DATA SCIENTIST - PROJET 4

Anticipez les besoins en consommation de bâtiments

Arnaud CHOUX

Formation OPENCLASSROOMS Financement région IdF





Plan		
Ι	Présentation du projet	
II	Nettoyage	
III	Traitement	
IV	Conclusions	

I	Présentation du projet		
		1.	Objectifs
11	Nettoyage	2.	Compétences
III	Traitement		abordées
TX 7		3.	Dataset
IV	Conclusions		

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

 \rightarrow 3376 individus, 46
indicateurs

Plan

Ι	Présentation du projet	1.	Indicateurs catégoriels
II	Nettoyage	2.	Indicateurs quantitatifs
		3.	Indicateurs vides
III	Traitement	4.	Suppression d'éléments
IV	Conclusions	5.	Imputation

II.1. Indicateurs catégoriels

OSEBuildingID:

```
DataYear:
BuildingType:
PrimaryPropertyType:
PropertyName:
Address:
City:
State:
ZipCode:
TaxParcelIdentificationNumber:
CouncilDistrictCode:
Neighborhood:
ListOfAllPropertyUseTypes:
LargestPropertyUseType:
SecondLargestPropertyUseType:
ThirdLargestPropertyUseType:
YearsENERGYSTARCertified:
DefaultData:
Comments:
ComplianceStatus:
Outlier:
```

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:
                                        len([1])
                                        len([3376])
DataYear:
BuildingType:
                                        len([1460 1018 580 110
PrimaryPropertyType:
                                        len([987 564 293 256 187 173 139 133 105 91 77 71 53 45 40...])
                                        len([3 2 1])
PropertyName:
                                  3
Address:
                                        len([4 2 1])
City:
                                        len([3376])
                                        len([3376])
State:
                                        len([294 251 243 230 191 186 169 167 161 152 129 124 101 99 93...])
ZipCode:
                                 28
TaxParcelIdentificationNumber:
                                        len([8 5 4 3 2 1])
CouncilDistrictCode:
                                        len([1037 596 509 367 338 282 247])
Neighborhood:
                                        len([573 453 423 375 280 251 210 166 145 126 107 95 80 42 27...])
ListOfAllPropertyUseTypes:
                                        len([866 464 139 135 120 101 61 56 52 51
                                 36
                                        len([1667 498 199 139
LargestPropertyUseType:
                                 27
```

18

0

SecondLargestPropertyUseType:

ThirdLargestPropertyUseType:

YearsENERGYSTARCertified:

DefaultData:

ComplianceStatus:

Comments:

Outlier:

98 85 24

102

len([110 105 71 56 49 29 18 17 14 12 11

len([976 215 155 59 40

len([3263 113]) len([])

len([3211 113

len([23 9])

len([14 8 7 6 4 3 2 1])

99

33 18 17 14

48 44

54 46

13 12 11 10

26...])

32...])

5...])

II.1.2. Indicateurs sans info

```
for co in cols cat:
                                                                     1 valeur unique
                                                                                           → aucune info
     vcu = df.loc[:,co].value counts().unique()
                                                                100% de valeurs uniques
     indice de variance = len(vcu)
OSEBuildingID:
                                         len([1])
DataYear:
                                         len([3376])
BuildingType:
                                  8
                                         len([1460 1018 580 110
                                                                  98
                                                                       85
PrimaryPropertyType:
                                         len([987 564 293 256 187 173 139 133 105
                                                                                 91 77
                                                                                       71 53 45 40...])
                                         len([3 2 1])
PropertyName:
Address:
                                         len([4 2 1])
City:
                                         len([3376])
State:
                                         len([3376])
ZipCode:
                                         len([294 251 243 230 191 186 169 167 161 152 129 124 101 99 93...])
                                 28
TaxParcelIdentificationNumber:
                                         len([8 5 4 3 2 1])
CouncilDistrictCode:
                                         len([1037 596 509 367 338 282 247])
Neighborhood:
                                         len([573 453 423 375 280 251 210 166 145 126 107
                                                                                        95 80 42
                                                                                                    27...])
ListOfAllPropertyUseTypes:
                                         len([866 464 139 135 120 101 61 56 52
                                 36
                                                                                    48
                                                                                        44
                                                                                                    26...])
                                                                 102
LargestPropertyUseType:
                                         len([1667 498 199 139
                                                                       99
                                                                                     54
                                                                                          46
SecondLargestPropertyUseType:
                                         len([976 215 155 59
                                                             40
                                                                 33 18
                                                                            14
ThirdLargestPropertyUseType:
                                         len([110 105 71 56 49 29 18
                                 18
                                                                        17
                                                                            14
                                                                                12
                                                                                                     5...])
YearsENERGYSTARCertified:
                                  8
                                         len([14 8 7 6 4 3 2 1])
DefaultData:
                                         len([3263 113])
Comments:
                                  0
                                         len([])
ComplianceStatus:
                                         len([3211 113
                                  4
Outlier:
                                         len([23 9])
```

II.1.3. Indicateurs presque sans info

