



202 EMISSION ANALYSIS

Seattle City



PLAN

- Introduction / Demo
- Exploration Analysis (EDA)
- Préprocessing
- Dashboard BI
- Models & interpretations
- Conclusion

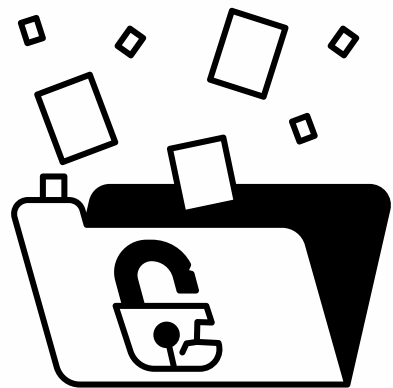
Projet

La collectivité d'une ville fait appel à notre agence spécialisée dans l'analyse des émissions CO2 des bâtiments. Le président de la collectivité souhaite mettre en place une stratégie à 4 ans pour réduire les émissions de gaz à effet de serre dans la ville de Seattle.



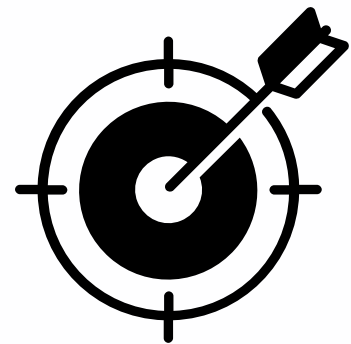
Exploration des données

Présentation du dataset



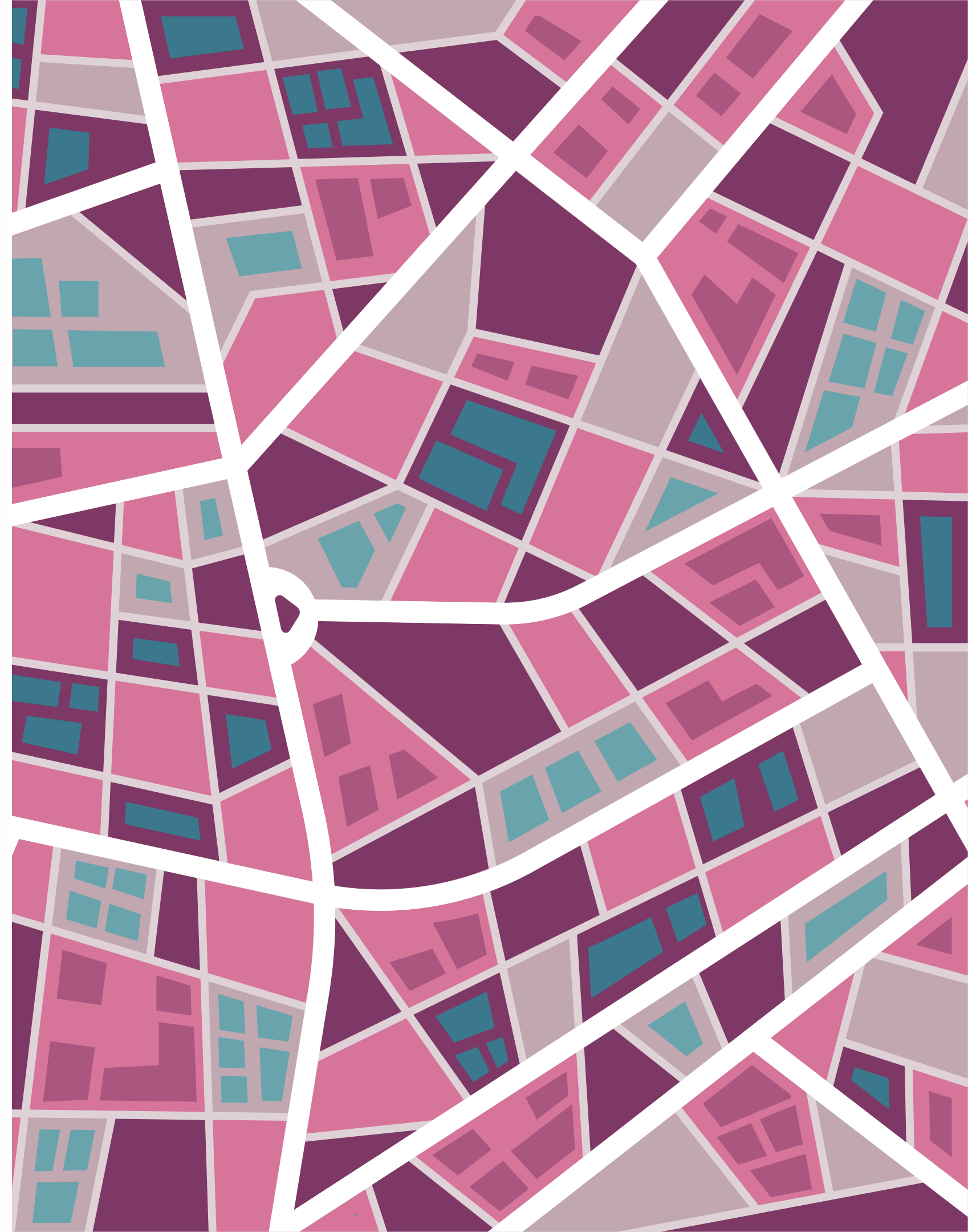
Dataset :

- 3376 Rows
- 46 Features



Target :

- TotalGHGEmissions





Analyse des données (EDA)

