DATA SCIENTIST - PROJET 4

Anticipez les besoins en consommation de bâtiments

Arnaud CHOUX

Formation OPENCLASSROOMS
Financement région IdF

Septembre 2022

Plan

Plan

I.1. Objectifs

1. Prédire les émissions de CO2 et la consommation totale d'énergie de bâtiments non destinés à l'habitation pour lesquels les relevés sur place, coûteux, n'ont pas encore été réalisés.

2. Au cours de cette étude: Tester différents modèles de prédiction afin de répondre au mieux à la problématique.

3. Finalement, se servir de cette étude pour évaluer l'importance de l'indicateur "ENERGY STAR Score" qui est fastidieux à calculer et dont on aimerait donc se passer s'il influe peu sur nos prédictions.

I.2. Compétences abordées

- collecter des data (sans phase de recherche car la source est donnée)

- comprendre ce que représente en pratique chacun des indicateurs

- nettoyer la data

- fit et predict par régression

- comparer l'importance des indicateurs (shap)

I.3. Dataset

BuildingType	PrimaryPropertyType	PropertyName	Address	City	State	ZipCode	TaxParcelIdentificationNumber	CouncilDistrictCode	Neighborhood	Latitude	Longitude
NonResidential	Hotel	Mayflower park hotel	405 Olive way	Seattle	WA	98101.0	0659000030	7	DOWNTOWN	47.61220	-122.33799
NonResidential	Hotel	Paramount Hotel	724 Pine street	Seattle	WA	98101.0	0659000220	7	DOWNTOWN	47.61317	-122.33393
NonResidential	Hotel	5673-The Westin Seattle	1900 5th Avenue	Seattle	WA	98101.0	0659000475	7	DOWNTOWN	47.61393	-122.33810
NonResidential	Hotel	HOTEL MAX	620 STEWART ST	Seattle	WA	98101.0	0659000640	7	DOWNTOWN	47.61412	-122.33664
NonResidential	Hotel	WARWICK SEATTLE HOTEL (ID8)	401 LENORA ST	Seattle	WA	98121.0	0659000970	7	DOWNTOWN	47.61375	-122.34047

$\rightarrow$ 3376 individus, 46indicateurs

Plan

II.1. Indicateurs catégoriels

```
OSEBuildingID:                  1      len([1])
DataYear:                       1      len([3376])
BuildingType:                   8      len([1460 1018  580  110   98   85   24    1])
PrimaryPropertyType:           22      len([987 564 293 256 187 173 139 133 105  91  77  71  53  45  40...])
PropertyName:                   3      len([3 2 1])
Address:                        3      len([4 2 1])
City:                           1      len([3376])
State:                          1      len([3376])
ZipCode:                       28      len([294 251 243 230 191 186 169 167 161 152 129 124 101  99  93...])
TaxParcelIdentificationNumber:  6      len([8 5 4 3 2 1])
CouncilDistrictCode:            7      len([1037  596  509  367  338  282  247])
Neighborhood:                  19      len([573 453 423 375 280 251 210 166 145 126 107  95  80  42  27...])
ListOfAllPropertyUseTypes:     36      len([866 464 139 135 120 101  61  56  52  51  48  44  43  40  26...])
LargestPropertyUseType:        27      len([1667  498  199  139  102   99   77   71   54   46   41   32...])
SecondLargestPropertyUseType:  21      len([976 215 155  59  40  33  18  17  14  13  12  11  10   8   7...])
ThirdLargestPropertyUseType:   18      len([110 105  71  56  49  29  18  17  14  12  11   9   8   6   5...])
YearsENERGYSTARCertified:       8      len([14  8  7  6  4  3  2  1])
DefaultData:                    2      len([3263  113])
Comments:                       0      len([])
ComplianceStatus:               4      len([3211  113   37   15])
Outlier:                        2      len([23  9])
```

II.1.1. Mon « Indice de variance »

```
for co in cols_cat:
    vcu = df.loc[:,co].value_counts().unique()
    indice_de_variance = len(vcu)
```

```
OSEBuildingID:                    1     len([1])
DataYear:                         1     len([3376])
BuildingType:                     8     len([1460 1018  580  110   98   85   24    1])
PrimaryPropertyType:             22     len([987 564 293 256 187 173 139 133 105  91  77  71  53  45  40...])
PropertyName:                     3     len([3 2 1])
Address:                          3     len([4 2 1])
City:                             1     len([3376])
State:                            1     len([3376])
ZipCode:                         28     len([294 251 243 230 191 186 169 167 161 152 129 124 101  99  93...])
TaxParcelIdentificationNumber:    6     len([8 5 4 3 2 1])
CouncilDistrictCode:              7     len([1037  596  509  367  338  282  247])
Neighborhood:                    19     len([573 453 423 375 280 251 210 166 145 126 107  95  80  42  27...])
ListOfAllPropertyUseTypes:       36     len([866 464 139 135 120 101  61  56  52  51  48  44  43  40  26...])
LargestPropertyUseType:          27     len([1667  498  199  139  102   99   77   71   54   46   41   32...])
SecondLargestPropertyUseType:    21     len([976 215 155  59  40  33  18  17  14  13  12  11  10   8   7...])
ThirdLargestPropertyUseType:     18     len([110 105  71  56  49  29  18  17  14  12  11   9   8   6   5...])
YearsENERGYSTARCertified:         8     len([14  8  7  6  4  3  2  1])
DefaultData:                      2     len([3263  113])
Comments:                         0     len([])
ComplianceStatus:                 4     len([3211  113   37   15])
Outlier:                          2     len([23  9])
```

II.1.2. Indicateurs sans info

```
for co in cols_cat:
    vcu = df.loc[:,co].value_counts().unique()
    indice_de_variance = len(vcu)
```

1 valeur unique
100% de valeurs uniques
} → aucune info

```
OSEBuildingID:                    1     len([1])
DataYear:                         1     len([3376])
BuildingType:                     8     len([1460 1018  580  110   98   85   24    1])
PrimaryPropertyType:             22     len([987 564 293 256 187 173 139 133 105  91  77  71  53  45  40...])
PropertyName:                     3     len([3 2 1])
Address:                          3     len([4 2 1])
City:                             1     len([3376])
State:                            1     len([3376])
ZipCode:                         28     len([294 251 243 230 191 186 169 167 161 152 129 124 101  99  93...])
TaxParcelIdentificationNumber:    6     len([8 5 4 3 2 1])
CouncilDistrictCode:              7     len([1037  596  509  367  338  282  247])
Neighborhood:                    19     len([573 453 423 375 280 251 210 166 145 126 107  95  80  42  27...])
ListOfAllPropertyUseTypes:       36     len([866 464 139 135 120 101  61  56  52  51  48  44  43  40  26...])
LargestPropertyUseType:          27     len([1667  498  199  139  102   99   77   71   54   46   41   32...])
SecondLargestPropertyUseType:    21     len([976 215 155  59  40  33  18  17  14  13  12  11  10   8   7...])
ThirdLargestPropertyUseType:     18     len([110 105  71  56  49  29  18  17  14  12  11   9   8   6   5...])
YearsENERGYSTARCertified:         8     len([14  8  7  6  4  3  2  1])
DefaultData:                      2     len([3263  113])
Comments:                         0     len([])
ComplianceStatus:                 4     len([3211  113   37   15])
Outlier:                          2     len([23  9])
```

→ suppression des indices < 2

II.1.3. Indicateurs presque sans info

```
for co in cols_cat:
    vcu = df.loc[:,co].value_counts().unique()
    indice_de_variance = len(vcu)
```

1 valeur unique
100% de valeurs uniques
} → aucune info

```
OSEBuildingID:                    1      len([1])
DataYear:                         1      len([3376])
BuildingType:                     8      len([1460 1018  580  110   98   85   24    1])
PrimaryPropertyType:             22      len([987 564 293 256 187 173 139 133 105  91  77  71  53  45  40...])
PropertyName:                     3      len([3 2 1])
Address:                          3      len([4 2 1])
City:                             1      len([3376])
State:                            1      len([3376])
ZipCode:                         28      len([294 251 243 230 191 186 169 167 161 152 129 124 101  99  93...])
TaxParcelIdentificationNumber:    6      len([8 5 4 3 2 1])
CouncilDistrictCode:              7      len([1037  596  509  367  338  282  247])
Neighborhood:                    19      len([573 453 423 375 280 251 210 166 145 126 107  95  80  42  27...])
ListOfAllPropertyUseTypes:       36      len([866 464 139 135 120 101  61  56  52  51  48  44  43  40  26...])
LargestPropertyUseType:          27      len([1667  498  199  139  102   99   77   71   54   46   41   32...])
SecondLargestPropertyUseType:    21      len([976 215 155  59  40  33  18  17  14  13  12  11  10   8   7...])
ThirdLargestPropertyUseType:     18      len([110 105  71  56  49  29  18  17  14  12  11   9   8   6   5...])
YearsENERGYSTARCertified:         8      len([14  8  7  6  4  3  2  1])
DefaultData:                      2      len([3263  113])
Comments:                         0      len([])
ComplianceStatus:                 4      len([3211  113   37   15])
Outlier:                          2      len([23  9])
```

→ suppression car tous uniques à 4 ou 5 voisins près.