```
for co in cols cat:
                                                                    1 valeur unique
                                                                                          → aucune info
     vcu = df.loc[:,co].value counts().unique()
                                                                100% de valeurs uniques
     indice de variance = len(vcu)
BuildingType:
                                  8
                                        len([1460 1018 580 110
                                                                  98
                                                                       85
PrimaryPropertyType:
                                        len([987 564 293 256 187 173 139 133 105
                                                                                91 77
                                                                                      71 53 45 40...])
PropertyName:
                                  3
                                        len([3 2 1])
                                        len([4 2 1])
Address:
                                  3
                                         len([294 251 243 230 191 186 169 167 161 152 129 124 101 99 93...])
ZipCode:
                                 28
TaxParcelIdentificationNumber:
                                        len([8 5 4 3 2 1])
CouncilDistrictCode:
                                         len([1037 596 509 367 338 282 247])
Neighborhood:
                                        len([573 453 423 375 280 251 210 166 145 126 107
                                                                                                   27...])
ListOfAllPropertyUseTypes:
                                        len([866 464 139 135 120 101 61 56 52
                                 36
                                                                                    48
                                                                                                   26...1)
LargestPropertyUseType:
                                        len([1667 498 199 139
                                                                102
                                                                       99
                                                                                     54
                                                                                                   32...])
SecondLargestPropertyUseType:
                                        len([976 215 155 59
                                                             40
                                                                 33 18
                                                                            14
ThirdLargestPropertyUseType:
                                        len([110 105 71 56 49 29 18
                                 18
                                                                        17
                                                                           14
                                                                               12
YearsENERGYSTARCertified:
                                  8
                                        len([14 8 7 6 4 3 2 1])
DefaultData:
                                         len([3263 113])
ComplianceStatus:
                                         len([3211 113
                                  4
Outlier:
                                        len([23 9])
```

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

ZipCode: len([294 251 243 230 191 186 169 167 161 152 129 124 101 99 93...]) CouncilDistrictCode: len([1037 596 509 367 338 282 247]) Neighborhood: len([573 453 423 375 280 251 210 166 145 126 107 95 80 42 27...]) 102 99 54 46

LargestPropertyUseType: 27

len([1667 498 199 139

len([23 9])

98

32...])

 \rightarrow BuildingType et PrimaryPropertyType semblent importants pour la régression.

Outlier:

→ dans l'idéal: traitement des [... 1] pour éviter handle_unknown dans le OHE.

II.2. Indicateurs quantitatifs: corrélations

110	Ч	ua	TIU	100	JUUJ	IID	. •	,OI	10	100	JIO	,111	,											
1.00	0.03	-0.03	0.03	0.03	0.03	0.02	0.03	0.01	-0.05	-0.03	0.04	0.03	0.03	-0.00	0.04	0.03	0.02	0.02	0.02	0.03	0.03	0.02	0.03	0.03
-0.03	0.40	1.00	0.21	0.25	0.07	0.47	0.36	0.02	0.15	0.02	0.14	0.22	-0.00	0.42	-0.04	0.34	0.08	0.03	0.04	0.25	0.01	0.03	0.07	0.29
0.03	0.18	0.07	0.51	0.29	1.00	0.39	0.18	0.02	0.02	-0.10	0.73	0.65	0.26	0.06	0.49	0.20	0.03	0.18	0.18	0.29	0.26	0.06	1.00	0.73
0.01	0.02	-0.02	0.02	0.02	0.02	-0.05	-0.02	1.00	0.12	0.08	0.03	-0.11	-0.02	-0.00	0.04	0.02	-0.02	-0.00	-0.00	-0.02	-0.01	0.02	0.02	-0.04
-0.05	0.10	0.15	0.03	0.04	0.02	0.20	0.08	0.12	1.00	0.03	0.01	0.09	-0.03	0.18	-0.15	0.07	-0.02	0.04	0.04	0.04	-0.02	0.02	0.02	0.07
-0.03	0.07	0.02	0.09	0.06	0.10	0.08	0.06	0.08	0.03	1.00	-0.10	-0.01	-0.35	0.05	-0.27	0.06	-0.04	-0.31	-0.31	-0.06	-0.34	0.00	-0.10	-0.09
0.04	0.53	0.14	0.86	0.69	0.73	0.51	0.55	-0.03	0.01	-0.10	1.00	0.68	0.27	0.09	0.47	0.58	0.68	0.22	0.23	0.69	0.29	0.41	0.73	0.86
0.03	0.52	0.22	0.75	0.68	0.65	0.66	0.54	-0.11	0.09	-0.01	0.68	1.00	0.13	0.21	0.33	0.46	0.04	0.14	0.14	0.68	0.13	0.00	0.65	0.75
0.03	0.04	-0.00	0.27	0.25	0.26	0.08	0.03	0.02	-0.03	-0.35	0.27	0.13	1.00	0.09	0.75	0.03	0.09	0.94	0.93	0.25	0.99	0.01	0.26	0.39
0.04	0.02	-0.04	0.31	0.18	0.49	0.11	0.03	0.04	-0.15	-0.27	0.47	0.33	0.75	-0.04	1.00	0.05	0.19	0.53	0.52	0.18	0.73	0.03	0.49	0.43
0.03	0.97	0.34	0.84	0.88	0.20	0.77	0.98	-0.02	0.07	0.06	0.58	0.46	0.03	0.30	0.05	1.00	0.50	0.03	0.06	0.88	0.06	0.76	0.20	0.39
0.02	0.44	0.08	0.60	0.55	0.03	0.26	0.46	0.02	-0.02	-0.04	0.68	0.04	0.09	0.01	0.19	0.50	1.00	0.08	0.09	0.55	0.11	0.40	0.03	0.47
0.02	0.05	0.03	0.27	0.29	0.18	0.10	0.04	0.00	0.04	-0.31	0.22	0.14	0.94	0.13	0.53	0.03	0.08	1.00	0.99	0.29	0.94	0.00	0.18	0.39
0.02	0.08	0.04	0.30	0.32	0.18	0.11	0.07	0.00	0.04	-0.31	0.23	0.14	0.93	0.13	0.52	0.06	0.09	0.99	1.00	0.32	0.94	0.03	0.18	0.39
0.03	0.85	0.25	0.96	1.00	0.29	0.63	0.86	-0.02	0.04	-0.06	0.69	0.68	0.25	0.22	0.18	0.88	0.55	0.29	0.32	1.00	0.29	0.74	0.29	0.59
0.03	0.07	0.01	0.30	0.29	0.26	0.10	0.06	0.01	-0.02	-0.34	0.29	0.13	0.99	0.10	0.73	0.06	0.11	0.94	0.94	0.29	1.00	0.03	0.26	0.40
0.02	0.69	-0.03	0.69	0.74	0.06	0.11	0.73	0.02	-0.02	-0.00	0.41	0.00	0.01	-0.00	0.03	0.76	0.40	0.00	0.03	0.74	0.03	1.00	0.06	0.09
0.03	0.18	0.07	0.51	0.29	1.00	0.39	0.18	0.02	0.02	-0.10	0.73	0.65	0.26	0.06	0.49	0.20	0.03	0.18	0.18	0.29	0.26	0.06	1.00	0.73
0.03	0.40	0.29	0.72	0.59	0.73	0.63	0.38	0.04	0.07	-0.09	0.86	0.75	0.39	0.24	0.43	0.39	0.47	0.39	0.39	0.59	0.40	0.09	0.73	1.00
Φ	-	S	<u> </u>	=	<u></u>	⋖	· ·	Φ	=	0)	S	۷.	6	מ	>-	A	=	f)	9	=	-G	S	<u>-</u>	3
tud	Tota	000	(Btu	Btu	rms	GF)	s)g(s	tud	Bui	COL	ion	GF.	ın/sı	ķ	nsit	GF.	BE C	ta/n	ın/sı	W	s/n:	ling	Btu	SiteEnergyUseWN(kBtu)
ngi	ΈĀ	ofFI	se(k	ţX,	the	уре	į	Lati	/ear	RS	niss	уре	KB†	Par	nte	уре	se(k	<u>8</u>	Ϋ́B	ty(k	KB.	Plin	as(k	ž
2	\$	perc	Ñ	trici	as(SeT	Bu		_	STA	ЭË	SeT	Š	ЭFА	nsl	SeT	Ĭ.	Š	ij	trici	Ĭ	ofB	<u>3</u>	Ne%
	bei	<u>m</u>	Jerg	Elec.	ra G	Ž	3FA			ĠΥξ	뜻	Ž	2	Ę	SSic	λČ	tear	5	ceE	Elec	iteE	per	ıturs	yUs
	Æ	Z	eEn	ш	atnı	pert	ĵţ.			ER	otal(pert	ite	edc	Ë	pert	∇	ceE	your	ш	S	m J	Ra	erg
			ŝ		Z	Pro	obe.			H	5	Pro	S	Ĕ	호	Pro		oni	(I)			_		끭
						st	ፚ					st			Ō	st		S						Sit.
																								1