Catégories de données



La localisation



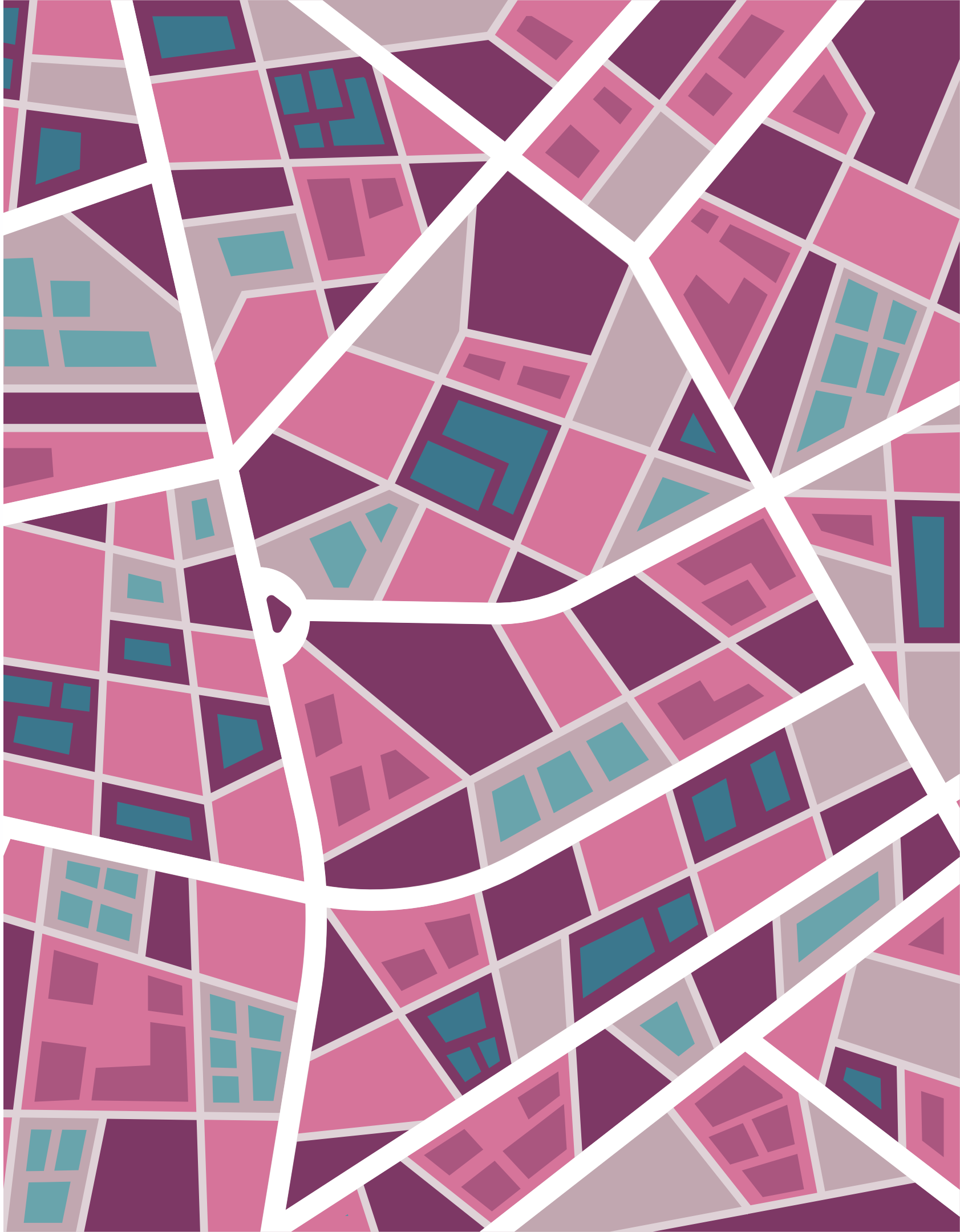
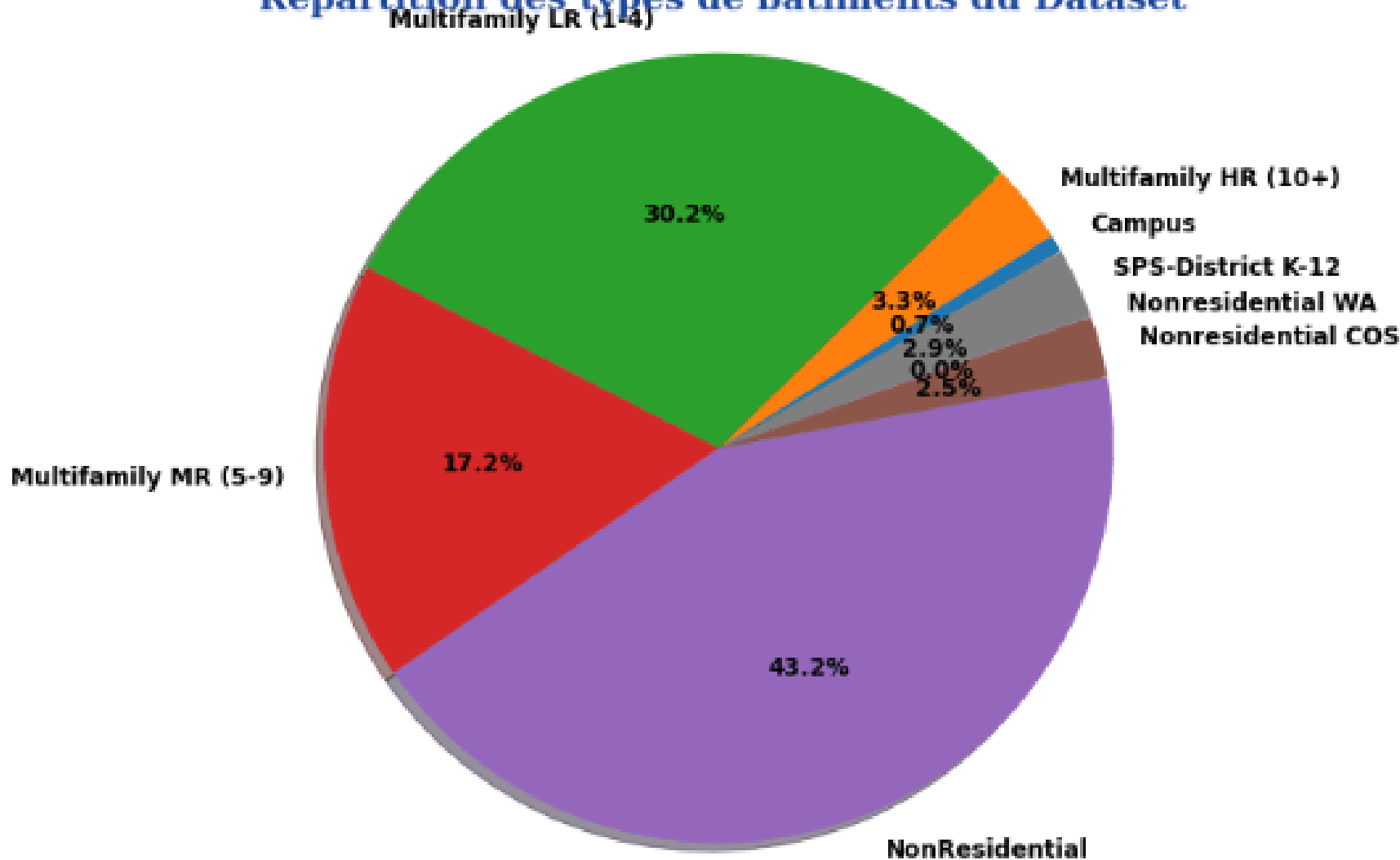
La consommation d'énergie



