Anticipez les besoins en consommation de bâtiments

...

14 avril 2023

Yoann Poupart

Problématique: prédire les émissions de CO2 et la consommation totale d'énergie de bâtiments non destinés à l'habitation.

Problématique: prédire les émissions de CO2 et la consommation totale d'énergie de bâtiments non destinés à l'habitation.

- Exploration des données

Problématique: prédire les émissions de CO2 et la consommation totale d'énergie de bâtiments non destinés à l'habitation.

- Exploration des données
- Feature engineering

Problématique: prédire les émissions de CO2 et la consommation totale d'énergie de bâtiments non destinés à l'habitation.

- Exploration des données
- Feature engineering
- Modélisation

Problématique: prédire les émissions de CO2 et la consommation totale d'énergie de bâtiments non destinés à l'habitation.

- Exploration des données
- Feature engineering
- Modélisation
- Résultats

Problématique: prédire les émissions de CO2 et la consommation totale d'énergie de bâtiments non destinés à l'habitation.

- Exploration des données
- Feature engineering
- Modélisation
- Résultats
- Conclusion

Exploration des données

Présentation des données

0	OSEBuildingID	3376 non-null	int64
1	DataYear	3376 non-null	int64
2	BuildingType	3376 non-null	object
3	PrimaryPropertyType	3376 non-null	object

•••

12	Latitude	3376 non-null	float64
13	Longitude	3376 non-null	float64
14	YearBuilt	3376 non-null	int64
15	NumberofBuildings	3368 non-null	float64
16	NumberofFloors	3376 non-null	int64
17	PropertyGFATotal	3376 non-null	int64
18	PropertyGFAParking	3376 non-null	int64
19	PropertyGFABuilding(s)	3376 non-null	int64
20	ListOfAllPropertyUseTypes	3367 non-null	object

•••

Présentation des données

0	OSEBuildingID	3376 non-null	int64
1	DataYear	3376 non-null	int64
2	BuildingType	3376 non-null	object
3	PrimaryPropertyType	3376 non-null	object

Identification

...

12	Latitude	3376 non-null	float64
13	Longitude	3376 non-null	float64
14	YearBuilt	3376 non-null	int64
15	NumberofBuildings	3368 non-null	float64
16	NumberofFloors	3376 non-null	int64
17	PropertyGFATotal	3376 non-null	int64
18	PropertyGFAParking	3376 non-null	int64
19	PropertyGFABuilding(s)	3376 non-null	int64
20	ListOfAllPropertyUseTypes	3367 non-null	object

...

Présentation des données

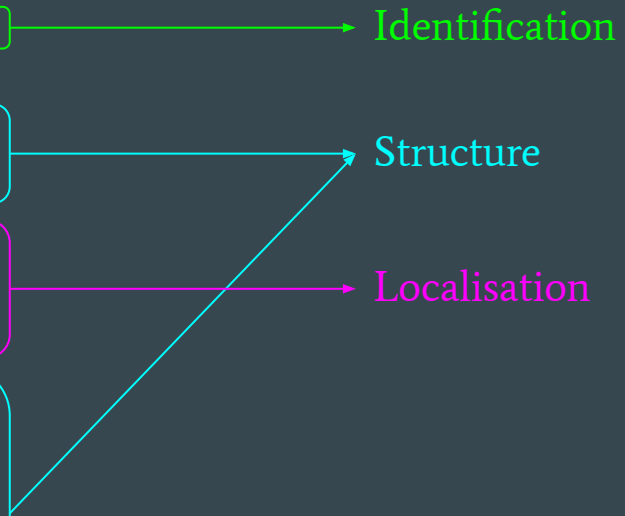
0	OSEBuildingID	3376 non-null	int64
1	DataYear	3376 non-null	int64
2	BuildingType	3376 non-null	object
3	PrimaryPropertyType	3376 non-null	object
...			
12	Latitude	3376 non-null	float64
13	Longitude	3376 non-null	float64
14	YearBuilt	3376 non-null	int64
15	NumberofBuildings	3368 non-null	float64
16	NumberofFloors	3376 non-null	int64
17	PropertyGFATotal	3376 non-null	int64
18	PropertyGFAParking	3376 non-null	int64
19	PropertyGFABuilding(s)	3376 non-null	int64
20	ListOfAllPropertyUseTypes	3367 non-null	object
...			

Identification

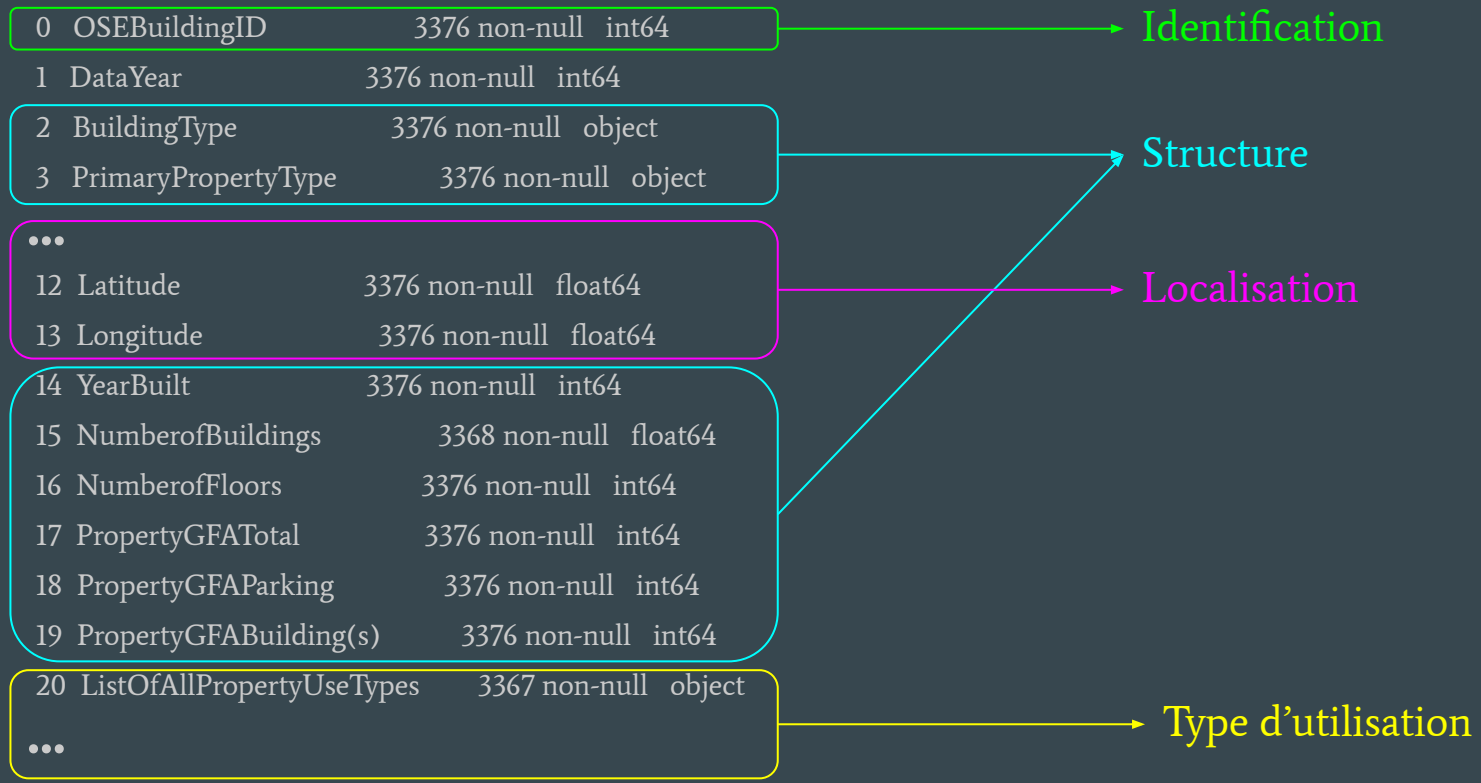
Structure

Présentation des données

0	OSEBuildingID	3376 non-null	int64
1	DataYear	3376 non-null	int64
2	BuildingType	3376 non-null	object
3	PrimaryPropertyType	3376 non-null	object
...			
12	Latitude	3376 non-null	float64
13	Longitude	3376 non-null	float64
14	YearBuilt	3376 non-null	int64
15	NumberofBuildings	3368 non-null	float64
16	NumberofFloors	3376 non-null	int64
17	PropertyGFATotal	3376 non-null	int64
18	PropertyGFAParking	3376 non-null	int64
19	PropertyGFABuilding(s)	3376 non-null	int64
20	ListOfAllPropertyUseTypes	3367 non-null	object
...			



Présentation des données



Présentation des données

```
27  YearsENERGYSTARCertified        119 non-null    object
 28  ENERGYSTARScore                 2533 non-null   float64
...
 33  SiteEnergyUse(kBtu)             3371 non-null   float64
 34  SiteEnergyUseWN(kBtu)           3370 non-null   float64
 35  SteamUse(kBtu)                  3367 non-null   float64
 36  Electricity(kWh)                3367 non-null   float64
 37  Electricity(kBtu)               3367 non-null   float64
 38  NaturalGas(therms)              3367 non-null   float64
 39  NaturalGas(kBtu)                3367 non-null   float64
...
 44  TotalGHGEmissions               3367 non-null   float64
 45  GHGEmissionsIntensity           3367 non-null   float64
```

Présentation des données

27	YearsENERGYSTARCertified	119 non-null	object
28	ENERGYSTARScore	2533 non-null	float64
•••			
33	SiteEnergyUse(kBtu)	3371 non-null	float64
34	SiteEnergyUseWN(kBtu)	3370 non-null	float64
35	SteamUse(kBtu)	3367 non-null	float64
36	Electricity(kWh)	3367 non-null	float64
37	Electricity(kBtu)	3367 non-null	float64
38	NaturalGas(therms)	3367 non-null	float64
39	NaturalGas(kBtu)	3367 non-null	float64
•••			
44	TotalGHGEmissions	3367 non-null	float64
45	GHGEmissionsIntensity	3367 non-null	float64

Pertinence à évaluer

Présentation des données

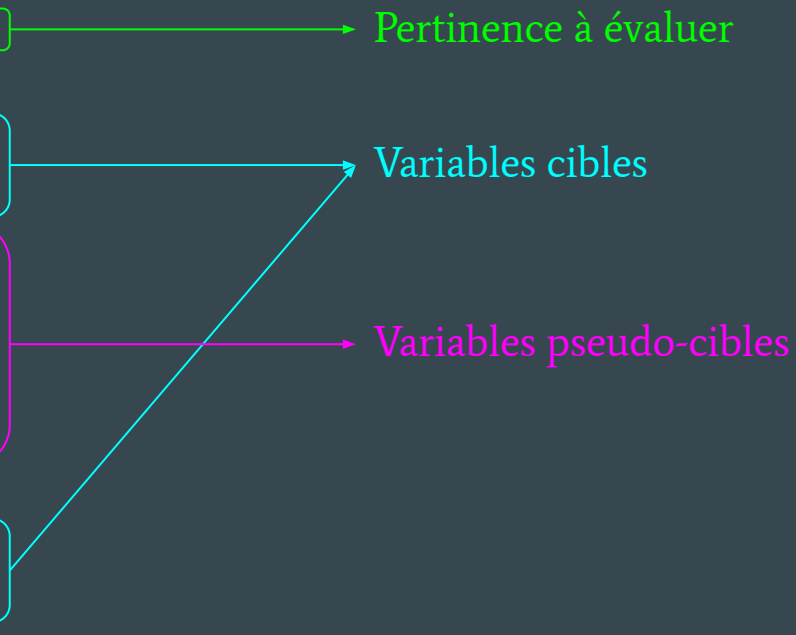
27	YearsENERGYSTARCertified	119 non-null	object
28	ENERGYSTARScore	2533 non-null	float64
...			
33	SiteEnergyUse(kBtu)	3371 non-null	float64
34	SiteEnergyUseWN(kBtu)	3370 non-null	float64
35	SteamUse(kBtu)	3367 non-null	float64
36	Electricity(kWh)	3367 non-null	float64
37	Electricity(kBtu)	3367 non-null	float64
38	NaturalGas(therms)	3367 non-null	float64
39	NaturalGas(kBtu)	3367 non-null	float64
...			
44	TotalGHGEmissions	3367 non-null	float64
45	GHGEmissionsIntensity	3367 non-null	float64

Pertinence à évaluer

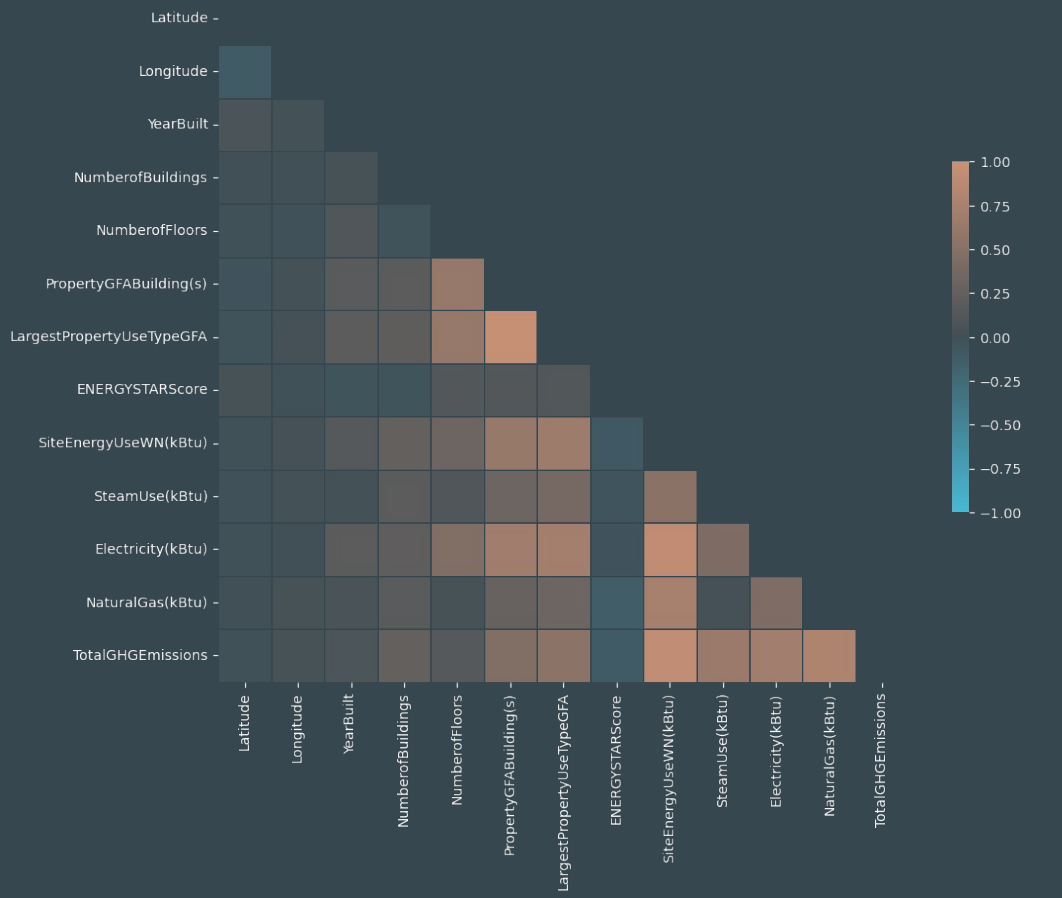
Variables cibles

Présentation des données

27	YearsENERGYSTARCertified	119 non-null	object
28	ENERGYSTARScore	2533 non-null	float64
•••			
33	SiteEnergyUse(kBtu)	3371 non-null	float64
34	SiteEnergyUseWN(kBtu)	3370 non-null	float64
35	SteamUse(kBtu)	3367 non-null	float64
36	Electricity(kWh)	3367 non-null	float64
37	Electricity(kBtu)	3367 non-null	float64
38	NaturalGas(therms)	3367 non-null	float64
39	NaturalGas(kBtu)	3367 non-null	float64
•••			
44	TotalGHGEmissions	3367 non-null	float64
45	GHGEmissionsIntensity	3367 non-null	float64