II.1.4. Indicateurs inintéressants

```
for co in cols_cat:
    vcu = df.loc[:,co].value_counts().unique()
    indice_de_variance = len(vcu)
```

```
OSEBuildingID:                  1       len([1])
DataYear:                       1       len([3376])
BuildingType:                   8       len([1460 1018  580  110   98   85   24    1])
PrimaryPropertyType:           22       len([987 564 293 256 187 173 139 133 105  91  77  71  53  45  40...])
PropertyName:                   3       len([3 2 1])
Address:                        3       len([4 2 1])
City:                           1       len([3376])
State:                          1       len([3376])
ZipCode:                       28       len([294 251 243 230 191 186 169 167 161 152 129 124 101  99  93...])
TaxParcelIdentificationNumber:  6       len([8 5 4 3 2 1])
CouncilDistrictCode:            7       len([1037  596  509  367  338  282  247])
Neighborhood:                  19       len([573 453 423 375 280 251 210 166 145 126 107  95  80  42  27...])
ListOfAllPropertyUseTypes:     36       len([866 464 139 135 120 101  61  56  52  51  48  44  43  40  26...])
LargestPropertyUseType:        27       len([1667  498  199  139  102   99   77   71   54   46   41   32...])
SecondLargestPropertyUseType:  21       len([976 215 155  59  40  33  18  17  14  13  12  11  10   8   7...])
ThirdLargestPropertyUseType:   18       len([110 105  71  56  49  29  18  17  14  12  11   9   8   6   5...])
YearsENERGYSTARCertified:       8       len([14  8  7  6  4  3  2  1])
DefaultData:                    2       len([3263  113])
Comments:                       0       len([])
ComplianceStatus:               4       len([3211  113   37   15])
Outlier:                        2       len([23  9])
```

→ suppression d'informations moins utiles, inutiles, redondantes

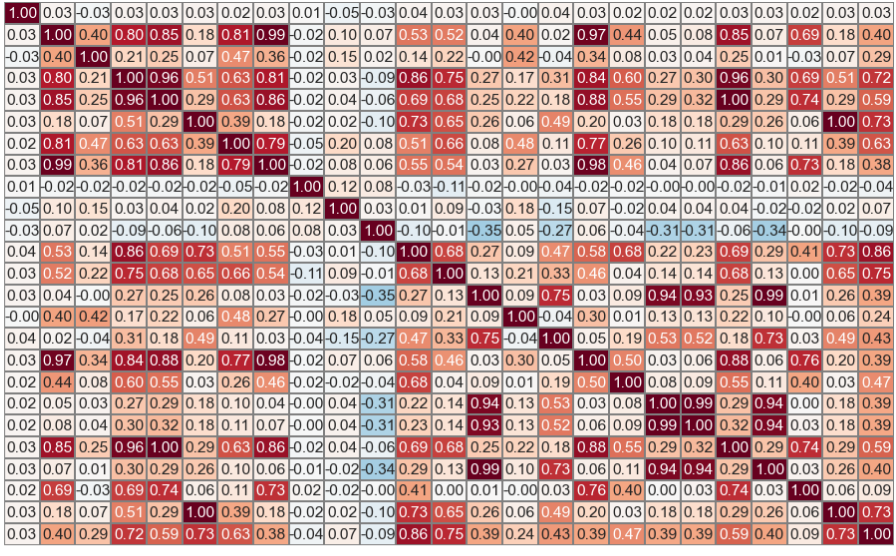
II.1.5. Indicateurs conservés

```
for co in cols_cat:
    vcu = df.loc[:,co].value_counts().unique()
    indice_de_variance = len(vcu)
```

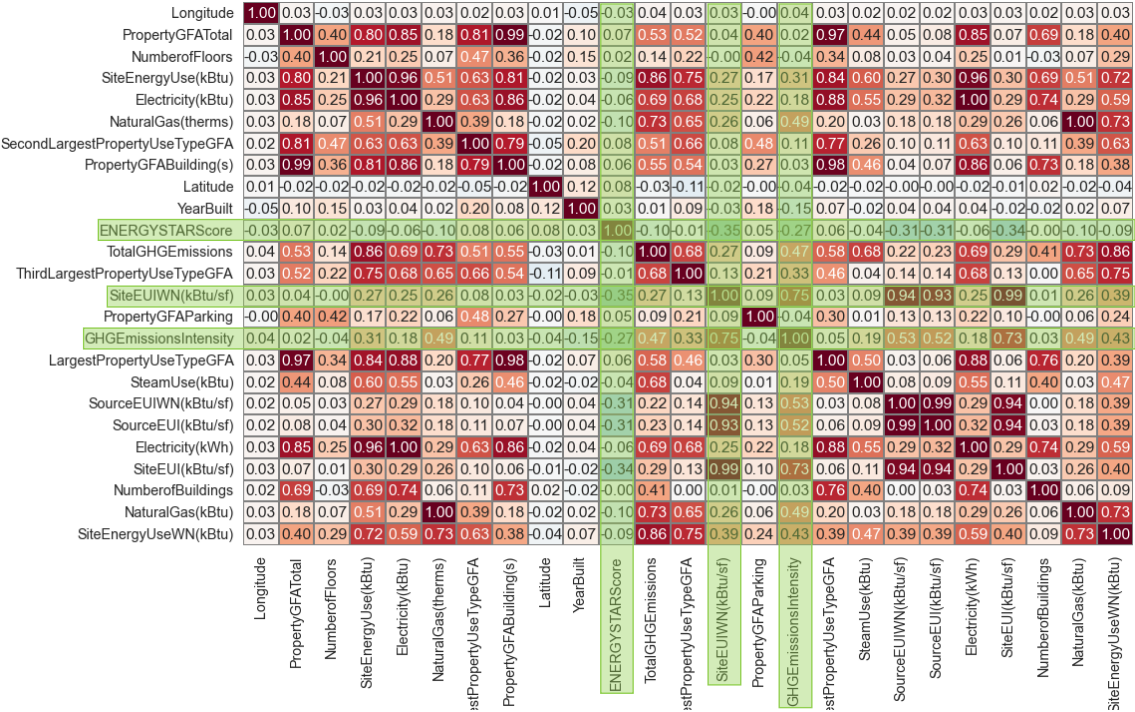
```
OSEBuildingID:                   1      len([1])
DataYear:                        1      len([3376])
BuildingType:                    8      len([1460 1018  580  110   98   85   24    1])
PrimaryPropertyType:            22      len([987 564 293 256 187 173 139 133 105  91  77  71  53  45  40...])
PropertyName:                    3      len([3 2 1])
Address:                         3      len([4 2 1])
City:                            1      len([3376])
State:                           1      len([3376])
ZipCode:                        28      len([294 251 243 230 191 186 169 167 161 152 129 124 101  99  93...])
TaxParcelIdentificationNumber:   6      len([8 5 4 3 2 1])
CouncilDistrictCode:             7      len([1037  596  509  367  338  282  247])
Neighborhood:                   19      len([573 453 423 375 280 251 210 166 145 126 107  95  80  42  27...])
ListOfAllPropertyUseTypes:      36      len([866 464 139 135 120 101  61  56  52  51  48  44  43  40  26...])
LargestPropertyUseType:         27      len([1667  498  199  139  102   99   77   71   54   46   41   32...])
SecondLargestPropertyUseType:   21      len([976 215 155  59  40  33  18  17  14  13  12  11  10   8   7...])
ThirdLargestPropertyUseType:    18      len([110 105  71  56  49  29  18  17  14  12  11   9   8   6   5...])
YearsENERGYSTARCertified:        8      len([14  8  7  6  4  3  2  1])
DefaultData:                     2      len([3263  113])
Comments:                        0      len([])
ComplianceStatus:                4      len([3211  113   37   15])
Outlier:                         2      len([23  9])
```

→ BuildingType et PrimaryPropertyType semblent importants pour la régression.
→ dans l'idéal: traitement des [… 1] pour éviter handle_unknown dans le OHE.