Les caractéristiques de l'appart

Analyse univariée

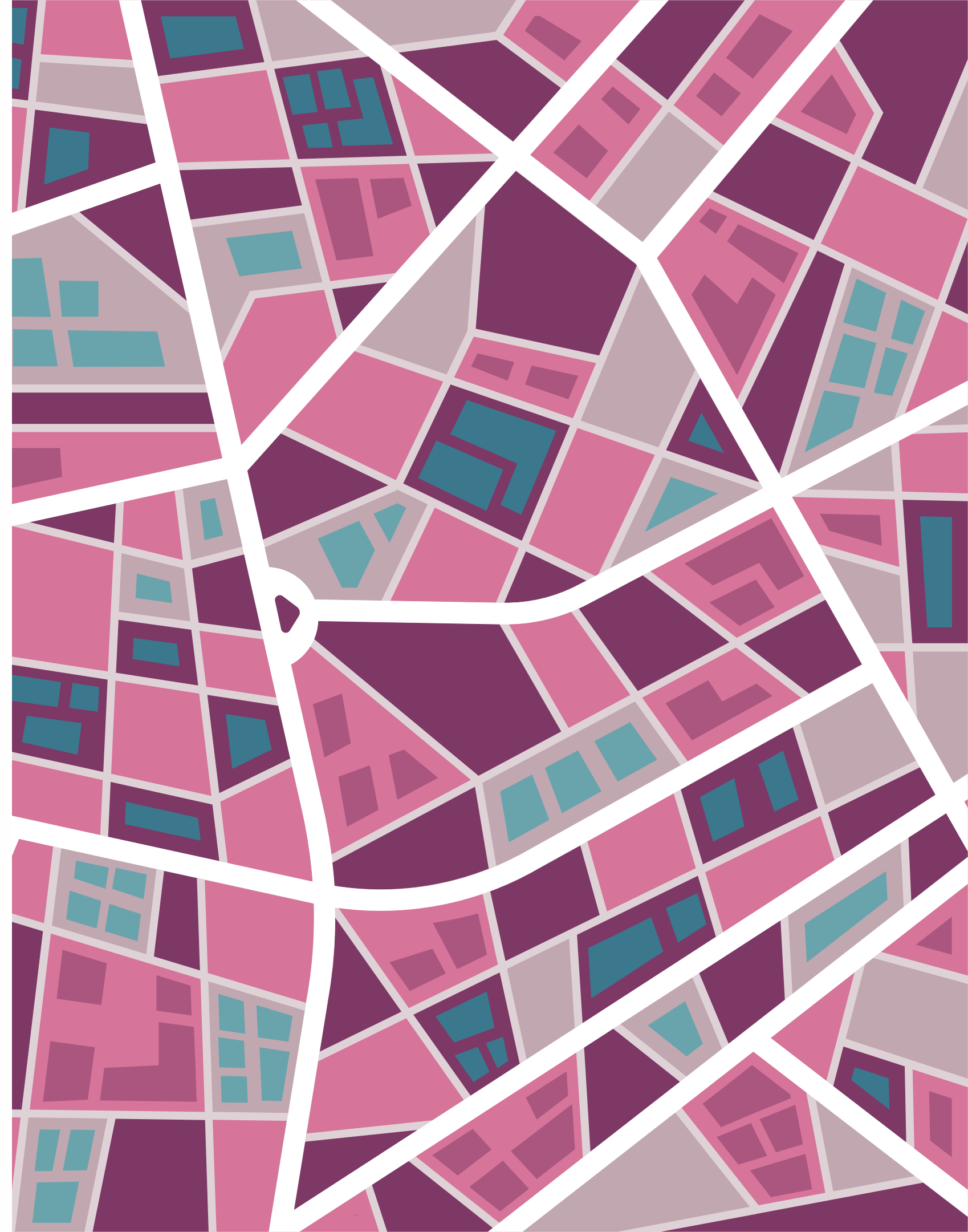
Répartition des types de bâtiments du Dataset



