

# Anticiper les besoins en consommation électrique de bâtiments

21/07/2021 - Parcours Data Scientist  
Sébastien Bourgeois

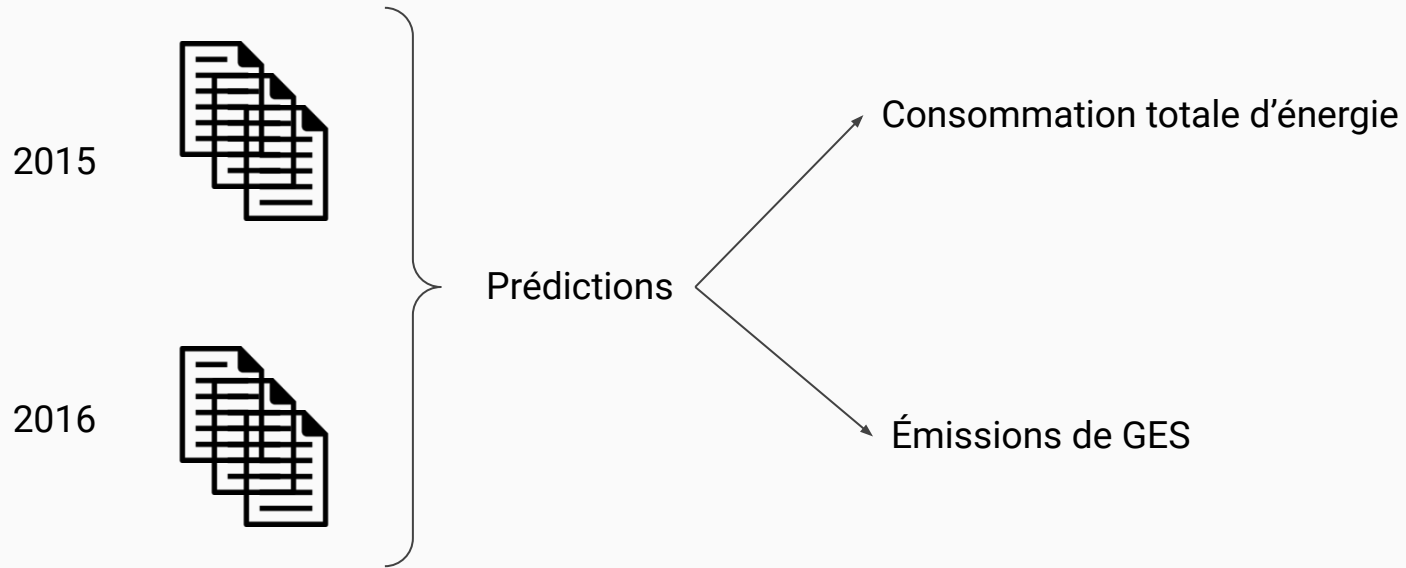


# Sommaire

1. Problématique
2. Nettoyage & exploration
3. Pistes de modélisation
4. Modèle final sélectionné

# 1. Problématique

*Problématique*



# 1. Problématique

*Problématique*



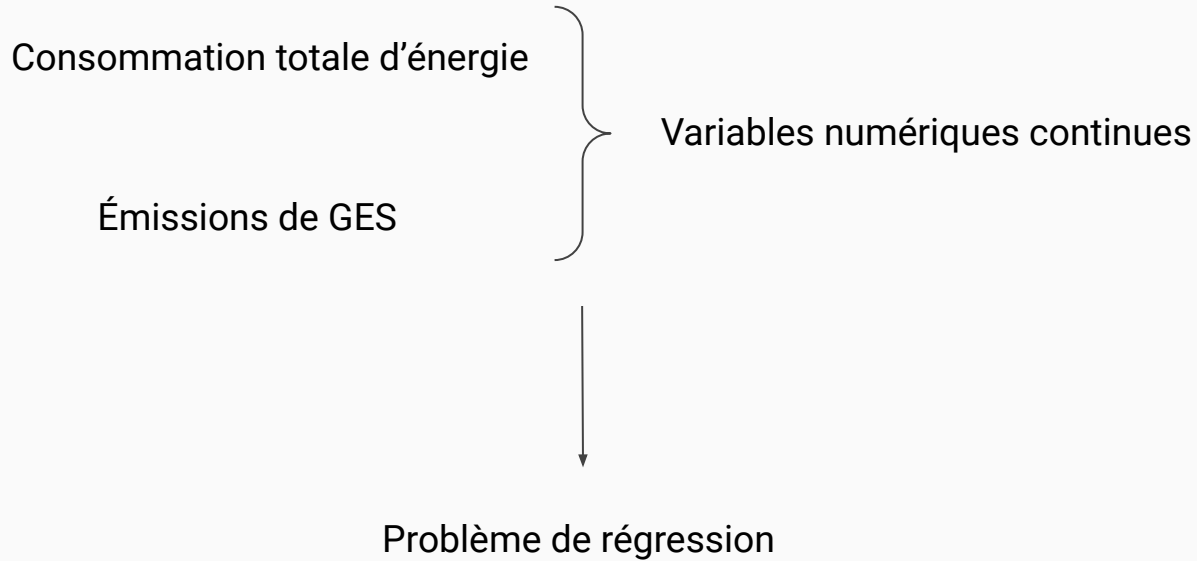
Évaluation



Prédictions des émissions de GES

