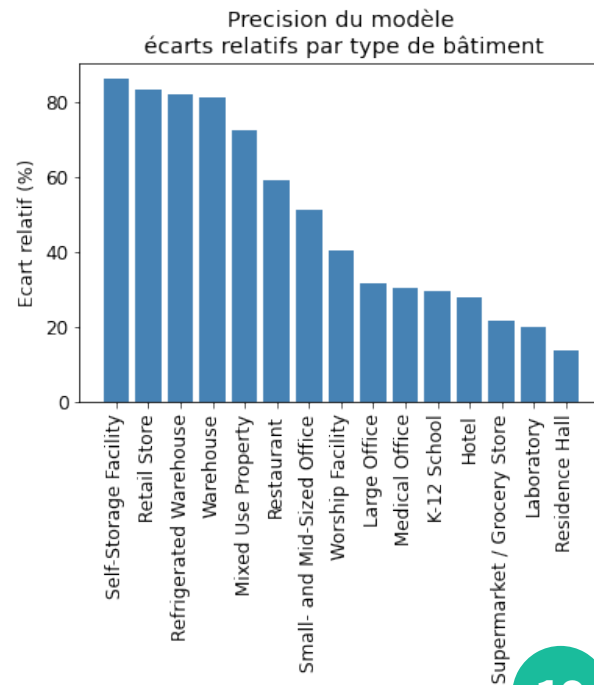
# Anticipez les besoins en consommation électrique de bâtiments

## Modélisation

### Résultat du modèle de consommation :

L'algorithme 'boosting du gradient' offre les meilleurs résultats :

- La précision varie de  $\pm 20\%$  à  $\pm 80\%$  suivant les bâtiments.

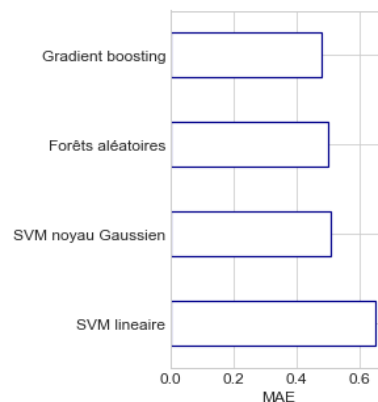
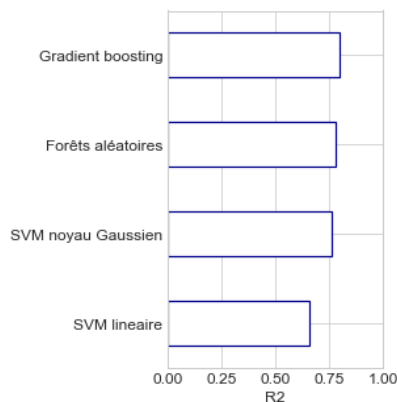


# Anticipez les besoins en consommation électrique de bâtiments

## Modélisation

### Résultat du modèle d'émission :

Pour évaluer l'influence de la variable EnergyStar Score, on réalise deux modèles : avec et sans cette variable. L'algorithme 'boosting du gradient' offre aussi les meilleurs résultats.



Avec EnergyStar Score

# Anticipez les besoins en consommation électrique de bâtiments

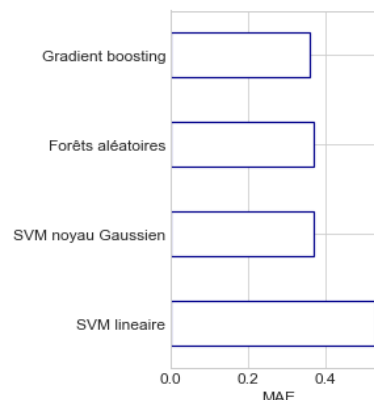
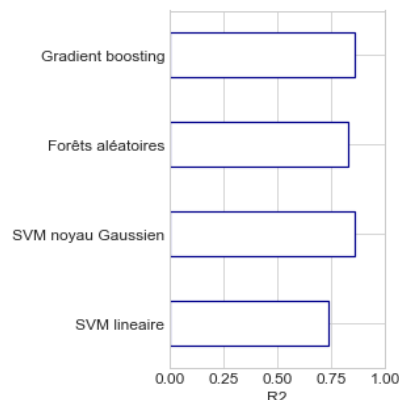
## Modélisation

### Résultat du modèle d'émission :

L'algorithme 'boosting du gradient' offre toujours les meilleurs résultats.

La présence de la variable EnergyStar score améliore la qualité du modèle et réduit les erreurs :

- Variation de R2 :  $\sim 0,1$  point
- Variation de la moyenne des écarts relatifs  $\sim 7$  %.



