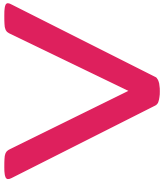
ANTICIPEZ LES BESOINS EN CONSOMMATION ELECTRIQUE DE BATIMENTS

Seattle

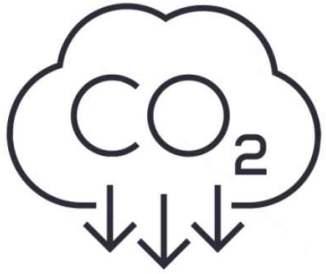
PROBLÉMATIQUE

_DONNÉES

Données: Chères



Energie

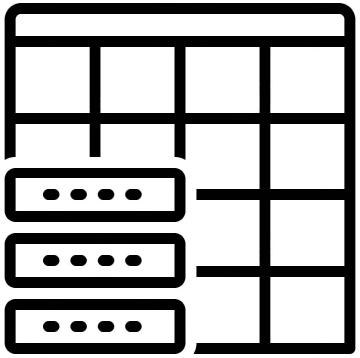


Emissions

OBJECTIF: Ville neutre en emission de CO2 en 2050

PROBLÉMATIQUE

_DATASET



2015-2016
50 COLONNES
3400 BATIMENTS

- Batiments: ID – Nom – Type – Catégorie – Dates
- Localisation: Latitude – Longitude – Ville – Quartier
- Structure: Etages – Nombre Batiments – Garage – Superficie
- Scoring: Energie Star
- Usages: PrimaryUse – SecondaryUse
- Relevés: Conso electrique – Gaz – Fuel – Emission CO2

NETTOYAGE

_DONNÉES

3400 BATIMENTS

- **Valeurs Manquantes:**
  - Suppression Colonnes > 90% NaN
  - Dropna sur les targets (Emission CO2 – Energy Use)
- **Harmonisation:**
  - Colonnes
  - Noms – Mise en forme
- **Valeurs Abérrantes: 200k**
  - Electricité et Emission CO2 < 0 (Bullitt Center)
  - Université de Washington
- **Assembler Dataset 2015 et 2016: Concat**
- **Valeurs Redondantes: sur l'ID en gardant le plus recent**

FEATURE ENGINEERING

_INDICATEURS

SIMPLIFICATION

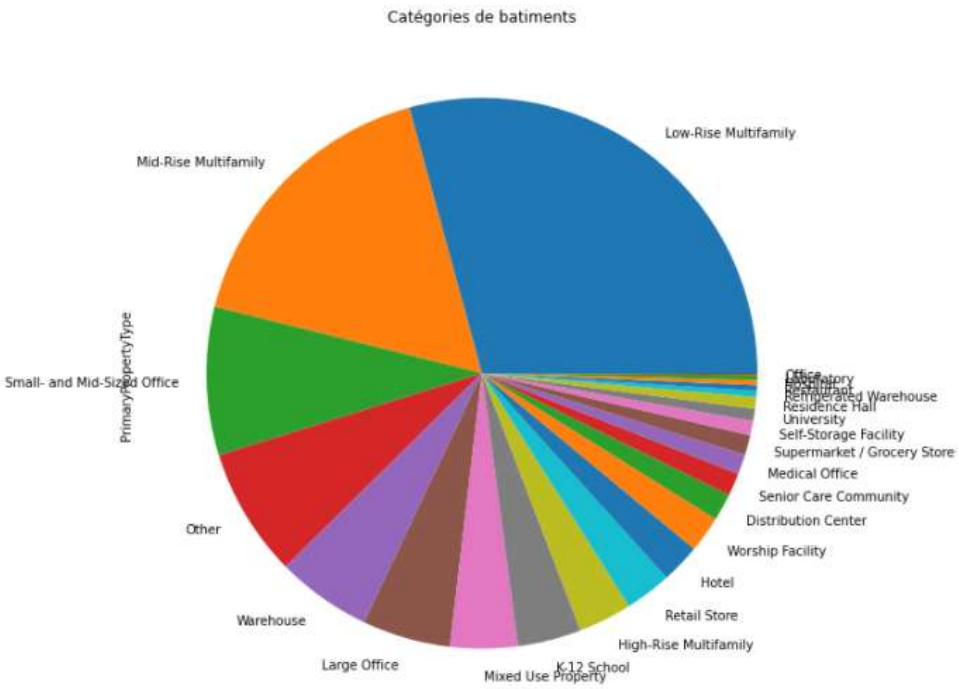
- **BuildingType: Reunion de sous-groupes**
  - NonResidential et NonResidential COS
  - NonResidential WA et Campus
  - SPS District K12 et Campus
- **PrimaryPropertyType : Reunion de sous-groupes**

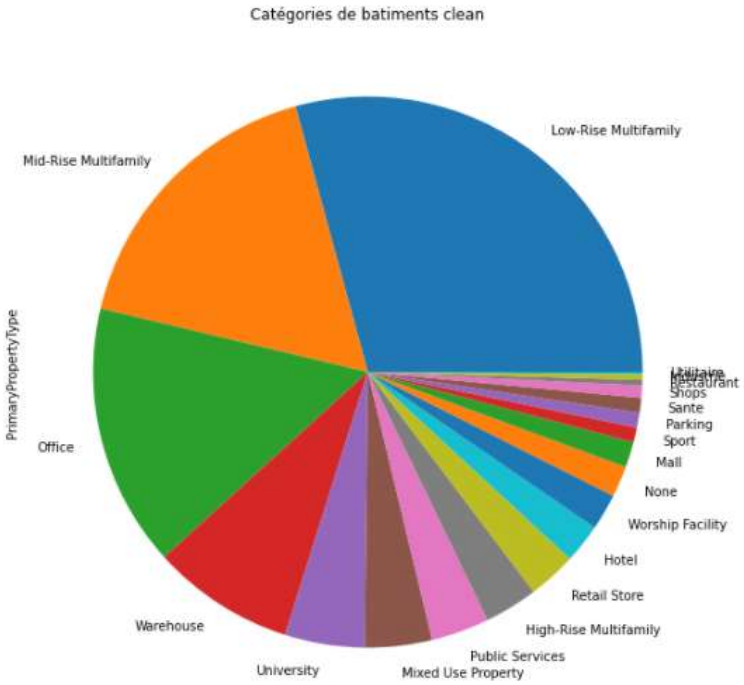
CRÉATION

- **Age: Année relevé – Année de construction**
- **Energy Rate: Consommation energetique / Superficie**
- **Booleen Parking**
- **Passage au log:**
  - Targets
  - Superficie