Exploration des données

Top 5 Catégorie des bâtiments - NonResidential

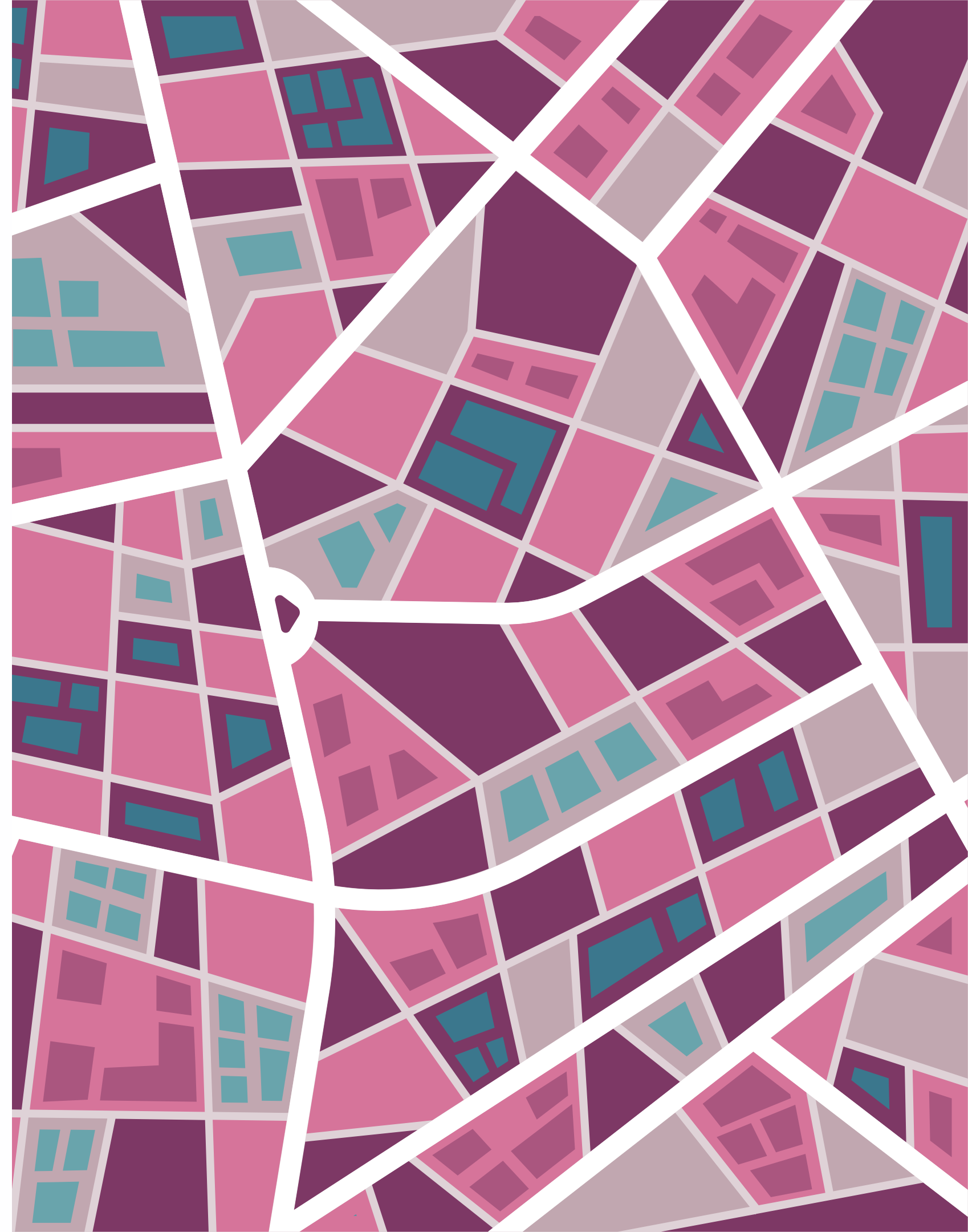
Small- and Mid-Sized Office	288
Warehouse	180
Large Office	168
Mixed Use Property	102
Retail Store	91
Hotel	76



Exploration des données

Top 5 Catégorie des bâtiments - Multifamily

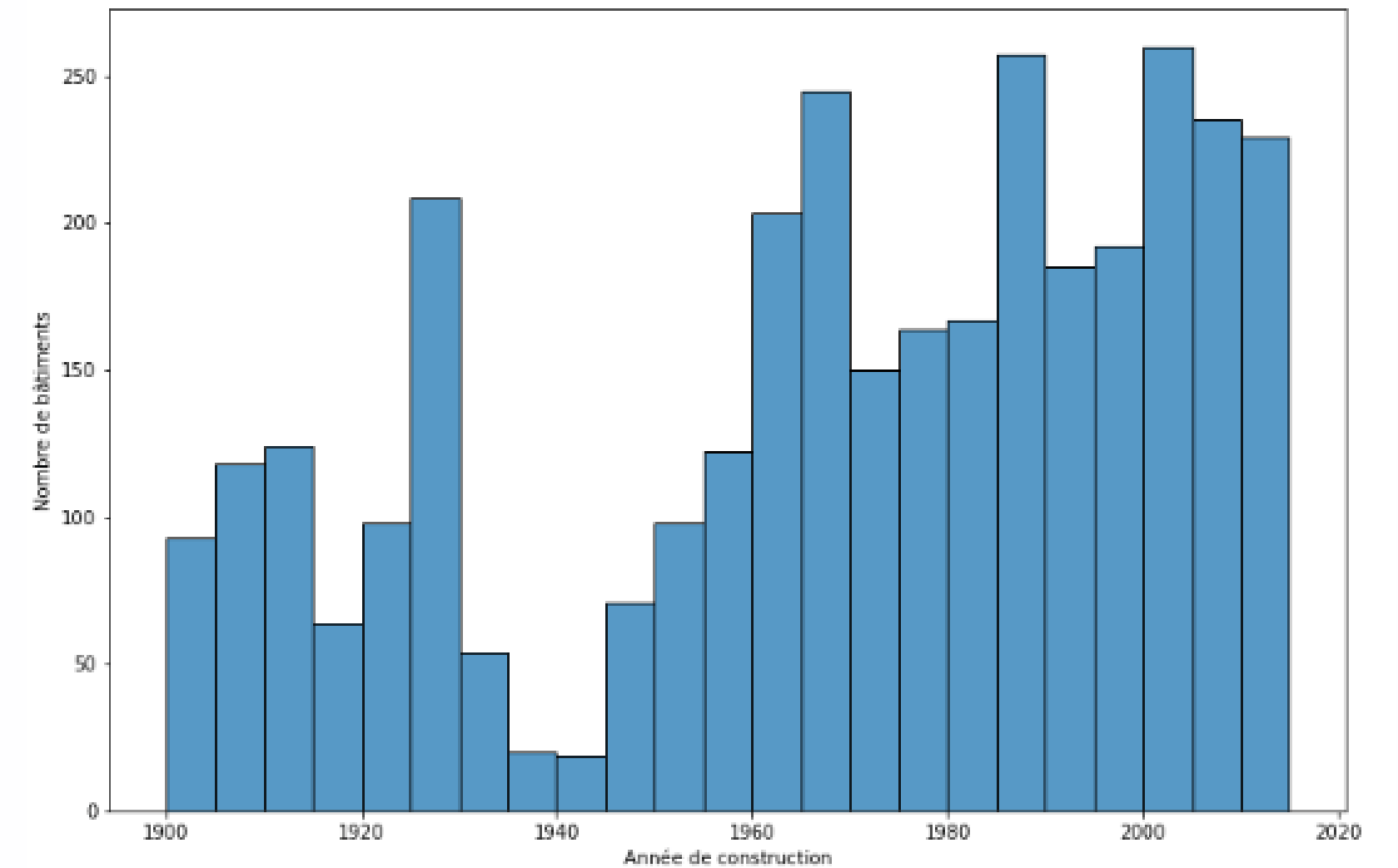
Low-Rise Multifamily	984
Senior Care Community	16
Mixed Use Property	11