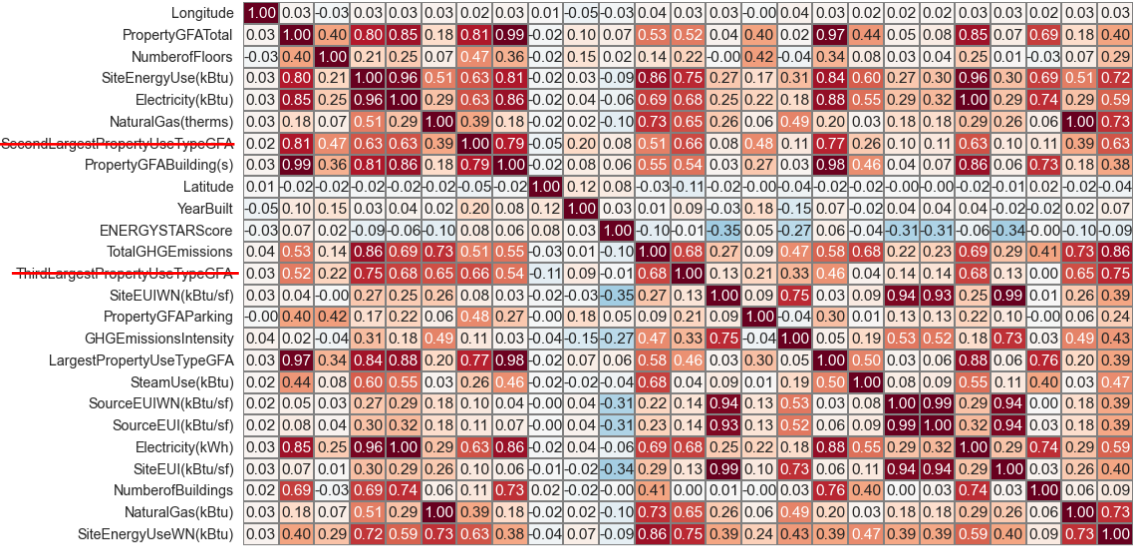
II.2. Indicateurs quantitatifs: corrélations