FEATURE ENGINEERING

_INDICATEURS

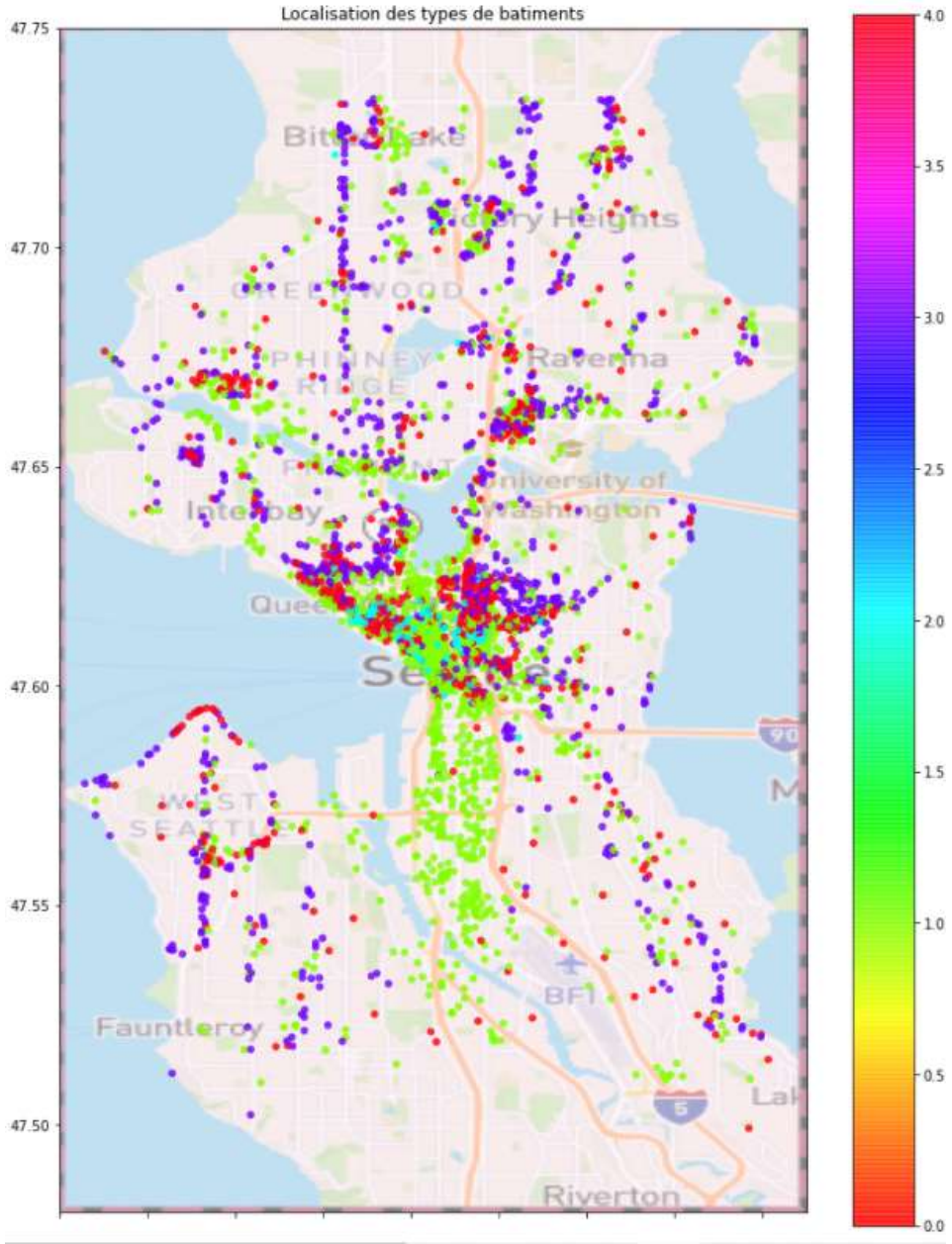




Seattle

VISUALISATION

_GEOGRAPHIQUE



MultiFamily MR

MultiFamily LR

MultiFamily HR

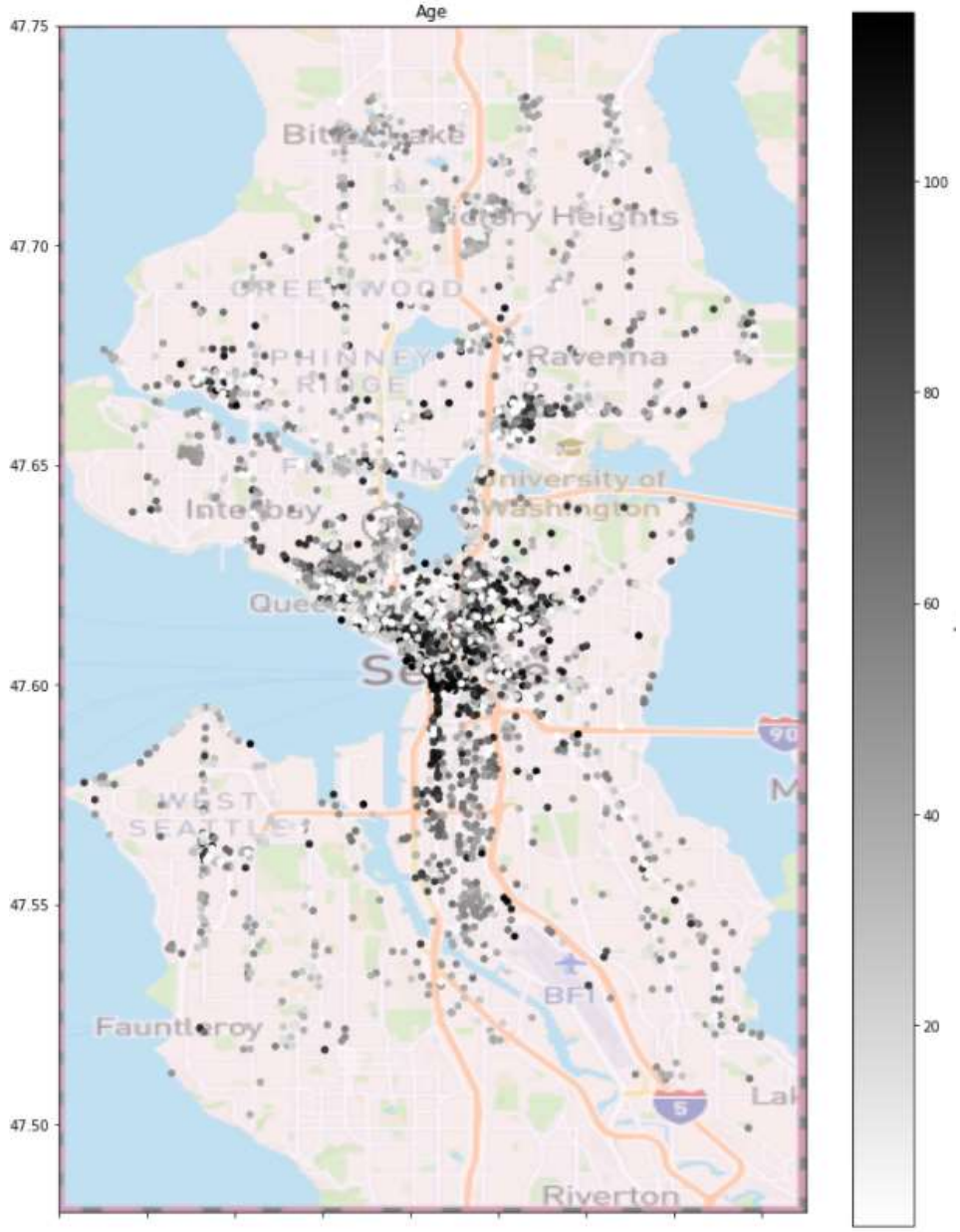
Non Residentiel

Campus

Seattle

VISUALISATION

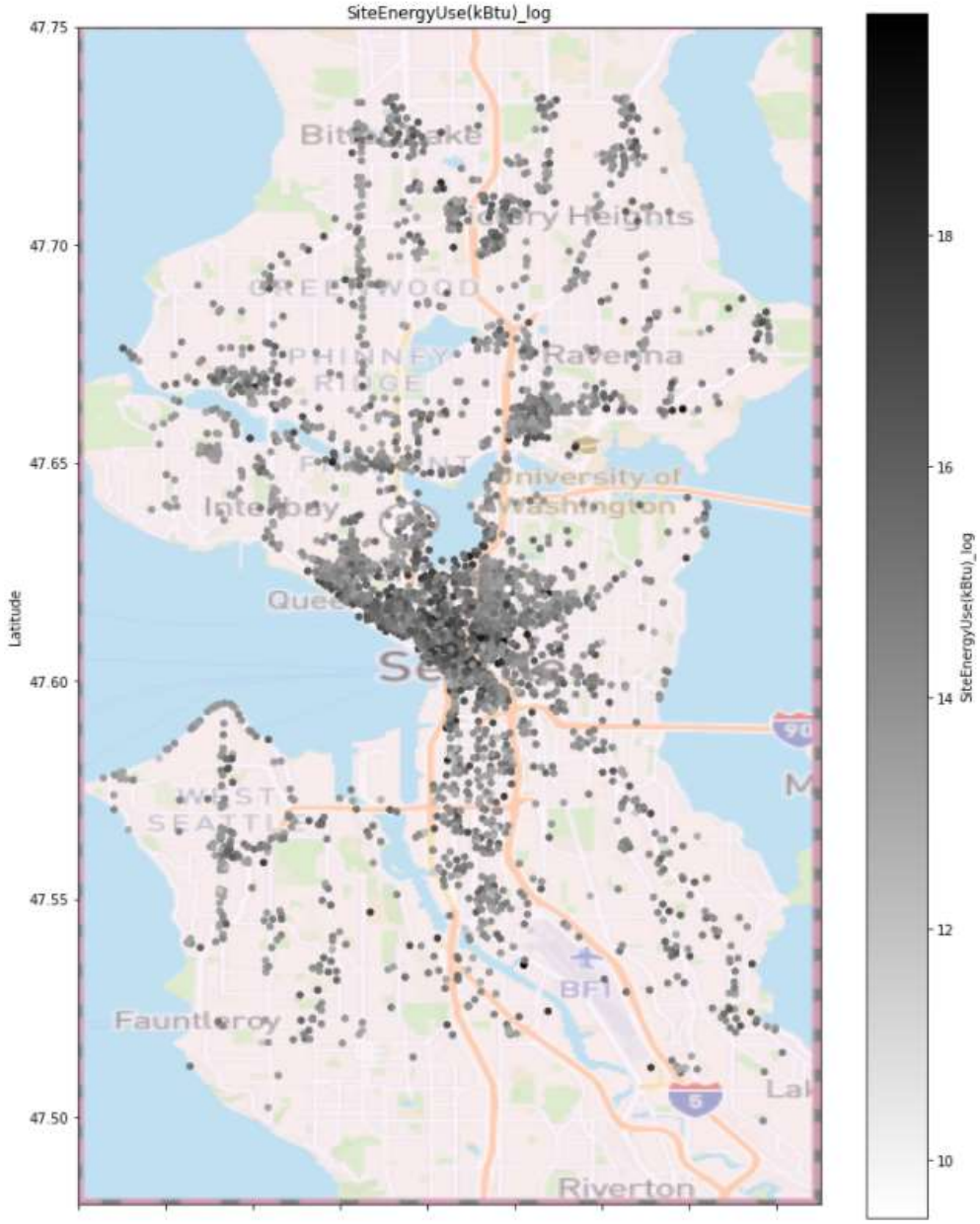
_GEOGRAPHIQUE



Seattle