Exploration des données

Analyse Univariée

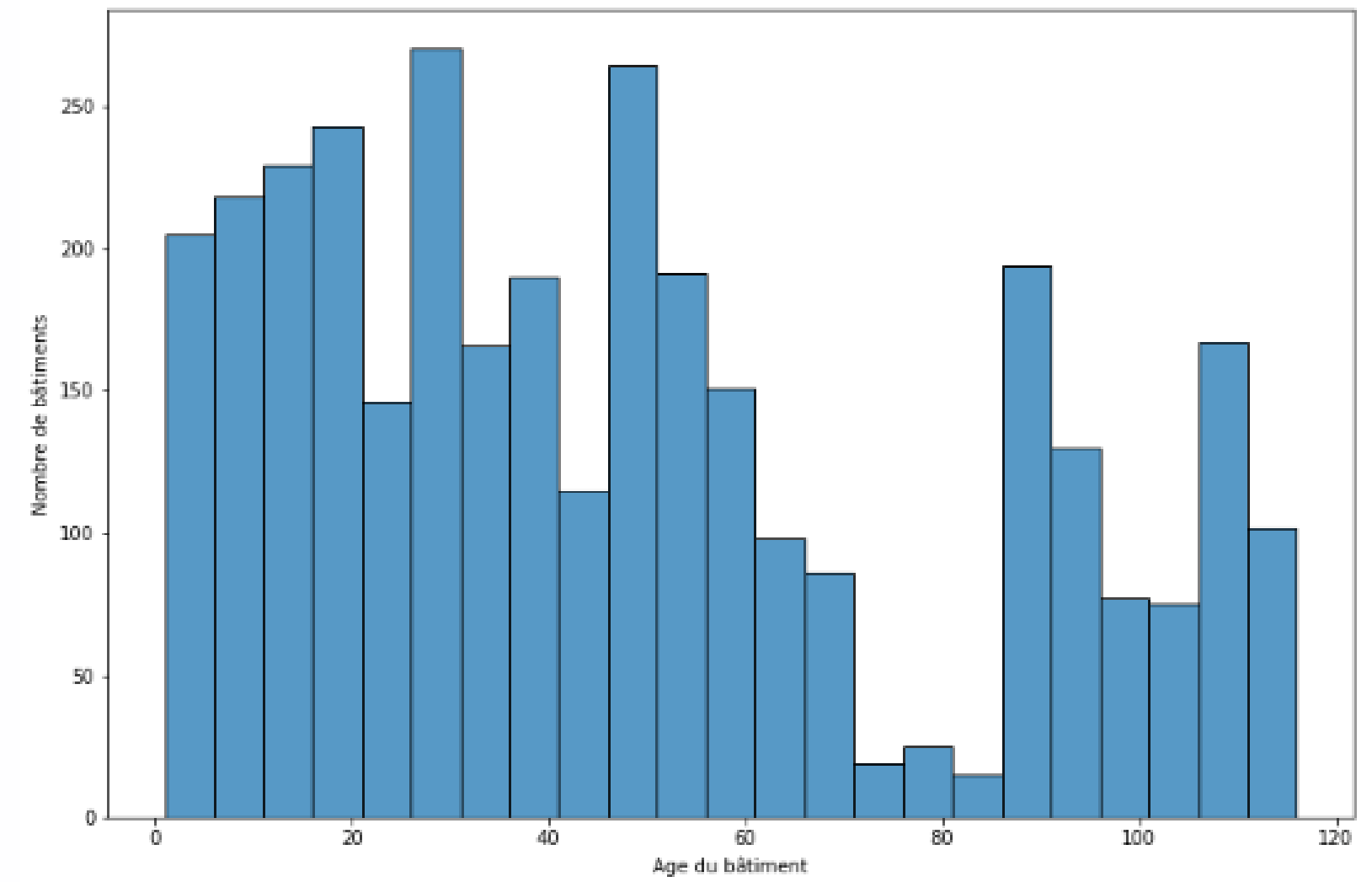
Distribution des années de construction des bâtiments