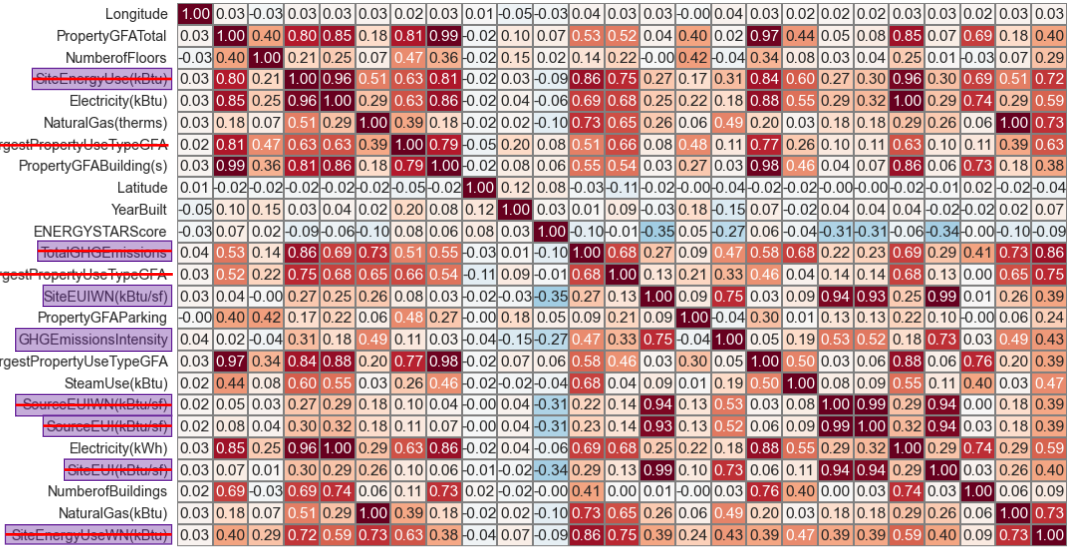
II.2.1. Targets et feature particulière



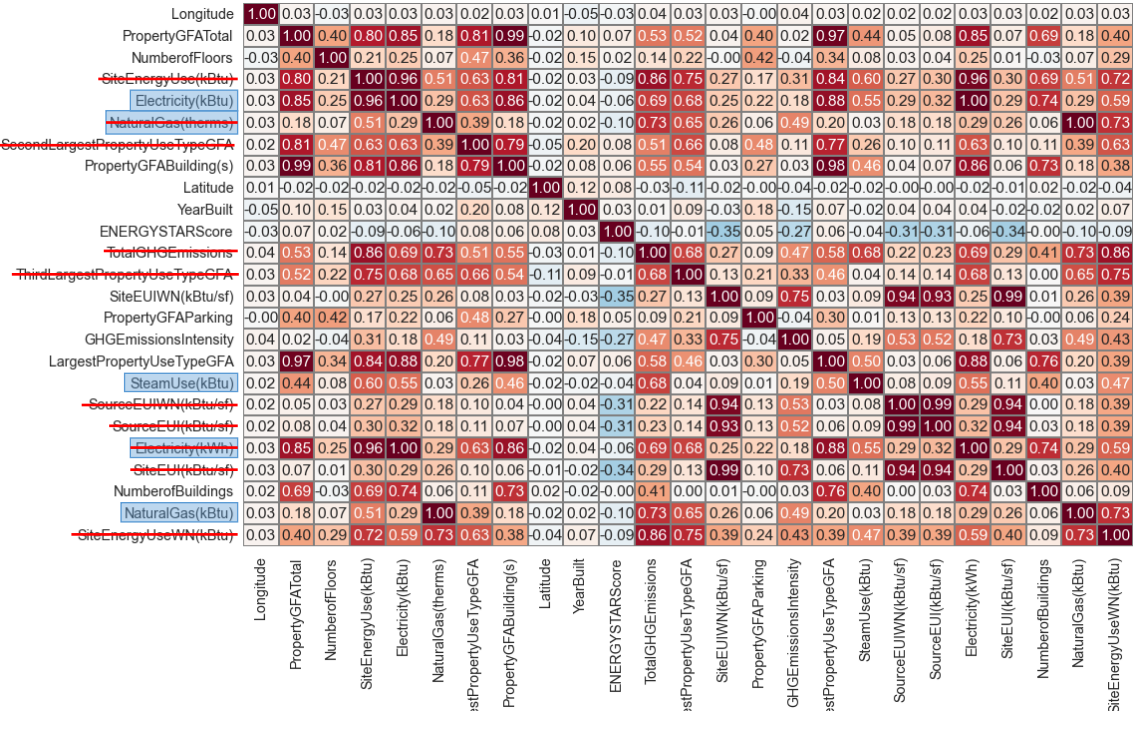
II.2.2. Features inintéressantes



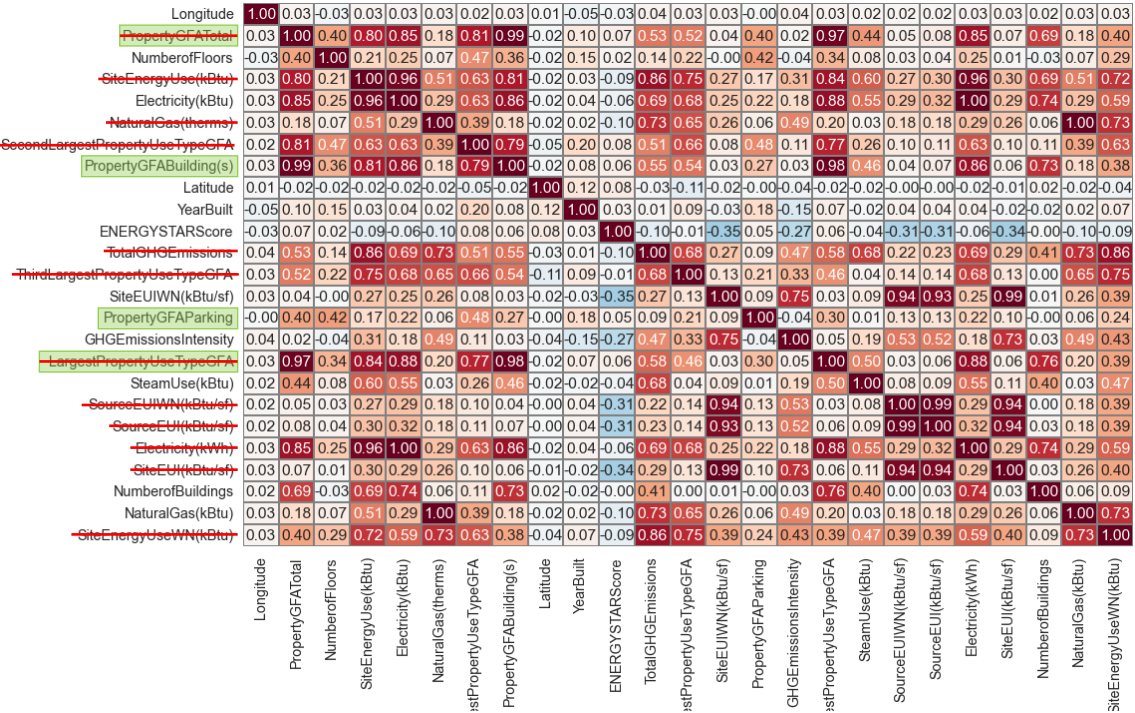
II.2.3. Features redondantes



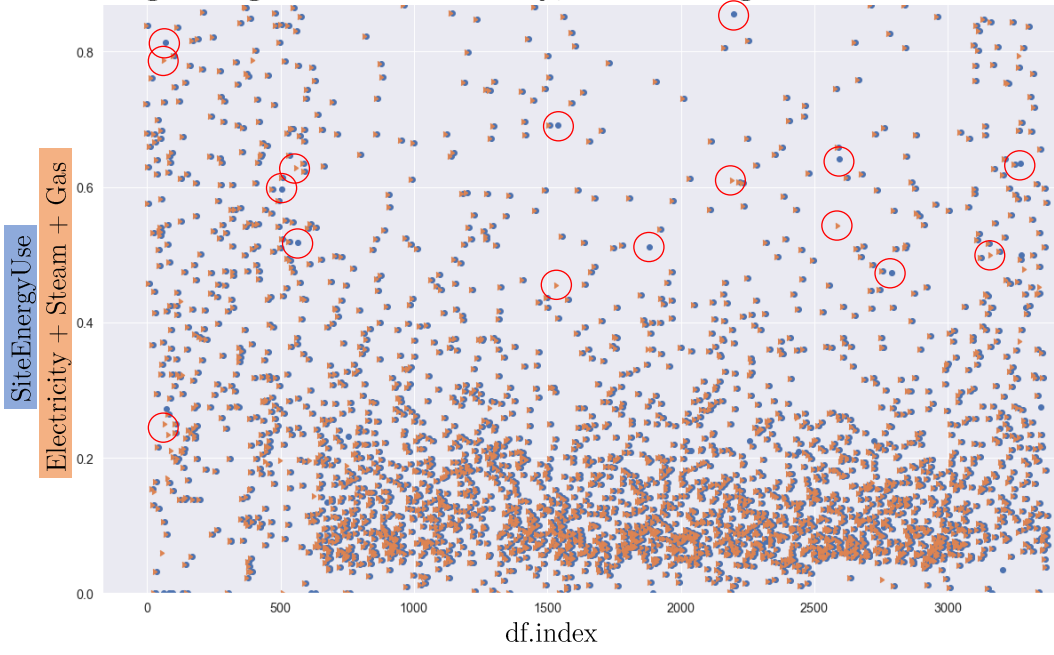
II.2.3. Features redondantes



II.2.3. Features redondantes

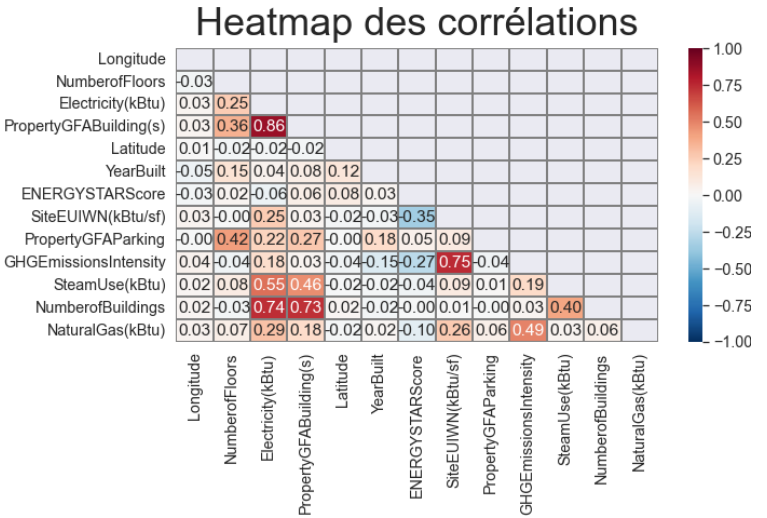


II.2.4. Leakage en gardant electricity, steam et gas ?



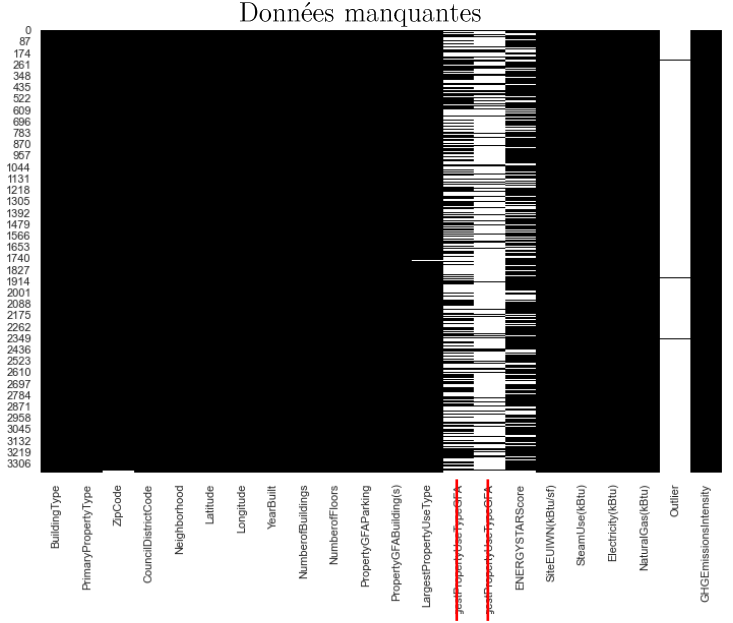
→ données identiques à qques exceptions près
→ SiteEUIWN ≈ (Electricity + Steam + Gas)/GFA

II.2.5. Corrélations



→ corrélations de {n_buildings, GFA_buildings et Electricity} et {GHGEI et SiteEUIWN} élevées mais pas alarmantes.

II.3. Indicateurs peu remplis



→ ENERGYSTARScore voulu.

→ Outlier va nous servir.

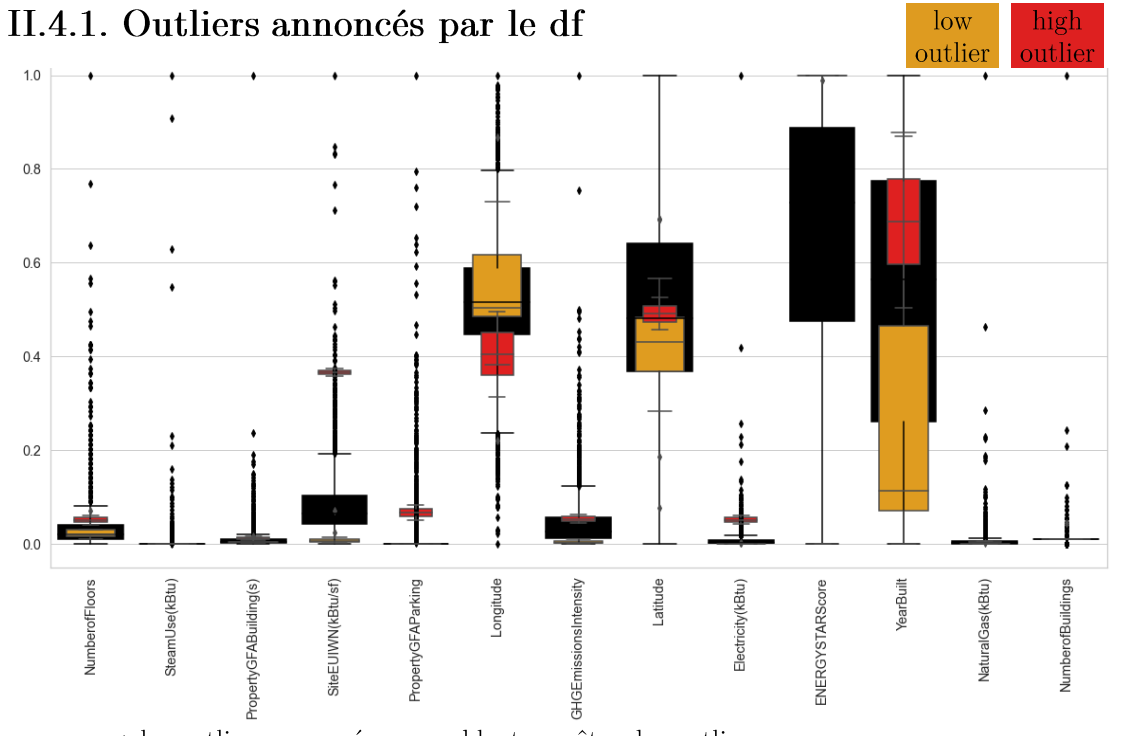
II.4. Suppression d'éléments

- Bâtiments non résidentiels (énoncé) → suppression de la moitié des lignes du df.
- Deux éléments supprimés car presque vides