VISUALISATION

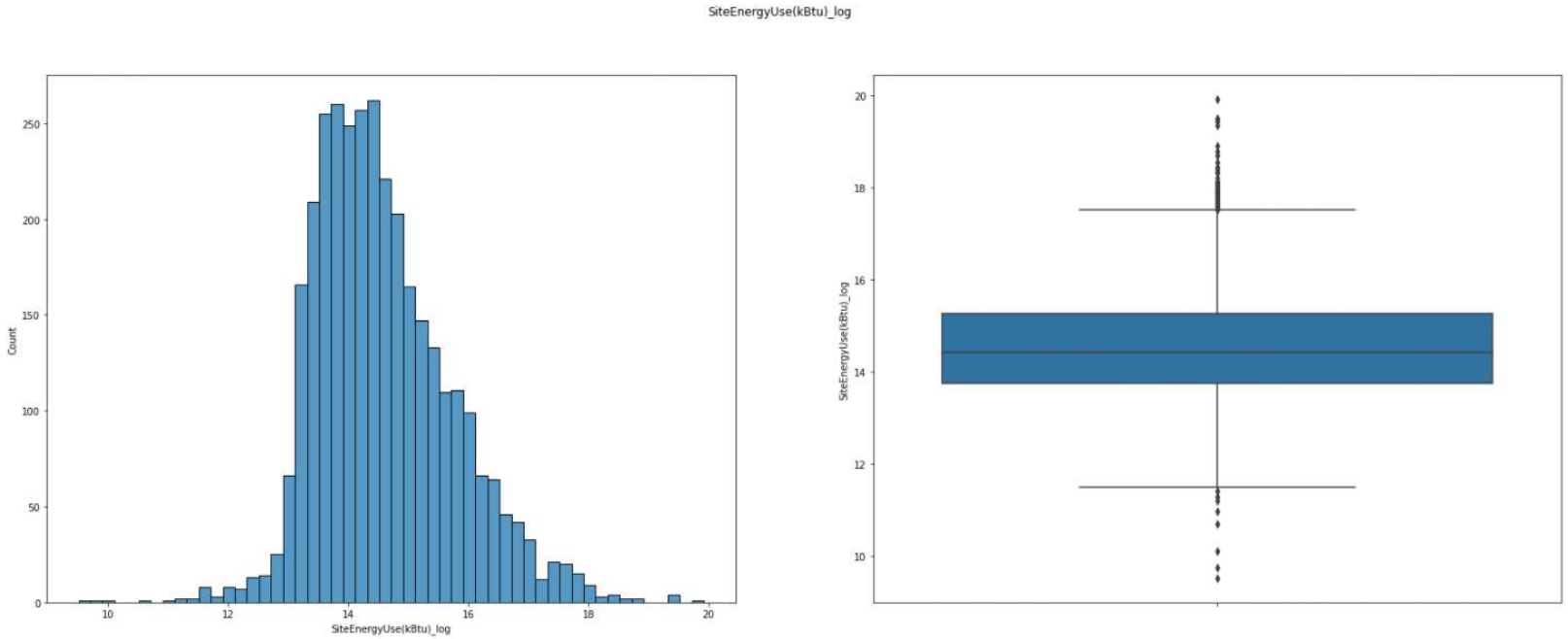
_GEOGRAPHIQUE



Seattle

ANALYSE
_UNIVARIÉ

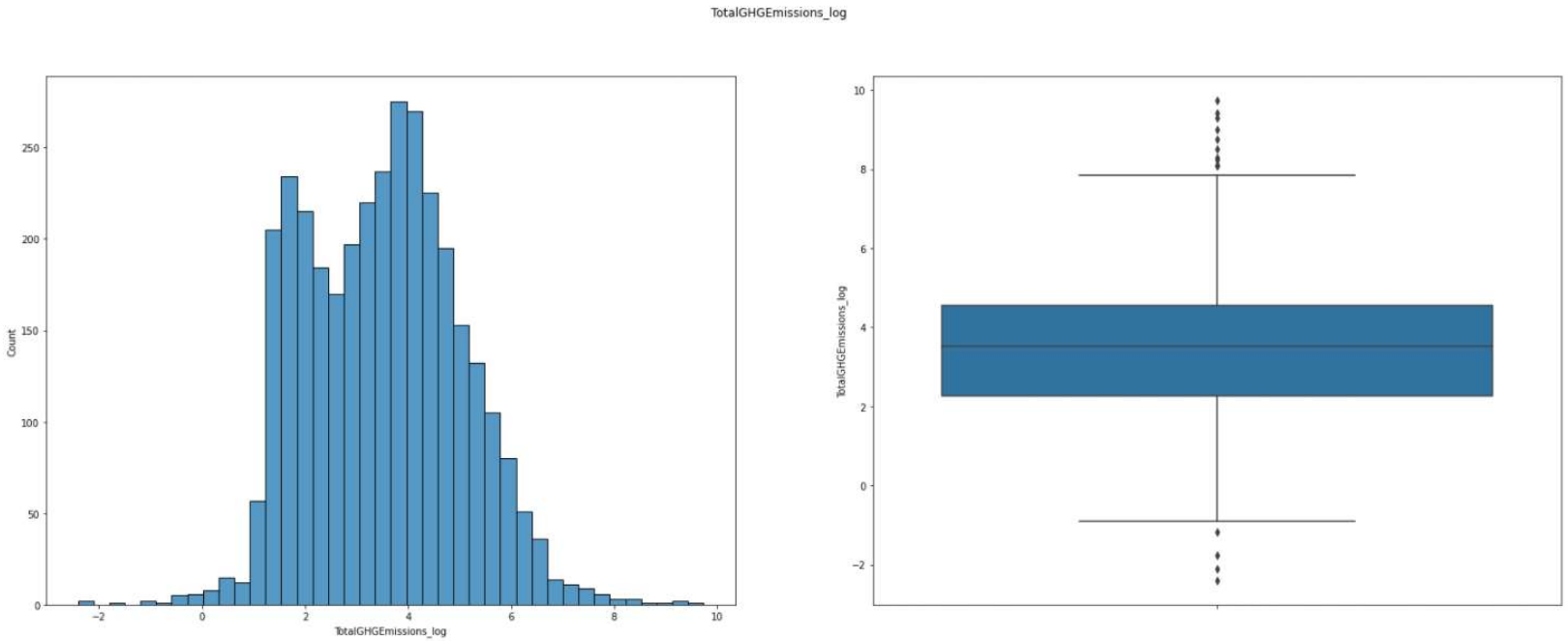
TARGET
Consommation energetique



Seattle

ANALYSE
_UNIVARIÉ

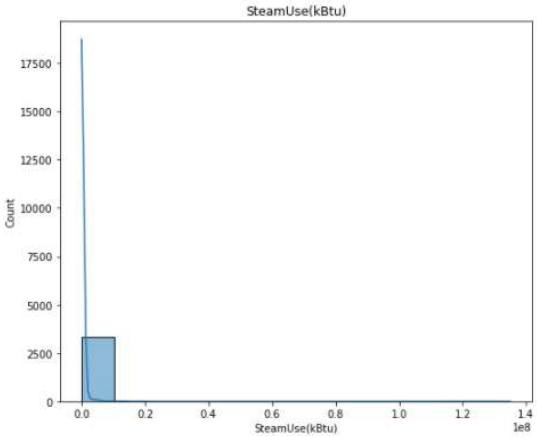
TARGET
Emissions CO2

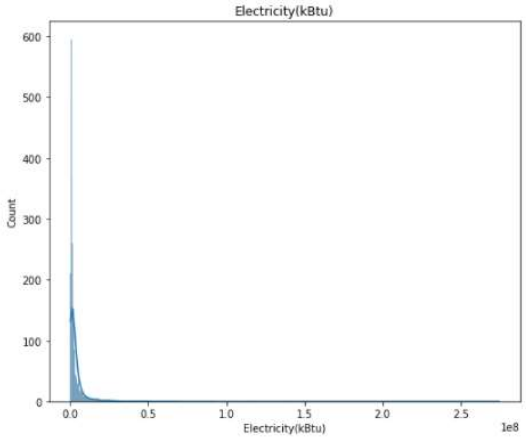


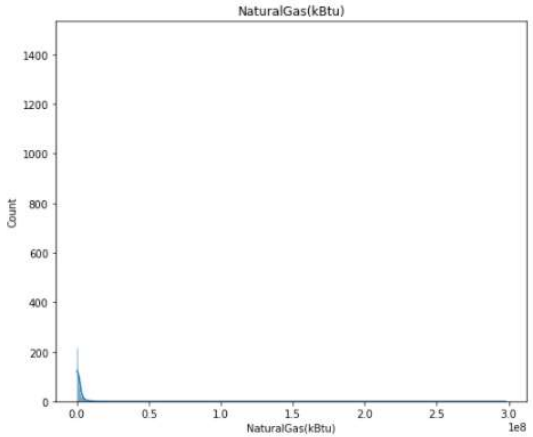