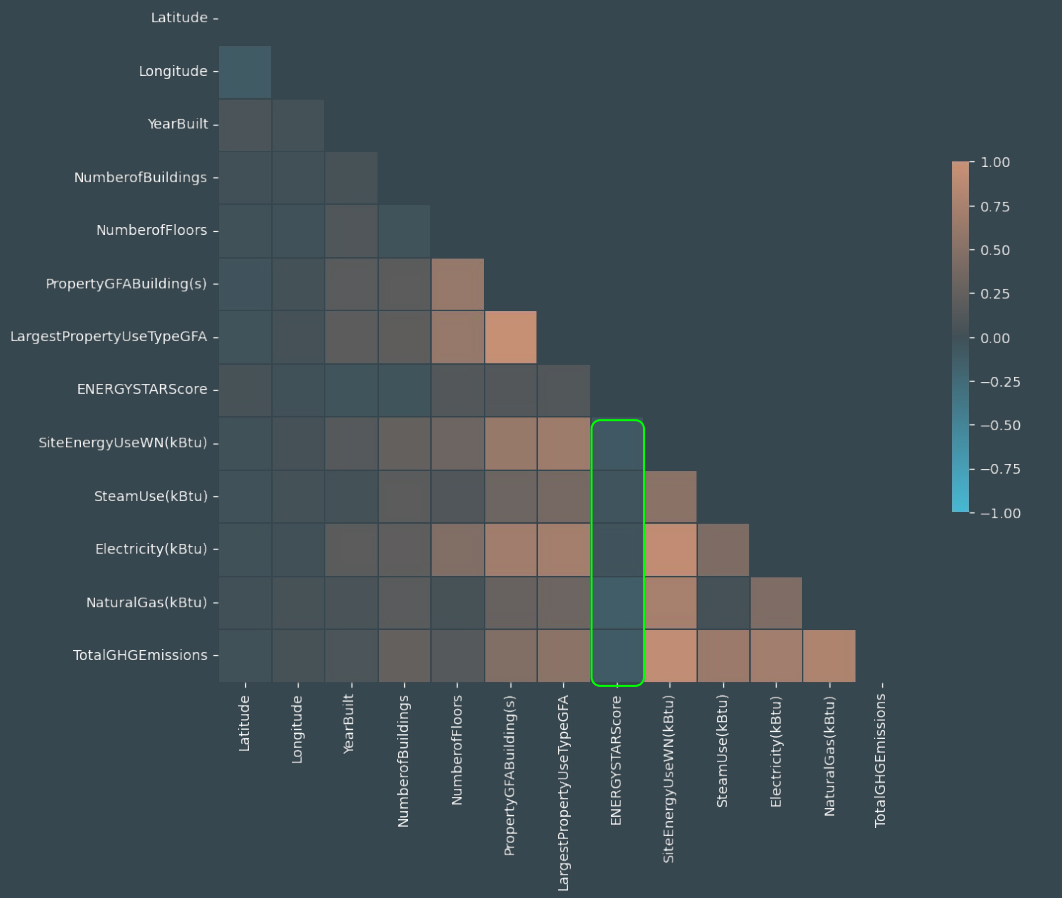


Corrélations



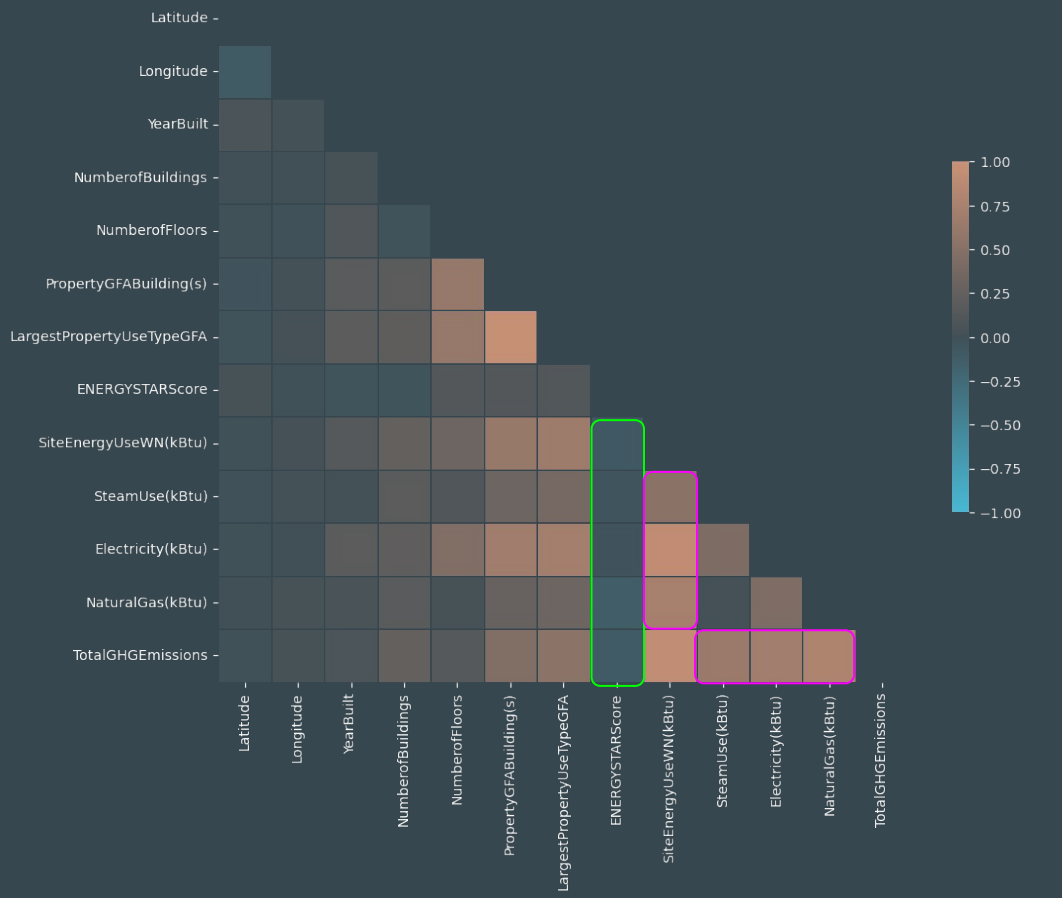
Corrélations

- Energie score informationnel



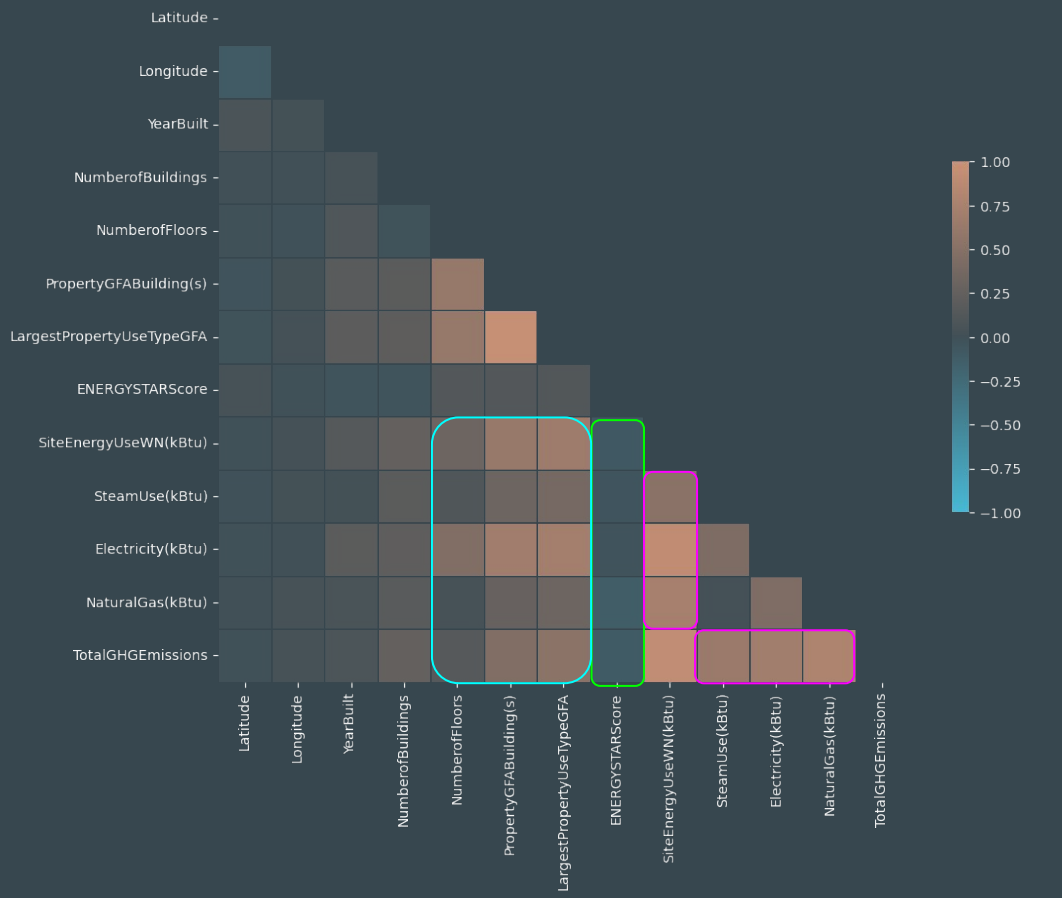
Corrélations

- Energie score informationnel
- Corrélations cibles/pseudo-cibles

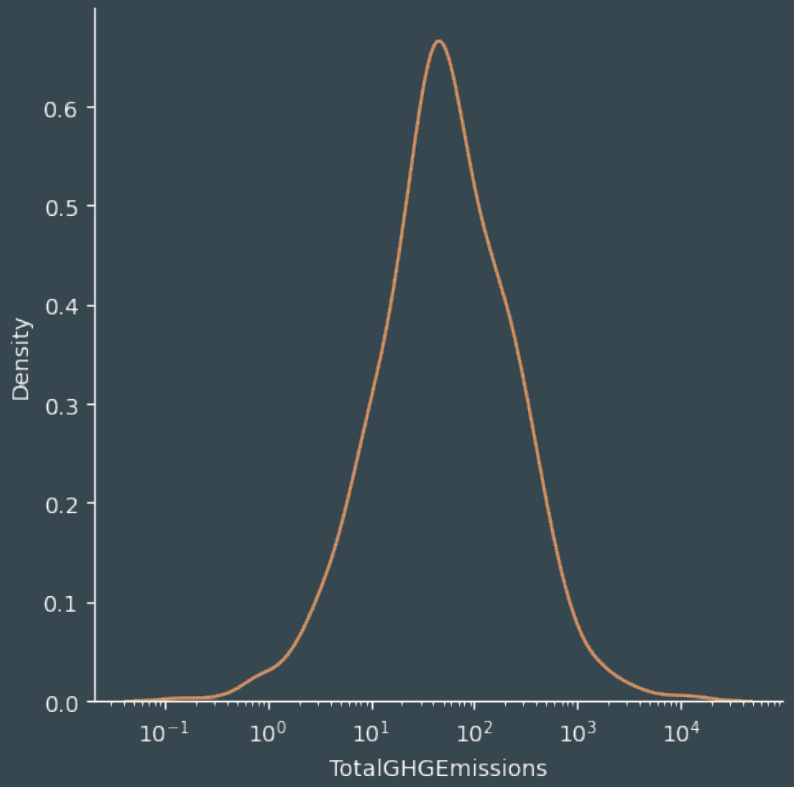


Corrélations

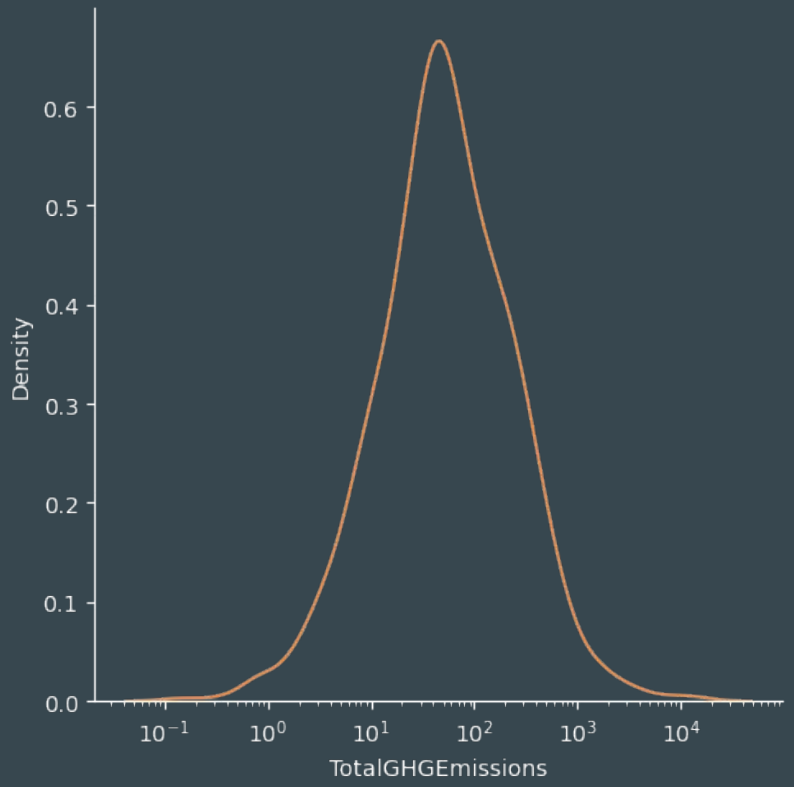
- Energie score informationnel
- Corrélations cibles/pseudo-cibles
- Données structurelles