# 1. Problématique

## *Interprétation*



# 1. Problématique

*Pistes de recherche envisagées*

Régression linéaire

Modèles non linéaires

Autres modèles linéaires

Modèles ensemblistes

## 2. Nettoyage & exploration

### *Nettoyage des données*



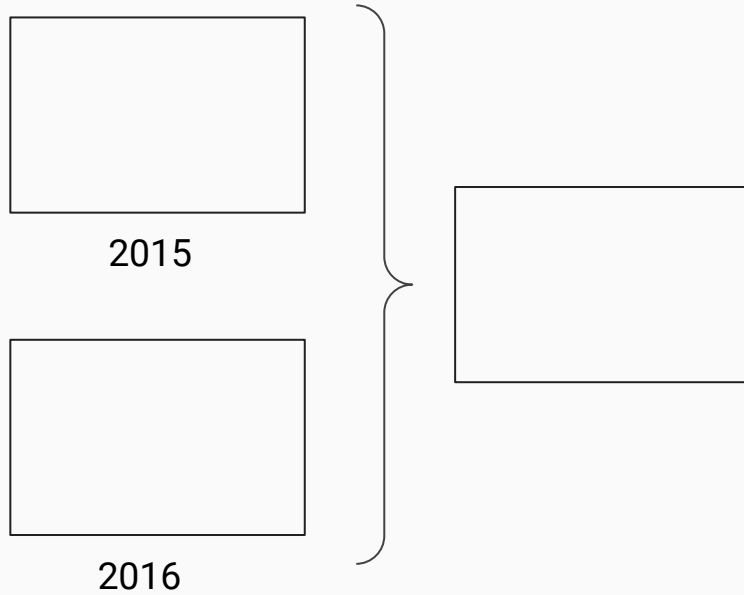
2015



2016

## 2. Nettoyage & exploration

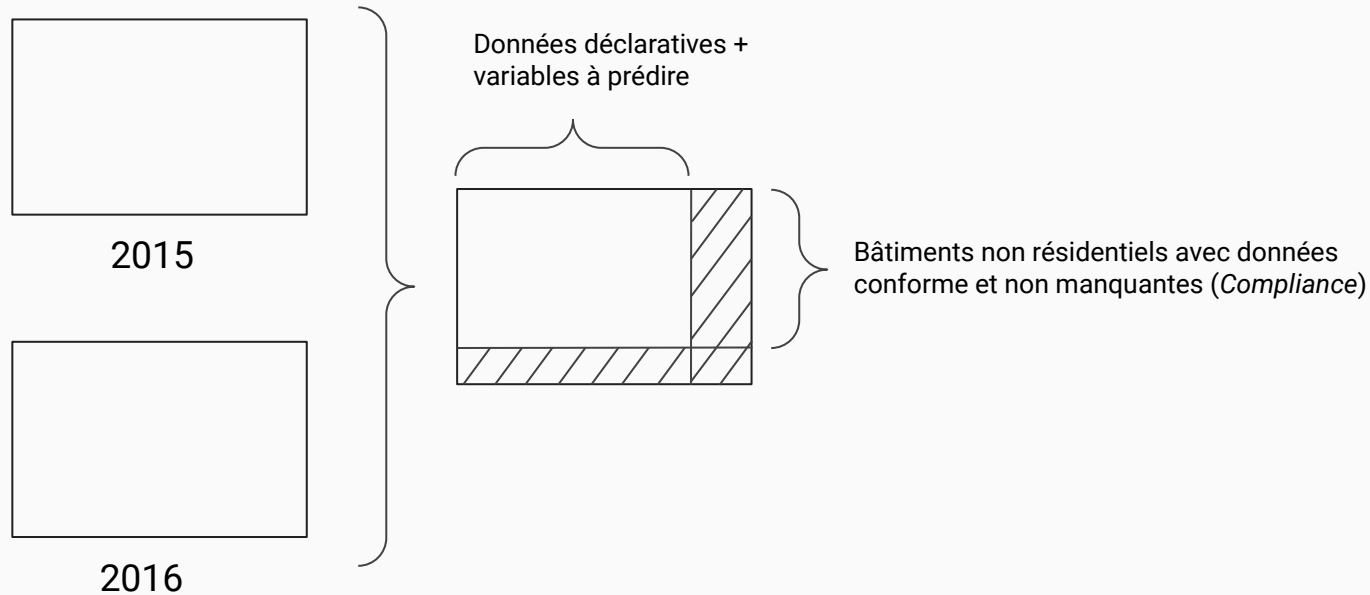
*Nettoyage des données*





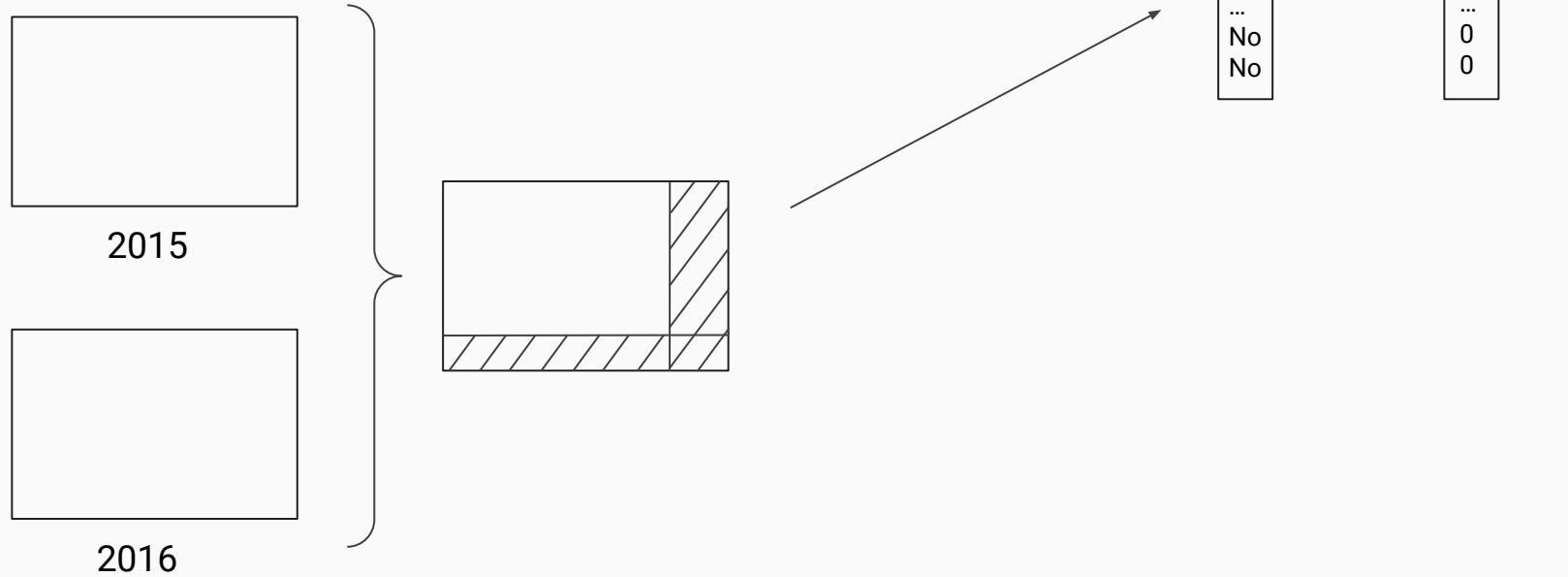