...Parking	PropertyGFABuilding(s)	LargestPropertyUseType	ENERGYSTARScore	SiteEUIWN(kBtu/sf)	SteamUse(kBtu)	Electricity(kBtu)	NaturalGas(kBtu)	Outlier	GHGEmissionsIntensity
0	63150	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
0	20760	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

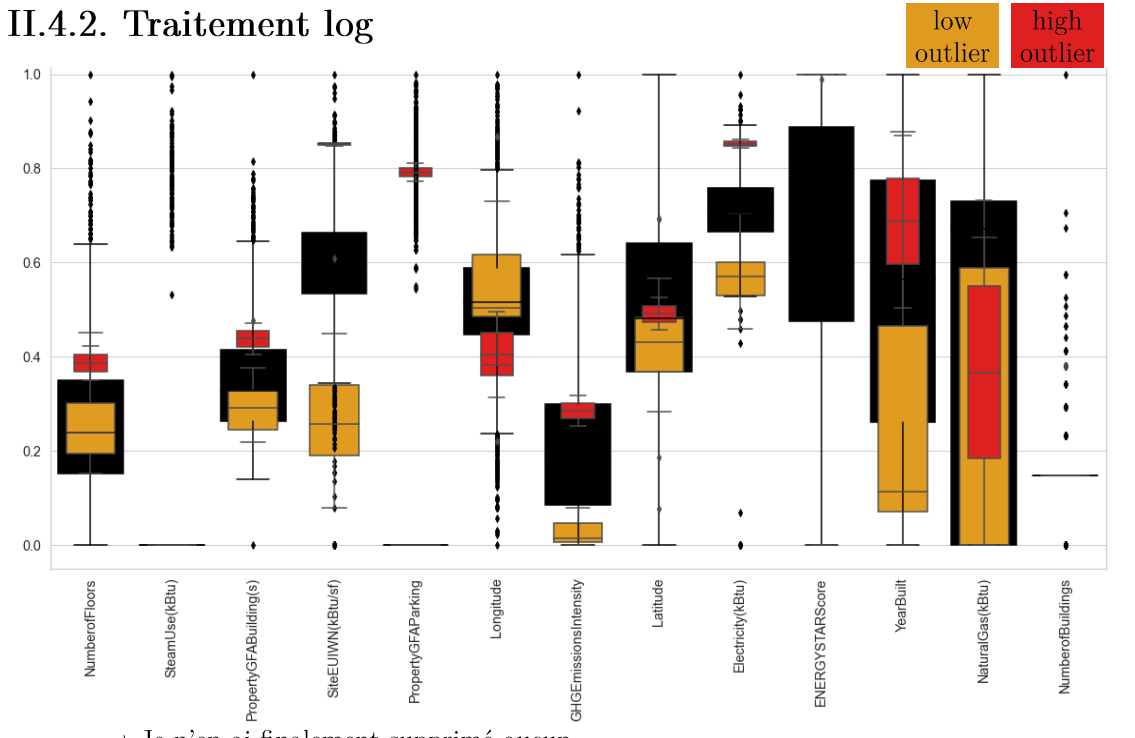
- Outliers

II.4.1. Outliers annoncés par le df



→ les outliers annoncés ne semblent pas être des outliers.
→ illisible, il faut certainement regarder en échelle log.

II.4.2. Traitement log



→ Je n'en ai finalement supprimé aucun.

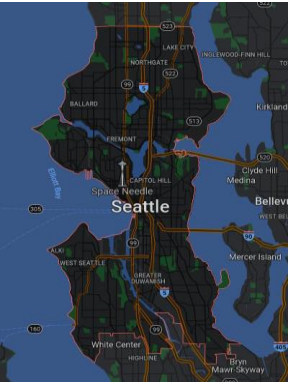
→ Faut-il garder ce traitement log pour la pipeline de régression ?

II.5. Imputation

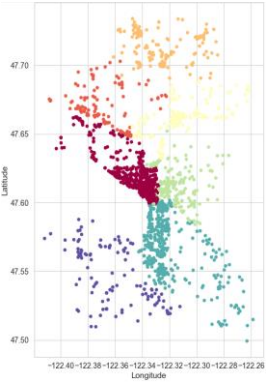
- Imputation de LargestPropertyType par most frequent (mode).
- Imputation de ZipCode par une constante (car OHE ensuite).
- Imputation de SiteEUIWN(kBtu/sf), Electricity(kBtu) et GHGEmissionsIntensity par knn.

Interlude

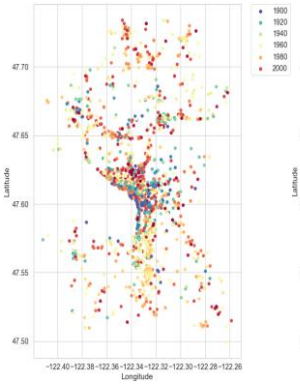
Plan de ville



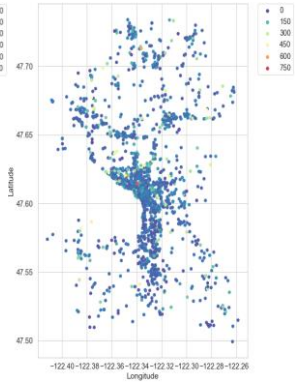
CouncilDistrictCode



YearBuilt



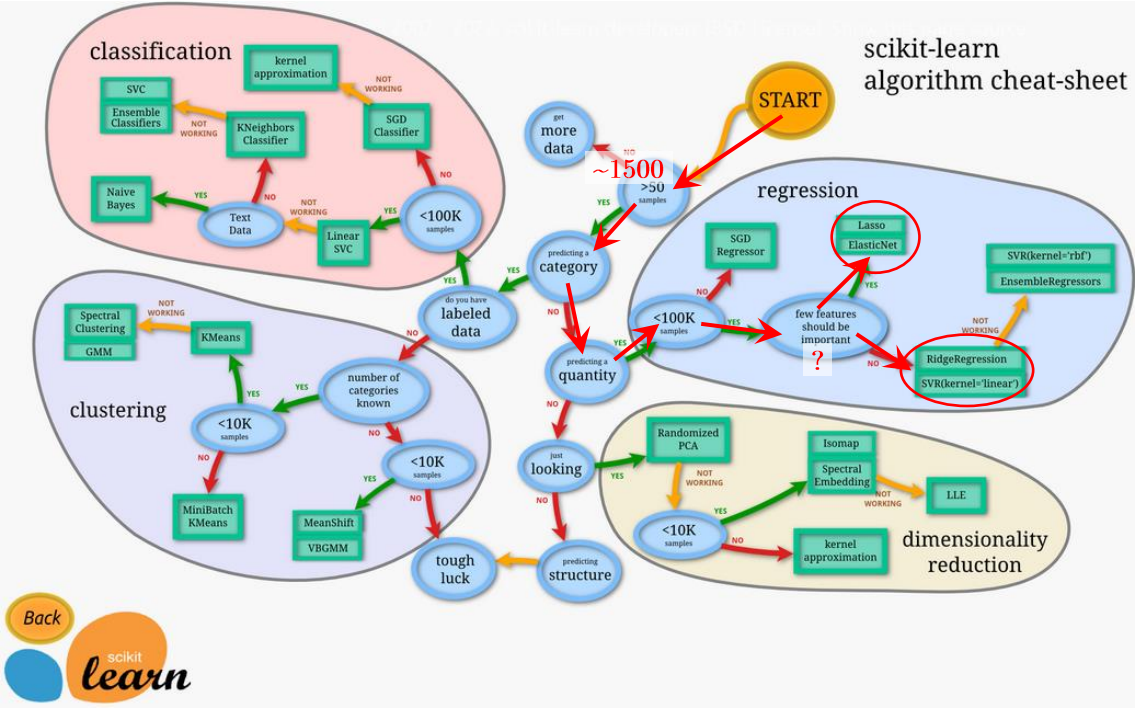
SiteEUIWN(kBtu/sf)



La géographie de la ville semble (encore une fois) peu influer sur nos targets.