Exploration des données

Analyse Univariée

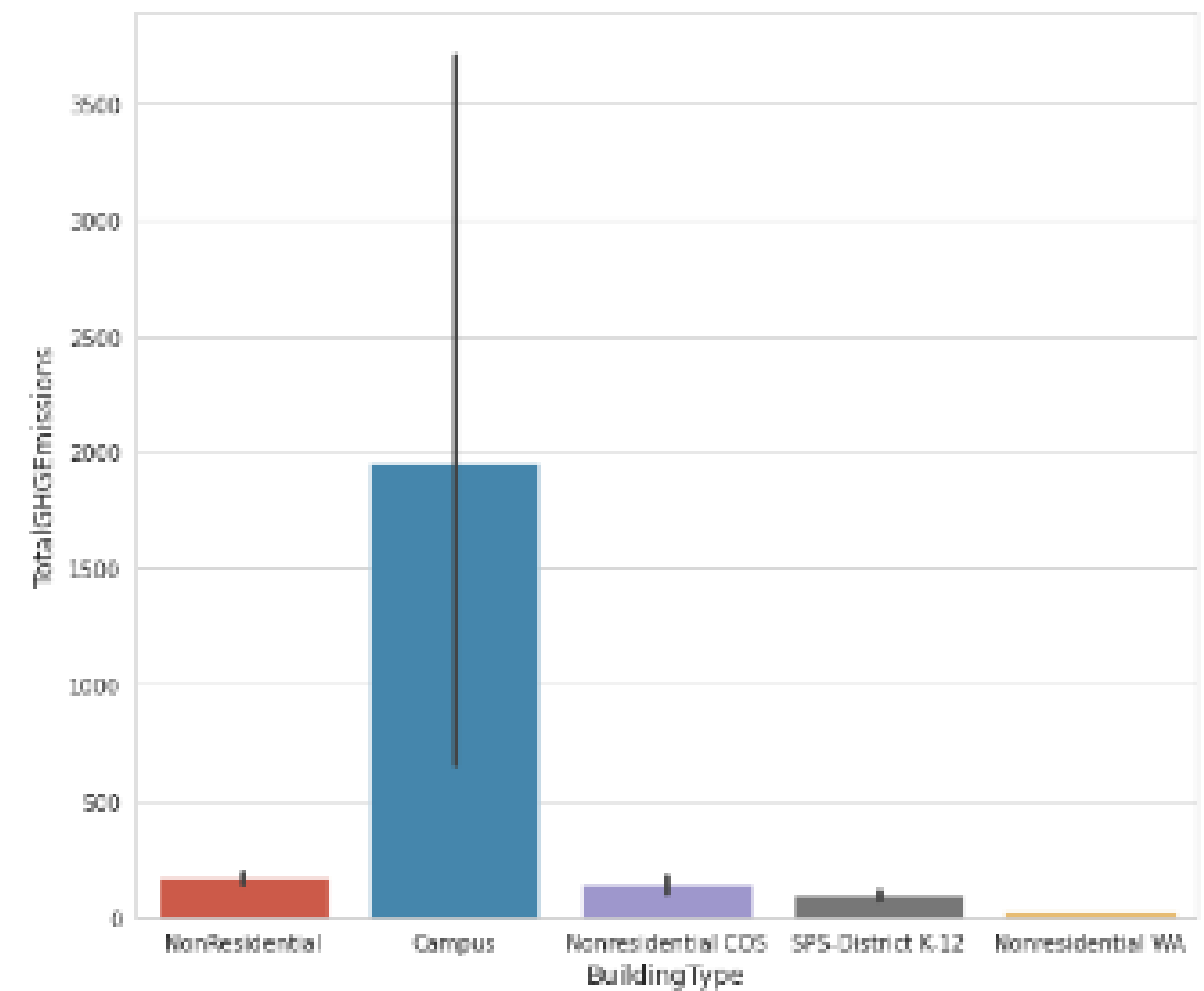
Distribution de l'âge des bâtiments





Exploration des données

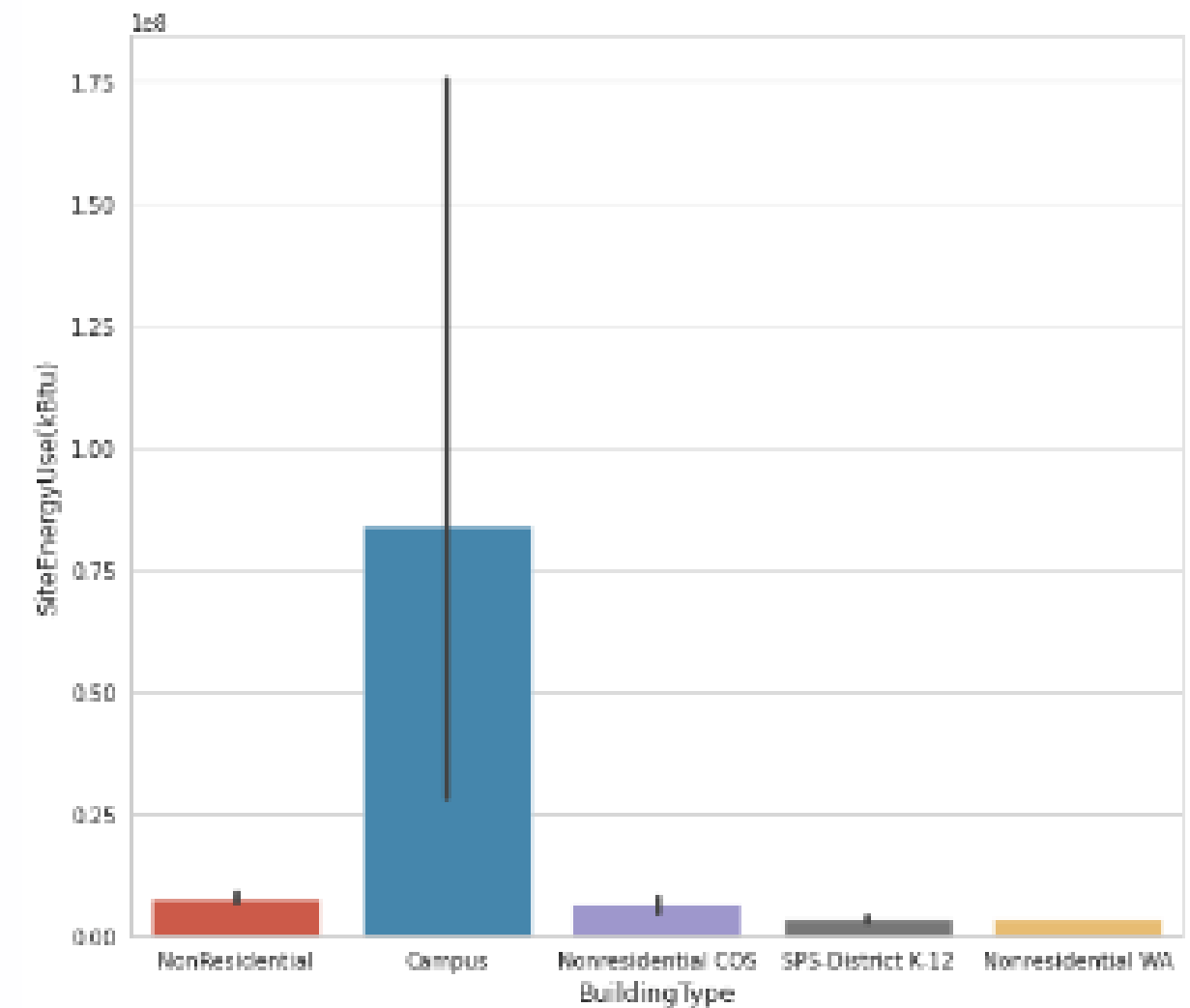
Analyse Multivariée