Seattle

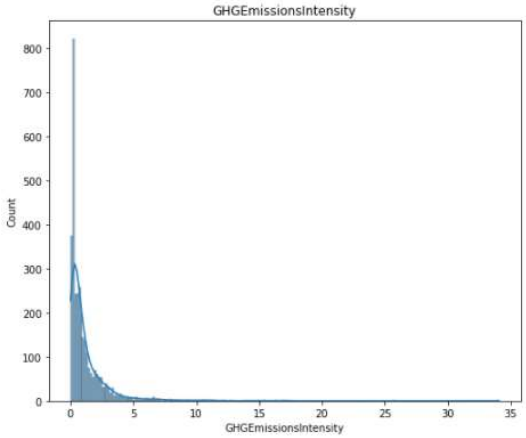
ANALYSE
_UNIVARIÉ

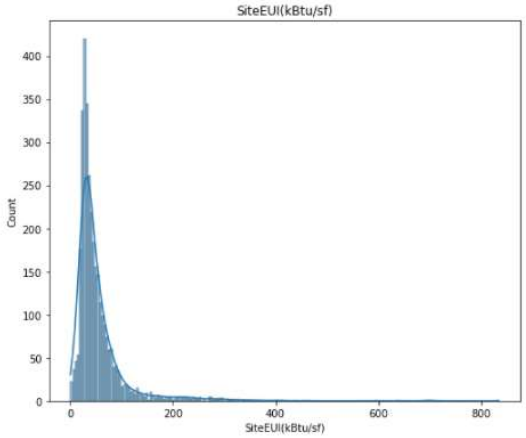
VARIABLES
Quantitatives

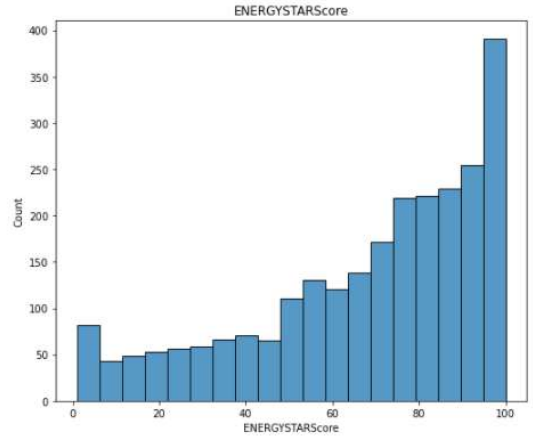








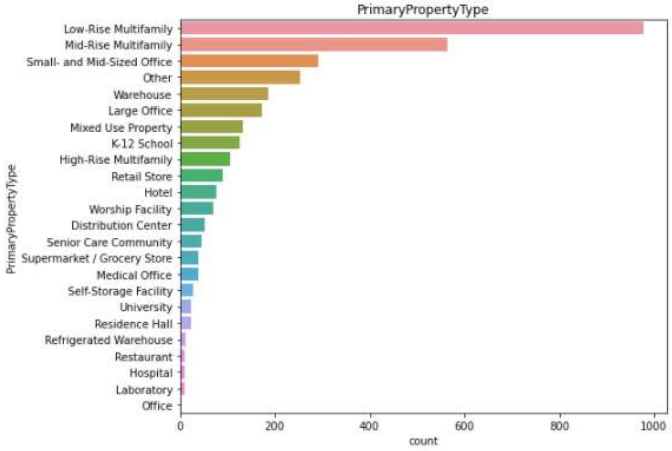


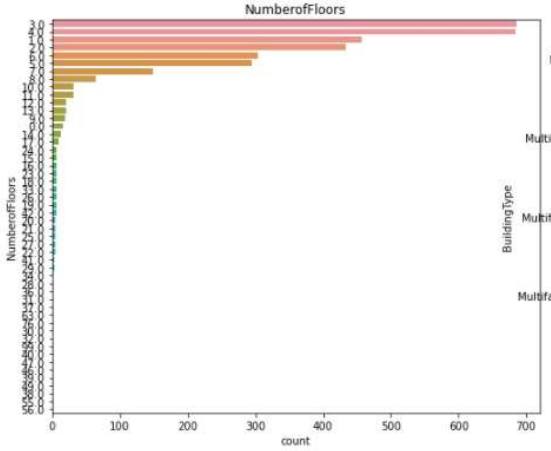


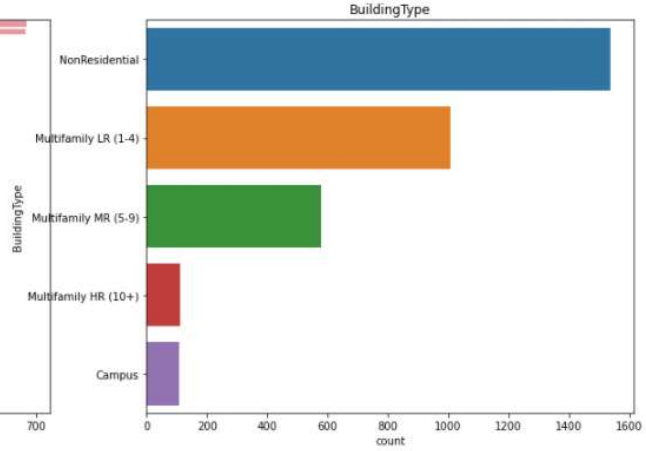