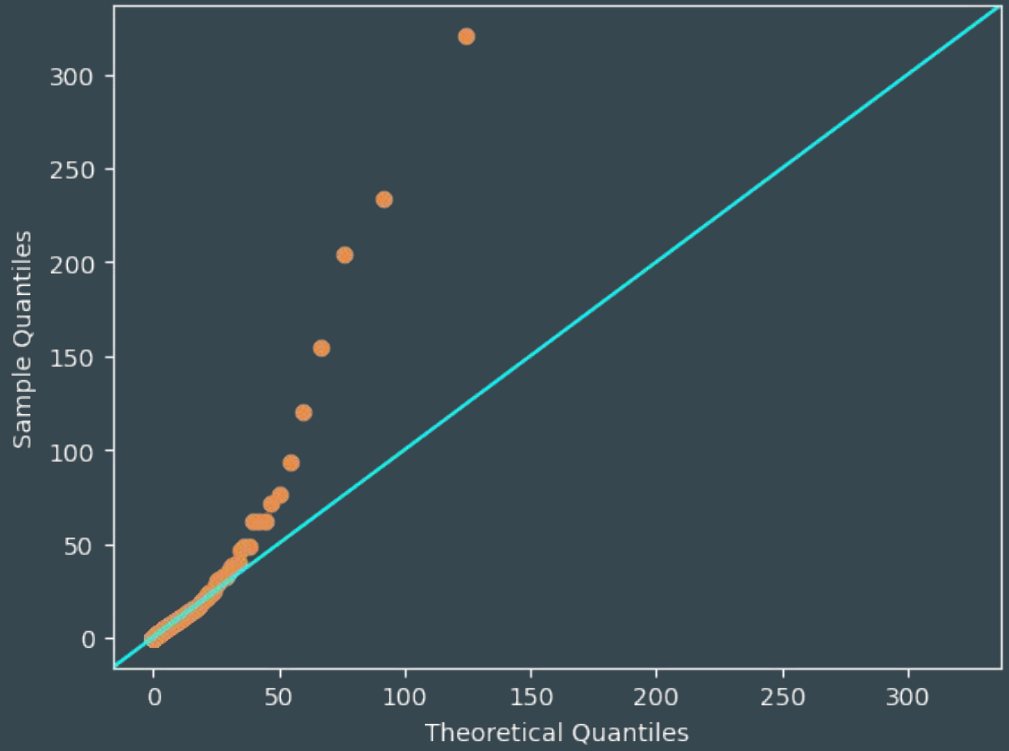


Distributions des émissions



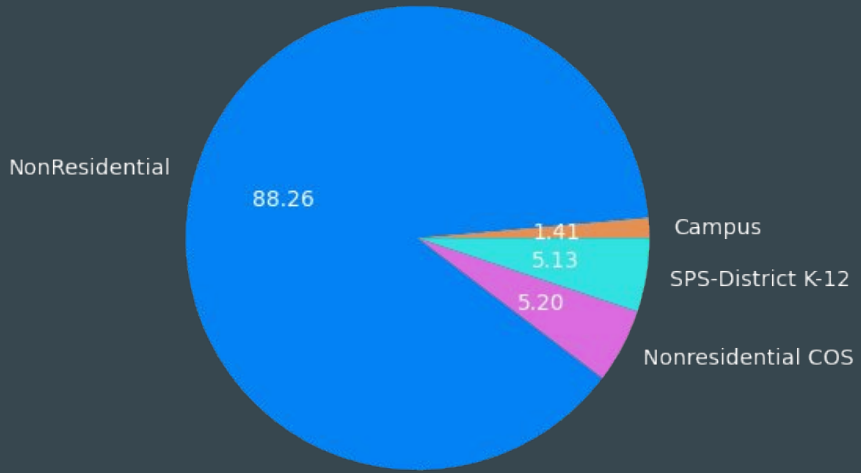
Distributions des émissions





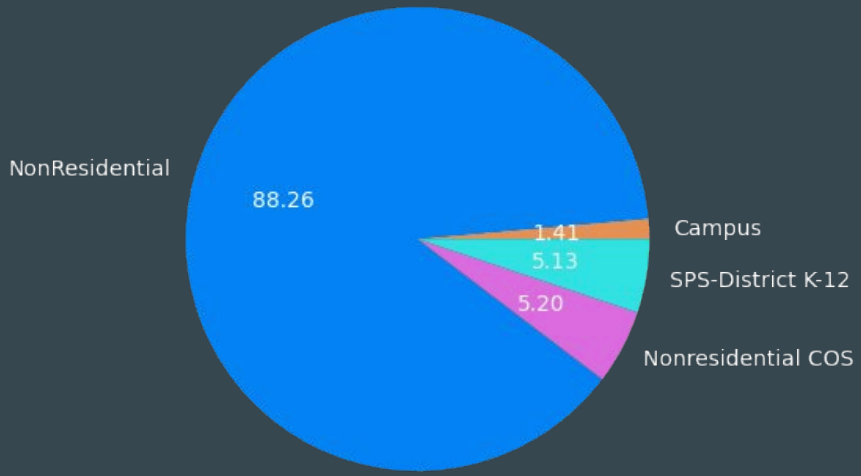
Distributions catégorielles

BuildingType

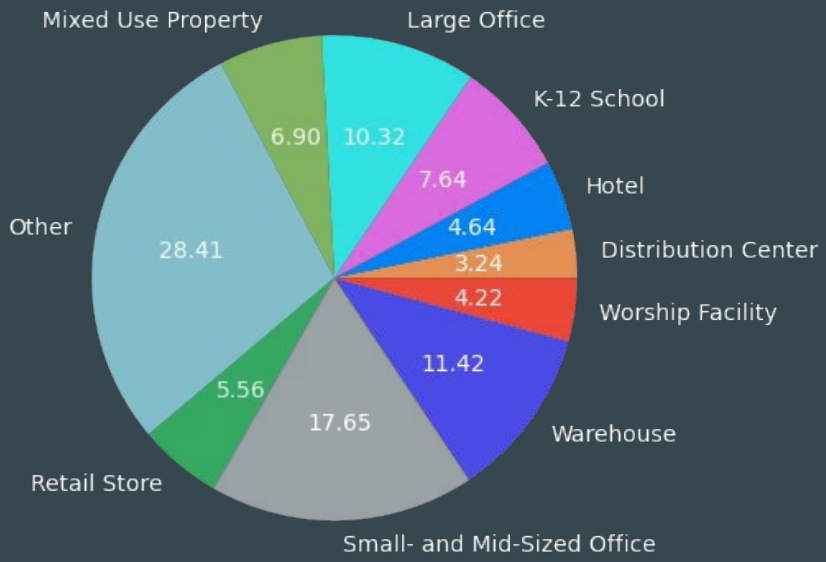


Distributions catégorielles

BuildingType

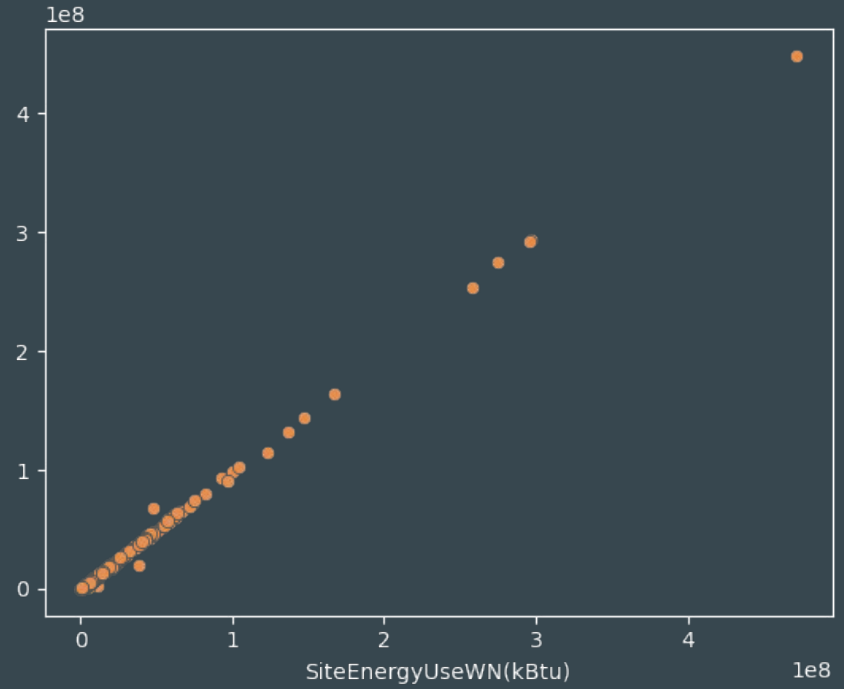


PrimaryPropertyType



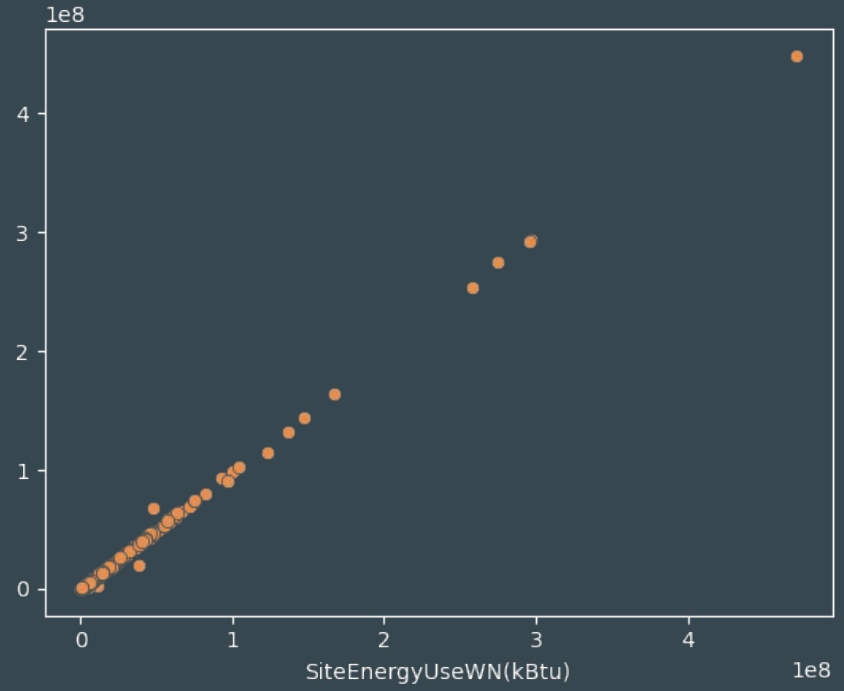
Dépendances linéaires

Somme des consommations

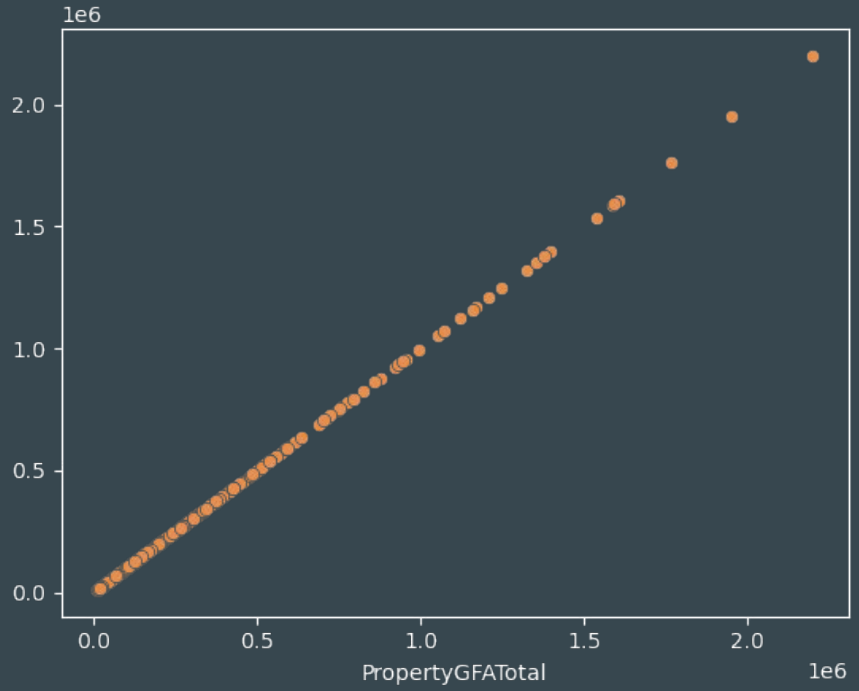


Dépendances linéaires

Somme des consommations



Somme des aires



Feature engineering

Démarche expérimentale

Démarche expérimentale

Version 1

- One hot encoding de “PrimaryPropertyType”
- Présence de “ENERGYSTARScore”

Démarche expérimentale

Version 1	• One hot encoding de “PrimaryPropertyType” • Présence de “ENERGYSTARScore”
Version 2	• Nullité de “ParkingGFA” • Steam ratio - Gas ratio - Elec ratio

Démarche expérimentale

Version 1	• One hot encoding de “PrimaryPropertyType” • Présence de “ENERGYSTARScore”
Version 2	• Nullité de “ParkingGFA” • Steam ratio - Gas ratio - Elec ratio
Version 3	• Fusion 2015 - 2016 • Text embedding

Ratio des consommations

35	SteamUse(kBtu)	3367 non-null	float64
36	Electricity(kWh)	3367 non-null	float64
37	Electricity(kBtu)	3367 non-null	float64
38	NaturalGas(therms)	3367 non-null	float64
39	NaturalGas(kBtu)	3367 non-null	float64

→ Variables pseudo-cibles

Ratio des consommations

35	SteamUse(kBtu)	3367 non-null	float64
36	Electricity(kWh)	3367 non-null	float64
37	Electricity(kBtu)	3367 non-null	float64
38	NaturalGas(therms)	3367 non-null	float64
39	NaturalGas(kBtu)	3367 non-null	float64

Variables pseudo-cibles

Ratio des consommations

35	SteamUse(kBtu)	3367 non-null	float64
36	Electricity(kWh)	3367 non-null	float64
37	Electricity(kBtu)	3367 non-null	float64
38	NaturalGas(therms)	3367 non-null	float64
39	NaturalGas(kBtu)	3367 non-null	float64

→ Variables pseudo-cibles

Ratios:

s ÷ (s+e+g)

g ÷ (s+e+g)

Ratio des consommations

35	SteamUse(kBtu)	3367 non-null	float64
36	Electricity(kWh)	3367 non-null	float64
37	Electricity(kBtu)	3367 non-null	float64
38	NaturalGas(therms)	3367 non-null	float64
39	NaturalGas(kBtu)	3367 non-null	float64

→ Variables pseudo-cibles

Ratios: $s \div (s+e+g)$ $g \div (s+e+g)$

/!\ Data leakage: $\lfloor r \times 10 \rfloor \div 10$

Présence/nullité d'une variable

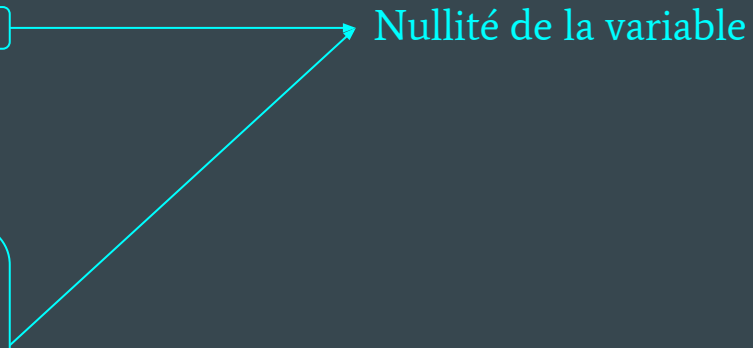
7 PropertyGFAParking 1637 non-null float64

...

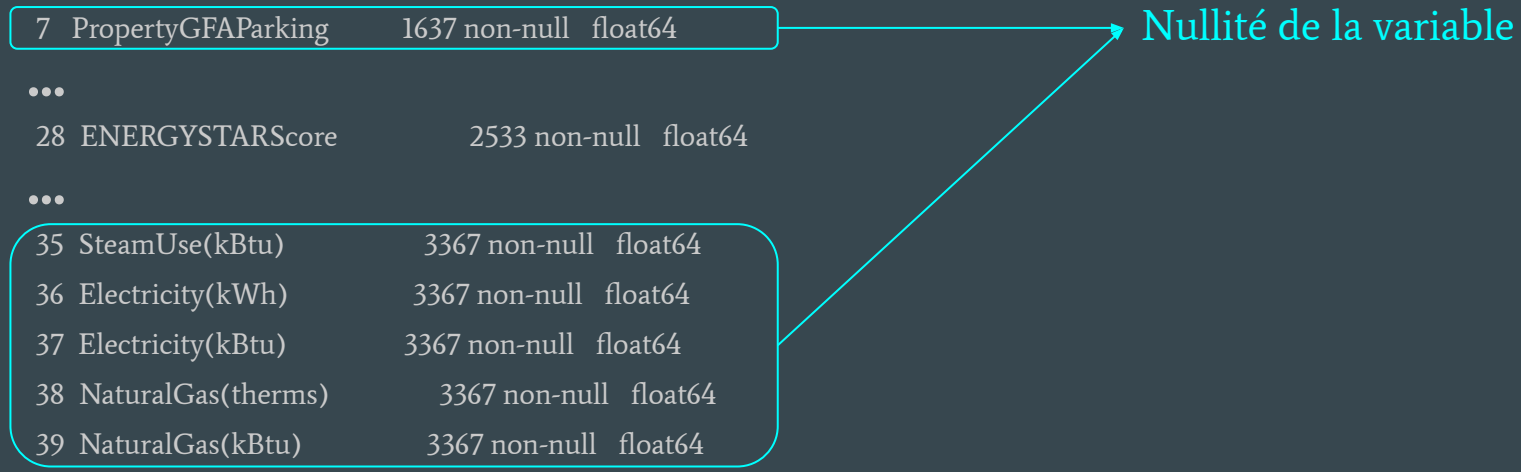
28 ENERGYSTARScore 2533 non-null float64

...