## 2. Nettoyage & exploration

### *Nettoyage des données*



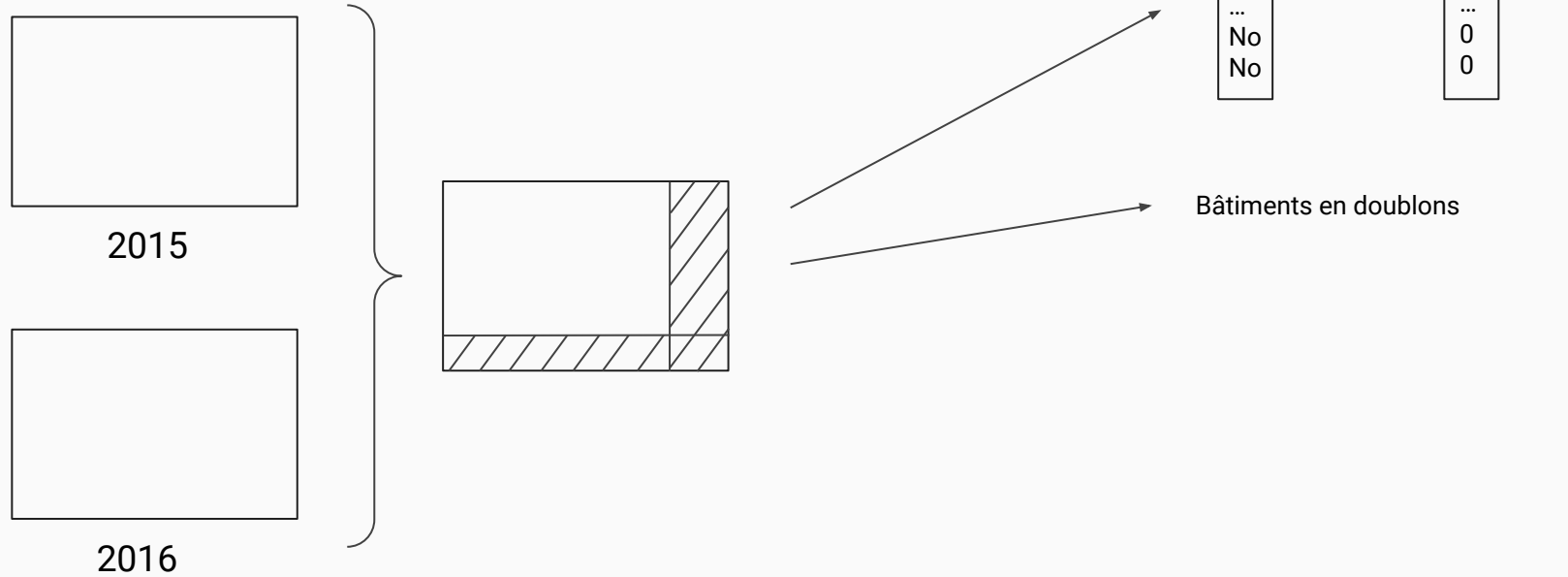
## 2. Nettoyage & exploration

### *Nettoyage des données*



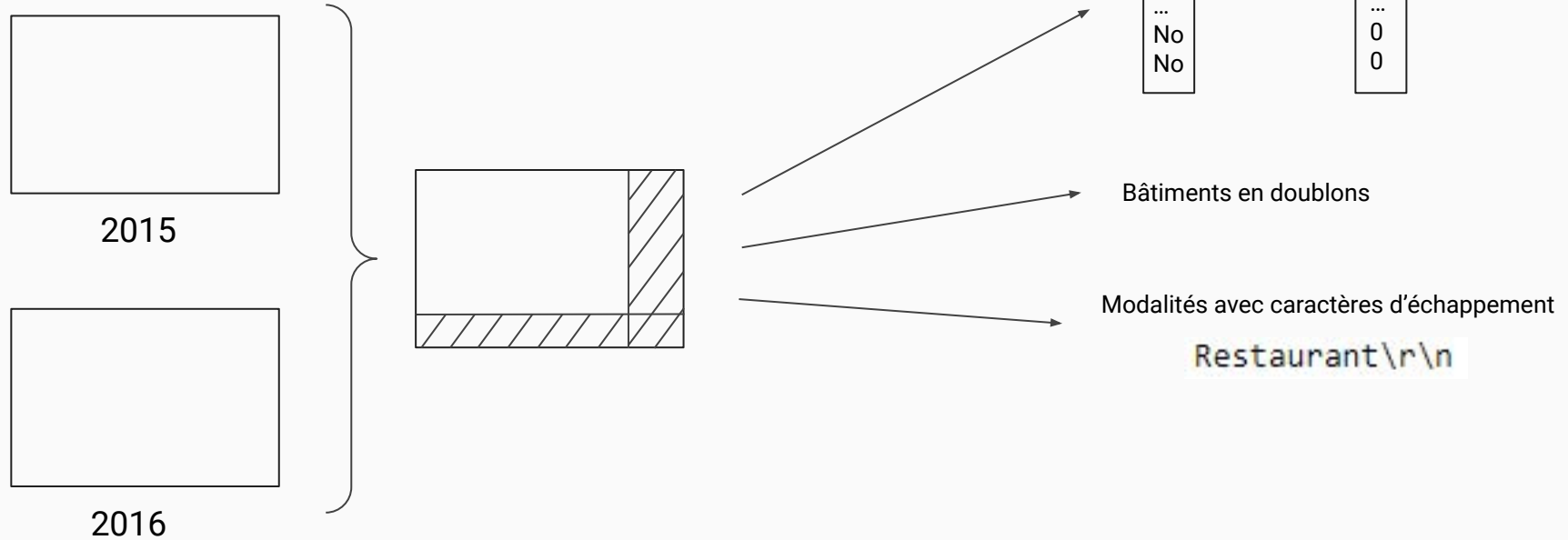
## 2. Nettoyage & exploration

### *Nettoyage des données*



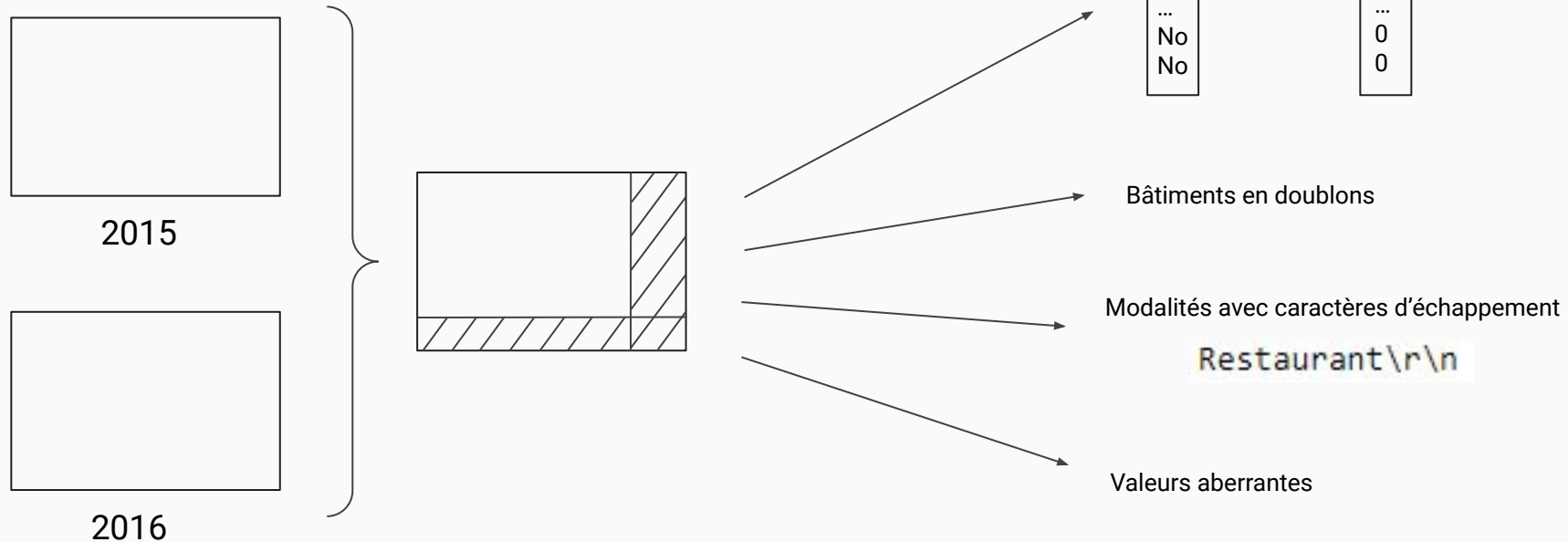
## 2. Nettoyage & exploration

### *Nettoyage des données*



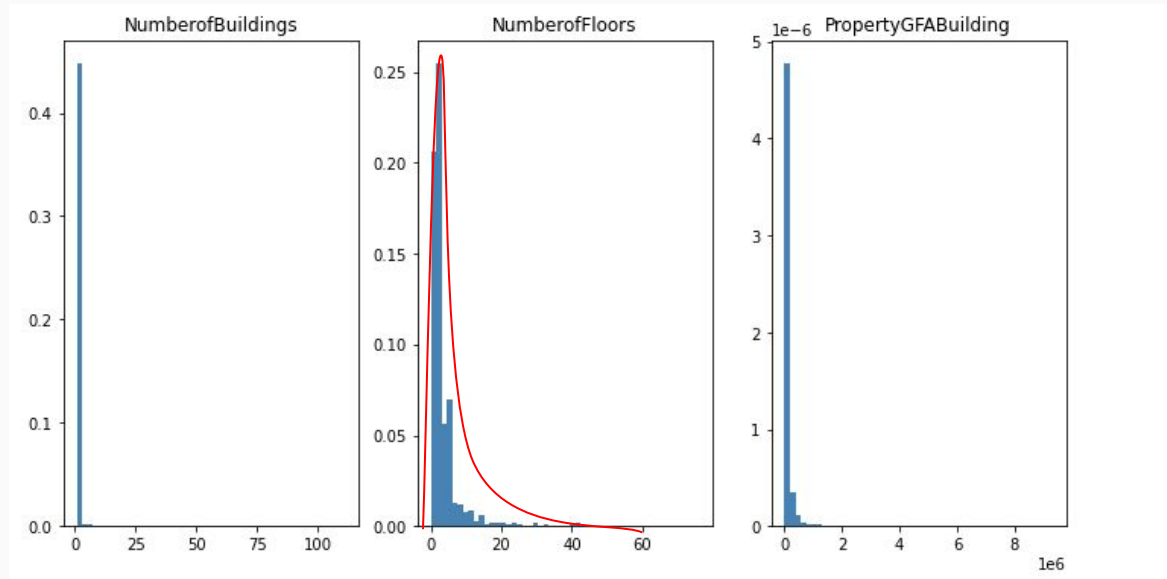