Plan

III.1.1. Identification du problème



III.1.2. Paramétrage

train_test_split()

```
encoders = [OrdinalEncoder(handle_unknown='use_encoded_value', unknown_value=-1), OneHotEncoder(sparse=False,handle_unknown='ig
scalers = [RobustScaler(), StandardScaler()]
```

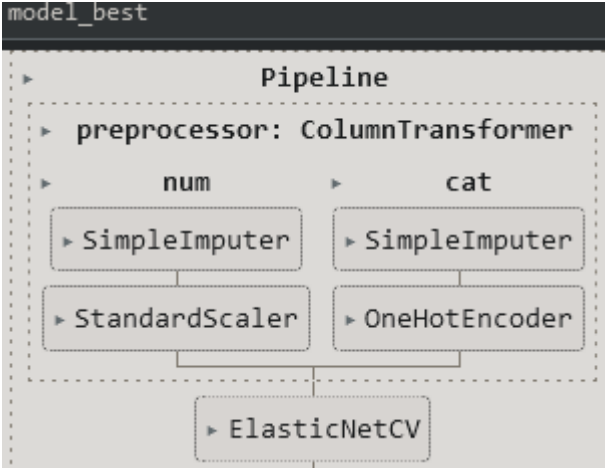
	estimator	param_names	param_nums
0	Ridge()	[ridge__alpha]	[[0.1, 0.3, 1, 3, 10]]
1	Lasso()	[lasso__alpha]	[[0.1, 0.3, 1, 3, 10]]
2	DummyRegressor()	[dummyregressor__strategy]	[[mean], [median]]
3	LinearRegression()	[linearregression__fit_intercept]	[[True]]
4	ElasticNetCV()	[elasticnetcv__l1_ratio, elasticnetcv__n_alphas]	[[0.1, 0.3, 1], [30, 100]]

III.1.3. Scoring

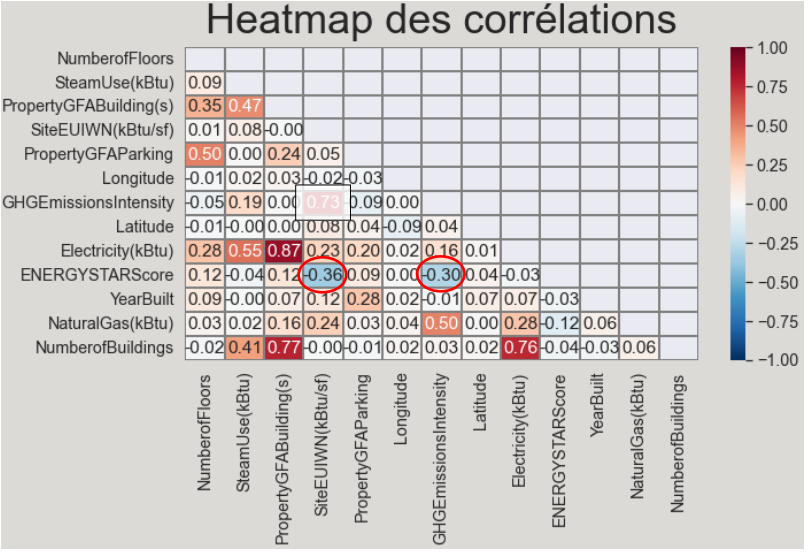
Estimator	best score	r2	mae	msqe	time
Ridge()	1.279617e-01	-1.670998e+00	4.305065e+01	1.303636e+04	0.078462
Lasso()	2.896550e-01	-1.192557e+00	4.271999e+01	1.070123e+04	0.079444
DummyRegressor()	-2.075314e-03	-4.763885e-04	4.547045e+01	4.883032e+03	0.089119
LinearRegression()	1.182682e-01	-1.674003e+00	4.305699e+01	1.305103e+04	0.088264
ElasticNetCV()	-6.414048e-03	-4.763885e-04	4.547045e+01	4.883032e+03	0.201605

```
time_1 = time_passed/(param_loops)
```

III.1.4. best_estimator_

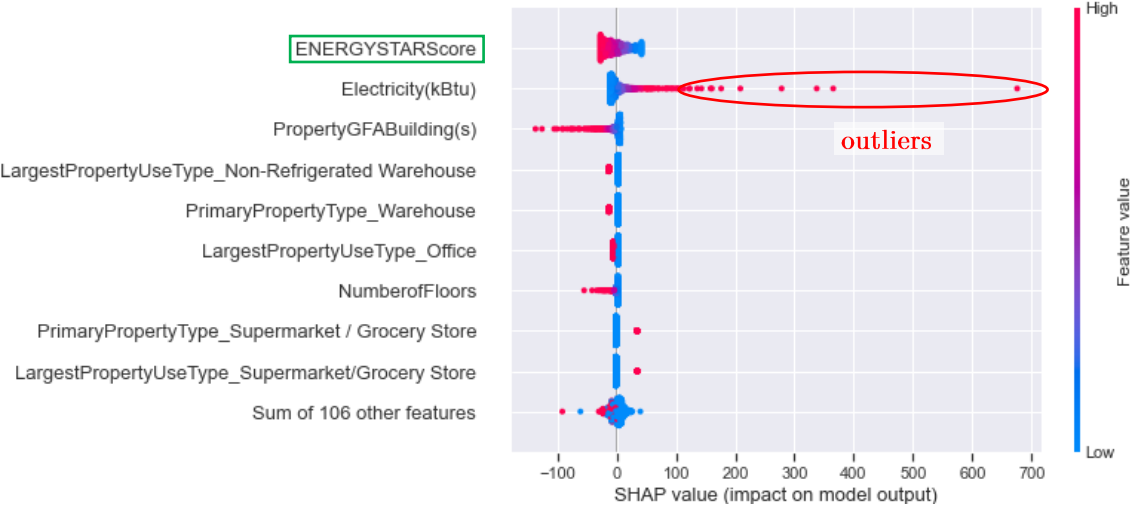


III.2.1. Corrélation



- Feature principale pour **SiteEUIWN(kBtu/sf)** et majeure pour GHGEmissionsIntensity.

III.2.2. shapley values



- Feature principale pour **SiteEUIWN(kBtu/sf)**

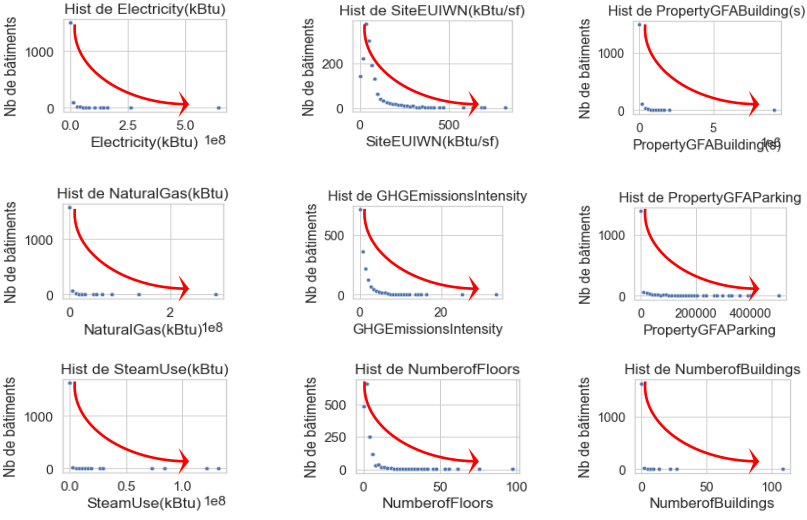
III.2.3. r^2 AVEC/SANS ENERGYSTARScore

```
r2_E-r2_0, mae_E-mae_0, mse_E-mse_0

(0.05465912749765389, -5.119671419303817, -300.27286418069525)
```

- Scores différents → feature non-négligeable
- r^2 supérieur, mae et mse inférieures → E* aide la prédiction

III.3.1. Nettoyages nécessaires



- Distribution exp des features
- r^2 très bas ($< .5$)
- affichage des shapley values illisible (outliers)

III.3.2. log

indicateur	skewness		skewness
NumberofFloors	5.106580	→	1.222773
SteamUse(kBtu)	18.518288	→	3.485820
PropertyGFABuilding(s)	4.667818	→	0.963639
SiteEUIWN(kBtu/sf)	4.685968	→	-0.835211
PropertyGFAParking	4.912275	→	1.360725
Longitude	0.030541		0.030541
GHGEmissionsIntensity	6.380721	→	1.213944
Latitude	0.174232		0.174232
Electricity(kBtu)	9.139249	→	-1.332960
ENERGYSTARScore	-0.775344		-0.775344
YearBuilt	-0.392232		-0.392232
NaturalGas(kBtu)	21.417662	→	-0.744709
NumberofBuildings	10.426625	→	2.711552

III.3.3. interquartile

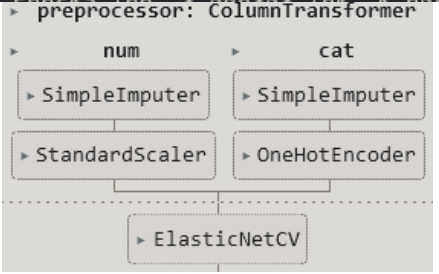
indicateur	# outliers bas	# outliers hauts
YearBuilt	0	0
Latitude	0	0
NaturalGas(kBtu)	0	0
PropertyGFABuilding(s)	1	23
SiteEUIWN(kBtu/sf)	48	33
Longitude	65	62
Electricity(kBtu)	2	2
ENERGYSTARScore	0	0
GHGEmissionsIntensity	0	7

- Suppression seulement des éléments qui sont toujours hors interquartiles même après le log.
- (suppressions selon Longitude pas forcément nécessaires)