Seattle

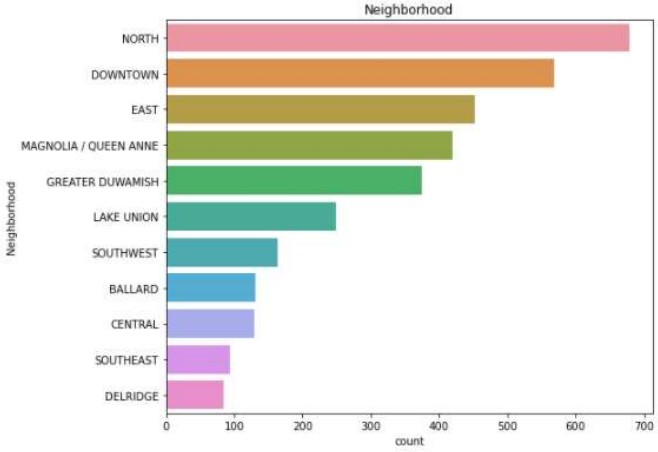
ANALYSE
_UNIVARIÉ

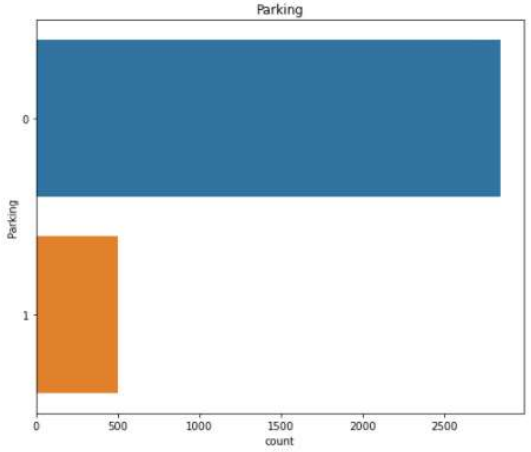
VARIABLES
Qualitatives

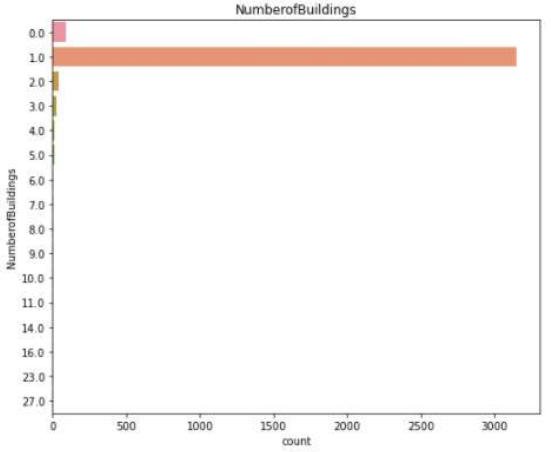






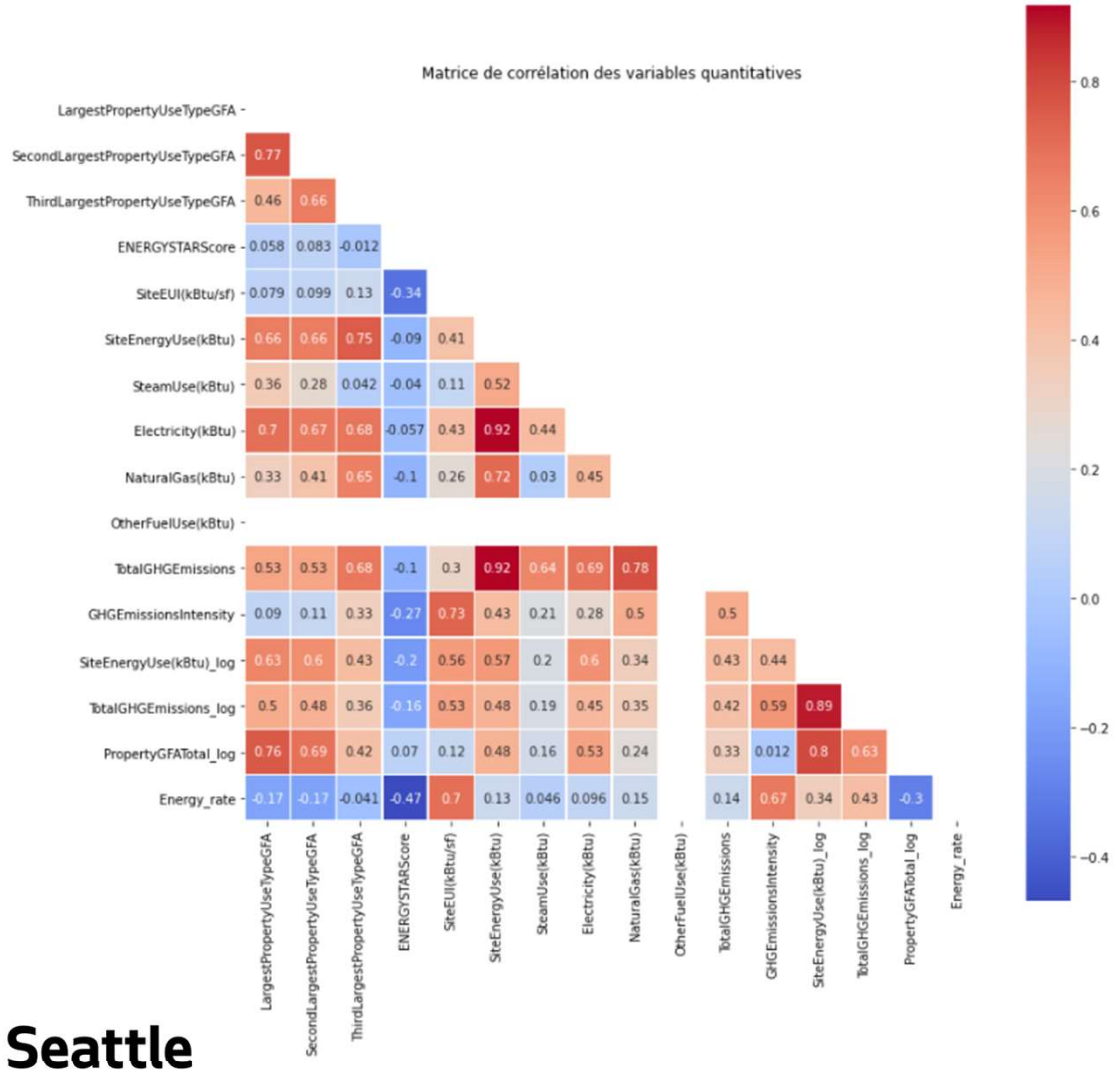






Seattle

ANALYSE
_BIVARIÉ



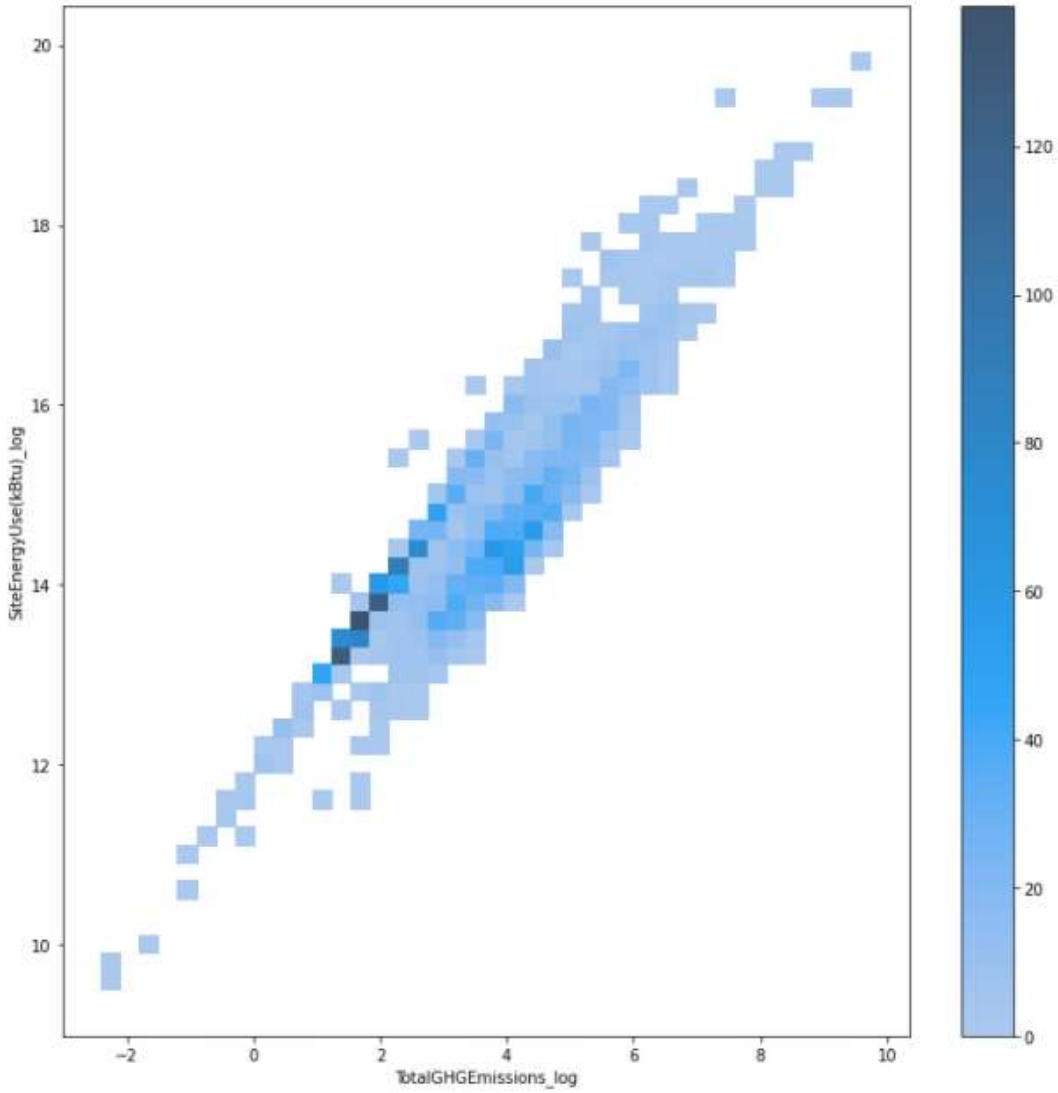
RELATION

Matrice de Corrélation