## 2. Nettoyage & exploration

### *Nettoyage des données*



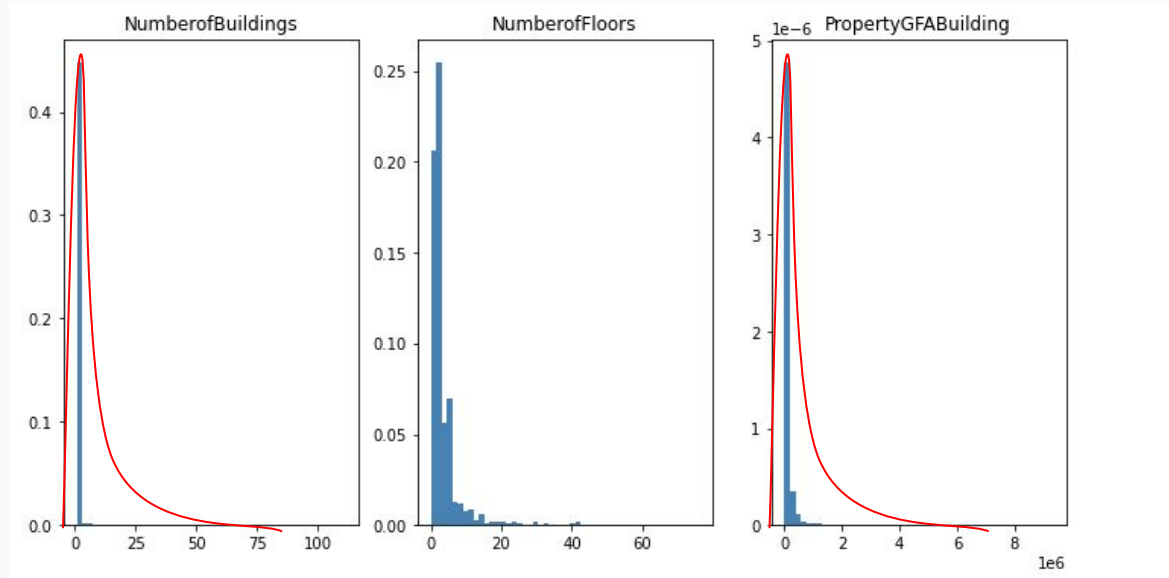
## 2. Nettoyage & exploration

### *Exploration*



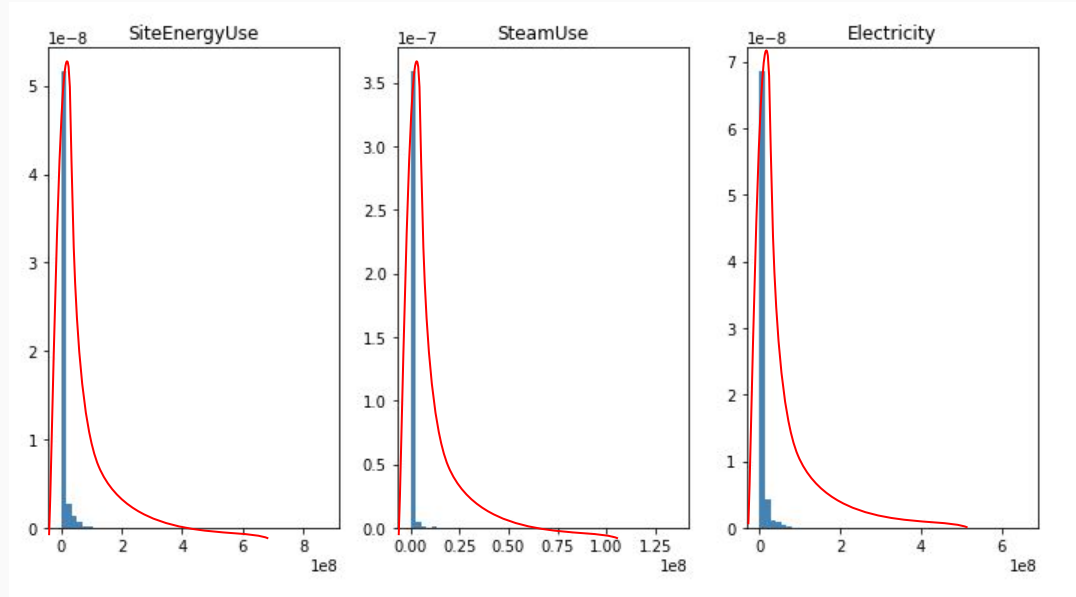
## 2. Nettoyage & exploration

### *Exploration*



## 2. Nettoyage & exploration

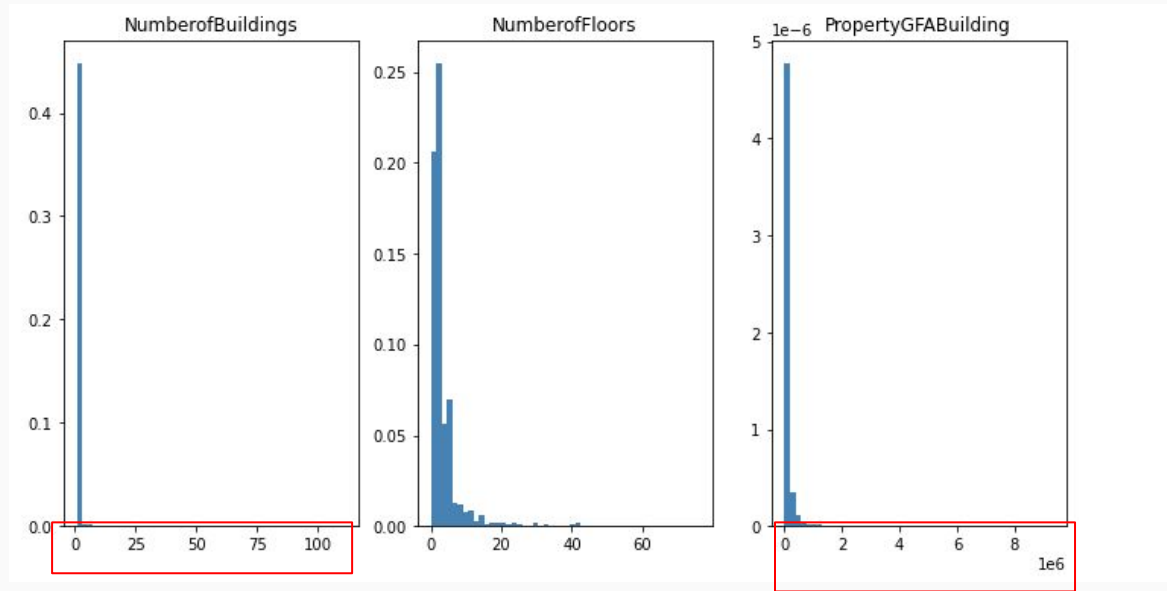
### *Exploration*





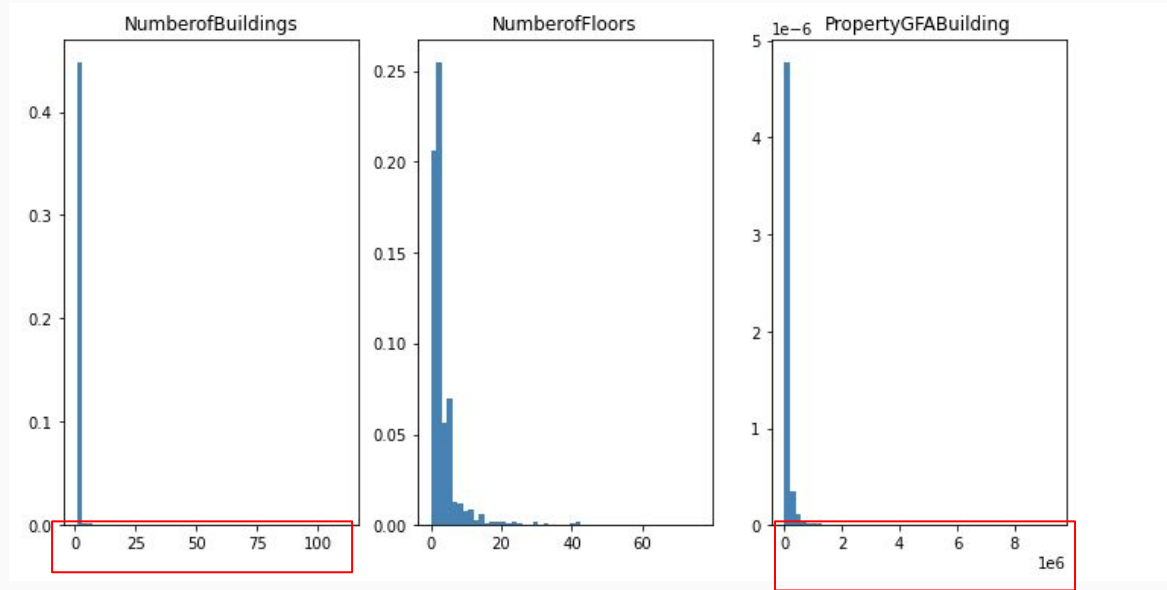
## 2. Nettoyage & exploration

### *Exploration*



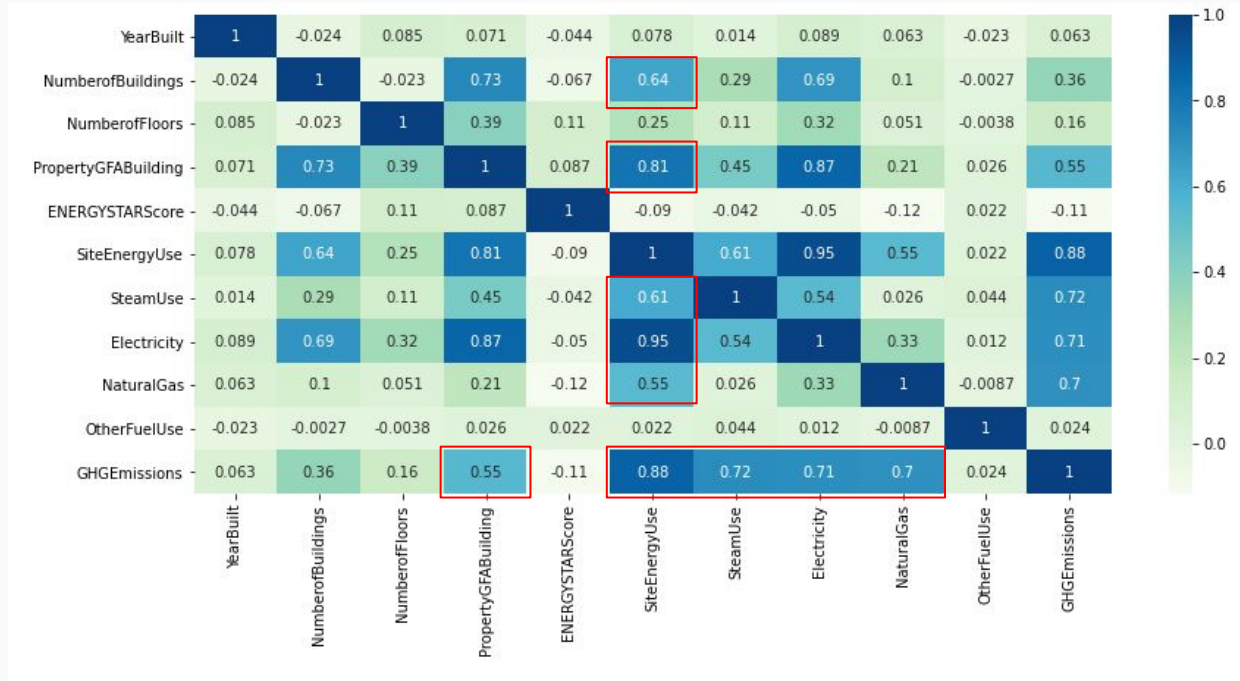