III.4.1. best_estimator_

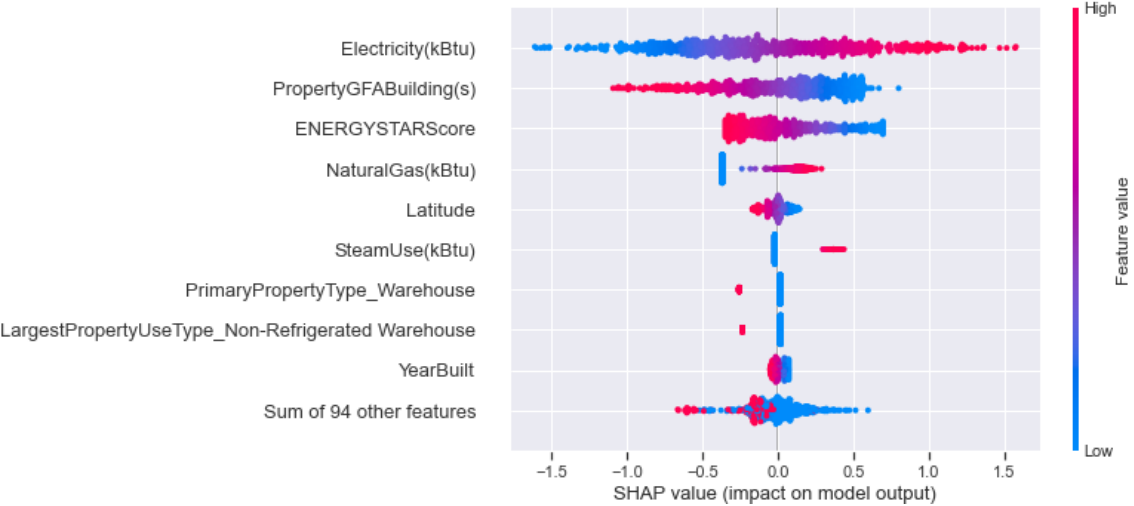
```
print(dfscores2[col_score] - dfscores[col_score])
```

	best score	r2	mae	msqe	time
0	5.456720e-01	2.220879e+00	-4.266704e+01	-1.303612e+04	-0.009930
1	8.160592e-02	1.537467e+00	-4.227454e+01	-1.070087e+04	-0.012977
2	-6.066227e-03	4.737353e-04	-4.491818e+01	-4.882494e+03	-0.022371
3	5.547913e-01	2.221856e+00	-4.267244e+01	-1.305078e+04	-0.018669
4	6.804202e-01	5.530310e-01	-4.508878e+01	-4.882791e+03	-0.009507
5	4.792426e-01	2.145800e+00	-4.280415e+01	-1.266811e+04	-0.006643



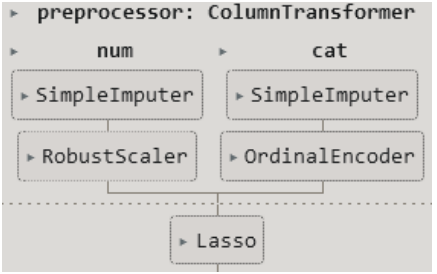
- Meilleurs scores dans les pipelines (sauf le Dummy)
- best_estimator_ similaire
- Shapley values mieux représentées

III.4.2. shapley values

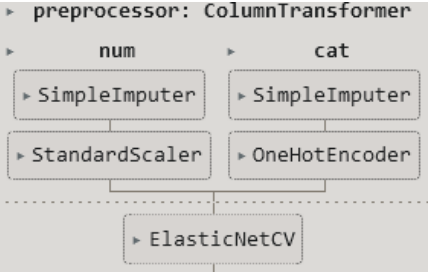


- E* n'apparaît plus comme la feature principale pour **SiteEUIWN(kBtu/sf)** mais reste majeure.

III.5.1 GHGEmissionsIntensity



nettoyage

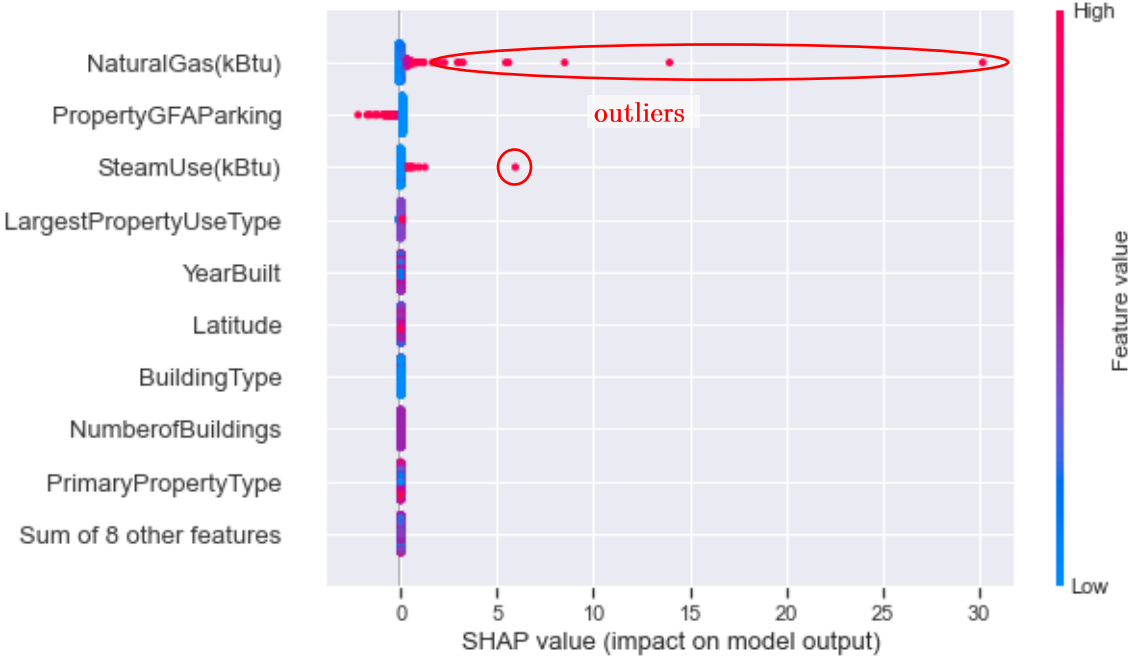



```
r2_E-r2_0, mae_E-mae_0, mse_E-mse_0

(0.0, 0.0, 0.0)
```

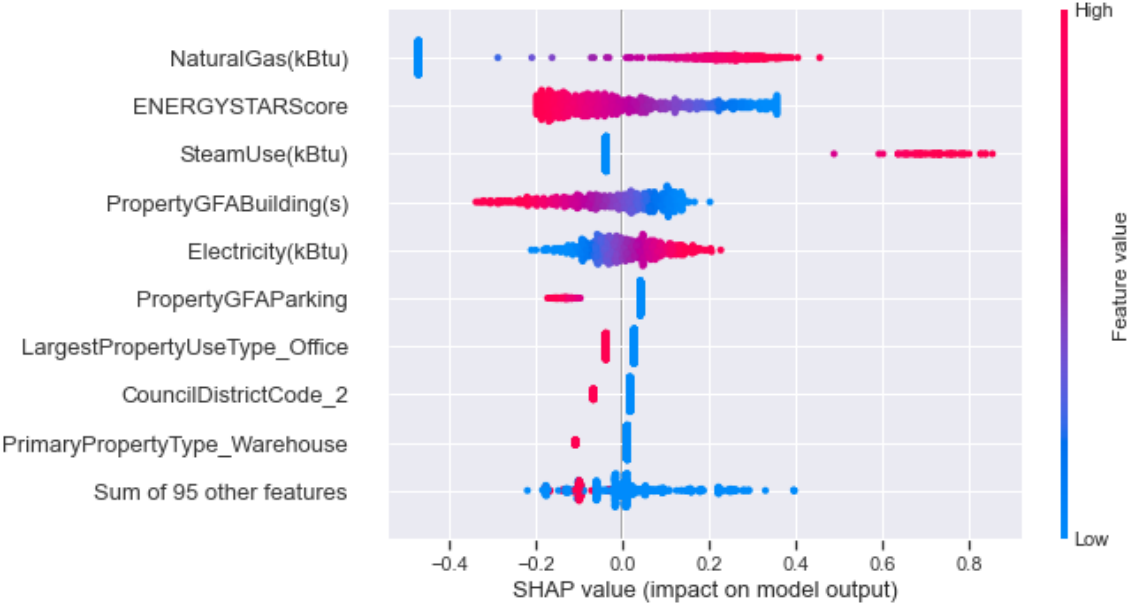
- Pour cette target je trouve une importance nulle de l'ENERGYSTARScore avant nettoyage {log + outliers} malgré ce qu'annonce le tableau de corrélations.
- Le best_estimator est modifié après nettoyage.