	1.00 0.03 -0.03 0.03 0.03 0.03 0.02 0.03 0.04 0.03 0.04 0.03 0.04 0.03 0.00 0.04 0.03 0.00 0.04 0.03 0.00	1.00 0.03 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.004 0.005 0.003 0.004 0.005	1.00 0.03 0.03 0.03 0.03 1.00 0.40 0.40 0.40 0.2 0.2 0.2 0.40 0.2 0.2 0.2 0.40 0.3 0.40 0.2 0.2 0.40 0.3 0.40 0.2 0.40 0.3 0.40 0.2 0.40 0.3 0.40 0.2 0.40 0.3 0.40 0.2 0.40 0.3 0.40 0.2 0.40 0.3 0.5 0.3 0.40 0.3 0.5 0.3 0.40 0.3 0.5 0.3 0.40 0.3 0.5 0.3 0.40 0.3 0.5 0.3 0.40 0.3 0.5 0.3 0.40 0.3 0.5 0.3 0.40 0.3 0.5 0.3 0.40 0.3 0.5 0.3 0.40 0.3 0.5 0.3 0.40 0.3 0.5 0.3 0.40 0.3 0.5 0.3 0.40 0.3 0.5 0.3 0.40 0.3 0.40 0.2 0.5 0.3 0.3 0.3 0.40 0.2 0.5 0.3 0.3 0.3 0.3 0.40 0.2 0.3 0.3 0.40 0.2 0.3 0.3 0.40 0.2 0.3 0.3 0.40 0.2 0.3 0.3 0.40 0.2 0.3 0.3 0.40 0.2 0.3 0.40 0.2 0.3 0.3 0.40 0.2 0.3 0.3 0.40 0.2 0.3 0.40 0.2 0.3 0.3 0.40 0.2 0.3 0.40 0.2 0.3 0.3 0.40 0.2 0.3 0.40 0.2 0.3 0.40 0.2 0.3 0.3 0.40 0.2 0.3 0.40 0.2 0.3 0.40 0.2 0.3 0.40 0.2 0.3 0.40 0.2 0.2 0.3 0.40 0.2 0.3 0.40 0.2 0.3 0.40 0.2 0.3 0.40 0.2 0.2 0.3 0.40 0.2 0.3 0.40 0.2 0.2 0.3 0.40 0.2 0.2 0.2 0.3 0.40 0.2 0.2 0.3 0.40 0.2 0.2 0.2 0.3 0.40 0.2 0.2 0.3 0.40 0.2 0.2 0.2 0.3 0.40 0.2 0.2 0.2 0.3 0.40 0.2 0.2 0.2 0.3 0.40 0.2 0.2 0.2 0.2 0.2 0.3 0.40 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.	1.00 0.03 0.03 0.03 0.03 0.03 0.03 0.00 0.40 0.80 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.05 0.	1.00 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.04 0.00 0.21 0.25 0.03 0.05 0.25 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.03 0.04 0.05 0.03 0.04 0.05 0.03 0.04 0.05 0.03 0.04 0.05 0.	1.00 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.04 0.00 0.05 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.03 0.04 0.05 0.	1.00 0.03 0.03 0.03 0.03 0.02 0.03 1.00 0.40 0.80 0.85 0.18 0.81 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.04 0.00 0.21 0.25 0.07 0.47 0.03 0.85 0.25 0.09 0.00 0.39 0.03 0.03 0.03 0.02 0.03 0.04 0.05 0.	1.00 0.03 0.03 0.03 0.03 0.02 0.03 0.03 1.00 0.40 0.80 0.85 0.18 0.81 0.99 0.03 0.04 1.00 0.25 0.07 0.47 0.36 0.03 0.05 0.25 0.07 0.47 0.36 0.03 0.05 0.25 0.06 0.05 0.63 0.85 0.38 0.85 0.25 0.06 1.00 0.29 0.63 0.86 0.03 0.18 0.07 0.51 0.63 0.86 0.30 0.39 0.18 0.07 0.51 0.03 0.18 0.07 0.03 0.18 0.07 0.03 0.18 0.07 0.03 0.18 0.07 0.03 0.18 0.07 0.03 0.18 0.07 0.03 0.18 0.07 0.02 0.05 0.02 0.05 0.02 0.05 0.02 0.05 0.02 0.05 0.02 0.05 0.02 0.05 0.02 0.05 0.	1.00 0.03 0.03 0.03 0.03 0.02 0.03 0.01 0.03 1.00 0.40 0.80 0.85 0.18 0.81 0.99 0.02 0.03 0.40 1.00 0.21 0.25 0.07 0.47 0.36 0.02 0.03 0.80 0.21 1.00 0.96 0.51 0.63 0.81 0.03 0.85 0.25 0.96 1.00 0.29 0.63 0.86 0.02 0.03 0.18 0.07 0.51 0.29 1.00 0.39 0.18 0.02 0.02 0.81 0.47 0.63 0.63 0.39 1.00 0.79 0.05 0.03 0.99 0.36 0.81 0.86 0.18 0.79 1.00 0.02 0.01 0.02 0.02 0.02 0.02 0.05 0.02 1.00 0.05 0.10 0.15 0.30 0.04 0.02 0.00 0.10 0.05 0.10 0.15 0.30 0.40 0.08 0.61 0.88 0.04 0.53 0.14 0.86 0.69 0.73 0.51 0.55 0.03 0.04 0.05 0.12 0.07 0.25 0.26 0.08 0.30 0.04 0.04 0.07 0.27 0.25 0.26 0.08 0.30 0.02 0.00 0.40 0.42 0.17 0.22 0.06 0.48 0.27 0.00 0.40 0.02 0.04 0.31 0.38 0.20 0.77 0.38 0.02 0.02 0.44 0.08 0.60 0.55 0.03 0.26 0.46 0.02 0.02 0.44 0.08 0.60 0.55 0.03 0.26 0.46 0.02 0.02 0.44 0.08 0.60 0.55 0.03 0.26 0.46 0.02 0.02 0.05 0.03 0.27 0.29 0.18 0.10 0.44 0.00 0.02 0.05 0.03 0.27 0.29 0.18 0.10 0.44 0.00 0.03 0.85 0.25 0.96 1.00 0.29 0.63 0.86 0.02 0.03 0.85 0.25 0.96 0.09 0.29 0.63 0.86 0.02 0.03 0.85 0.25 0.96 0.09 0.29 0.63 0.86 0.02 0.03 0.85 0.25 0.96 0.09 0.29 0.63 0.86 0.02 0.03 0.86 0.03 0.29 0.26 0.10 0.06 0.01 0.02 0.69 0.03 0.29 0.26 0.10 0.06 0.01 0.03 0.86 0.05 0.05 0.05 0.05 0.05 0.05 0.03 0.86 0.07 0.10 0.39 0.74 0.06 0.11 0.73 0.02 0.03 0.86 0.03 0.74 0.06 0.11 0.73 0.02 0.03 0.86 0.03 0.74 0.06 0.11 0.73 0.02 0.03 0.86 0.03 0.74 0.06 0.11 0.73 0.02 0.03 0.86 0.03 0.74 0.06 0.11 0.75 0.05 0.03 0.86 0.05 0.75 0.75 0.75 0.75 0.	1.00 0.03 0.03 0.03 0.03 0.02 0.03 0.01 0.05 0.03 1.00 0.40 0.80 0.85 0.18 0.81 0.99 0.02 0.10 0.03 0.04 1.00 0.21 0.25 0.07 0.47 0.36 0.02 0.15 0.03 0.80 0.21 0.25 0.07 0.47 0.36 0.02 0.15 0.03 0.85 0.25 0.96 1.00 0.29 0.63 0.86 0.02 0.04 0.03 0.18 0.07 0.25 0.63 0.86 0.02 0.04 0.03 0.18 0.07 0.25 0.05 0.39 0.18 0.02 0.02 0.02 0.03 0.18 0.02 0.02 0.03 0.18 0.02 0.02 0.03 0.18 0.02 0.02 0.03 0.18 0.02 0.02 0.03 0.18 0.02 0.02 0.03 0.18 0.02 0.02 0.05 0.02 0.05 0.02 0.05 0.02 0.05 0.02 0.05 0.02 0.05 0.02 0.05 0.02 0.05 0.02 0.05 0.02 0.05 0.02 0.05 0.03 0.12 0.05 0.03 0.04 0.02 0.05 0.06 0.10 0.05 0.03 0.04 0.02 0.05 0.05 0.05 0.05 0.03 0.01 0.05 0.03 0.04 0.02 0.05 0.	1.00 0.03 0.03 0.03 0.03 0.02 0.03 0.01 0.05 0.03 0.03 1.00 0.40 0.80 0.85 0.18 0.81 0.99 0.02 0.10 0.07 0.03 0.04 1.00 0.21 0.25 0.07 0.47 0.36 0.02 0.15 0.02 0.30 0.30 0.21 1.00 0.96 0.10 0.36 0.36 0.02 0.04 0.06 0.30 0.38 0.22 0.06 0.39 0.18 0.02 0.02 0.03 0.30 0.30 0.30 0.35 0.25 0.96 1.00 0.29 0.39 0.18 0.02 0.02 0.03 0.30 0.	1.00 0.03 0.03 0.03 0.03 0.02 0.03 0.01 0.05 0.03 0.04 0.03 0.05 0.08 0.85 0.18 0.81 0.99 0.02 0.10 0.07 0.53 0.03 0.04 0.00 0.21 0.25 0.07 0.47 0.36 0.02 0.15 0.02 0.14 0.03 0.85 0.25 0.96 0.02 0.63 0.86 0.02 0.15 0.02 0.15 0.03 0.03 0.35 0.25 0.96 1.00 0.29 0.63 0.86 0.02 0.04 0.06 0.99 0.03 0.18 0.02 0.02 0.01 0.73 0.03 0.18 0.07 0.53 0.03 0.18 0.07 0.53 0.03 0.18 0.07 0.05 0.20 0.04 0.05 0.05 0.03 0.18 0.07 0.05 0.20 0.04 0.05 0.	1.00 0.03 0.03 0.03 0.03 0.03 0.02 0.03 0.01 0.05 0.03 0.04 0.03 0.03 0.03 0.03 0.03 0.04 0.03 0.05 0.03 0.04 0.05 0.03 0.04 0.05 0.	0.03 1.00 0.40 0.80 0.85 0.18 0.81 0.99 0.02 0.10 0.07 0.53 0.52 0.04 0.03 0.40 1.00 0.21 0.25 0.07 0.47 0.36 0.02 0.15 0.02 0.14 0.22 0.00 0.03 0.80 0.21 1.00 0.96 0.51 0.63 0.81 0.02 0.05 0.09 0.86 0.75 0.27 0.03 0.85 0.25 0.96 1.00 0.29 0.63 0.86 0.02 0.02 0.04 0.06 0.69 0.68 0.25 0.03 0.03 0.18 0.07 0.51 0.29 1.00 0.39 0.18 0.02 0.02 0.04 0.06 0.69 0.68 0.25 0.03 0.09 0.81 0.47 0.63 0.63 0.39 1.00 0.79 0.05 0.20 0.08 0.05 0.55 0.54 0.03 0.03 0.99 0.36 0.81 0.86 0.18 0.79 1.00 0.02 0.08 0.06 0.55 0.54 0.03 0.01 0.02 0.02 0.02 0.02 0.02 0.02 0.02	1.00 0.03 0.03 0.03 0.03 0.03 0.02 0.03 0.01 0.05 0.03 0.04 0.03 0.03 0.04 0.03 0.04 0.03 0.05 0.04 0.03 0.05 0.04 0.03 0.05 0.04 0.03 0.05 0.04 0.03 0.05 0.04 0.03 0.05 0.04 0.03 0.05 0.04 0.03 0.05 0.04 0.03 0.05 0.04 0.03 0.05 0.04 0.03 0.05 0.04 0.03 0.05 0.	1.00 0.03 0.03 0.03 0.03 0.03 0.02 0.03 0.01 0.05 0.03 0.04 0.03 0.03 0.00 0.04 0.03 1.00 0.04 0.02 0.05 0.08 0.08 0.08 0.08 0.08 0.02 0.05 0.02 0.16 0.02 0.16 0.02 0.04 0.06 0.02 0.04 0.03 0.03 0.00 0.04 0.03 0.03 0.00 0.04 0.03 0.03 0.00 0.04 0.03 0.03 0.00 0.04 0.03 0.03 0.05 0.02 0.04 0.06 0.05 0.	1.00	1.00	1.00 0.03 0.03 0.03 0.03 0.03 0.03 0.02 0.03 0.01 0.05 0.03 0.04 0.03 0.03 0.00 0.04 0.03 0.02 0.02 0.03 0.04 0.03 0.03 0.00 0.04 0.03 0.02 0.02 0.03 0.04 0.03 0.05 0.04 0.05 0.	1.00 0.03 0.03 0.03 0.03 0.03 0.03 0.02 0.03 0.01 0.05 0.03 0.04 0.03 0.03 0.00 0.04 0.03 0.02 0.02 0.02 0.02 0.03 0.03 0.04 0.03 0.05 0.05 0.08 0.08 0.08 0.08 0.08 0.08 0.08 0.08 0.08 0.08 0.08 0.08 0.03 0.04 0.03 0.02 0.02 0.02 0.02 0.02 0.03 0.03 0.04 0.03 0.02 0.05 0.08 0.03 0.04 0.03 0.04 0.03 0.04 0.04 0.04 0.05 0.08 0.03 0.04 0.03 0.04 0.05 0.08 0.03 0.04 0.05 0.08 0.03 0.04 0.05 0.08 0.05 0.	1.00	1.00 0.03 0.03 0.03 0.03 0.03 0.03 0.02 0.03 0.01 0.05 0.03 0.04 0.03 0.03 0.00 0.04 0.03 0.02 0.02 0.03 0.	1.00 0.03 0.03 0.03 0.03 0.03 0.03 0.02 0.03 0.01 0.05 0.03 0.04 0.03 0.02 0.04 0.03 0.02 0.02 0.02 0.03 0.03 0.02 0.03 0.03 0.02 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.04 0.03 0.04 0.05 0.08 0.05 0.05 0.03 0.05 0.03 0.04 0.05 0.	1.00

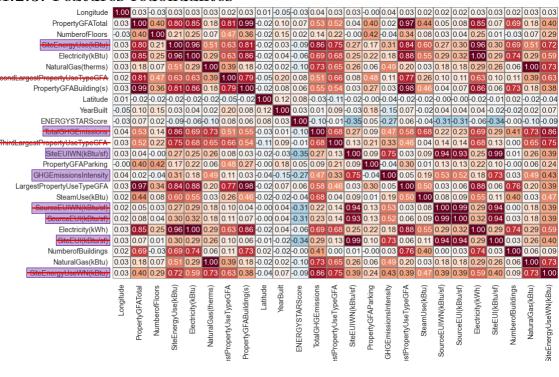
II.2.1. Targets et feature particulière

11.2.1. Targe	, C	U J	LUa	U	шС	·Ρ	aı	OI	Ju.	110	ıc														
Longitude	1.00	0.03	-0.03	0.03	0.03	0.03	0.02	0.03	0.01	-0.05	-0.03	0.04	0.03	0.03	-0.00	0.04	0.03	0.02	0.02	0.02	0.03	0.03	0.02	0.03	0.03
PropertyGFATotal	0.03	1.00	0.40	0.80	0.85	0.18	0.81	0.99	0.02	0.10	0.07			0.04	0.40	0.02	0.97	0.44	0.05	0.08	0.85	0.07	0.69	0.18	0.40
NumberofFloors	-0.03	0.40	1.00	0.21	0.25	0.07	0.47	0.36	0.02	0.15	0.02	0.14	0.22	-0.00	0.42	0.04	0.34	0.08	0.03	0.04	0.25	0.01	-0.03	0.07	0.29
SiteEnergyUse(kBtu)	0.03	0.80	0.21	1.00	0.96		0.63	0.81	0.02	0.03	-0.09	0.86	0.75	0.27	0.17	0.31	0.84	0.60	0.27	0.30	0.96	0.30	0.69	0.51	0.72