- **Consommation:**
  - Emissions CO2
- **Superficie:**
  - Consommation
  - Emissions CO2
- **Consommation:**
  - Electricité
  - Vapeur
  - Gaz

Seattle

ANALYSE
_BIVARIÉ



RELATION
Targets

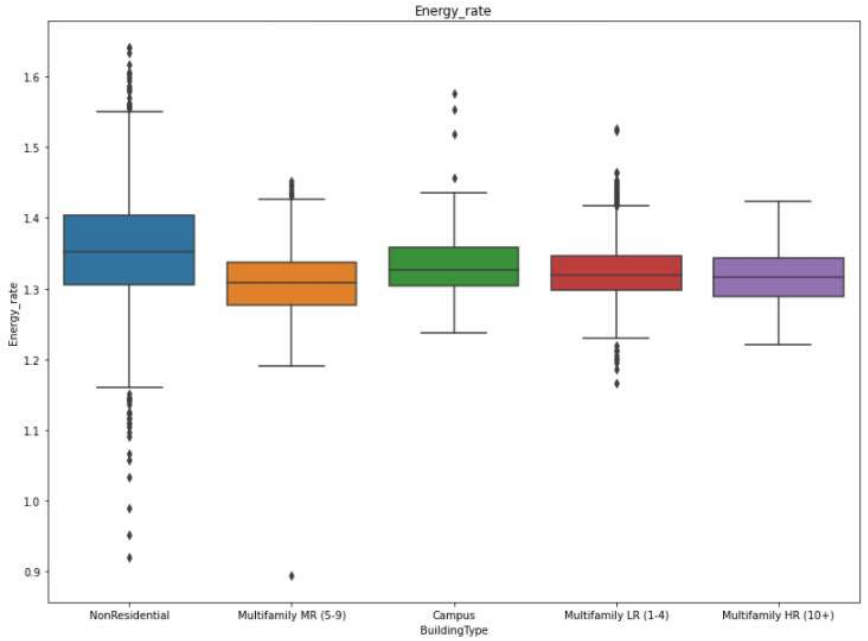
Seattle

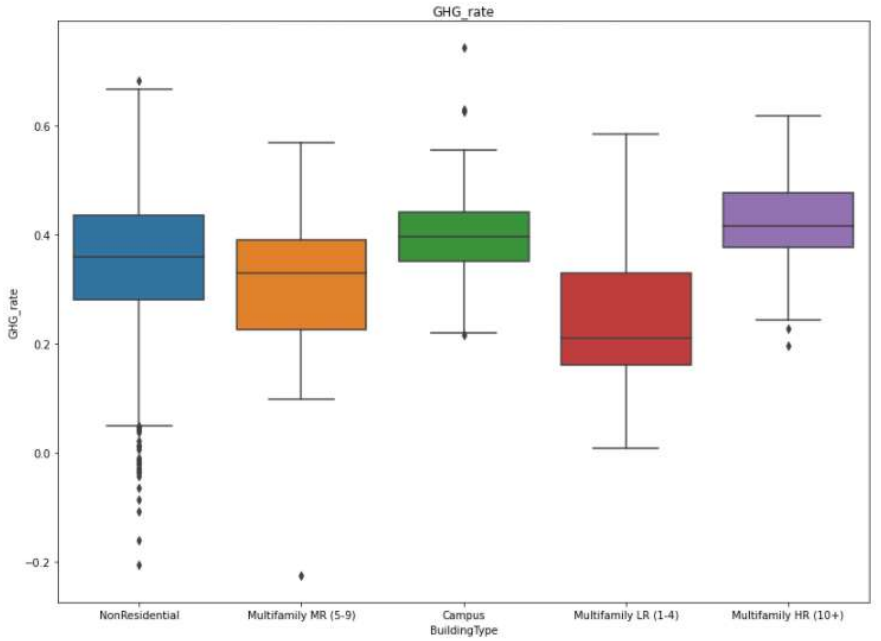
ANALYSE
_BIVARIÉ

RELATION

Anova

RELATION Energy_rate et GHG_rate / TYPE BATIMENTS





	Source	SS	DF	MS	F	p-unc	np2
0	BuildingType	1.041553	4	0.260388	60.655009	1.581702e-49	0.06774
1	Within	14.334128	3339	0.004293	NaN	NaN	NaN

	Source	SS	DF	MS	F	p-unc	np2
0	BuildingType	8.296306	4	2.074076	163.833547	3.201679e-128	0.164066
1	Within	42.270593	3339	0.012660	NaN	NaN	NaN