## 2. Nettoyage & exploration

### *Exploration*



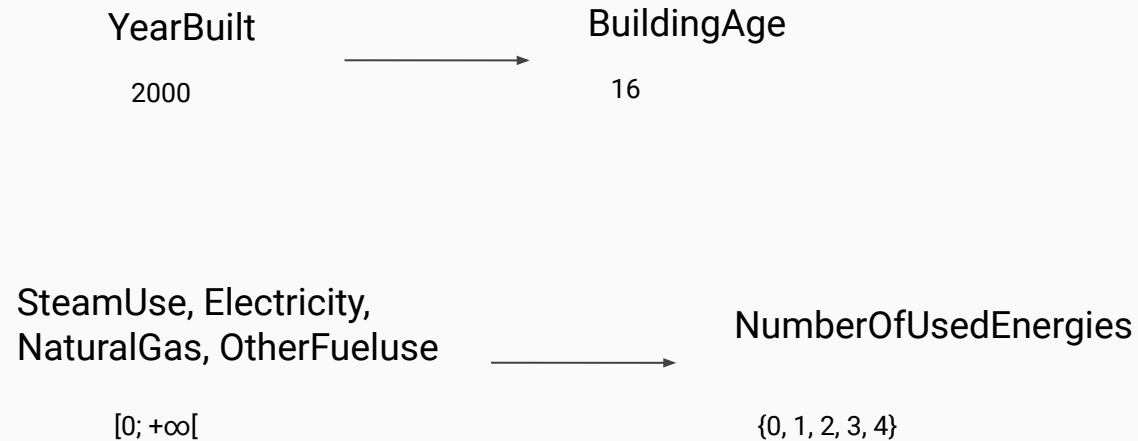
## 2. Nettoyage & exploration

### Exploration



## 2. Nettoyage & exploration

### *Features engineering*



## 2. Nettoyage & exploration

### *Features engineering*

Features & variables cibles —————> Standardisation & passage au log

PrimaryPropertyType  
"Small- and Mid-Sized Office"  
"Large Office" —————> GatheredPrimaryPropertyType  
"Office"