Electricity(kBtu)	0.03	0.85	0.25	0.96	1.00	0.29	0.63	0.86	0.02	0.04	-0.06	0.69	0.68	0.25	0.22	0.18	0.88		0.29	0.32	1.00	0.29	0.74	0.29	0.59
NaturalGas(therms)	0.03	0.18	0.07		0.29	1.00	0.39	0.18	0.02	0.02	-0.10	0.73	0.65	0.26	0.06		0.20	0.03	0.18	0.18	0.29	0.26	0.06	1.00	0.73
SecondLargestPropertyUseTypeGFA	0.02	0.81		0.63	0.63	0.39	1.00	0.79	0.05	0.20	0.08		0.66	0.08		0.11	0.77	0.26	0.10	0.11	0.63	0.10	0.11	0.39	0.63
PropertyGFABuilding(s)	0.03	0.99	0.36	0.81	0.86	0.18	0.79	1.00	0.02	0.08	0.06			0.03	0.27	0.03	0.98		0.04	0.07	0.86	0.06	0.73	0.18	0.38
Latitude	0.01	-0.02	-0.02	-0.02	-0.02	0.02	-0.05	-0.02	1.00	0.12	0.08	-0.03	-0.11	-0.02	-0.00	0.04	-0.02	-0.02	-0.00	0.00	-0.02	-0.01	0.02	-0.02	0.04
YearBuilt	-0.05	0.10	0.15	0.03	0.04	0.02	0.20	0.08	0.12	1.00	0.03	0.01	0.09	-0.03	0.18	0.15	0.07	-0.02	0.04	0.04	0.04	-0.02	0.02	0.02	0.07
ENERGYSTARScore	-0.03	0.07	0.02	-0.09	-0.06	-0.10	0.08	0.06	0.08	0.03	1.00	-0.10	-0.01		0.05	0.27	0.06	-0.04	-0.31	-0.31	-0.06	-0.34	-0.00	0.10	-0.09
TotalGHGEmissions	0.04	0.53	0.14	0.86	0.69	0.73	0.51	0.55	0.03	0.01	-0.10	1.00	0.68	0.27	0.09		0.58	0.68	0.22	0.23	0.69	0.29	0.41	0.73	0.86
ThirdLargestPropertyUseTypeGFA	0.03		0.22	0.75	0.68	0.65	0.66	0.54	-0.11	0.09	-0.01	0.68	1.00	0.13	0.21	0.33		0.04	0.14	0.14	0.68	0.13	0.00	0.65	0.75
SiteEUIWN(kBtu/sf)	0.03	0.04	-0.00	0.27	0.25	0.26	0.08	0.03	-0.02	-0.03		0.27	0.13	1.00	0.09		0.03	0.09	0.94	0.93	0.25	0.99	0.01	0.26	0.39
PropertyGFAParking	-0.00	0.40	0.42	0.17	0.22	0.06	0.48	0.27	0.00	0.18	0.05	0.09	0.21	0.09	1.00	0.04	0.30	0.01	0.13	0.13	0.22	0.10	0.00	0.06	0.24
GHGEmissionsIntensity	0.04	0.02	-0.04	0.31	0.18		0.11	0.03	-0.04	-0.15	-0.27		0.33		-0.04	1.00	0.05	0.19			0.18	0.73	0.03		0.43
LargestPropertyUseTypeGFA	0.03	0.97	0.34	0.84	0.88	0.20	0.77	0.98	0.02	0.07	0.06	0.58		0.03	0.30	0.05	1.00	0.50	0.03	0.06	0.88	0.06	0.76	0.20	0.39
SteamUse(kBtu)	0.02	0.44	0.08	0.60	0.55	0.03	0.26	0.46	0.02	-0.02	-0.04	0.68	0.04	0.09	0.01	0.19	0.50	1.00	0.08	0.09	0.55	0.11	0.40	0.03	0.47
SourceEUIWN(kBtu/sf)	0.02	0.05	0.03	0.27	0.29	0.18	0.10	0.04	0.00	0.04	-0.31	0.22	0.14	0.94	0.13	0.53	0.03	0.08	1.00	0.99	0.29	0.94	0.00	0.18	0.39
SourceEUI(kBtu/sf)	0.02	0.08	0.04	0.30	0.32	0.18	0.11	0.07	0.00	0.04	-0.31	0.23	0.14	0.93	0.13		0.06	0.09	0.99	1.00	0.32	0.94	0.03	0.18	0.39
Electricity(kWh)	0.03	0.85	0.25	0.96	1.00	0.29	0.63	0.86	0.02	0.04	-0.06	0.69	0.68	0.25	0.22	0.18	0.88	0.55	0.29	0.32	1.00	0.29	0.74	0.29	0.59
SiteEUI(kBtu/sf)	0.03	0.07	0.01	0.30	0.29	0.26	0.10	0.06	0.01	-0.02	-0.34	0.29	0.13	0.99	0.10	0.73	0.06	0.11	0.94	0.94	0.29	1.00	0.03	0.26	0.40
NumberofBuildings	0.02	0.69	-0.03	0.69	0.74	0.06	0.11	0.73	0.02	-0.02	-0.00	0.41	0.00	0.01	-0.00	0.03	0.76	0.40	0.00	0.03	0.74	0.03	1.00	0.06	0.09
NaturalGas(kBtu)	0.03	0.18	0.07		0.29	1.00	0.39	0.18	0.02	0.02	-0.10	0.73	0.65	0.26	0.06		0.20	0.03	0.18	0.18	0.29	0.26	0.06	1.00	0.73
SiteEnergyUseWN(kBtu)	0.03	0.40	0.29	0.72	0.59	0.73	0.63	0.38	0.04	0.07	-0.09	0.86	0.75	0.39	0.24	0.43	0.39	0.47	0.39	0.39	0.59	0.40	0.09	0.73	1.00
	d)	_	"	_	$\widehat{}$	$\widehat{}$	_	$\overline{}$	(I)	<u> </u>	0)	· · ·	_			>	_	$\overline{}$			$\overline{}$		ro.		
	Longitude	PropertyGFATotal	NumberofFloors	Bfu	Electricity(kBtu)	rms	GF/	g(s	Latitude	YearBuilt	COLE	io	GF/	u/sf	ķi	nsit	GF.	SteamUse(kBtu)	n/st	n/sf	Electricity(kWh)	SiteEUI(kBtu/sf)	ing	NaturalGas(kBtu)	Btn
	ngi	FA	Ĕ	še(k	×,	the	/be/	i <u>B</u>	ati	ear	RS	Jiss	/be	kBţ	Par	ntei	/be	e(k	Æ	Ŕ	ξ	ξ	Pin	s(k	ž
	೨	t R	Derc	χ	Lici	as(eT)	Bui		>	STA	넜	SeT,	Ň	ΈĀ	lsu	SeT.	Ju S	×	Ĕ	trici	Ĕ	용	<u>I</u>	e W
		per	Ĕ	erg	Ject	<u>a</u>	λΩ	ЭFА			37.8	¥	λČ	$\stackrel{>}{\sim}$	TŞG	ssio	λŠ	ean	\geq	GeE	Ject	iteE	per	tura	νΩs
		P	Ž	SiteEnergyUse(kBtu)	ш	NaturalGas(therms)	stPropertyUseTypeGFA	PropertyGFABuilding(s)			ENERGYSTARScore	Fotal GHGEmissions 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1	stPropertyUseTypeGFA	SiteEUIWN(kBtu/sf)	PropertyGFAParking	GHGEmissionsIntensity	stPropertyUseTypeGFA	₩	SourceEUIWN(kBtu/sf)	SourceEUI(kBtu/sf)	ш	S	NumberofBuildings	Na	SiteEnergyUseWN(kBtu)
				SE		Ž	Pro!	obe				욘	امار	S	Ĕ	후	اما		onu	S			_		Ä
							sst	ď.					st			ठं	st		S						Site

II.2.2. Features inintéressantes

1.4.4. reature	55	1111	ш	·GI	65	Sa	116	CS																	
Longitude	1.00	0.03	-0.03	0.03	0.03	0.03	0.02	0.03	0.01	-0.05	-0.03	0.04	0.03	0.03	-0.00	0.04	0.03	0.02	0.02	0.02	0.03	0.03	0.02	0.03	0.03
	0.03																								
NumberofFloors	-0.03	0.40	1.00	0.21	0.25	0.07	0.47	0.36	0.02	0.15	0.02	0.14	0.22	-0.00	0.42	-0.04	0.34	0.08	0.03	0.04	0.25	0.01	0.03	0.07	0.29
		0.80																							
Electricity(kBtu)	0.03	0.85	0.25	0.96	1.00	0.29	0.63	0.86	0.02	0.04	-0.06	0.69	0.68	0.25	0.22	0.18	0.88	0.55	0.29	0.32	1.00	0.29	0.74	0.29	0.59
NaturalGas(therms)	0.03	0.18	0.07	0.51	0.29	1.00	0.39	0.18	-0.02	0.02	-0.10	0.73	0.65	0.26	0.06	0.49	0.20	0.03	0.18	0.18	0.29	0.26	0.06	1.00	0.73
endLargestPropertyUseTypeGFA	0.02					0.39							0.66										0.11		
PropertyGFABuilding(s)		0.99				_		_	$\overline{}$		$\overline{}$				0.27	_				-		_	0.73		
Latitude	0.01	-0.02	0.02	0.02	-0.02	0.02	-0.05	-0.02	1.00	0.12	0.08	-0.03	-0.11	-0.02	-0.00	0.04	0.02	-0.02	-0.00	-0.00	-0.02	-0.01	0.02	0.02	-0.04
YearBuilt	-0.05	0.10	0.15	0.03	0.04	0.02	0.20	0.08	0.12	1.00	0.03	0.01	0.09	-0.03	0.18	-0.15	0.07	-0.02	0.04	0.04	0.04	-0.02	0.02	0.02	0.07
		0.07																							
TotalGHGEmissions		0.53																							
hirdLargestPropertyUseTypeGFA-	0.03	0.52	0.22	0.75	0.68	0.65	0.66		-0.11	0.09	-0.01	0.68	1.00	0.13	0.21	0.33	0.46	0.04	0.14	0.14	0.68	0.13	0.00	0.65	0.75
SiteEUIWN(kBtu/sf)		0.04																							
PropertyGFAParking	-0.00	0.40	0.42	0.17	0.22	0.06	0.48	0.27	-0.00	0.18	0.05	0.09	0.21	0.09	1.00	0.04	0.30	0.01	0.13	0.13	0.22	0.10	0.00	0.06	0.24
GHGEmissionsIntensity	0.04	0.02	0.04	0.31	0.18	0.49	0.11	0.03	0.04	-0.15	-0.27	0.47	0.33	0.75	-0.04	1.00	0.05	0.19		0.52	0.18	0.73	0.03	0.49	0.43
LargestPropertyUseTypeGFA	0.03	0.97	0.34	0.84	0.88	0.20	0.77	0.98	0.02	0.07	0.06	0.58	0.46	0.03	0.30	0.05	1.00	0.50	0.03	0.06	0.88	0.06	0.76	0.20	0.39
SteamUse(kBtu)	0.02	0.44	0.08	0.60	0.55	0.03	0.26	0.46	0.02	-0.02	-0.04	0.68	0.04	0.09	0.01	0.19	0.50	1.00	0.08	0.09	0.55	0.11	0.40	0.03	0.47
SourceEUIWN(kBtu/sf)	0.02	0.05	0.03	0.27	0.29	0.18	0.10	0.04	0.00	0.04	-0.31	0.22	0.14	0.94	0.13	0.53	0.03	0.08	1.00	0.99	0.29	0.94	0.00	0.18	0.39
SourceEUI(kBtu/sf)	0.02	0.08	0.04	0.30	0.32	0.18	0.11	0.07	0.00	0.04	-0.31	0.23	0.14	0.93	0.13	0.52	0.06	0.09	0.99	1.00	0.32	0.94	0.03	0.18	0.39
Electricity(kWh)	0.03	0.85	0.25	0.96	1.00	0.29	0.63	0.86	0.02	0.04	-0.06	0.69	0.68	0.25	0.22	0.18	0.88	0.55	0.29	0.32	1.00	0.29	0.74	0.29	0.59
SiteEUI(kBtu/sf)	0.03	0.07	0.01	0.30	0.29	0.26	0.10	0.06	0.01	-0.02	-0.34	0.29	0.13	0.99	0.10	0.73	0.06	0.11	0.94	0.94	0.29	1.00	0.03	0.26	0.40
NumberofBuildings	0.02	0.69	-0.03	0.69	0.74	0.06	0.11	0.73	0.02	-0.02	-0.00	0.41	0.00	0.01	-0.00	0.03	0.76	0.40	0.00	0.03	0.74	0.03	1.00	0.06	0.09
NaturalGas(kBtu)	0.03	0.18	0.07	0.51	0.29	1.00	0.39	0.18	0.02	0.02	-0.10	0.73	0.65	0.26	0.06	0.49	0.20	0.03	0.18	0.18	0.29	0.26	0.06	1.00	0.73
SiteEnergyUseWN(kBtu)	0.03	0.40	0.29	0.72	0.59	0.73	0.63	0.38	0.04	0.07	-0.09	0.86	0.75	0.39	0.24	0.43	0.39	0.47	0.39	0.39	0.59	0.40	0.09	0.73	1.00
	<u>o</u>	<u></u>	ço	<u> </u>	Î	(2)	⋖	(8)	е	#	Ф	S	×	£		-£	A	<u> </u>	÷.	Œ.	<u></u>	£	S	Î	Î
	ongitude-	PropertyGFATotal	NumberofFloors	SiteEnergyUse(kBtu)	∃ectricity(kBtu)	NaturalGas(therms)	stPropertyUseTypeGFA	PropertyGFABuilding(s)	Latitude	YearBuilt	ENERGYSTARScore	TotalGHGEmissions	stPropertyUseTypeGFA	SiteEUIWN(kBtu/sf)	PropertyGFAParking	GHGEmissionsIntensity	sstPropertyUseTypeGFA	SteamUse(kBtu)	SourceEUIWN(kBtu/sf)	SourceEUI(kBtu/sf)	Electricity(kWh)	SiteEUI(kBtu/sf)	NumberofBuildings	NaturalGas(kBtu)	SiteEnergyUseWN(kBtu)
	oud	3FA	JJ0.	se(l	iţ.	(the	ype	į	Tat	Year	\RS	JIIS	ype	(KB	Pa	Inte	уре	se()	8	(KB	iţ.	(kB	ĕ	as(Š