35 SteamUse(kBtu) 3367 non-null float64
36 Electricity(kWh) 3367 non-null float64
37 Electricity(kBtu) 3367 non-null float64
38 NaturalGas(therms) 3367 non-null float64
39 NaturalGas(kBtu) 3367 non-null float64

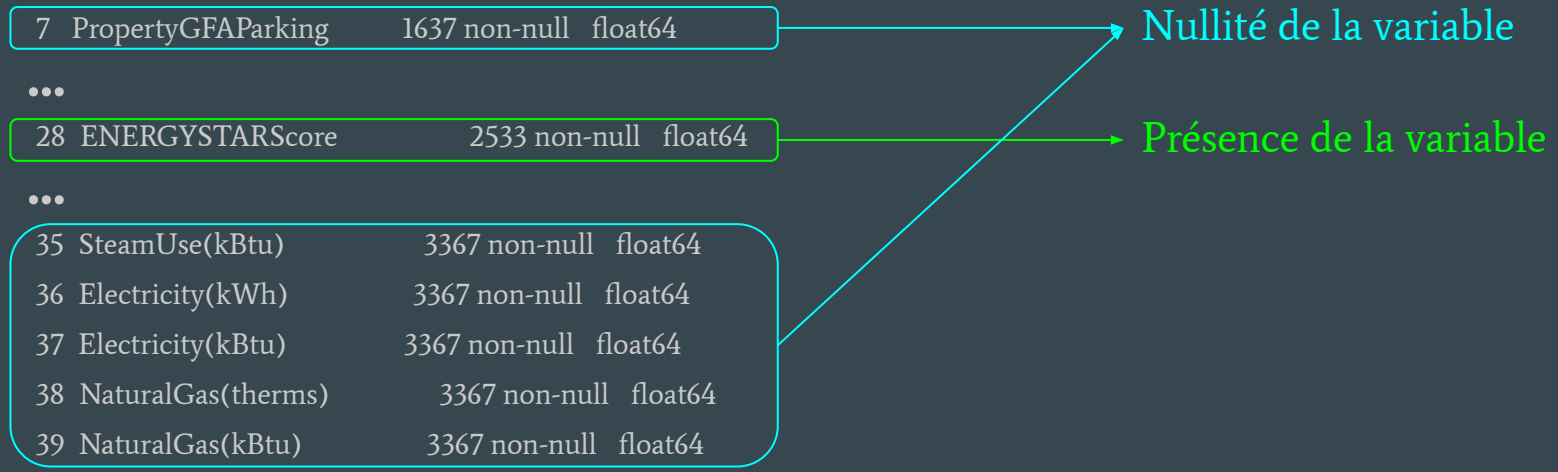


Présence/nullité d'une variable



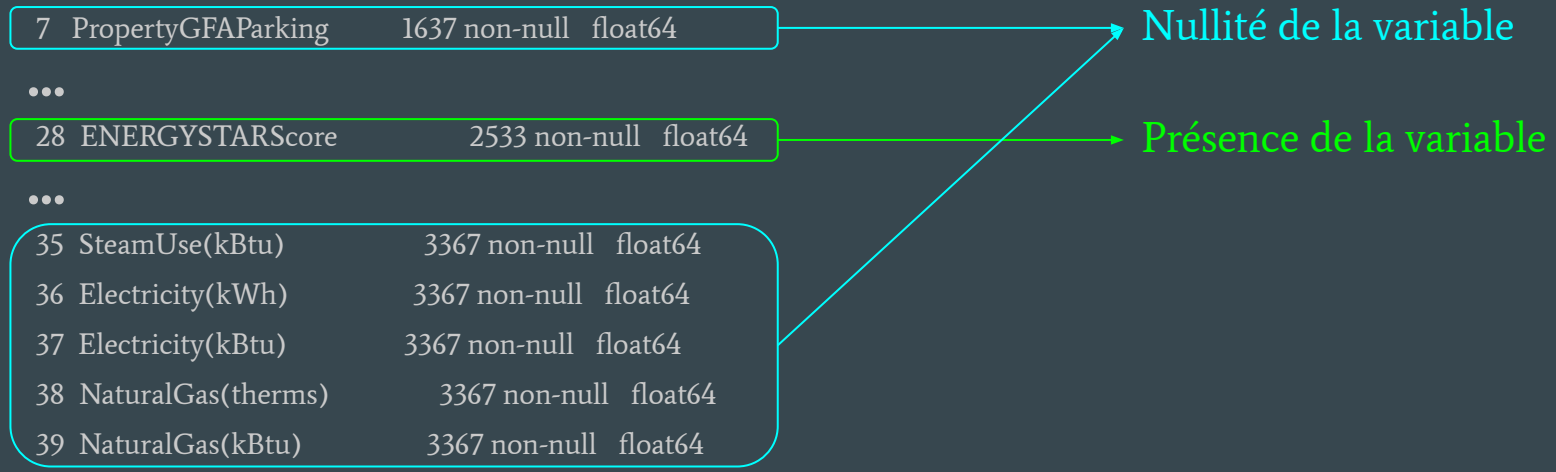
One hot encoding: $f \neq 0$

Présence/nullité d'une variable



One hot encoding: $f \neq 0$

Présence/nullité d'une variable

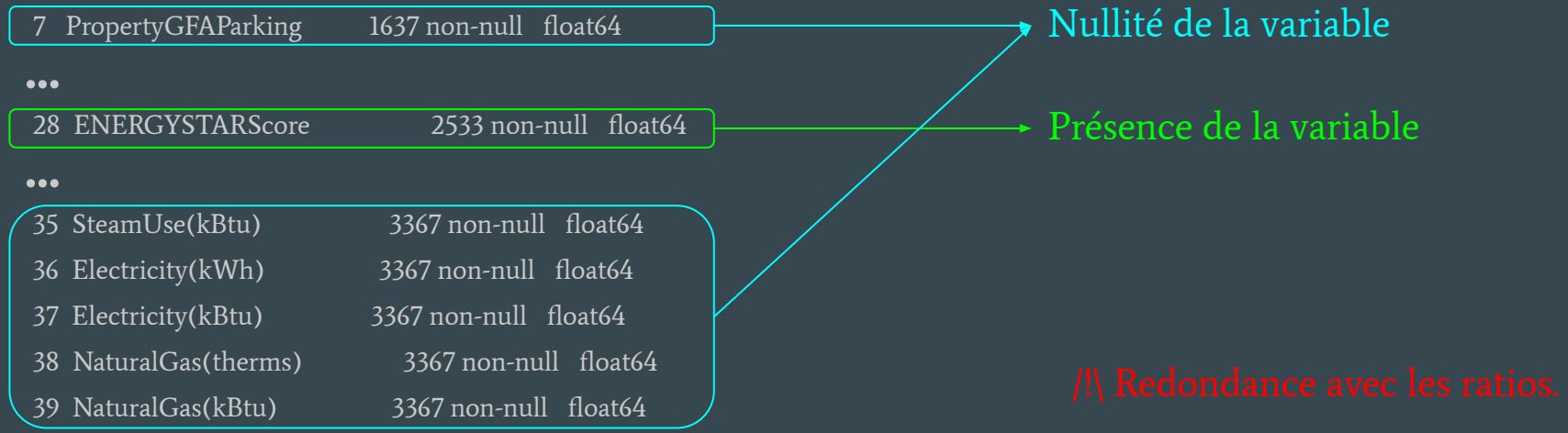


One hot encoding:

$f \neq 0$

$f \neq null$

Présence/nullité d'une variable



One hot encoding:

$f \neq 0$

$f \neq null$

One hot encoding

0 BuildingType 1637 non-null object
1 PrimaryPropertyType 1637 non-null object

Variable catégorielle

...

28 is_ENERGYSTARScore 1637 non-null object
29 is_SteamUse 1637 non-null object
30 is_NaturalGas 1637 non-null object
31 is_PropertyGFAParking 1637 non-null object

One hot encoding

0 BuildingType 1637 non-null object
1 PrimaryPropertyType 1637 non-null object

→ Variable catégorielle

...

28 is_ENERGYSTARScore 1637 non-null object
29 is_SteamUse 1637 non-null object
30 is_NaturalGas 1637 non-null object
31 is_PropertyGFAParking 1637 non-null object

→ Variable booléenne

One hot encoding

0 BuildingType 1637 non-null object
1 PrimaryPropertyType 1637 non-null object

→ Variable catégorielle

...

28 is_ENERGYSTARScore 1637 non-null object
29 is_SteamUse 1637 non-null object
30 is_NaturalGas 1637 non-null object
31 is_PropertyGFAParking 1637 non-null object

→ Variable booléenne

Encoding des valeurs: [“a”; “b”; “c”] → [(1, 0, 0); (0, 1, 0); (0, 0, 0)]

Valeurs uniques transformées en variables

Text embedding

20 ListOfAllPropertyUseTypes 3367 non-null object

Text embedding

20 ListOfAllPropertyUseTypes 3367 non-null object → Type d'utilisation

Équation sémantique: reine ≃ roi - homme + femme

Text embedding

20 ListOfAllPropertyUseTypes 3367 non-null object → Type d'utilisation

Équation sémantique:

reine ≃ roi - homme + femme

bibliothèque ≃ galerie - tableaux + livres

Text embedding

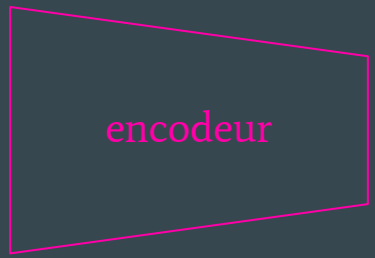
20 ListOfAllPropertyUseTypes 3367 non-null object → Type d'utilisation

Équation sémantique:

reine ≃ roi - homme + femme

bibliothèque ≃ galerie - tableaux + livres

Hotel, Parking, Restaurant



Text embedding

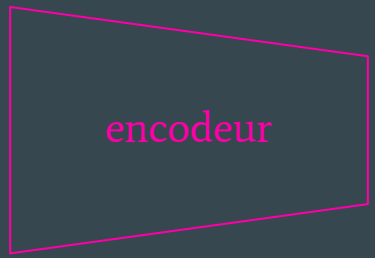
20 ListOfAllPropertyUseTypes 3367 non-null object → Type d'utilisation

Équation sémantique:

reine ≃ roi - homme + femme

bibliothèque ≃ galerie - tableaux + livres

Hotel, Parking, Restaurant



[0.5, …, -0.23] - (384)

Text embedding

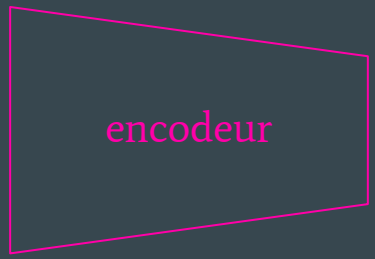
20 ListOfAllPropertyUseTypes 3367 non-null object → Type d'utilisation

Équation sémantique:

reine ≃ roi - homme + femme

bibliothèque ≃ galerie - tableaux + livres

Hotel, Parking, Restaurant



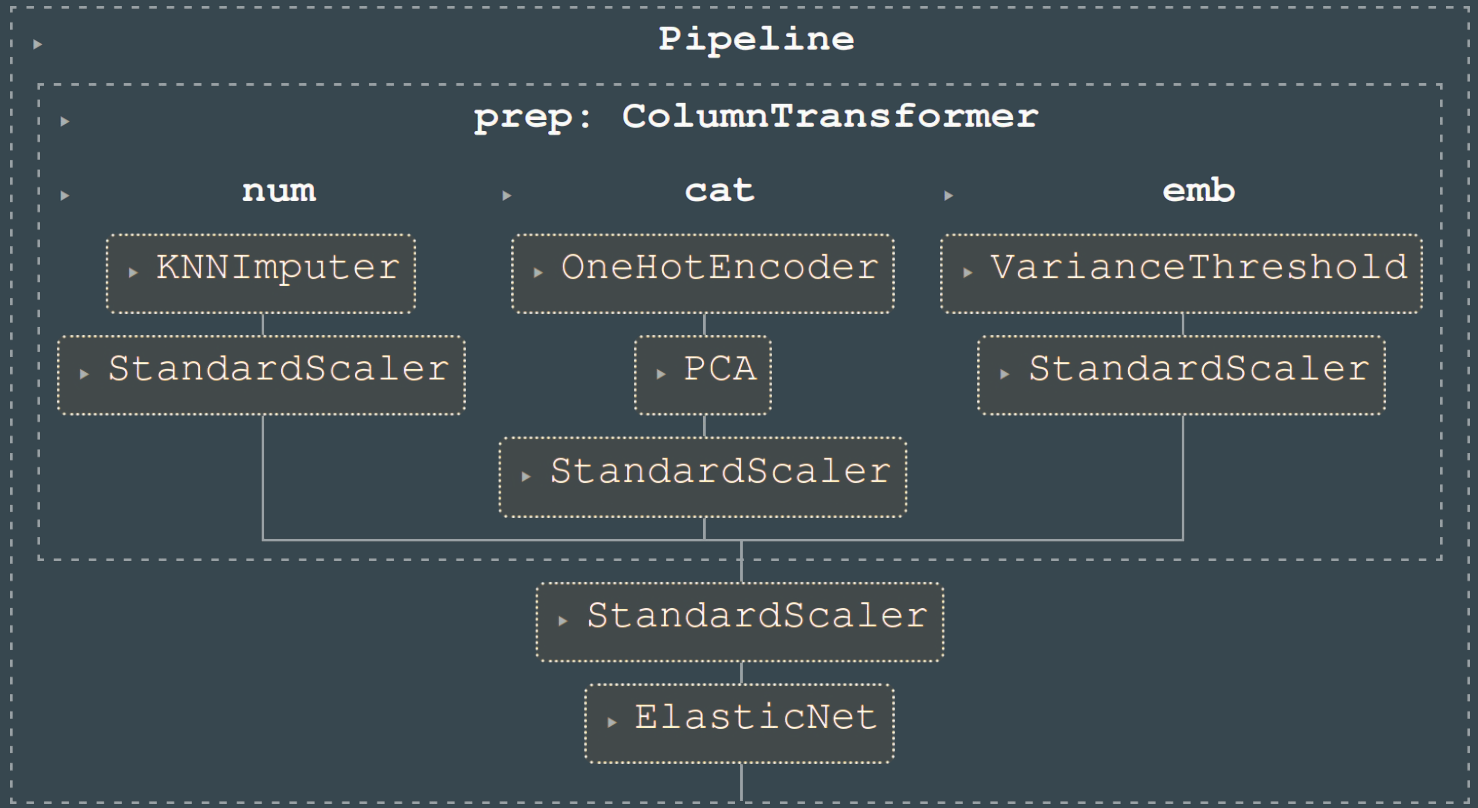
[0.5, …, -0.23] - (384)

Hypothèse:

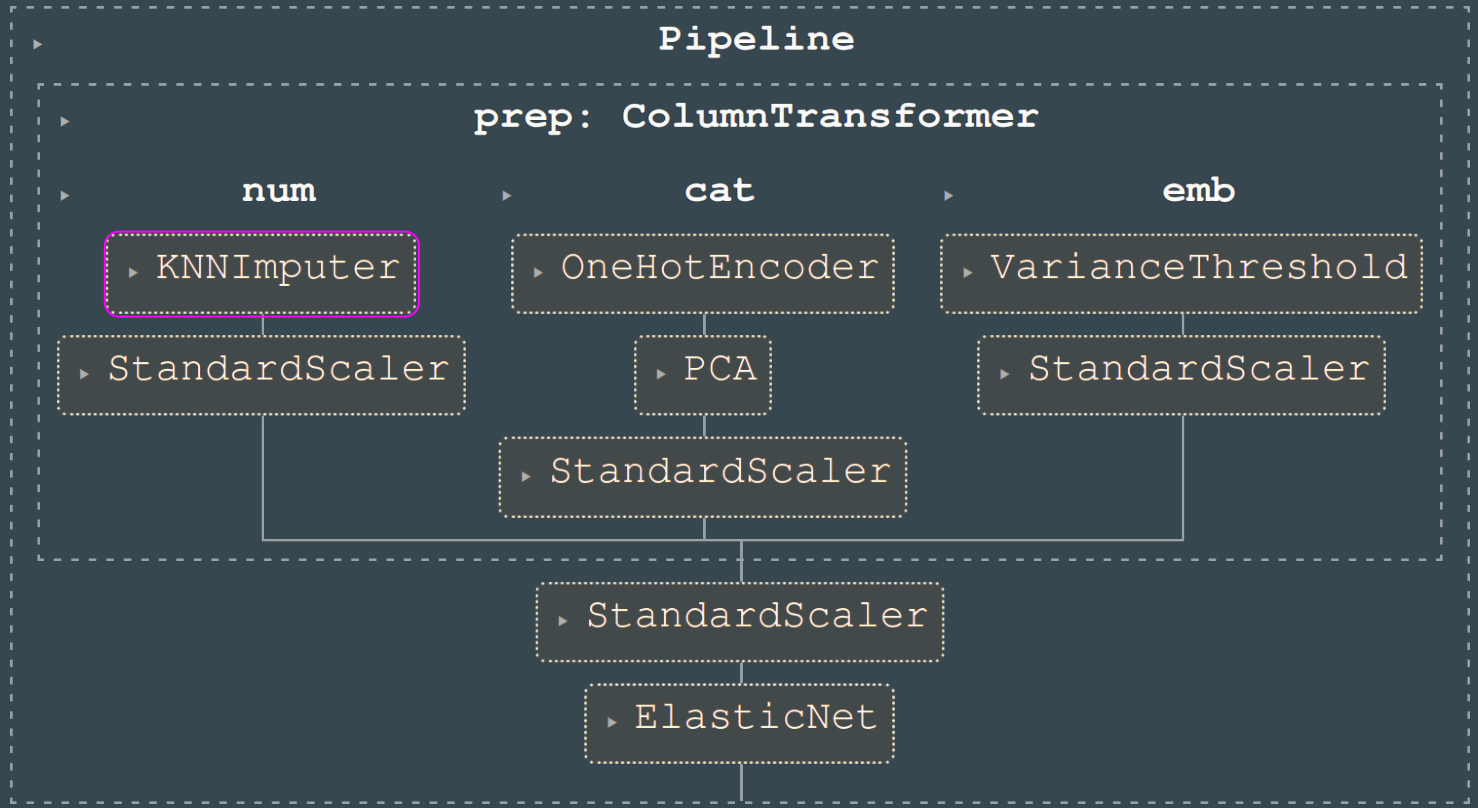
"Hotel, Parking, Restaurant" = "Hotel" + "Parking" + "Restaurant"