Sans EnergyStar Score

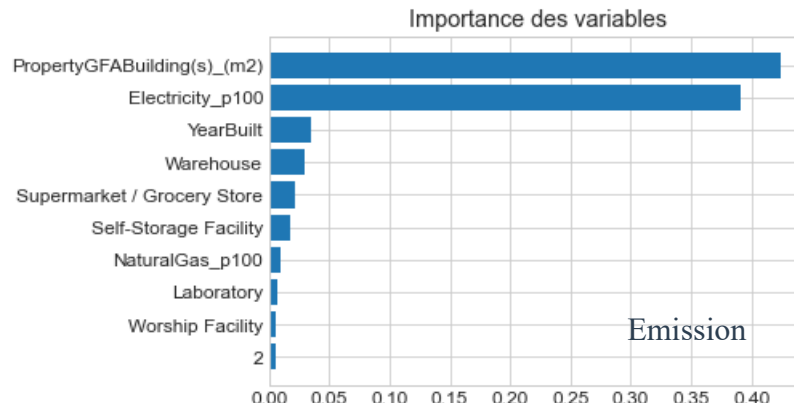
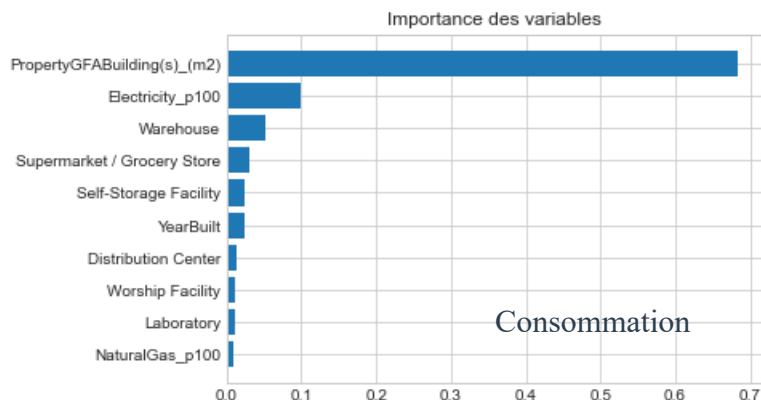
# Anticipez les besoins en consommation électrique de bâtiments

## Modélisation

### Importance des variables des modèles:

Les variables importantes sont la surface, la part d'électricité dans la consommation et le type de bâtiments (entrepôt, supermarché etc.).

- Il convient de limiter la surface pour réduire la consommation et les émissions.
- Le type/proportion d'énergie est important pour le niveaux d'émission de CO<sub>2</sub> ainsi que la performance énergétique des bâtiments.



# Anticipez les besoins en consommation électrique de bâtiments

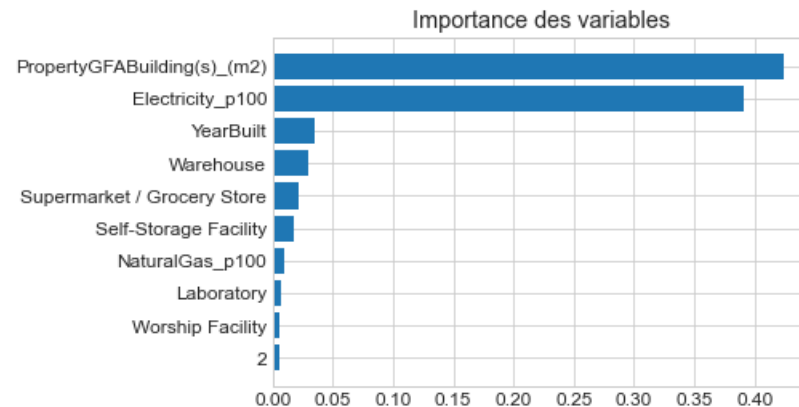
- Contexte
- Présentation des informations disponibles
- Étude du jeu de données
- Analyse
- Modélisation
- Conclusion et perspectives

# Anticipez les besoins en consommation électrique de bâtiments

## Conclusion et perspectives

L'algorithme retenu pour les modèles de consommation et d'émissions de gaz à effet de serre est le 'boosting du gradient'. Cet algorithme offre la meilleure précision avec un temps de calcul comparable aux autres algorithmes. L'influence de l'EnergyStar score dans le calcul des émissions est réel mais limité. Il peut ne pas être utilisé pour des premières estimations.

On a réalisé ici un premier modèle totalisant l'ensemble des comportements. Pour en améliorer la précision, il peut être intéressant de le scinder en plusieurs modèles spécifiques à des distributions plus restreintes. Enfin, pour inciter à certaines rénovations, une perspective d'étude intéressante, serait de pouvoir estimer l'impact de certains travaux ou du choix du type d'énergie sur les émissions polluantes.





# Annexe

Référence :

Page web du département energie à Seattle : [Energy benchmarking](#)



**Merci de votre attention !**