	ت	Ę.	per	JyC	tric	Sas	SeT	ABu		_	ST/	GEr	seT	Š	3FA	ons	seT	J.	N N	E	ctric	E	roff	alG	seV
		obe	E H	nerç	Elec	IIa	₽ P	ĠF,			ĠΥ	3	₽ F	E) L	ISS	φ.	tea		rce	Elec	Site	nbe	atru	Ν̈́
		ď	2	teEr	_	Jatu	ber	erty			H	otal	ber	iteE	obe.	Em	ber	S	CeE	Jon S	_	0,	Nun	Ž	Jerg
				20.		2	Pro	rope			Ξ	ĭ	Pro	(J)	ቯ	9	Pro		Jon J	0,			_		eEr.
							st	ď					ss			O	SS		0,						泛

II.2.3. Features redondantes



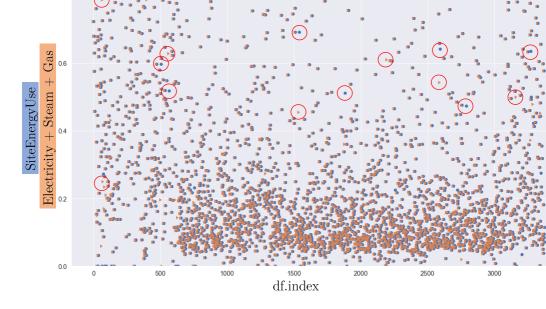
II.2.3. Features redondantes

ii.z.o. i caudi	,,,	10	uo	11(LOUI	LIUV																			
Longitude	1.00	0.03	-0.03	0.03	0.03	0.03	0.02	0.03	0.01	-0.05	-0.03	0.04	0.03	0.03	-0.00	0.04	0.03	0.02	0.02	0.02	0.03	0.03	0.02	0.03	0.03
PropertyGFATotal	0.03	1.00	0.40	0.80	0.85	0.18	0.81	0.99	-0.02	0.10	0.07			0.04	0.40	0.02	0.97	0.44	0.05	0.08	0.85	0.07	0.69	0.18	0.40
NumberofFloors	-0.03	0.40	1.00	0.21	0.25	0.07		0.36	-0.02	0.15	0.02	0.14	0.22	-0.00	0.42	-0.04	0.34	0.08	0.03	0.04	0.25	0.01	-0.03	0.07	0.29
-SiteEnergyUse(kBtu)					-				_	-											0.96				0.72
Electricity(kBtu)	0.03	0.85	0.25	0.96	1.00	0.29	0.63	0.86	-0.02	0.04	-0.06	0.69	0.68	0.25	0.22	0.18	0.88		0.29	0.32	1.00	0.29	0.74	0.29	0.59
NaturalCas(therms)	0.03	0.18	0.07		0.29	1.00	0.39	0.18	-0.02	0.02	-0.10	0.73	0.65	0.26	0.06		0.20	0.03	0.18	0.18	0.29	0.26	0.06	1.00	0.73
econdLargestPropertyUseTypeGFA	0.02	0.81		0.63	0.63	0.39	1.00	0.79	-0.05	0.20				0.08					0.10	0.11	0.63	0.10	0.11	0.39	0.63
PropertyGFABuilding(s)	0.03	0.99	0.36	0.81	0.86	0.18	0.79	1.00	-0.02	0.08	0.06			0.03	0.27	0.03	0.98		0.04	0.07	0.86	0.06	0.73	0.18	0.38
Latitude	0.01	-0.02	-0.02	-0.02	-0.02	-0.02	-0.05	-0.02	1.00	0.12	0.08	-0.03	-0.11	-0.02	-0.00	-0.04	-0.02	-0.02	-0.00	0.00	-0.02	-0.01	0.02	-0.02	-0.04
YearBuilt	-0.05	0.10	0.15	0.03	0.04	0.02	0.20	0.08	0.12	1.00	0.03	0.01	0.09	-0.03	0.18	-0.15	0.07	-0.02	0.04	0.04	0.04	-0.02	-0.02	0.02	0.07
ENERGYSTARScore	-0.03	0.07	0.02	-0.09	-0.06	-0.10	0.08	0.06	0.08	0.03	1.00	-0.10	-0.01	-0.35	0.05	-0.27	0.06	-0.04	-0.31	-0.31	-0.06	-0.34	-0.00	-0.10	-0.09
Total GHGEmissions	0.04		0.14	0.86	0.69	0.73			-0.03	0.01	-0.10	1.00	0.68	0.27	0.09		0.58	0.68	0.22	0.23	0.69	0.29	0.41	0.73	0.86
-ThirdLargestPropertyUseTypeGFA-	0.03		0.22	0.75	0.68	0.65	0.66		-0.11	0.09	-0.01	0.68	1.00	0.13	0.21	0.33		0.04	0.14	0.14	0.68	0.13	0.00	0.65	0.75
SiteEUIWN(kBtu/sf)	0.03	0.04	-0.00	0.27	0.25	0.26	0.08	0.03	-0.02	-0.03	-0.35	0.27	0.13	1.00	0.09	0.75	0.03	0.09	0.94	0.93	0.25	0.99	0.01	0.26	0.39
PropertyGFAParking	-0.00	0.40	0.42	0.17	0.22	0.06		0.27	-0.00	0.18	0.05	0.09	0.21	0.09	1.00	-0.04	0.30	0.01	0.13	0.13	0.22	0.10	-0.00	0.06	0.24
GHGEmissionsIntensity	0.04	0.02	-0.04	0.31	0.18		0.11	0.03	0.04	-0.15	-0.27		0.33	0.75	-0.04	1.00	0.05	0.19			0.18	0.73	0.03		0.43
LargestPropertyUseTypeGFA	0.03	0.97	0.34	0.84	0.88	0.20	0.77	0.98	-0.02	0.07	0.06			0.03	0.30	0.05	1.00		0.03	0.06	0.88	0.06	0.76	0.20	0.39
SteamUse(kBtu)	0.02	0.44	0.08			0.03	0.26	0.46	-0.02	-0.02	-0.04	0.68	0.04	0.09	0.01	0.19	0.50	1.00	0.08	0.09		0.11	0.40	0.03	
-SourceEUIWN(kBtu/sf)-	0.02	0.05	0.03	0.27	0.29	0.18	0.10	0.04	0.00	0.04	-0.31	0.22	0.14	0.94	0.13		0.03	0.08	1.00	0.99	0.29	0.94	0.00	0.18	0.39
SourceEUI(kBtu/sf)	0.02	0.08	0.04	0.30	0.32	0.18	0.11	0.07	-0.00	0.04	-0.31	0.23	0.14	0.93	0.13		0.06	0.09	0.99	1.00	0.32	0.94	0.03	0.18	0.39
Electricity(kWh)	0.03	0.85	0.25	0.96	1.00	0.29	0.63	0.86	-0.02	0.04	-0.06	0.69	0.68	0.25	0.22	0.18	0.88		0.29	0.32	1.00	0.29	0.74	0.29	0.59
SiteEUI(kBtu/sf)	0.03	0.07	0.01	0.30	0.29	0.26	0.10	0.06	-0.01	-0.02	-0.34	0.29	0.13	0.99	0.10	0.73	0.06	0.11	0.94	0.94	0.29	1.00	0.03	0.26	0.40
NumberofBuildings	0.02	0.69	-0.03	0.69	0.74	0.06	0.11	0.73	0.02	-0.02	-0.00	0.41	0.00	0.01	-0.00	0.03	0.76	0.40	0.00	0.03	0.74	0.03	1.00	0.06	0.09
NaturalGas(kBtu)	0.03	0.18	0.07		0.29	1.00	0.39	0.18	-0.02	0.02	-0.10	0.73	0.65	0.26	0.06		0.20	0.03	0.18	0.18	0.29	0.26	0.06	1.00	0.73
-SiteEnergyUseWN(kBtu)-	0.03	0.40	0.29	0.72	0.59	0.73	0.63	0.38	-0.04	0.07	-0.09	0.86	0.75	0.39	0.24	0.43	0.39		0.39	0.39	0.59	0.40	0.09	0.73	1.00
	ongitude	3FATotal	NumberofFloors	se(kBtu)	∃ectricity(kBtu)	(therms)	ypeGFA	(s)guiplir	Latitude	YearBuilt	ARScore	missions	ypeGFA	(kBtu/sf)	AParking	Intensity	ypeGFA	SteamUse(KBtu)	(kBtu/sf)	(kBtu/sf)	Electricity(kWh)	SiteEUI(kBtu/sf)	Suildings	as(kBtu)	VN(kBtu)
	ב	PropertyGFATotal	Number	SiteEnergyUse(kBtu)	Electric	NaturalGas(therms)	stPropertyUseTypeGFA	PropertyGFABuilding(s)			ENERGYSTARScore	TotalGHGEmissions	stPropertyUseTypeGFA	SiteEUIWN(kBtu/sf)	PropertyGFAParking	GHGEmissionsIntensity	*stPropertyUseTypeGFA	SteamU	SourceEUIWN(kBtu/sf)	SourceEUI(kBtu/sf)	Electric	SiteEUI	NumberofBuildings	NaturalGas(kBtu)	SiteEnergyUseWN(kBtu)

II.2.3. Features redondantes

1.00	0.03	-0.03	0.03	0.03	0.03	0.02	0.03	0.01	-0.05	-0.03	0.04	0.03	0.03	-0.00	0.04	0.03	0.02	0.02	0.02	0.03	0.03	0.02	0.03	0.03
0.03	1.00	0.40	0.80	0.85	0.18																	_		
-0.03	0.40	1.00	0.21	0.25	0.07						$\overline{}$								-			$\overline{}$		0.29
0.03	0.80	0.21	1.00	0.96		0.63	0.81	-0.02	0.03	-0.09	0.86	0.75	0.27	0.17	0.31	0.84	0.60	0.27	0.30	0.96	0.30	0.69	0.51	0.72
0.03	0.85	0.25	0.96	_	_			-														_		0.59
0.03	0.18	0.07	0.51	0.29	1.00	0.39	0.18	-0.02	0.02	-0.10										0.29	0.26	0.06	1.00	0.73
0.02	0.81	0.47	0.63																_			2111		
$\overline{}$			$\overline{}$	_	-			-		$\overline{}$					$\overline{}$				_		$\overline{}$			
0.01	-0.02	-0.02	-0.02	-0.02	0.02	-0.05	-0.02	1.00	0.12	0.08	-0.03	-0.11	-0.02	-0.00	-0.04	-0.02	-0.02	-0.00	0.00	-0.02	-0.01	0.02	-0.02	-0.04
-0.05	0.10	0.15	0.03	0.04	0.02	0.20	0.08	0.12	1.00	0.03	0.01	0.09	-0.03	0.18	-0.15	0.07	-0.02	0.04	0.04	0.04	-0.02	0.02	0.02	0.07
-0.03			$\overline{}$		$\overline{}$				-		-	$\overline{}$						_				-		
0.04									-					$\overline{}$				_	_					0.86
0.03		_	$\overline{}$							-		_	$\overline{}$	$\overline{}$	_		0.04	0.14	0.14	0.68	0.13	0.00	0.65	0.75
0.03	0.04	-0.00	0.27	0.25	0.26	0.08	0.03	-0.02	-0.03	-0.35	0.27	0.13	1.00	0.09	0.75	0.03	0.09	0.94	0.93	0.25	0.99	0.01	0.26	0.39
-0.00	0.40	0.42	0.17	0.22	$\overline{}$			$\overline{}$		-	0.09	0.21	0.09	1.00	-0.04	0.30	0.01	0.13	0.13	0.22	0.10	-0.00	0.06	0.24
												0.33	0.75	-0.04	1.00	0.05	0.19			0.18	0.73	0.03	0.49	0.43
0.03	0.97	0.34	0.84	0.88	0.20	0.77	0.98	-0.02	0.07	0.06	0.58	0.46	0.03	0.30	0.05	1.00	0.50	0.03	0.06	0.88	0.06	0.76	0.20	0.39
																0.50	1.00	0.08	0.09	0.55	0.11	0.40	0.03	0.47
															0.53	0.03	0.08	1.00	0.99	0.29	0.94	0.00	0.18	0.39
0.02	0.08	0.04	0.30	0.32	0.18	0.11	0.07	-0.00	0.04	-0.31	0.23	0.14	0.93	0.13	0.52	0.06	0.09	0.99	1.00	0.32	0.94	0.03	0.18	0.39
0.03	0.85	0.25	0.96	1.00	0.29	0.63	0.86	-0.02	0.04	-0.06	0.69	0.68	0.25	0.22	0.18	0.88	0.55	0.29	0.32	1.00	0.29	0.74	0.29	0.59
0.03	0.07	0.01	0.30	0.29	0.26	0.10	0.06	-0.01	-0.02	-0.34	0.29	0.13	0.99	0.10	0.73	0.06	0.11	0.94	0.94	0.29	1.00	0.03	0.26	0.40
0.02	0.69	-0.03	0.69	0.74	0.06	0.11	0.73	0.02	-0.02	-0.00	0.41	0.00	0.01	-0.00	0.03	0.76	0.40	0.00	0.03	0.74	0.03	1.00	0.06	0.09
0.03	0.18	0.07		0.29	1.00	0.39	0.18	-0.02	0.02	-0.10	0.73	0.65	0.26	0.06	0.49	0.20	0.03	0.18	0.18	0.29	0.26	0.06	1.00	0.73
0.03	0.40	0.29	0.72	0.59	0.73	0.63	0.38	-0.04	0.07	-0.09	0.86	0.75	0.39	0.24	0.43	0.39	0.47	0.39	0.39	0.59	0.40	0.09	0.73	1.00
Longitude	PropertyGFATotal	NumberofFloors	SiteEnergyUse(kBtu)	Electricity(kBtu)	NaturalGas(therms)	stPropertyUseTypeGFA	PropertyGFABuilding(s)	Latitude	YearBuilt	ENERGYSTARScore	TotalGHGEmissions	*stPropertyUseTypeGFA	SiteEUIWN(kBtu/sf)	PropertyGFAParking	GHGEmissionsIntensity	₃stPropertyUseTypeGFA	SteamUse(kBtu)	SourceEUIWN(kBtu/sf)	SourceEUI(kBtu/sf)	Electricity(kWh)	SiteEUI(kBtu/sf)	NumberofBuildings	NaturalGas(kBtu)	SiteEnergyUseWN(kBtu)