Modélisation

Structure du pipeline

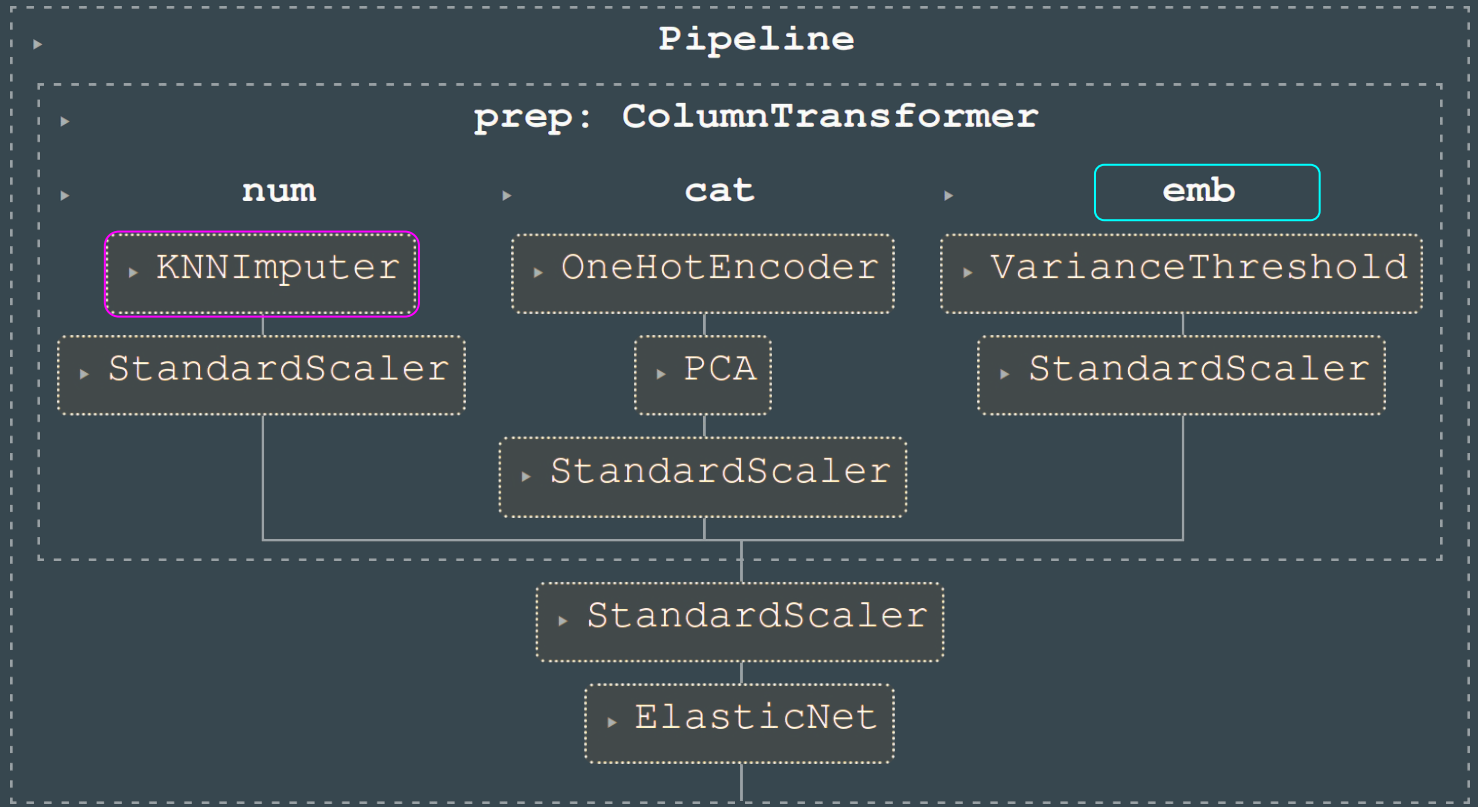


Structure du pipeline



Energy score

Structure du pipeline

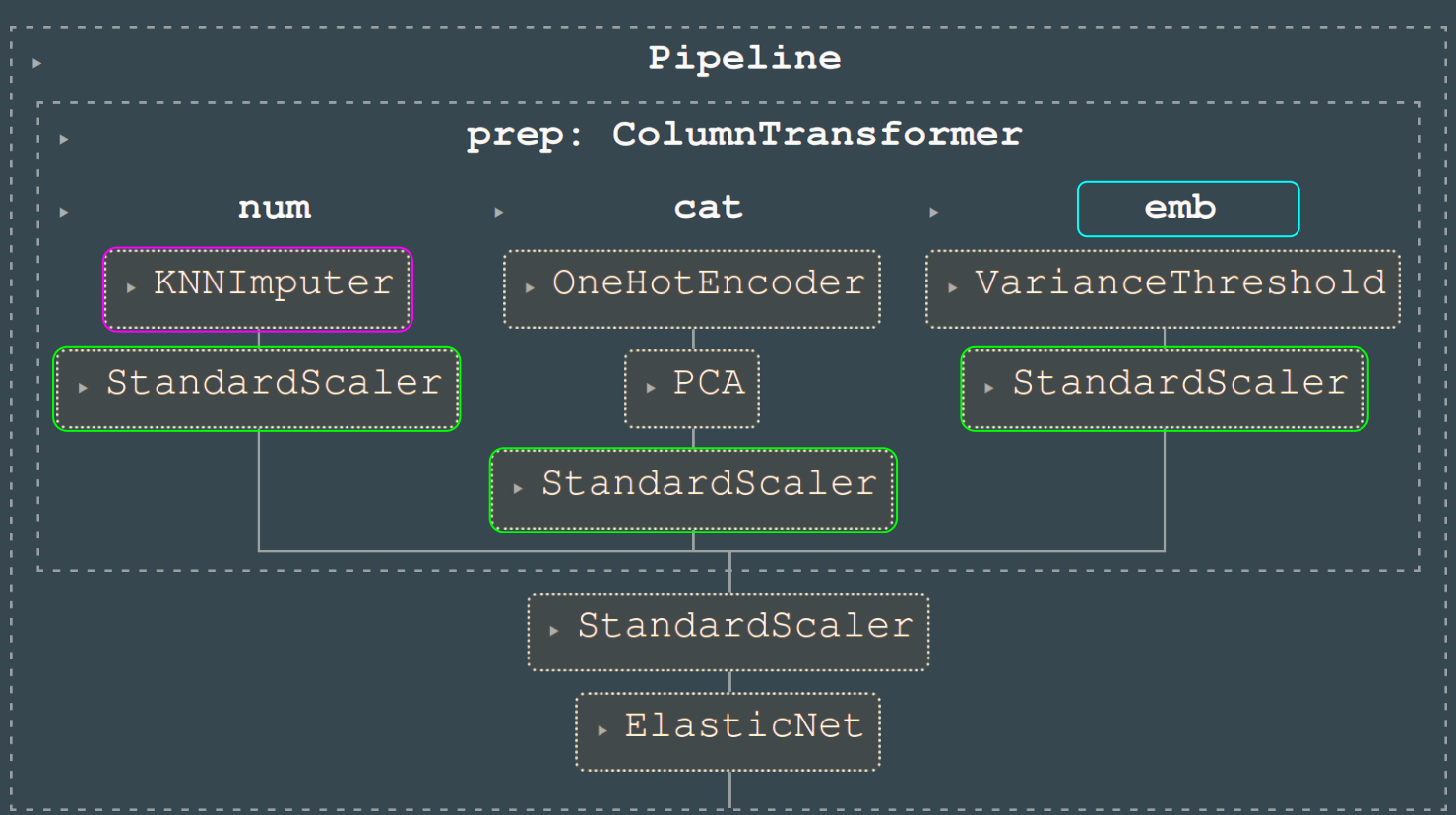


Energy score

Post PCA

700’000 → 100’000

Structure du pipeline



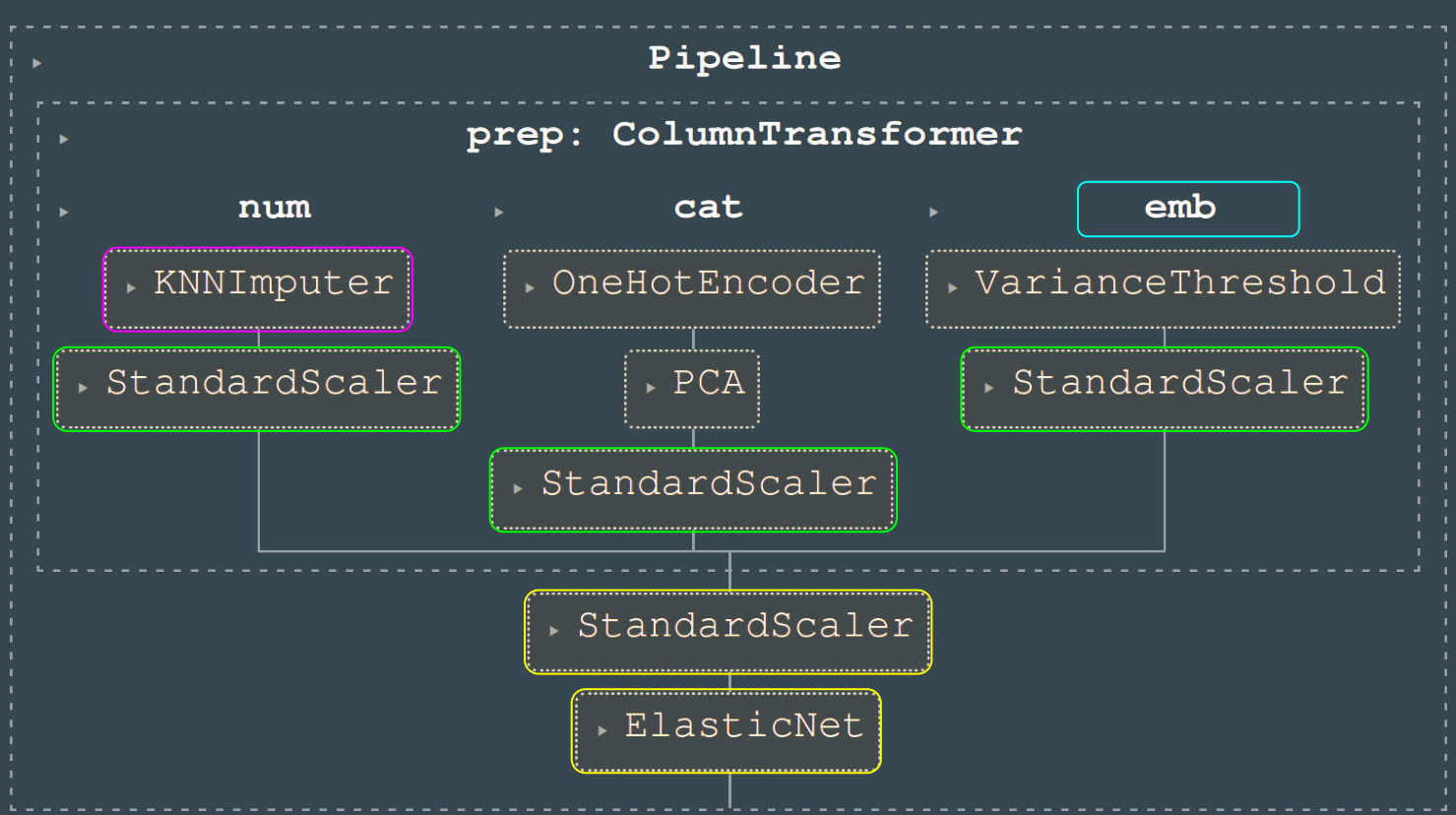
Energy score

Post PCA

700’000 → 100’000

À ajuster

Structure du pipeline



Energy score

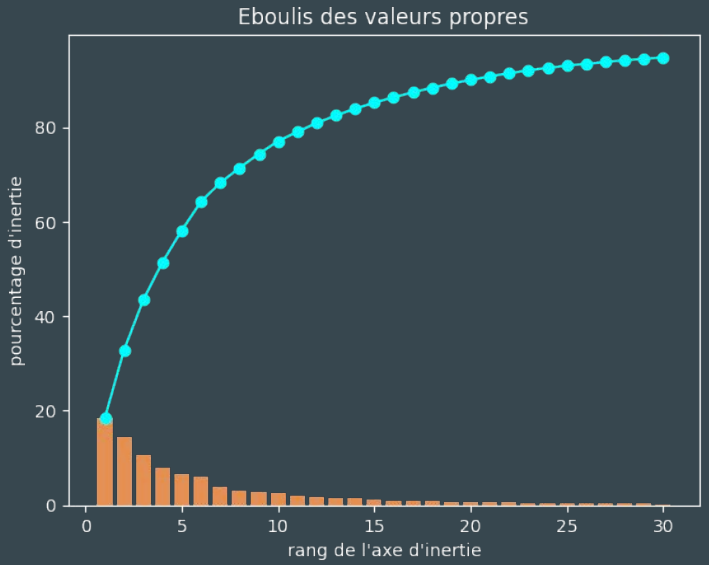
Post PCA

700’000 → 100’000

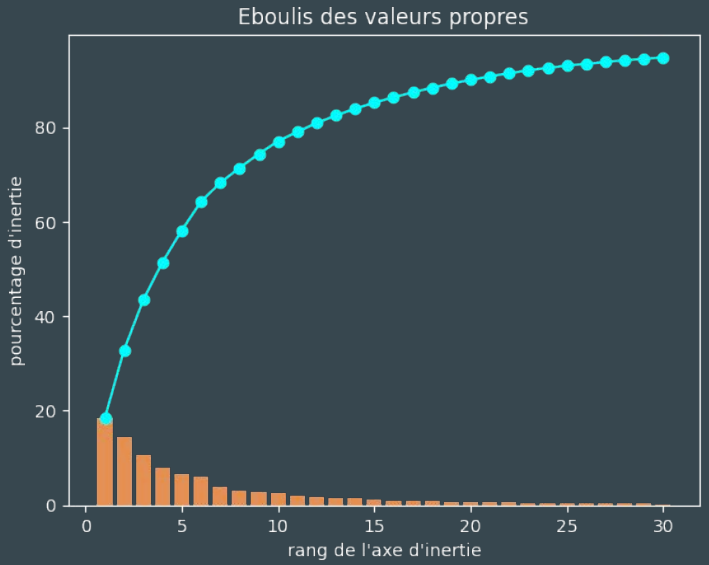
À ajuster

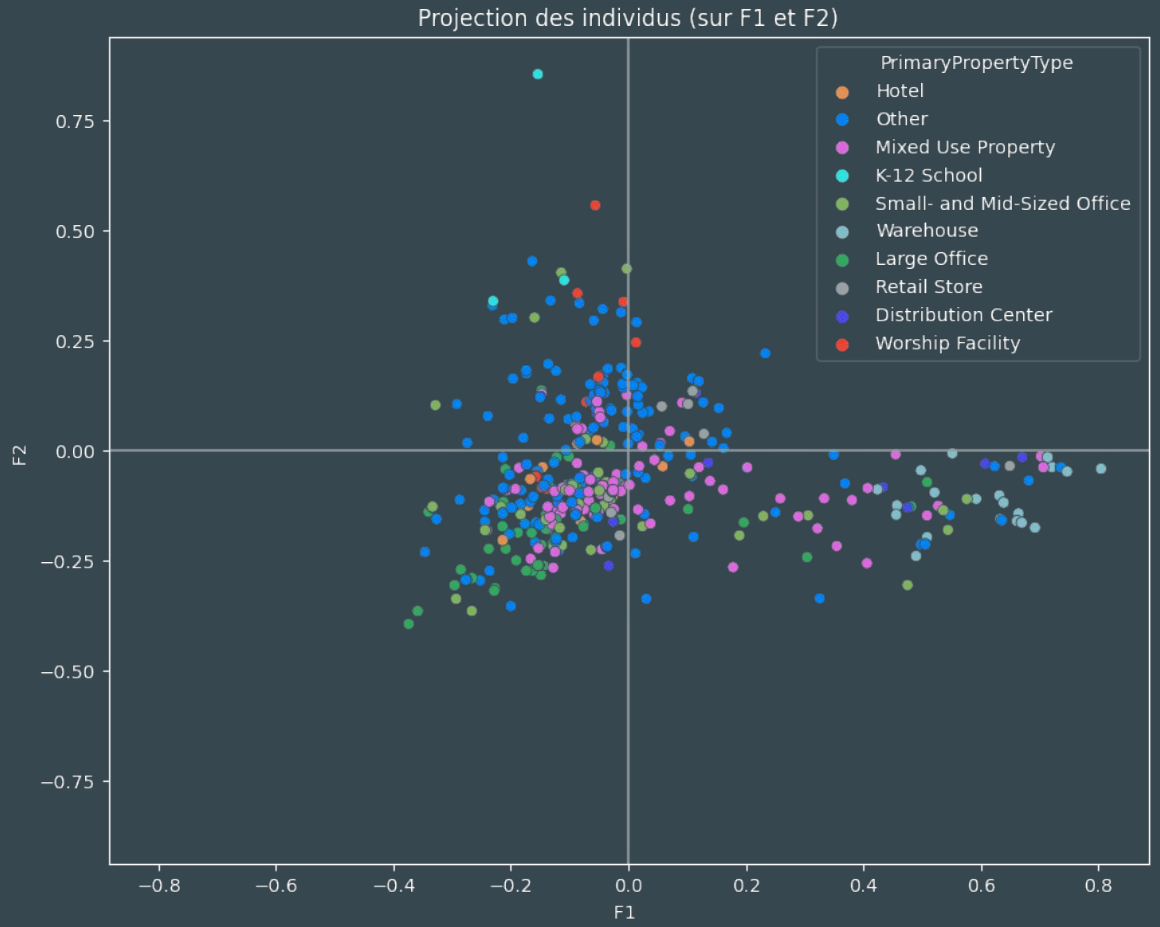
Prédiction

Embeddings PCA

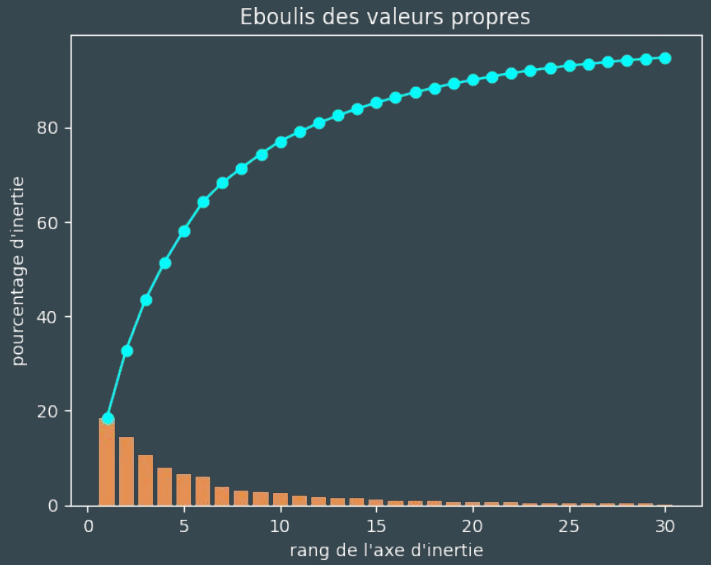


Embeddings PCA

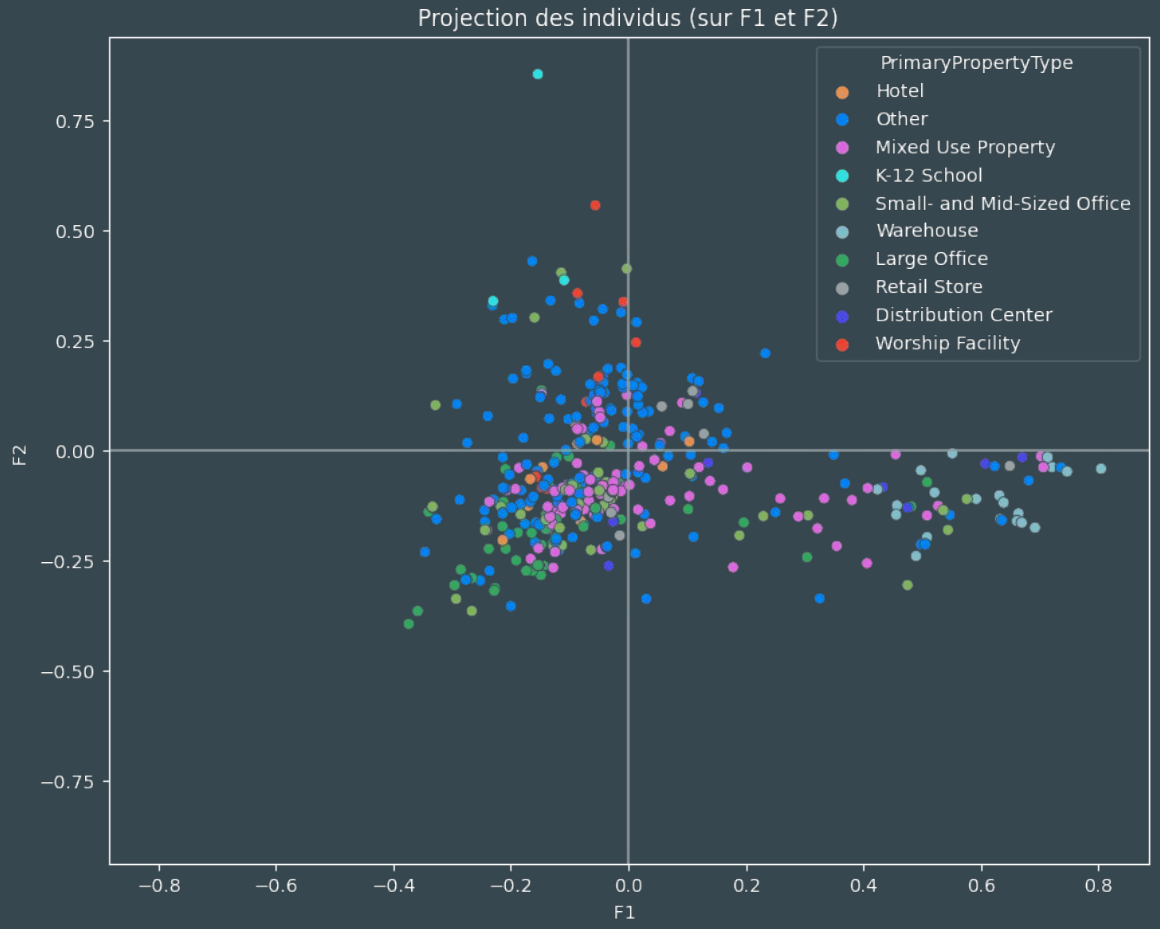




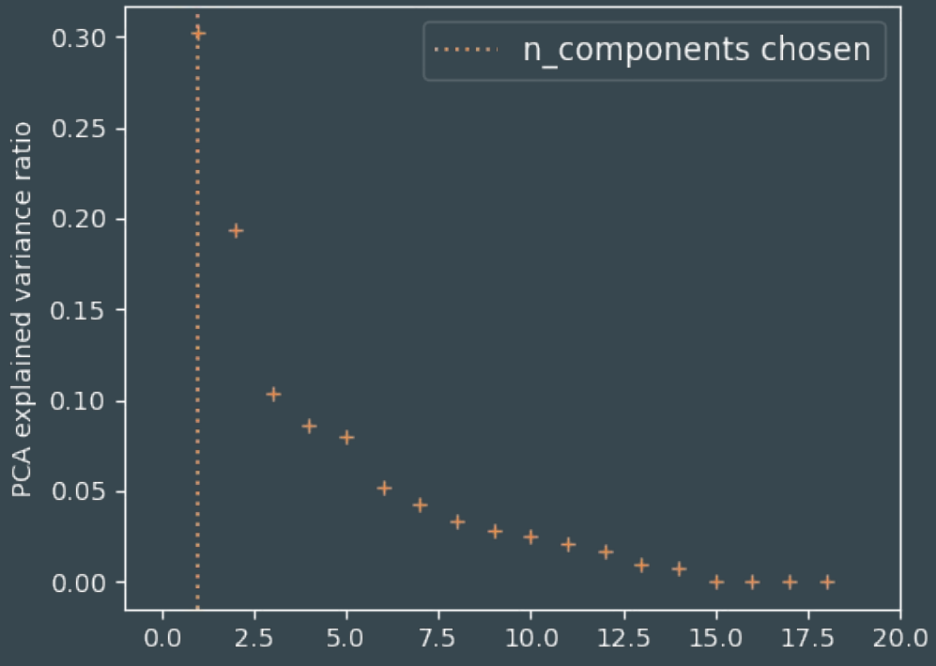
Embeddings PCA



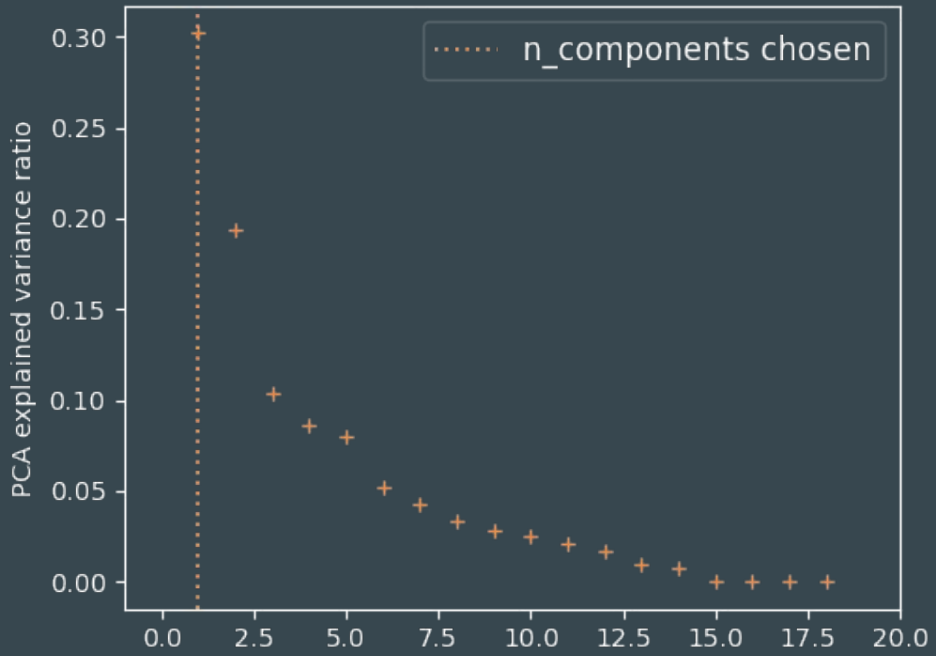
LR	Train	Test
15 Composantes	R^2: 0.079	R^2: 0.094
30	R^2: 0.131	R^2: 0.226