## 2. Nettoyage & exploration

### *Features engineering*

#### Imputation

NumberOfFloors → Mode

PropertyGFABuilding → Médiane

#### Encodage variables catégorielles

BuildingType → BuildingType\_Campus  
BuildingType\_NonResidential  
BuildingType\_NonResidential\_COS  
BuildingType\_SPS-District\_K-12

## 2. Nettoyage & exploration

### *Features engineering*

#### Imputation

NumberOfFloors → Mode

PropertyGFABuilding → Médiane

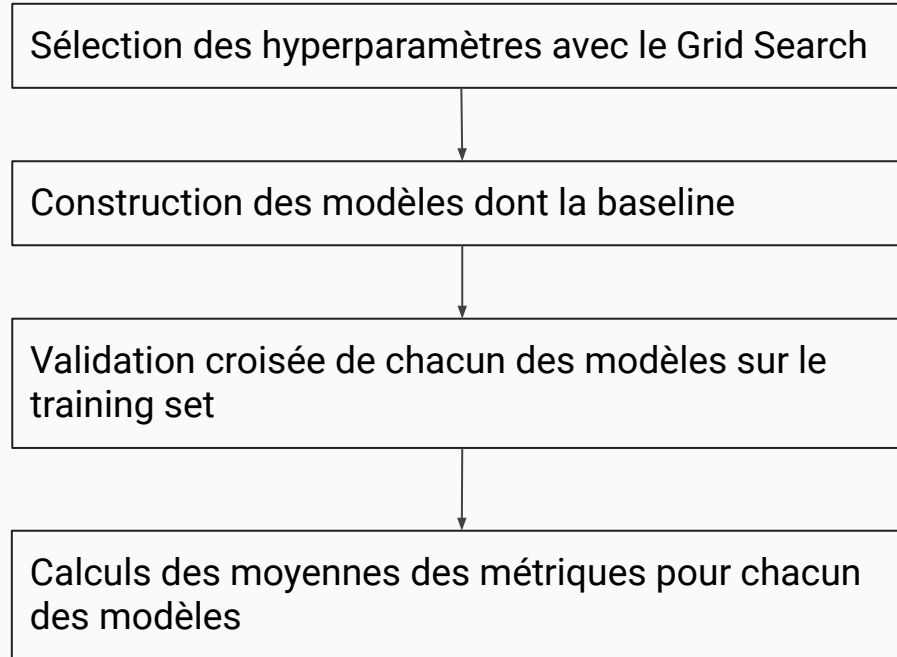
#### Encodage variables catégorielles

BuildingType → BuildingType\_Campus  
BuildingType\_NonResidential  
BuildingType\_NonResidential\_COS  
BuildingType\_SPS-District\_K-12

0
1
0
...
0
1
1
0

### 3. Pistes de modélisation de la consommation d'énergie

#### *Méthodologie*





### 3. Pistes de modélisation de la consommation d'énergie

#### *Modèles simples*

- 1    Baseline        Prédiction par la moyenne
- 2    Régression linéaire classique