	0.03 -0.03 0.03 0.03 0.02 0.03 0.01 -0.05 -0.03 0.04 0.03 0.04 0.03 0.02 0.02 0.02 0.03 0.03 0.03	0.03 1.00 0.03 0.40 0.03 0.80 0.03 0.85 0.03 0.99 0.01 0.02 0.05 0.10 0.03 0.52 0.03 0.04 0.00 0.40 0.40	0.03 1.00 0.40 -0.03 0.40 1.00 0.03 0.85 0.25 0.03 0.81 0.47 0.03 0.99 0.36 0.01 0.02 0.02 0.05 0.10 0.15 0.03 0.41 0.00 0.04 0.02 0.24 0.03 0.97 0.34 0.04 0.02 0.04 0.05 0.04 0.02 0.04 0.00 0.00 0.04 0.02 0.04 0.00 0.00 0.04 0.02 0.04 0.00 0.00 0.05 0.03 0.05 0.03 0.02 0.08 0.04 0.03 0.05 0.25 0.03 0.07 0.01 0.00 0.03 0.05 0.25 0.03 0.07 0.01 0.00 0.03 0.05 0.05 0.03 0.03 0.05 0.05 0.05 0 0.05 0.03 0.05 0.05 0.05 0 0.05 0.05 0.05 0.05 0.05 0	0.03 1.00 0.40 0.80 0.03 0.40 1.00 0.21 0.03 0.85 0.25 0.96 0.03 0.81 0.47 0.63 0.03 0.99 0.36 0.81 0.01 0.02 0.02 0.02 0.05 0.10 0.15 0.03 0.03 0.07 0.02 0.02 0.04 0.53 0.14 0.86 0.03 0.52 0.22 0.75 0.03 0.04 0.04 0.17 0.04 0.02 0.04 0.31 0.03 0.97 0.34 0.84 0.02 0.04 0.03 0.27 0.02 0.03 0.27 0.04 0.30 0.02 0.04 0.03 0.27 0.02 0.04 0.03 0.27 0.02 0.05 0.03 0.27 0.02 0.05 0.03 0.27	0.03 1.00 0.40 0.80 0.85 0.03 0.40 1.00 0.21 0.25 0.03 0.85 0.25 0.96 1.00 0.03 0.85 0.25 0.96 1.00 0.03 0.81 0.47 0.63 0.63 0.03 0.99 0.36 0.81 0.86 0.01 0.02 0.02 0.02 0.02 0.05 0.10 0.15 0.03 0.04 0.04 0.05 0.01 0.00 0.06 0.06 0.03 0.77 0.02 0.02 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.07 0.21 0.02 0.06 0.06 0.06 0.06 0.06 0.05 0.06 0.06 0.05 0.06 0.05 0.06 0.05 0.06 0.05 0.06 0.05	0.03 1.00 0.40 0.80 0.85 0.18 0.03 0.40 1.00 0.21 0.25 0.07 0.03 0.85 0.25 0.96 1.00 0.29 0.02 0.03 0.85 0.25 0.96 1.00 0.29 1.00 0.02 0.81 0.47 0.63 0.63 0.39 0.03 0.99 0.36 0.81 0.36 0.18 0.01 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.03 0.04 -0.03 0.04 0.02 -0.09 -0.06 0.01 0.02 0.09 -0.06 0.01 0.02 0.09 -0.09 0.02 0.09 0.09 0.09 0.09 0.09 0.09 0.09 0.09 0.09 0.09 0.09 0.09 0.09	0.03 1.00 0.40 0.80 0.85 0.18 0.81 0.03 0.40 1.00 0.21 0.25 0.07 0.47 0.03 0.85 0.25 0.96 1.00 0.96 0.51 0.63 0.03 0.85 0.25 0.96 100 0.29 0.63 0.03 0.81 0.47 0.63 0.63 0.39 1.00 0.03 0.99 0.36 0.81 0.86 0.18 0.79 0.01 0.02 0.02 0.02 0.02 0.02 0.05 0.03 0.10 0.15 0.03 0.04 0.02 0.02 0.03 0.10 0.15 0.03 0.04 0.02 0.05 0.03 0.07 0.02 0.09 0.06 0.10 0.08 0.04 0.53 0.14 0.88 0.69 0.73 0.51 0.03 0.04 0.02 0.02 0.06	0.03 1.00 0.40 0.80 0.85 0.18 0.81 0.99 0.03 0.40 1.00 0.21 0.25 0.07 0.47 0.36 0.03 0.85 0.25 0.96 1.00 0.95 0.51 0.63 0.81 0.03 0.85 0.25 0.96 1.00 0.92 0.63 0.86 0.03 0.81 0.47 0.63 0.63 0.39 0.18 0.07 0.51 0.09 0.18 0.02 0.81 0.47 0.63 0.63 0.39 1.00 0.79 0.03 0.79 1.00 0.09 0.05 0.01 0.07 0.01 0.02	0.03 1.00 0.40 0.80 0.85 0.18 0.81 0.99 0.02 0.03 0.40 1.00 0.21 0.25 0.07 0.47 0.36 0.02 0.03 0.80 0.21 1.00 0.96 0.51 0.63 0.81 0.02 0.03 0.85 0.25 0.96 1.00 0.29 0.63 0.86 0.02 0.03 0.85 0.25 0.96 1.00 0.29 0.63 0.86 0.02 0.03 0.81 0.47 0.63 0.81 0.92 1.00 0.39 0.81 0.02 0.03 0.89 0.36 0.81 0.86 0.18 0.79 1.00 0.02 0.01 0.02 0.02 0.02 0.02 0.05 0.02 1.00 0.05 0.10 0.15 0.03 0.04 0.02 0.00 0.8 0.12 0.03 0.04 0.05 0.08 0.12 0.03 0.04 0.05 0.08 0.12 0.03 0.04 0.03 0.04 0.05 0.08 0.05 0.08 0.03 0.04 0.03 0.04 0.05 0.08 0.05 0.08 0.03 0.04 0.00 0.05 0.05 0.05 0.05 0.05 0.05	0.03	0.03 1.00 0.40 0.80 0.85 0.18 0.81 0.99 -0.02 0.10 0.07 -0.03 0.40 1.00 0.21 0.25 0.07 0.47 0.36 -0.02 0.15 0.02 0.03 0.85 0.25 0.96 1.00 0.95 0.51 0.63 0.81 -0.02 0.03 -0.99 0.03 0.85 0.25 0.96 1.00 0.29 1.03 0.86 -0.02 0.04 -0.06 0.03 0.81 0.63 0.39 1.00 0.79 -0.05 0.02 0.08 0.03 0.99 0.36 0.81 0.80 0.18 0.02 0.08 0.06 0.00 0.01 0.02 0.02 -0.02 -0.02 -0.02 0.02 1.00 0.12 0.08 0.03 0.04 0.02 0.05 0.02 1.00 0.02 0.00 0.00 0.02 0.00 0.03	0.03 1.00 0.40 0.80 0.85 0.18 0.91 0.99 0.02 0.10 0.07 0.53 0.03 0.40 1.00 0.21 0.25 0.07 0.47 0.36 0.02 0.15 0.02 0.15 0.02 0.15 0.02 0.15 0.02 0.15 0.02 0.15 0.02 0.15 0.02 0.15 0.02 0.03 0.03 0.05 0.00 0.09 0.03 0.81 0.07 0.51 0.02 0.03 0.18 0.07 0.51 0.02 0.03 0.18 0.02 0.02 0.01 0.73 0.02 0.02 0.01 0.73 0.02 0.02 0.03 0.04 0.02 0.02 0.08 0.05 0.01 0.03 0.04 0.02 0.03 0.04 0.02 0.03 0.04 0.02 0.03 0.01 0.03 0.01 0.03 0.01 0.03 0.02 0.03 0.01 0.03 0.	0.03	0.03 1.00 0.40 0.80 0.85 0.18 0.81 0.99 -0.02 0.10 0.07 0.53 0.52 0.04 0.03 0.40 1.00 0.21 0.25 0.07 0.47 0.36 0.02 0.15 0.02 0.14 0.22 0.00 0.03 0.80 0.21 1.00 0.96 0.51 0.63 0.81 0.02 0.03 0.09 0.86 0.75 0.27 0.03 0.85 0.25 0.96 1.00 0.29 0.63 0.86 0.02 0.04 0.06 0.69 0.68 0.25 0.03 0.03 0.08 0.07 0.51 0.29 1.00 0.39 0.18 0.02 0.02 0.10 0.73 0.65 0.26 0.03 0.03 0.03 0.03 0.03 0.63 0.83 1.00 0.79 0.05 0.20 0.08 0.51 0.66 0.08 0.03 0.03 0.03 0.03 0.03 0.03 0.03	0.03 1.00 0.40 0.80 0.85 0.18 0.81 0.99 0.02 0.10 0.07 0.53 0.52 0.04 0.40 0.03 0.80 0.21 1.00 0.96 0.51 0.63 0.86 0.02 0.15 0.02 0.14 0.22 0.00 0.42 0.03 0.85 0.25 0.96 1.00 0.29 0.63 0.86 0.02 0.04 0.06 0.69 0.68 0.25 0.22 0.03 0.30 0.85 0.25 0.96 1.00 0.29 0.63 0.86 0.02 0.04 0.06 0.69 0.68 0.25 0.22 0.03 0.30 0.80 0.27 0.07 0.51 0.29 1.00 0.39 0.18 0.02 0.02 0.08 0.51 0.66 0.66 0.60 0.60 0.60 0.60 0.60 0.6	0.03	0.03	0.03 1.00 0.40 0.80 0.85 0.18 0.81 0.99 0.02 0.10 0.07 0.53 0.52 0.04 0.40 0.02 0.97 0.44 0.03 0.40 1.00 0.21 0.25 0.07 0.47 0.36 0.02 0.15 0.02 0.14 0.22 0.00 0.42 0.04 0.34 0.88 0.03 0.80 0.21 1.00 0.96 0.51 0.63 0.81 0.02 0.03 0.09 0.86 0.75 0.27 0.17 0.31 0.84 0.68 0.03 0.85 0.25 0.96 1.00 0.29 0.63 0.85 0.02 0.04 0.06 0.69 0.65 0.25 0.27 0.17 0.31 0.84 0.68 0.03 0.85 0.25 0.96 1.00 0.29 0.63 0.85 0.02 0.02 0.04 0.06 0.69 0.65 0.25 0.27 0.17 0.31 0.84 0.65 0.03 0.18 0.07 0.51 0.29 1.00 0.39 0.18 0.02 0.02 0.04 0.06 0.69 0.65 0.25 0.25 0.06 0.49 0.20 0.03 0.03 0.30 0.03 0.63 0.39 1.00 0.79 0.05 0.20 0.08 0.51 0.66 0.03 0.48 0.11 0.77 0.26 0.03 0.99 0.36 0.81 0.86 0.83 0.99 1.00 0.79 0.05 0.20 0.08 0.06 0.55 0.54 0.03 0.27 0.03 0.98 0.46 0.01 0.02 0.02 0.02 0.02 0.02 0.02 0.02	0.03	0.03	0.03	0.03	0.03	0.03 0.40 1.00 0.21 0.25 0.07 0.47 0.36 0.02 0.15 0.02 0.14 0.22 0.00 0.42 0.04 0.34 0.08 0.30 0.04 0.25 0.01 0.03 0.07 0.03 0.07 0.03 0.07 0.03 0.08 0.21 1.00 0.96 0.51 0.53 0.81 0.02 0.03 0.09 0.66 0.75 0.27 0.17 0.31 0.84 0.60 0.27 0.30 0.96 0.30 0.96 0.51 0.03 0.85 0.25 0.96 1.00 0.29 0.53 0.86 0.02 0.04 0.06 0.69 0.68 0.25 0.26 0.06 0.49 0.20 0.03 0.18 0.10 0.29 0.74 0.29 0.03 0.18 0.07 0.51 0.29 1.00 0.39 0.18 0.02 0.02 0.10 0.73 0.65 0.26 0.06 0.49 0.20 0.03 0.18 0.18 0.29 0.26 0.06 0.06 0.09 0.10 0.20 0.18 0.18 0.29 0.26 0.06 0.06 0.09 0.10 0.20 0.10 0.20 0.10 0.20 0.20 0.20

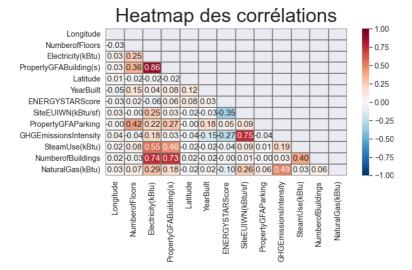
II.2.4. Leakage en gardant electricity, steam et gas? 0.8



 $[\]rightarrow$ données identiques à ques exceptions près

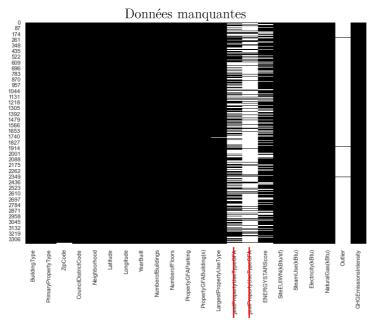
 $[\]rightarrow$ SiteEUIWN \approx (Electricity + Steam + Gas)/GFA

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



- \rightarrow ENERGYSTARScore voulu.
- \rightarrow Outlier va nous servir.

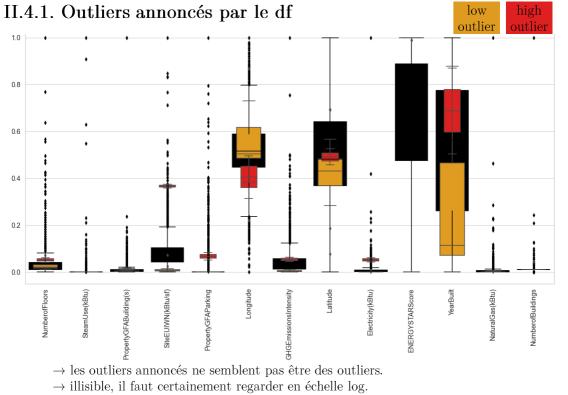
II.4. Suppression d'éléments

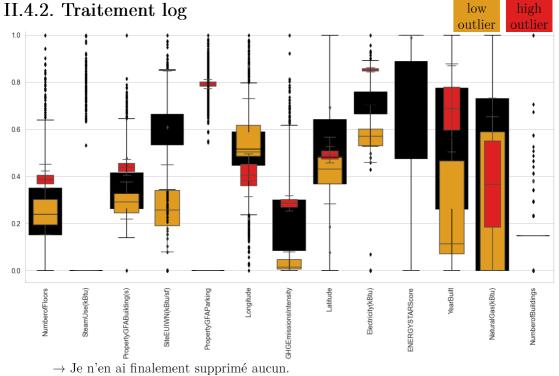
- Bâtiments non résidentiels (énoncé) \rightarrow suppression de la moitié des lignes du df.
- Deux éléments supprimés car presque vides

Do		лешение вирр	riffes car p	resque viac	,,,				
PropertyGFABuild	ding(s)	LargestPropertyUseType	ENERGYSTARScore	SiteEUIWN(kBtu/sf)	SteamUse(kBtu)	Electricity(kBtu)	NaturalGas(kBtu)	Outlier	GHGEmissionsInten
	63150	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

NaN

- Outliers



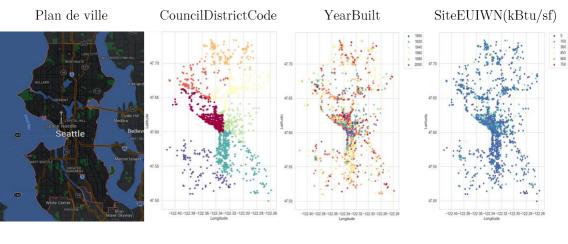


 \rightarrow Faut-il garder ce traitement log pour la pipeline de régression ?

II.5. Imputation

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

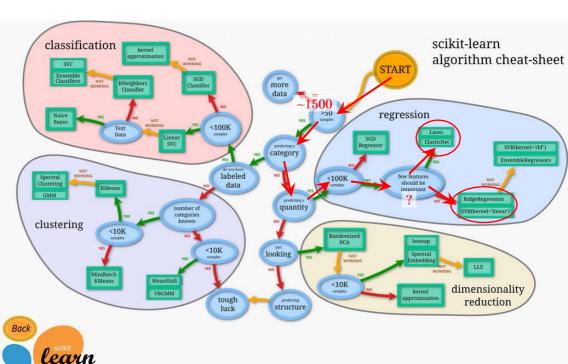
Interlude



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

Plan I Présentation du projet 1. best estimator Importance de l'ENERGYSTARScore Nettoyage Plus de nettoyage **Traitement** best estimator bis Conclusions Empreinte carbone

III.1.1. Identification du problème



III.1.2. Paramétrage

scalers = [RobustScaler(), StandardScaler()]

train_test_split()