III.5.2. shapley values



- ENERGYSTARScore négligeable avant nettoyage

III.5.3. shapley values



- ENERGYSTARScore majeur après nettoyage

CCL

- La meilleure pipeline est trouvée.
- Preuves que l'ENERGYSTARScore est indispensable aux prédictions.
- Note: $1^{\text{ère}}$ recherche de meilleure pipeline sur tous éléments mais sans E*.

Questions ?

V. Approfondissement suite à la soutenance

V.1. Recommandations ajoutées

V.1.1. Correction d'un data leakage

SteamUse(y/n)	Electricity(y/n)	NaturalGas(y/n)
True	True	True
False	True	True
True	True	True
False	True	True

V. Approfondissement suite à la soutenance

V.1. Recommandations ajoutées

V.1.2. Ajout d'estimateurs non-linéaires

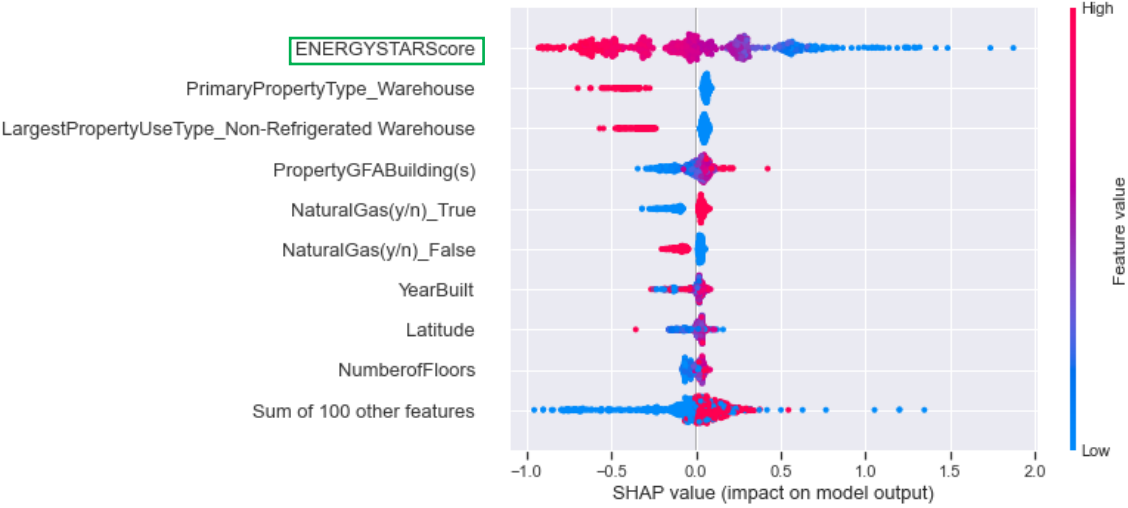
	estimator	param_names	param_nums
0	Ridge()	[ridge__alpha]	[[0.1, 0.3, 1, 3, 10]]
1	Lasso()	[lasso__alpha]	[[0.1, 0.3, 1, 3, 10]]
2	DummyRegressor()	[dummyregressor__strategy]	[[mean], [median]]
3	LinearRegression()	[linearregression__fit_intercept]	[[True]]
4	ElasticNetCV()	[elasticnetcv__l1_ratio, elasticnetcv__n_alphas]	[[0.1, 0.3, 1], [30, 100]]
5	RandomForestRegressor()	[randomforestregressor__max_depth, randomfores...	[[2], [0]]
6	GradientBoostingRegressor()	[gradientboostingregressor__random_state]	[[0]]
7	AdaBoostRegressor()	[adaboostregressor__n_estimators, adaboostregr...	[[30, 100], [0]]

V.2. Résultats

V.2.1. best_estimator_ (SiteEUIWN)

Encoder	Scaler	Estimator	best score
OneHotEncoder(handle_unknown='ignore', sparse=...	StandardScaler()	Ridge()	0.59
OneHotEncoder(handle_unknown='ignore', sparse=...	RobustScaler()	ElasticNetCV()	0.60
OneHotEncoder(handle_unknown='ignore', sparse=...	RobustScaler()	GradientBoostingRegressor()	0.60
OrdinalEncoder(handle_unknown='use_encoded_val...	StandardScaler()	GradientBoostingRegressor()	0.60
OneHotEncoder(handle_unknown='ignore', sparse=...	StandardScaler()	ElasticNetCV()	0.60
OrdinalEncoder(handle_unknown='use_encoded_val...	RobustScaler()	GradientBoostingRegressor()	0.60
OneHotEncoder(handle_unknown='ignore', sparse=...	StandardScaler()	GradientBoostingRegressor()	0.61

V.2.2. shapley values (SiteEUIWN)



- Feature principale pour **SiteEUIWN(kBtu/sf)**

V.2.3. r^2 avec/sans E*score (SiteEUIWN)

r^2 avec E*score r^2 sans E*score

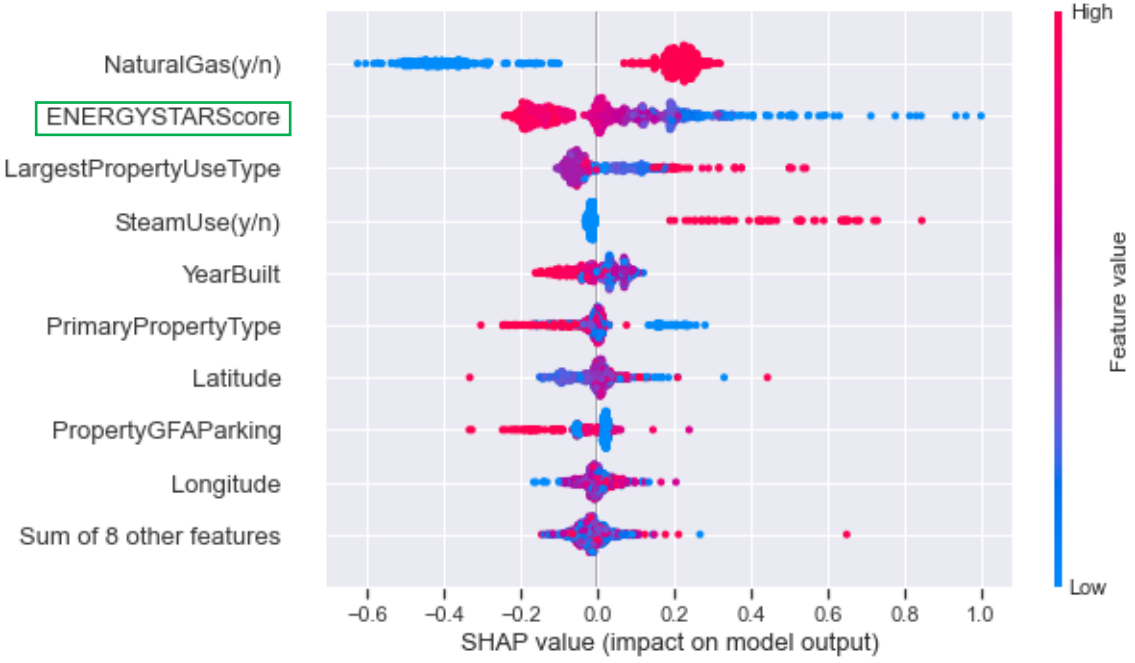
```
0.6122466721288635, 0.3000463312617414
```

- E*score indispensable

V.2.4. best_estimator_ (GHGEI)

Encoder	Scaler	Estimator	best score
OneHotEncoder(handle_unknown='ignore', sparse=...	RobustScaler()	Ridge()	0.50
OneHotEncoder(handle_unknown='ignore', sparse=...	RobustScaler()	ElasticNetCV()	0.51
OneHotEncoder(handle_unknown='ignore', sparse=...	StandardScaler()	ElasticNetCV()	0.51
OneHotEncoder(handle_unknown='ignore', sparse=...	StandardScaler()	GradientBoostingRegressor()	0.52
OrdinalEncoder(handle_unknown='use_encoded_val...	StandardScaler()	GradientBoostingRegressor()	0.52
OneHotEncoder(handle_unknown='ignore', sparse=...	RobustScaler()	GradientBoostingRegressor()	0.52
OrdinalEncoder(handle_unknown='use_encoded_val...	RobustScaler()	GradientBoostingRegressor()	0.52

V.2.5. shapley values (GHGEI)



- Feature importante pour **GHGEI**

V.2.6. r^2 avec/sans E*score (GHGEI)

r^2 avec E*score r^2 sans E*score

```
0.5176801383113236, 0.3851714287519793
```

- E*score indispensable

CCL

- Un important data leakage solutionné.
- La meilleure pipeline est trouvée, avec davantage d'estimateurs.
- Preuves que l'ENERGYSTARScore est indispensable aux prédictions.

Questions ?