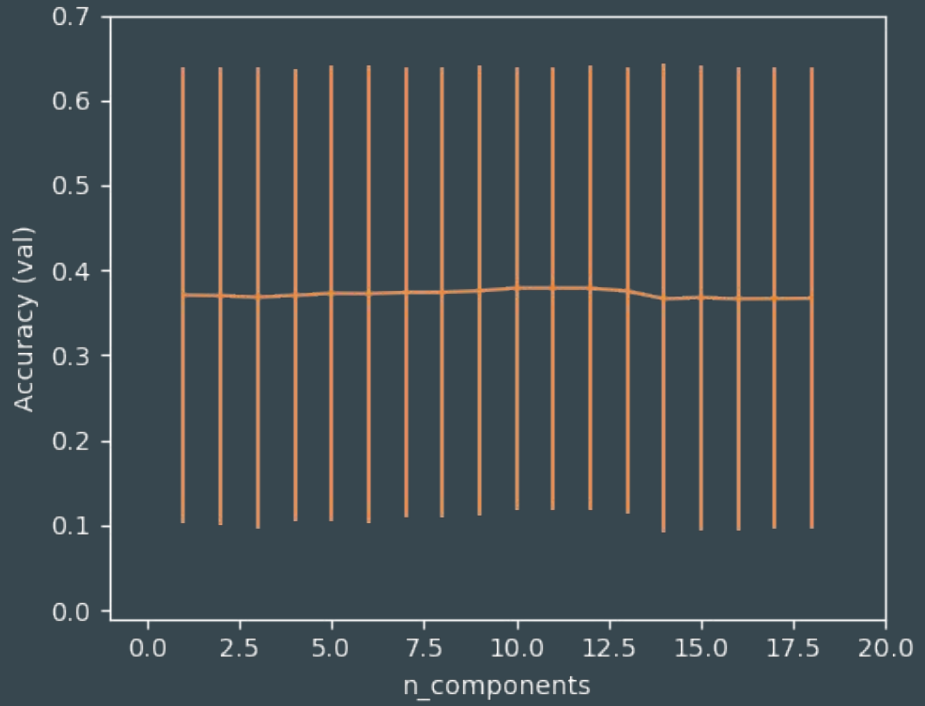


Categorical PCA

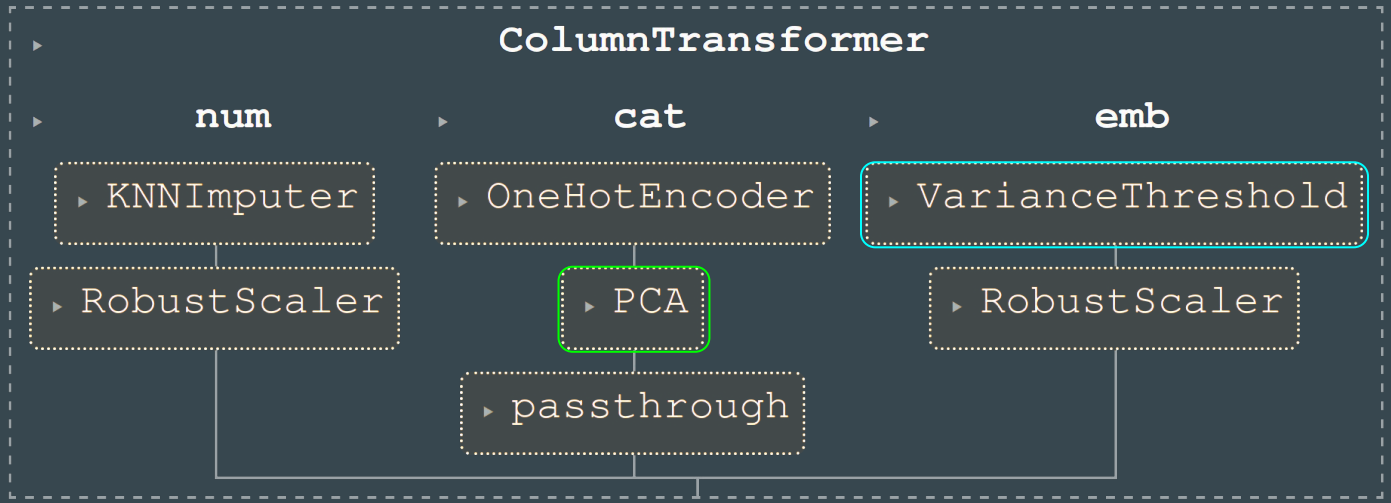


Categorical PCA





Structure du pipeline



Threshold = 0

N_components=1

Résultats

Baselines énergie

	Train	Test
Dummy (mean)	R^2: 0.0	R^2: -0.001
Dummy (Q:0.8)	R^2: -0.012	R^2: -0.021

Baselines énergie

	Train	Test
Dummy (mean)	R^2: 0.0	R^2: -0.001
Dummy (Q:0.8)	R^2: -0.012	R^2: -0.021
LR (S:surface)	R^2: 0.4617	R^2: 0.502

Baselines énergie

	Train	Test
Dummy (mean)	R^2: 0.0	R^2: -0.001
Dummy (Q:0.8)	R^2: -0.012	R^2: -0.021
LR (S:surface)	R^2: 0.4617	R^2: 0.502
LR (S+year+#floor…)	R^2: 0.522	R^2: 0.298
LR (S+embeddings)	R^2: 0.547	R^2: 0.575

Résultats énergie

	Train	Test
ElasticNet	R^2: 0.417	R^2: 0.546

L1-ratio = 0.1
Alpha = 3.2

Résultats énergie

	Train	Test
ElasticNet	R^2: 0.417	R^2: 0.546
Ridge	R^2: 0.548	R^2: 0.643

Alpha = 10^3

Résultats énergie

	Train	Test
ElasticNet	R^2: 0.417	R^2: 0.546
Ridge	R^2: 0.548	R^2: 0.643
SVR (racine)	R^2: 0.822	R^2: 0.815
SVR (log)	R^2: 0.743	R^2: 0.656

$C = 10^4$
$Epsilon = 10^2$

Résultats énergie

	Train	Test
ElasticNet	R^2: 0.417	R^2: 0.546
Ridge	R^2: 0.548	R^2: 0.643
SVR (racine)	R^2: 0.822	R^2: 0.815
SVR (log)	R^2: 0.743	R^2: 0.656

C = 1
Epsilon = 0.01

Résultats énergie

	Train	Test
RandomForest (RF)	R^2: 0.730	R^2: 0.658
RF (racine)	R^2: 0.734	R^2: 0.714

Estimators = 200
Depth = 25
Samples = 7

Résultats énergie

	Train	Test
RandomForest (RF)	R^2: 0.730	R^2: 0.658
RF (racine)	R^2: 0.734	R^2: 0.714

Estimators = 150
Depth = 15
Samples = 5

Résultats énergie

	Train	Test
RandomForest (RF)	R^2: 0.730	R^2: 0.658
RF (racine)	R^2: 0.734	R^2: 0.714
XGB	R^2: 0.998	R^2: 0.492
XGB (racine)	R^2: 0.970	R^2: 0.679

Estimators = 150
Depth = 15
Gamma = 0.1

Résultats énergie

	Train	Test
RandomForest (RF)	R^2: 0.730	R^2: 0.658
RF (racine)	R^2: 0.734	R^2: 0.714
XGB	R^2: 0.998	R^2: 0.492
XGB (racine)	R^2: 0.970	R^2: 0.679

Estimators = 200
Depth = 25
Gamma = 0.01

Baselines émissions

	Train	Test
Dummy (mean)	R^2: 0.0	R^2: -0.001
Dummy (Q:0.8)	R^2: -0.001	R^2: -0.000
LR (S:surface)	R^2: 0.236	R^2: 0.331
LR (S+year+#floor…)	R^2: 0.294	R^2: 0.409
LR (S+embeddings)	R^2: 0.396	R^2: 0.421

Baselines émissions

	Train	Test
Dummy (mean)	R^2: 0.0	R^2: -0.001
Dummy (Q:0.8)	R^2: -0.001	R^2: -0.000
LR (S:surface)	R^2: 0.236	R^2: 0.331
LR (S+year+#floor…)	R^2: 0.294	R^2: 0.409
LR (S+embeddings)	R^2: 0.396	R^2: 0.421

Résultats émissions

	Train	Test
ElasticNet	R^2: 0.440	R^2: 0.422
Ridge	R^2: 0.543	R^2: 0.544

L1-ratio = 10^{-3}
Alpha = 1

Résultats émissions

	Train	Test
ElasticNet	R^2: 0.440	R^2: 0.422
Ridge	R^2: 0.543	R^2: 0.544

Alpha = 10^2

Résultats émissions

	Train	Test
ElasticNet	R^2: 0.440	R^2: 0.422
Ridge	R^2: 0.543	R^2: 0.544
SVR	R^2: 0.963	R^2: 0.856
SVR (racine)	R^2: 0.971	R^2: 0.909

$C = 10^4$
Epsilon = 1

Résultats émissions

	Train	Test
ElasticNet	R^2: 0.440	R^2: 0.422
Ridge	R^2: 0.543	R^2: 0.544
SVR	R^2: 0.963	R^2: 0.856
SVR (racine)	R^2: 0.971	R^2: 0.909

C = 100
Epsilon = 0.01

Résultats émissions

	Train	Test
RandomForest (RF)	R^2: 0.680	R^2: 0.501
RF (racine)	R^2: 0.827	R^2: 0.570

Estimators =250
Depth = 15
Samples = 5

Résultats émissions

	Train	Test
RandomForest (RF)	R^2: 0.680	R^2: 0.501
RF (racine)	R^2: 0.827	R^2: 0.570

Estimators = 200
Depth = 15
Samples = 3

Résultats émissions

	Train	Test
RandomForest (RF)	R^2: 0.680	R^2: 0.501
RF (racine)	R^2: 0.827	R^2: 0.570
XGB	R^2: 0.995	R^2: 0.846
XGB (racine)	R^2: 1.000	R^2: 0.855

Estimators = 200
Depth = 5
Gamma = 0.1

Résultats émissions

	Train	Test
RandomForest (RF)	R^2: 0.680	R^2: 0.501
RF (racine)	R^2: 0.827	R^2: 0.570
XGB	R^2: 0.995	R^2: 0.846
XGB (racine)	R^2: 1.000	R^2: 0.855

Estimators = 300
Depth = 15
Gamma = 0.01

Conclusion

Meilleurs modèles

Énergie	Train	Test	Fit time
SVR (racine)	R^2: 0.85 ± 0.04 MAE: 143.10^4 ± 9.10^4	R^2: 0.78 ± 0.09 MAE: 29.10^5 ± 5.10^5	0.58 ± 0.07 s
RF (racine)	R^2: 0.93 ± 0.01 MAE: 126.10^4 ± 4.10^4	R^2: 0.6 ± 0.2 MAE: 32.10^5 ± 5.10^5	3.4 ± 0.4 s

Meilleurs modèles

Énergie	Train	Test	Fit time
SVR (racine)	R^2: 0.85 ± 0.04 MAE: 143.10^4 ± 9.10^4	R^2: 0.78 ± 0.09 MAE: 29.10^5 ± 5.10^5	0.58 ± 0.07 s
RF (racine)	R^2: 0.93 ± 0.01 MAE: 126.10^4 ± 4.10^4	R^2: 0.6 ± 0.2 MAE: 32.10^5 ± 5.10^5	3.4 ± 0.4 s

Meilleurs modèles

Énergie	Train	Test	Fit time
SVR (racine)	R^2: 0.85 ± 0.04 MAE: 143.10^4 ± 9.10^4	R^2: 0.78 ± 0.09 MAE: 29.10^5 ± 5.10^5	0.58 ± 0.07 s
RF (racine)	R^2: 0.93 ± 0.01 MAE: 126.10^4 ± 4.10^4	R^2: 0.6 ± 0.2 MAE: 32.10^5 ± 5.10^5	3.4 ± 0.4 s

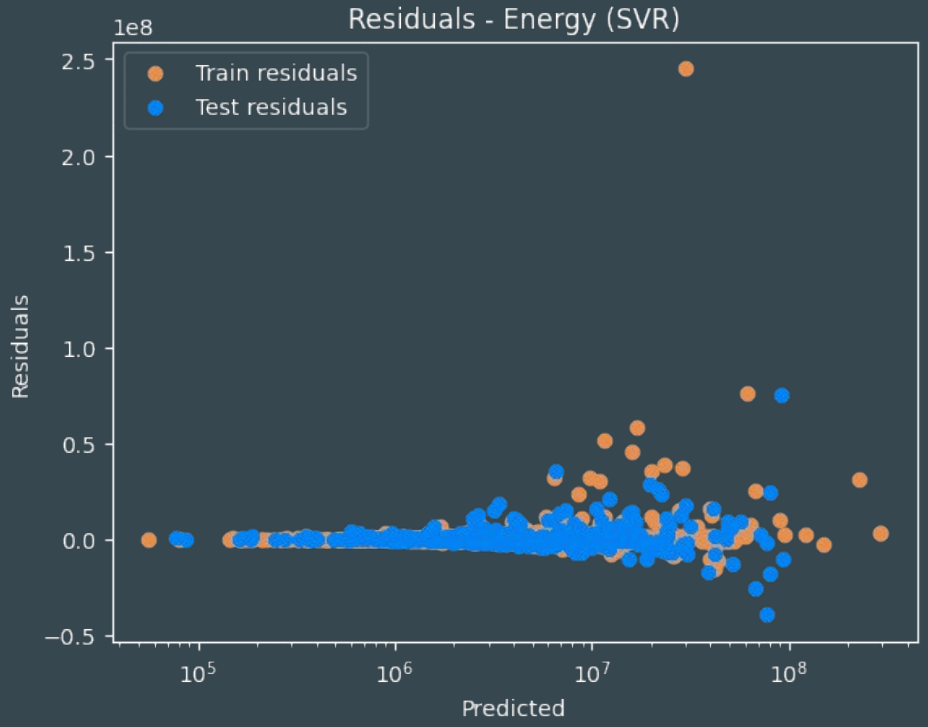
Émissions	Train	Test	Fit time
SVR (racine)	R^2: 0.977 ± 0.003 MAE: 17.1 ± 0.8	R^2: 0.87 ± 0.05 MAE: 65 ± 10	0.75 ± 0.02 s
XGB (racine)	R^2: 1.000 ± 0.001 MAE: 1.0 ± 0.1	R^2: 0.7 ± 0.2 MAE: 79 ± 9	4 ± 1 s

Meilleurs modèles

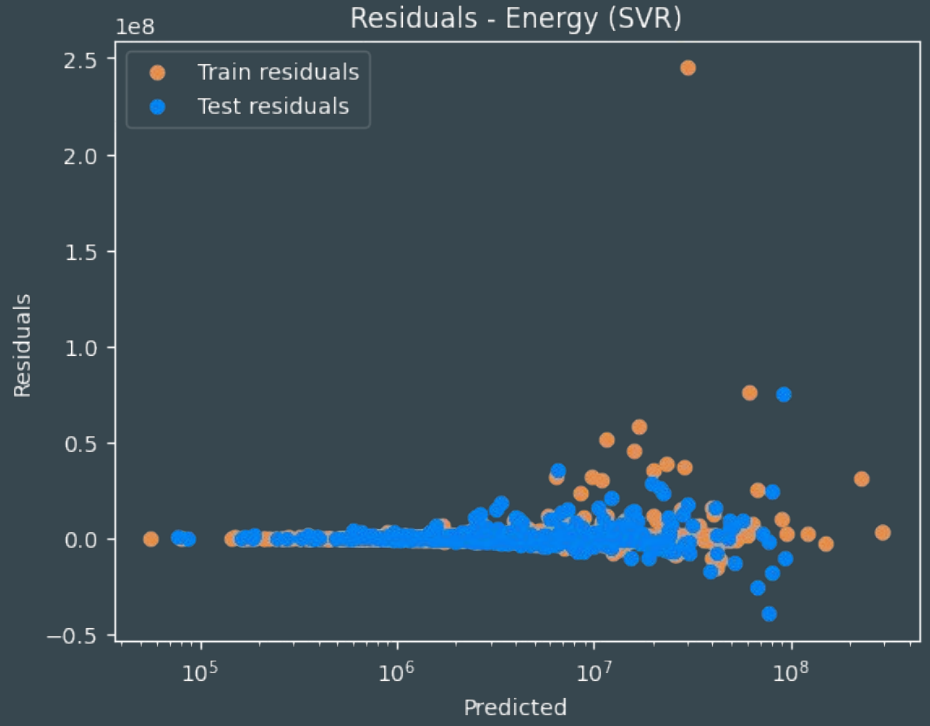
Énergie	Train	Test	Fit time
SVR (racine)	R^2: 0.85 ± 0.04 MAE: 143.10^4 ± 9.10^4	R^2: 0.78 ± 0.09 MAE: 29.10^5 ± 5.10^5	0.58 ± 0.07 s
RF (racine)	R^2: 0.93 ± 0.01 MAE: 126.10^4 ± 4.10^4	R^2: 0.6 ± 0.2 MAE: 32.10^5 ± 5.10^5	3.4 ± 0.4 s