Répartition d'émmissions de CO2 en fonction du type de bâtiment

Exploration des données

Analyse Multivariée



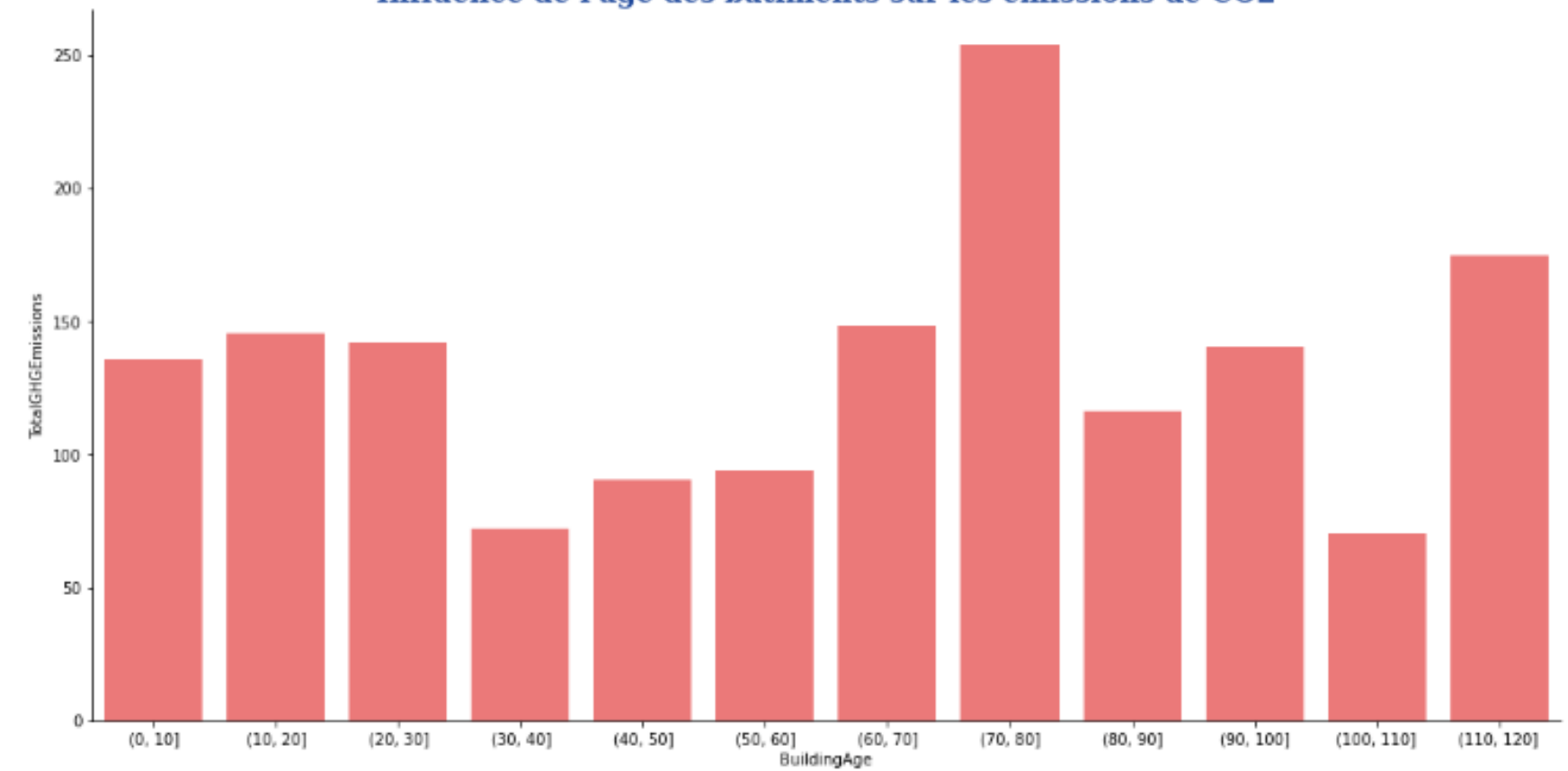
Répartition d'énergie consommées en fonction du type de bâtiment