```
estimator
                                                                                 param_nums
                                                        param_names
0
               Ridge()
                                                         [ridge_alpha]
                                                                             [[0.1, 0.3, 1, 3, 10]]
               Lasso()
                                                         [lasso alpha]
                                                                             [[0.1, 0.3, 1, 3, 10]]
   DummyRegressor()
                                                                            [[mean], [median]]
                                          [dummyregressor_strategy]
3
    LinearRegression()
                                                                                        [[True]]
                                        [linearregression_fit_intercept]
        ElasticNetCV() [elasticnetcv_l1_ratio, elasticnetcv_n_alphas] [[0.1, 0.3, 1], [30, 100]]
```

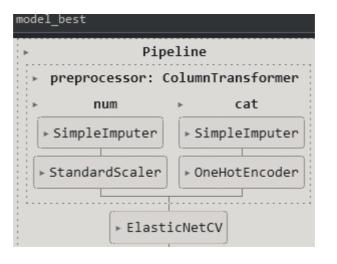
encoders = [OrdinalEncoder(handle_unknown='use_encoded_value', unknown_value=-1), OneHotEncoder(sparse=False,handle_unknown='ig

III.1.3. Scoring

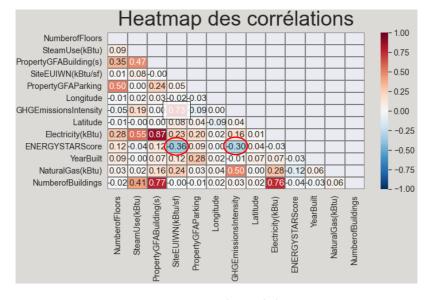
time	msqe	mae	r2	best score	Estimator
0.078462	1.303636e+04	4.305065e+01	-1.670998e+00	1.279617e-01	Ridge()
0.079444	1.070123e+04	4.271999e+01	-1.192557e+00	2.896550e-01	Lasso()
0.089119	4.883032e+03	4.547045e+01	-4.763885e-04	-2.075314e-03	DummyRegressor()
0.088264	1.305103e+04	4.305699e+01	-1.674003e+00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

time_1 = time_passed/(param_loops)

III.1.4. best_estimator_

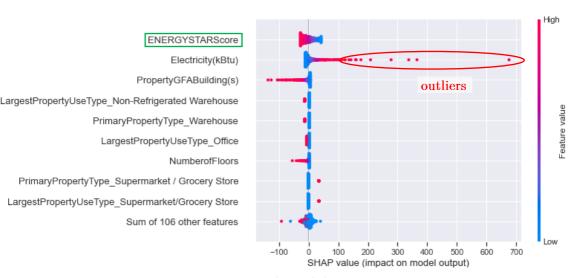


III.2.1. Corrélation



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

III.2.2. shapley values



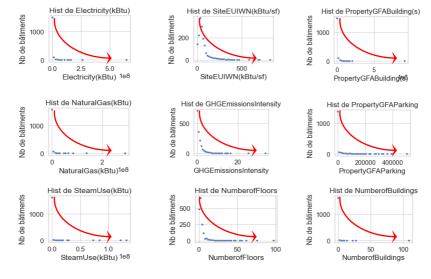
Feature principale pour SiteEUIWN(kBtu/sf)

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)
```