### 3. Pistes de modélisation de la consommation d'énergie

*Modèles plus complexes*

#### 3 Elastic Net

Gridsearch → alpha & l1\_ratio

```
ElasticNet(alpha=0.5455594781168515, l1_ratio=0)
```

### 3. Pistes de modélisation de la consommation d'énergie

#### *Modèles plus complexes*

#### 4 Random Forest

Gridsearch



- profondeur maximum
- nombre maximum de features
- nombre d'arbres

```
RandomForestRegressor(max_depth=20, max_features=10, n_estimators=250)
```

### 3. Pistes de modélisation de la consommation d'énergie

#### *Modèles plus complexes*

#### 5 Gradient Boosting

Gridsearch



- learning rate
- profondeur maximum
- nombre d'arbres

```
GradientBoostingRegressor(learning_rate=0.025, max_depth=4, n_estimators=400)
```

### 3. Pistes de modélisation de la consommation d'énergie

#### Résultats

	R2	MAE	RMSE	Temps
<b>baseline</b>	-0.0572	7.31694e+06	2.80556e+07	0.00559549
<b>regression_lineaire</b>	-1.22419e+27	4.44761e+19	8.29692e+20	0.291057
<b>elastic_net</b>	-1.45934e+06	1.54112e+09	2.86614e+10	0.152254
<b>gradient_boosting</b>	0.5438	4.18717e+06	1.9683e+07	1.8827
<b>random_forest</b>	0.4696	4.21628e+06	2.13886e+07	2.76984

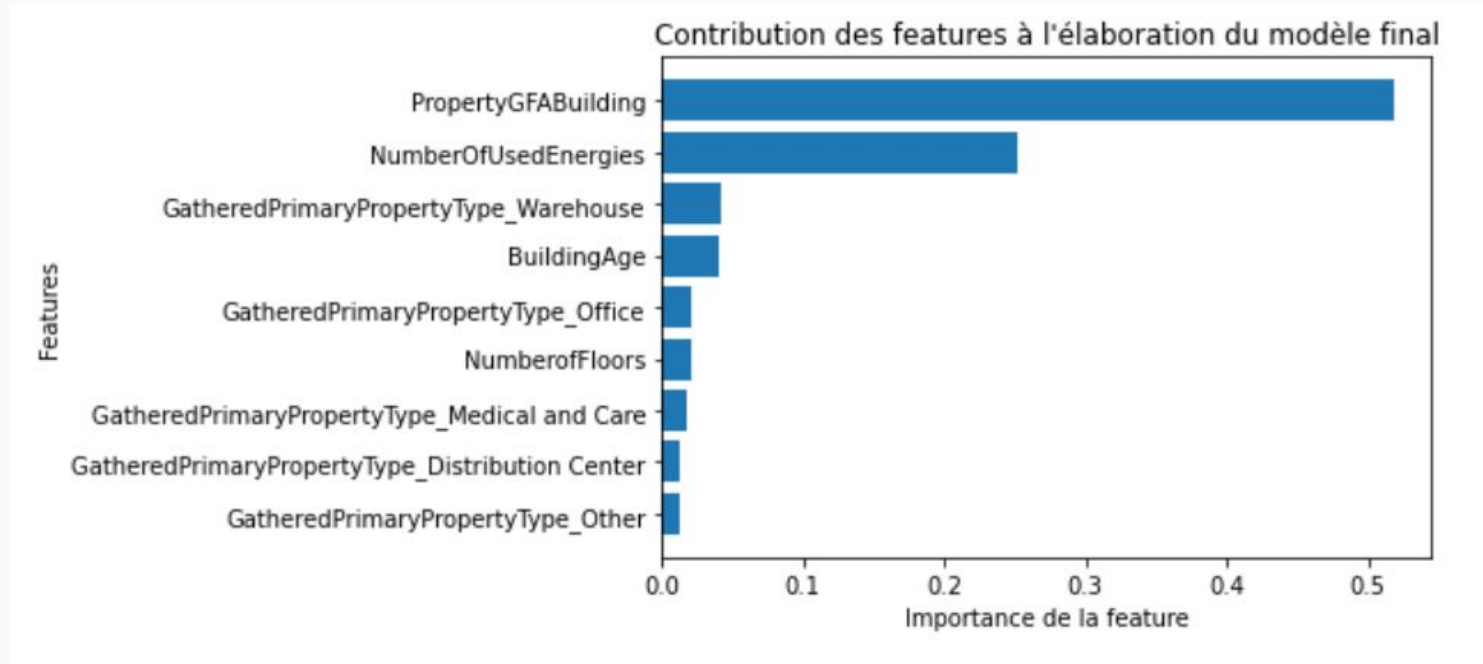
## 4. Modèle final sélectionné

*Gradient Boosting : résultats sur le jeu de test*

	Consommation d'énergie	Émissions de GES
$R^2$	0.847	0.721
MAE	3 251 609.56	99.18
RMSE	7 877 011.31	424.86

## 4. Modèle final sélectionné

*Gradient Boosting : features importance*



## 4. Modèle final sélectionné

*Gradient Boosting : pertinence de l'ajout de l'Energy Star Score*

	Émissions de GES Sans Energy Star Score	Émissions de GES Avec Energy Star Score
R <sup>2</sup>	0.721	0.905
MAE	99.18	79.20
RMSE	424.86	239.90



## 4. Modèle final sélectionné

*Gradient Boosting : pertinence de l'ajout de l'Energy Star Score*

	Gain de l'ajout de l'Energy Star Score	} 30% en moyenne
R <sup>2</sup>	26%	
MAE	20%	
RMSE	44%	

Conclusion : en ajoutant l'Energy Star Score dans le modèle, ses performances sont améliorées de 30% en moyenne. Cependant, étant donné la complexité de calculer cette donnée et le fait que le modèle initial avait déjà de bonnes performances. Il n'est pas recommandé d'utiliser cette variable.

# Questions/Réponses

Fin.