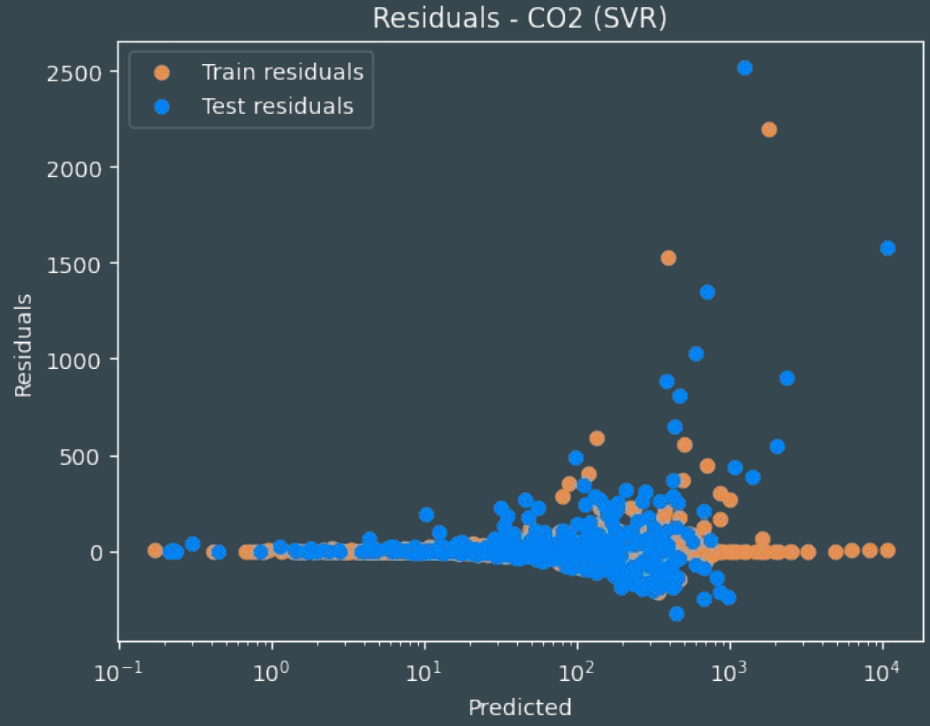
Émissions	Train	Test	Fit time
SVR (racine)	R^2: 0.977 ± 0.003 MAE: 17.1 ± 0.8	R^2: 0.87 ± 0.05 MAE: 65 ± 10	0.75 ± 0.02 s
XGB (racine)	R^2: 1.000 ± 0.001 MAE: 1.0 ± 0.1	R^2: 0.7 ± 0.2 MAE: 79 ± 9	4 ± 1 s

Prédictions résiduelles



Prédictions résiduelles





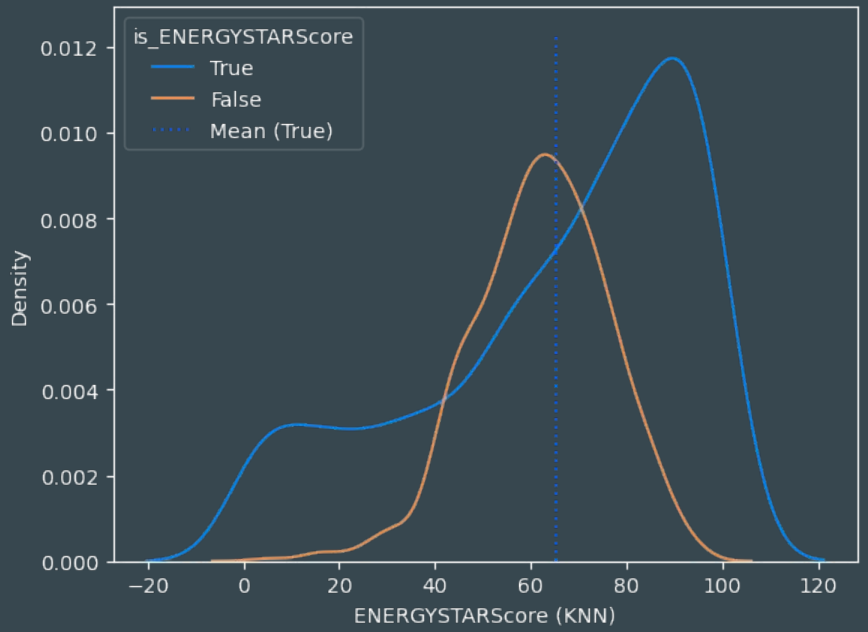
Energy star score

Énergie	Train	Test	Fit time
LR (S)	R^2: 0.477 ± 0.02 MAE: 47 ± 3.10^5	R^2: 0.4 ± 0.1 MAE: 48.10^5 ± 7.10^5	3 ± 1 ms
LR (S+is_ES)	R^2: 0.48 ± 0.02 MAE: 47 ± 2.10^5	R^2: 0.43 ± 0.09 MAE: 48 ± 4.10^5	3 ± 1 ms
LR (S+ES)	R^2: 0.490 ± 0.009 MAE: 47 ± 3.10^5	R^2: 0.44 ± 0.08 MAE: 48.10^5 ± 6.10^5	3 ± 1 ms
LR (S+ES+is_ES)	R^2: 0.49 ± 0.04 MAE: 47 ± 3.10^5	R^2: 0.47 ± 0.08 MAE: 48 ± 7.10^5	3 ± 1 ms

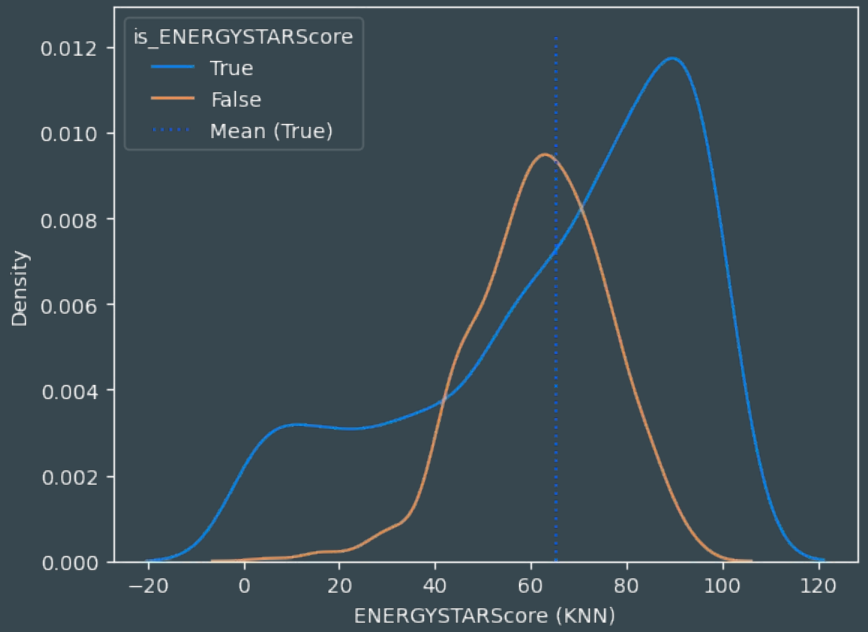
Energy star score

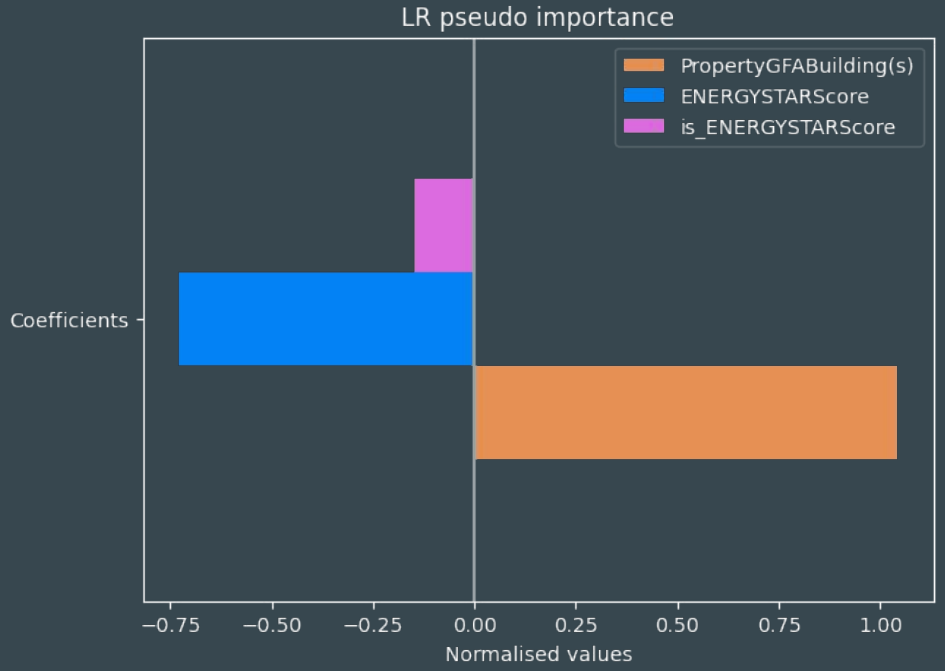
Énergie	Train	Test	Fit time
LR (S)	R^2: 0.477 ± 0.02 MAE: 47 ± 3.10^5	R^2: 0.4 ± 0.1 MAE: 48.10^5 ± 7.10^5	3 ± 1 ms
LR (S+is_ES)	R^2: 0.48 ± 0.02 MAE: 47 ± 2.10^5	R^2: 0.43 ± 0.09 MAE: 48 ± 4.10^5	3 ± 1 ms
LR (S+ES)	R^2: 0.490 ± 0.009 MAE: 47 ± 3.10^5	R^2: 0.44 ± 0.08 MAE: 48.10^5 ± 6.10^5	3 ± 1 ms
LR (S+ES+is_ES)	R^2: 0.49 ± 0.04 MAE: 47 ± 3.10^5	R^2: 0.47 ± 0.08 MAE: 48 ± 7.10^5	3 ± 1 ms

Energy star score



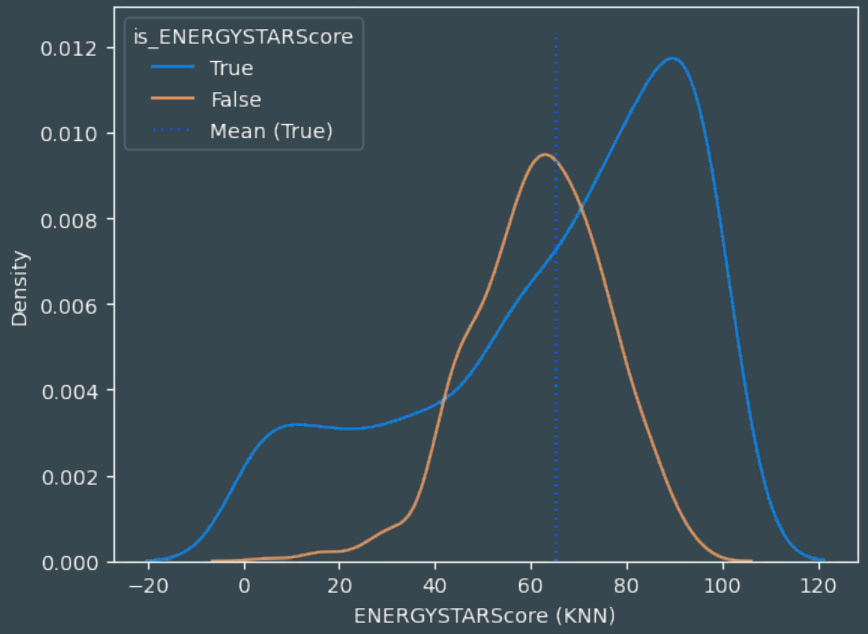
Energy star score

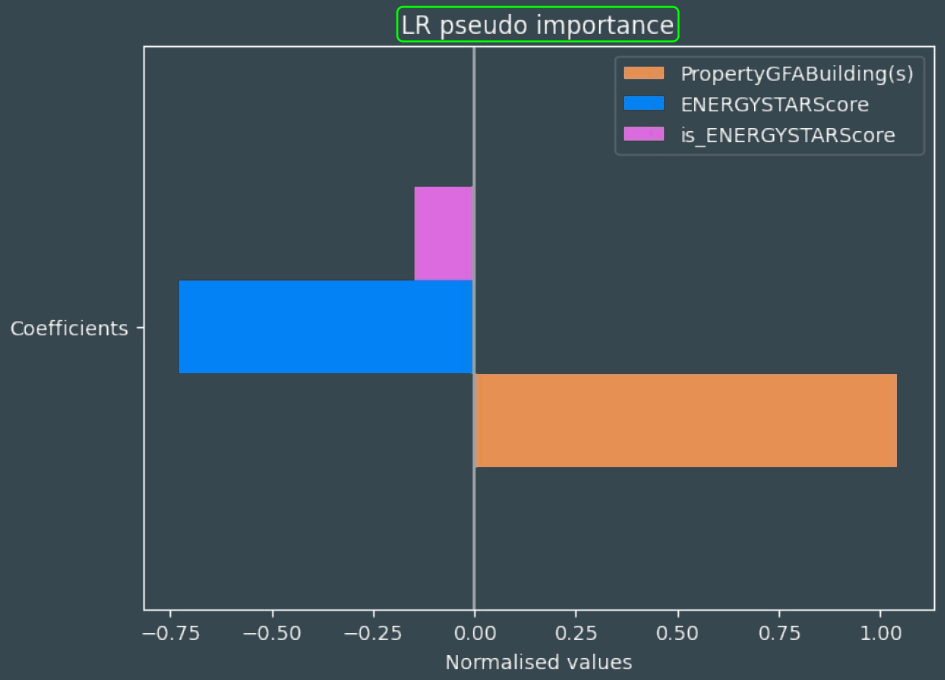




Energy star score

$$(C_{var} \times M_{var}) \div M_E$$





General conclusion

Analyse rétrospective

- Propreté des données
  - ➡ Quelques outliers (2015 pire)

General conclusion

Analyse rétrospective

- Propreté des données
  ➡ Quelques outliers (2015 pire)

- Energie score
  ➡ Contient de l'information pertinente
  ➡ Pas de motif caché

General conclusion

Analyse rétrospective

- Propreté des données
  ➡ Quelques outliers (2015 pire)

- Energie score
  ➡ Contient de l'information pertinente
  ➡ Pas de motif caché

- Feature importance
  ➡ Valeurs de shapley

General conclusion

Analyse rétrospective

- Propreté des données
  ➡ Quelques outliers (2015 pire)
- Energie score
  ➡ Contient de l'information pertinente
  ➡ Pas de motif caché
- Feature importance
  ➡ Valeurs de shapley

Axes d'amélioration

- Text embeddings
  ➡ Ajouter toutes les variables catégorielles
  ➡ Ajuster la PCA dans la modélisation

General conclusion

Analyse rétrospective

- Propreté des données
  ➡ Quelques outliers (2015 pire)
- Energie score
  ➡ Contient de l'information pertinente
  ➡ Pas de motif caché
- Feature importance
  ➡ Valeurs de shapley

Axes d'amélioration

- Text embeddings
  ➡ Ajouter toutes les variables catégorielles
  ➡ Ajuster la PCA dans la modélisation
- Analyse résiduelle
  ➡ Inspecter les différences de distribution
  ➡ Raffinement par boosting

Merci de votre attention.

...

Des questions ?