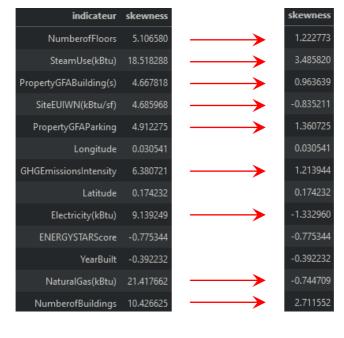
- Scores différents \rightarrow feature non-négligeable
- r^2 supérieur, mae et mse inférieures $\rightarrow E^*$ aide la prédiction

III.3.1. Nettoyages nécessaires



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

$III.3.2. \log$



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
${\sf GHGEmissionsIntensity}$	0	7

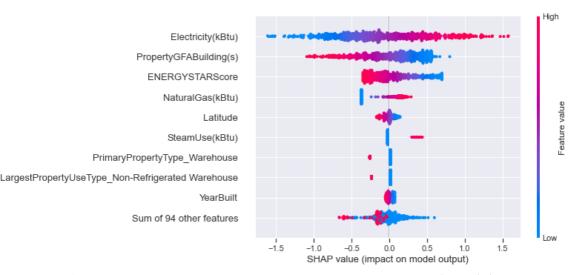
- Suppression seulement des éléments qui sont toujours hors interquartiles même après le log.
 - (suppressions selon Longitude pas forcément nécessaires)

III.4.1. best_estimator_

```
print(dfscores2[col score] - dfscores[col score])
     best score
                                                                  time
                            r2
                                          mae
                                                       msge
   5.456720e-01 2.220879e+00 -4.266704e+01 -1.303612e+04 -0.009930
   8.160592e-02
                  1.537467e+00 -4.227454e+01 -1.070087e+04 -0.012977
  -6.066227e-03
                  4.737353e-04 -4.491818e+01 -4.882494e+03 -0.022371
   5.547913e-01
                  2.221856e+00 -4.267244e+01 -1.305078e+04 -0.018669
   6.804202e-01
                  5.530310e-01 -4.508878e+01 -4.882791e+03 -0.009507
   4.792426e-01
                  2.145800e+00 -4.280415e+01 -1.266811e+04 -0.006643
                     preprocessor: ColumnTransformer
                                           cat
                           num
                     ▶ SimpleImputer
                                     ▶ SimpleImputer
                     ▶ StandardScaler
                                     ▶ OneHotEncoder
                             ▶ ElasticNetCV
```

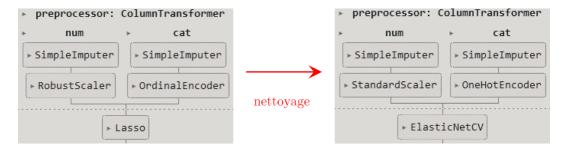
- Meilleurs scores dans les pipelines (sauf le Dummy)
- best_estimator_ similaire
- Shapley values mieux représentées

III.4.2. shapley values



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

III.5.1 GHGEmissionsIntensity



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

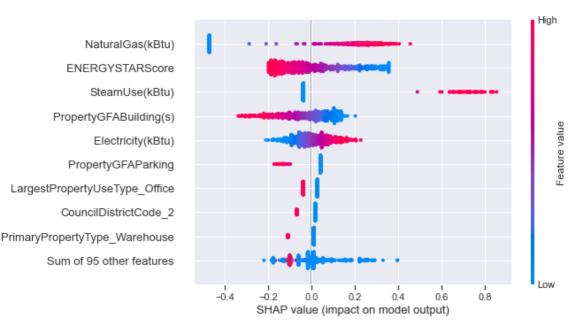
- Pour cette target je trouve une importance nulle de l'ENERGYSTARScore avant nettoyage {log + outliers} malgré ce qu'annonce le tableau de corrélations.
 - Le best estimator est modifié après nettoyage.

III.5.2. shapley values



- ENERGYSTARScore négligeable avant nettoyage

III.5.3. shapley values



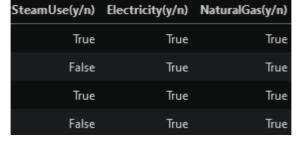
- ENERGYSTARScore majeur après nettoyage

CCL

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



- V. Approfondissement suite à la soutenance
- V.1. Recommandations ajoutées
- V.1.1. Correction d'un data leakage



V. Approfondissement suite à la soutenanceV.1. Recommandations ajoutéesV.1.2. Ajout d'estimateurs non-linéaires

estimator

RandomForestRegressor()

AdaBoostRegressor()

GradientBoostingRegressor()

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	Flastic NetCV/	[elasticnetov 1 ratio elasticnetov n alphas]	[[0.1.0.3.1] [30.100]]

[randomforestregressor_max_depth, randomfores...

[adaboostregressor_n_estimators, adaboostregr...

[gradientboostingregressor_random_state]

param_nums

[[2], [0]]

[[30, 100], [0]]

[[0]]

param_names

V.2. Résultats V.2.1. best_estimator_ (SiteEUIWN)

OrdinalEncoder(handle_unknown='use_encoded_val...

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

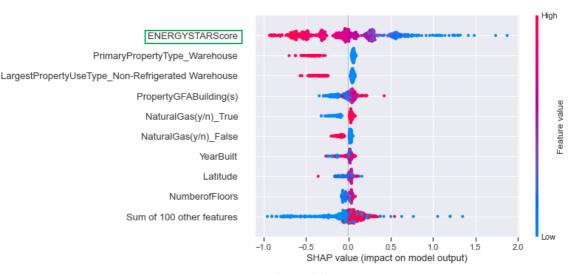
OneHotEncoder(handle_unknown='ignore', sparse=... StandardScaler() GradientBoostingRegressor()

RobustScaler() GradientBoostingRegressor()

0.60

0.61

V.2.2. shapley values (SiteEUIWN)



- Feature principale pour SiteEUIWN(kBtu/sf)

r² avec E*score r² sans E*score

3/21200033, 0.300040331201/41-

- E*score indispensable

V.2.4. best_estimator_ (GHGEI)

Encoder

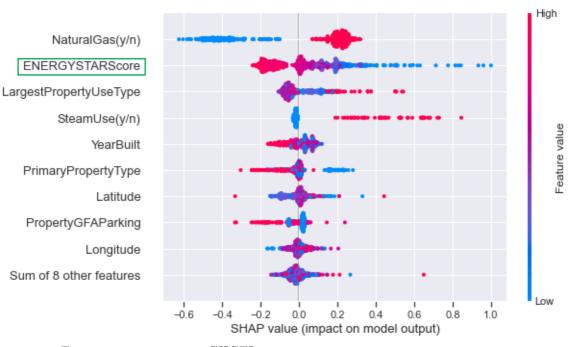
OneHotEncoder(handle_unknown='ignore', sparse=	RobustScaler()	Ridge()	0.50
OneHotEncoder(handle_unknown='ignore', sparse=	RobustScaler()	ElasticNetCV()	0.51
OneHotEncoder(handle_unknown='ignore', sparse=	StandardScaler()	ElasticNetCV()	0.51
OneHotEncoder(handle_unknown='ignore', sparse=	StandardScaler()	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

Scaler

Estimator

best score

V.2.5. shapley values (GHGEI)



- Feature importante pour GHGEI

r² avec E*score r² 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.