Exploration des données

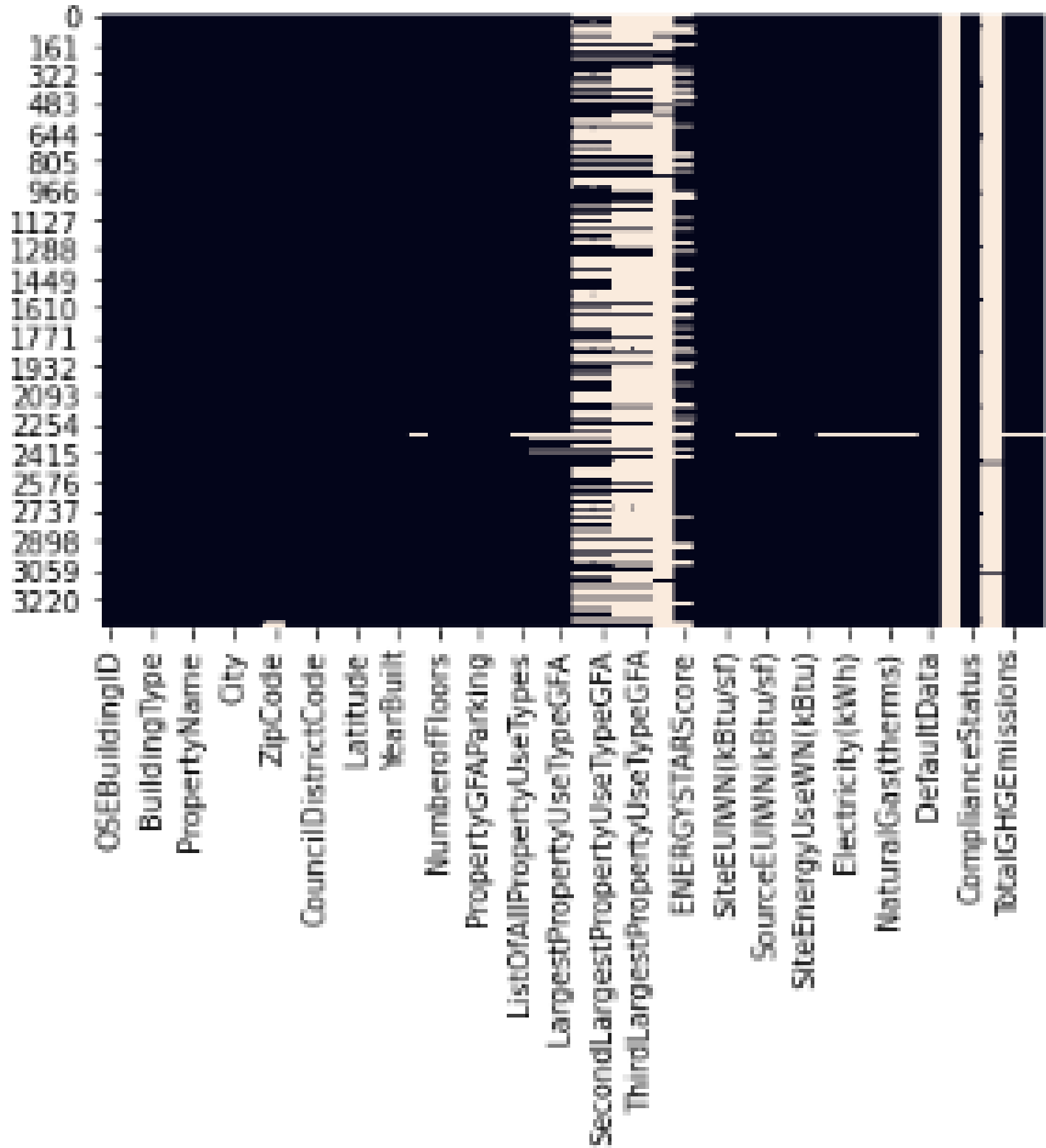
Analyse Multivariée

Influence de l'âge des bâtiments sur les émissions de CO2





Missing values





Missing values

	Total	Percent
Comments	3378	100.000000
Outlier	3344	99.052133
YearsENERGYSTARCertified	3257	96.475118
ThirdLargestPropertyUseType	2780	82.345972
ThirdLargestPropertyUseTypeGFA	2780	82.345972
SecondLargestPropertyUseType	1697	50.266588
SecondLargestPropertyUseTypeGFA	1697	50.266588
ENERGYSTARScore	843	24.970379
LargestPropertyUseTypeGFA	20	0.592417
LargestPropertyUseType	20	0.592417
ZipCode	16	0.473934
ListOfAllPropertyUseTypes	9	0.266588
SourceEUIWN(kBtu/sf)	9	0.266588
SourceEUI(kBtu/sf)	9	0.266588
Electricity(kWh)	9	0.266588
Electricity(kBtu)	9	0.266588
NaturalGas(therms)	9	0.266588
NaturalGas(kBtu)	9	0.266588
TotalGHGEmissions	9	0.266588
SteamUse(kBtu)	9	0.266588
GHGEmissionsIntensity	9	0.266588
NumberofBuildings	8	0.238967
SiteEUI(kBtu/sf)	7	0.207346
SiteEUIWN(kBtu/sf)	6	0.177725

Neffoyage des données

Features drop - raw

```
: df.min()
```

OSEBuildingID	1
BuildingType	Campus
PrimaryPropertyType	Distribution Center
PropertyName	#4706 Bitterlake
TaxParcelIdentificationNumber	0000000000
CouncilDistrictCode	1
Neighborhood	BALLARD
Latitude	47.49917
Longitude	-122.41425
YearBuilt	1900
NumberofFloors	0
PropertyGFAParking	0
ListofAllPropertyUseTypes	Adult Education
LargestPropertyUseType	Adult Education
SecondLargestPropertyUseType	Adult Education
SecondLargestPropertyUseTypeGFA	0.0
SourceEUI(kBtu/sf)	0.0
SiteEnergyUseWN(kBtu)	0.0
SteamUse(kBtu)	0.0
Electricity(kBtu)	-115417.0
DefaultData	False
ComplianceStatus	Compliant
TotalGHGEmissions	-0.8
GHGEmissionsIntensity	-0.02
dtype:	object

Neiŕoyage des don  es

Features drop - columns

'Electricity(kWh)','NaturalGas(therms)','SourceEUI
WN(kBtu/sf)','SiteEUIWN(kBtu/sf)','PropertyGFABui
lding(s)','SiteEnergyUse(kBtu)','SiteEUI(kBtu/sf)','La
rgestPropertyUseTypeGFA','PropertyGFATotal','Nu
mberofBuildings','NaturalGas(kBtu)'



Sélection des features

Corrélation entre features

	feature_1	feature_2	correlation
666	Electricity(kBtu)	Electricity(kWh)	1.000000
639	Electricity(kWh)	Electricity(kBtu)	1.000000
697	NaturalGas(therms)	NaturalGas(kBtu)	1.000000
724	NaturalGas(kBtu)	NaturalGas(therms)	1.000000
521	SourceEUIWN(kBtu/sf)	SourceEUI(kBtu/sf)	0.994317
494	SourceEUI(kBtu/sf)	SourceEUIWN(kBtu/sf)	0.994317
463	SiteEUIWN(kBtu/sf)	SiteEUI(kBtu/sf)	0.993549
436	SiteEUI(kBtu/sf)	SiteEUIWN(kBtu/sf)	0.993549
317	PropertyGFABuilding(s)	PropertyGFATotal	0.989823
263	PropertyGFATotal	PropertyGFABuilding(s)	0.989823
320	PropertyGFABuilding(s)	LargestPropertyUseTypeGFA	0.978422



Sélection des features

Feature permutation

	feature	feature importance
13	SiteEnergyUseWN(kBtu)	98.8645
18	GHGEmissionsIntensity	49.8671
9	LargestPropertyUseType	48.9185
1	PrimaryPropertyType	41.7929
12	SourceEUI(kBtu/sf)	35.561
15	Electricity(kBtu)	22.0977
7	NumberofFloors	21.6372
14	SteamUse(kBtu)	16.2648
2	CouncilDistrictCode	7.81821
3	Neighborhood	6.9013
0	BuildingType	6.08641
10	SecondLargestPropertyUseType	1.96588
4	Latitude	1.83993
6	YearBuilt	1.80296
11	SecondLargestPropertyUseTypeGFA	0.422591
8	PropertyGFAParking	0.222365
17	ComplianceStatus	0.164566
5	Longitude	0.134397



Sélection des features

VIF

	vif_index	features
0	5.612186	CouncilDistrictCode
1	10.708534	Latitude
2	6.102476	Longitude
3	5.028966	YearBuilt
4	2.707548	NumberofFloors
5	1.523874	PropertyGFAParking
6	2.851095	SecondLargestPropertyUseTypeGFA
7	4.408256	SourceEUI(kBtu/sf)
8	8.104120	SiteEnergyUseWN(kBtu)
9	2.221709	SteamUse(kBtu)
10	3.003748	Electricity(kBtu)
11	1.054888	DefaultData
12	12.041810	TotalGHGEmissions



Sélection des features

10 features conservées :

- **BuildingType**
- **PrimaryPropertyType**
- **NumberOfFloors**
- **PropertyGFAParking**
- **LargestPropertyUseType**
- **SecondLargestPropertyUseType**
- **SecondLargestPropertyUseTypeGFA**
- **SourceEUI(kBtu/sf)**
- **SteamUse(kBtu)**
- **Electricity(kBtu)**



Model

lère itération

Linear Regression

Scoring : MAE 105 Tons CO2e



Model

2ème itération

Random Forest Regressor

Scoring : MAE 98 Tons CO₂e



Conclusion

Axes d'amélioration

- Learning Curves sur différentes catégories
- Features selection plus approfondie avec plusieurs itérations
- Déploiement sur Azure limitée
- Manque CI/CD Azure
- Interface graphique plus élaborée



Thank you! ✦