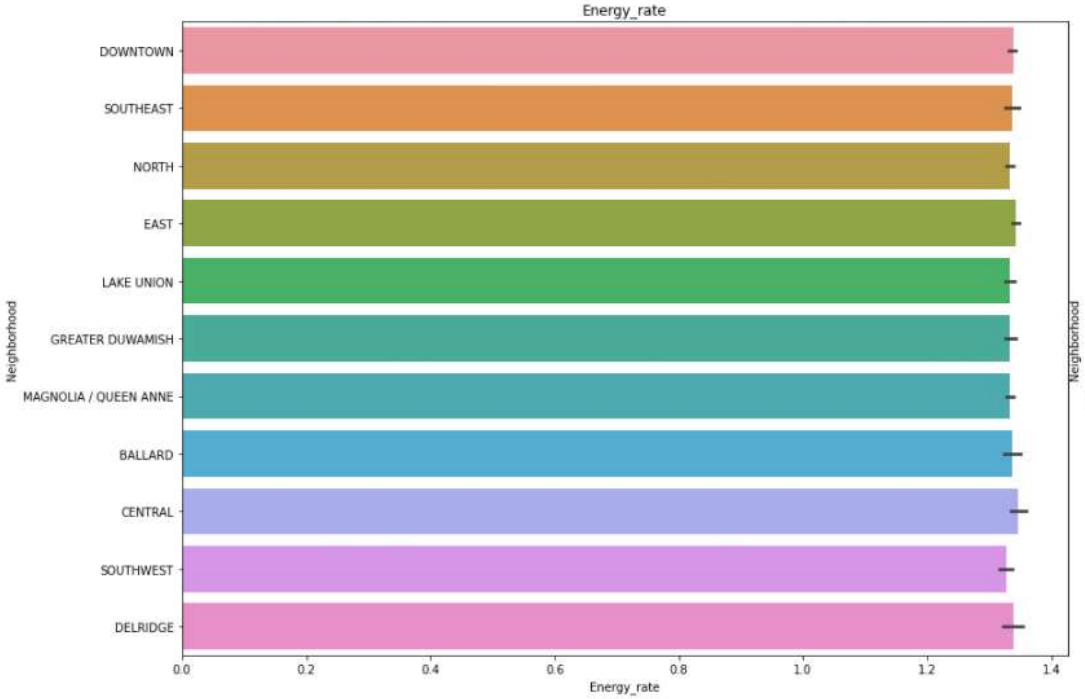
Seattle

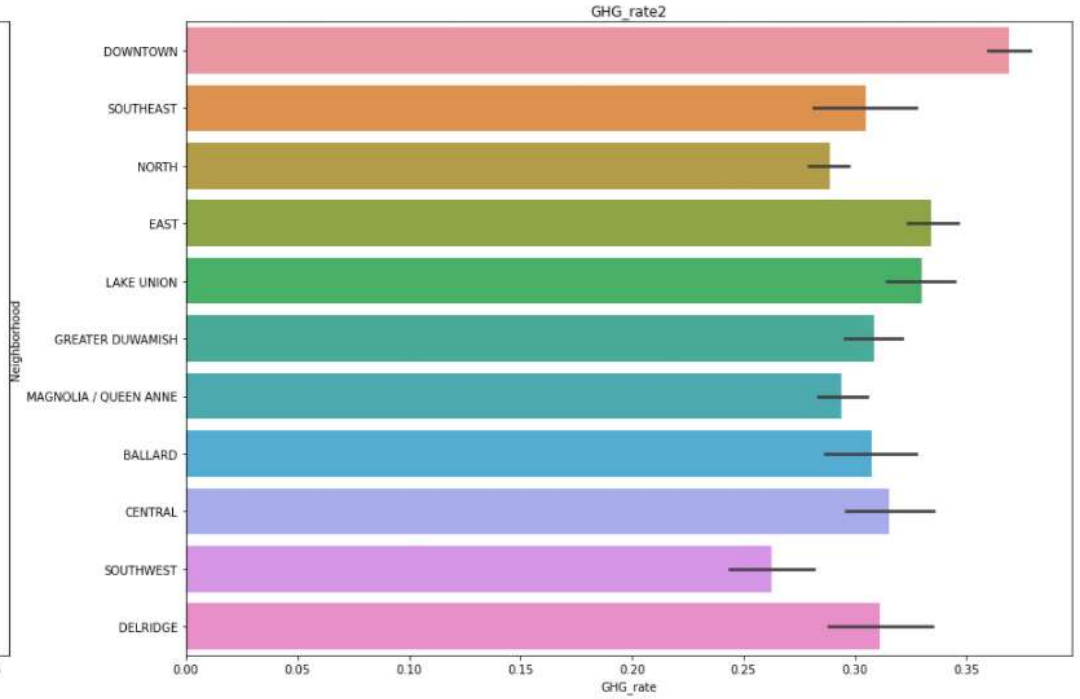
ANALYSE
_BIVARIÉ

RELATION

Anova

RELATION Energy_rate et GHG_rate / NEIGHBORHOOD





	Source	SS	DF	MS	F	p-unc	np2
0	Neighborhood	0.063088	10	0.006309	1.373203	0.186138	0.004103
1	Within	15.312593	3333	0.004594	NaN	NaN	NaN

	Source	SS	DF	MS	F	p-unc	np2
0	Neighborhood	3.016399	10	0.301640	21.143121	1.301375e-38	0.059652
1	Within	47.550499	3333	0.014267	NaN	NaN	NaN

Seattle

ANALYSE
_ORDRES GRANDEUR

RELATION
Anova

PrimaryPropertyType	
Hospital	5052.963000
Laboratory	849.424000
University	500.331250
Hotel	410.882727
Mixed Use Property	272.934848
Senior Care Community	269.472222
Medical Office	261.954872
Supermarket / Grocery Store	228.301282
High-Rise Multifamily	220.010381
Large Office	206.412312
Other	193.326245
Restaurant	154.452727
Residence Hall	114.675652
K-12 School	92.892240
Retail Store	88.940330
Mid-Rise Multifamily	64.325426
Distribution Center	50.225660
Warehouse	42.494492
Worship Facility	42.180704
Small- and Mid-Sized Office	40.577766
Refrigerated Warehouse	37.614167
Low-Rise Multifamily	26.591534
Self-Storage Facility	24.098929
Office	10.743333

Name: TotalGHGEmissions, dtype: float64

Emissions Consommation / Type

PrimaryPropertyType	
Hospital	1.415470e+08
Laboratory	2.629313e+07
University	1.885577e+07
Large Office	1.761539e+07
Hotel	1.255155e+07
Medical Office	1.197166e+07
Mixed Use Property	1.083570e+07
Supermarket / Grocery Store	1.003563e+07
Other	9.020341e+06
Senior Care Community	8.759823e+06
High-Rise Multifamily	8.665963e+06
Restaurant	4.696249e+06
Retail Store	4.694639e+06
Residence Hall	4.265060e+06
Refrigerated Warehouse	3.719428e+06
K-12 School	3.076938e+06
Mid-Rise Multifamily	2.830739e+06
Small- and Mid-Sized Office	2.632330e+06
Distribution Center	2.405360e+06
Warehouse	1.961577e+06
Low-Rise Multifamily	1.349221e+06
Worship Facility	1.171907e+06
Self-Storage Facility	9.321726e+05
Office	5.800745e+05

Name: SiteEnergyUse(kBtu), dtype: float64

Emissions CO2 / Type

MACHINE LEARNING
_TRAITEMENT

3 ENSEMBLES

- **Ensemble 1: 6 Features de base (Superficie_log, Vapeur, Gaz, Electricité, Age, Parking**
- **Ensemble 2: 8 Features (idem + Nombre d'étages et nombre de batiments) et la variables qualitatives PrimaryPropertyType**
- **Ensemble 3: Modèle 2 sans PrimaryPropertyType**

TARGETS

- **Consommation energetique: Au Log**
- **Emission de CO2: Avec et sans EnergyStar Score**

MACHINE LEARNING
_METHODOLOGIE

TRANSFORMEURS

- Quantitatives: StandardScaler et PolynomialFeature
- Catégorielles: OrdinalEncoder

ESTIMATEURS

- Base: Dummy
- Linéaire: LinearRegressor, SGDRegressor, Ridge, Lasso
- Ensembliste: RandomForestRegressor

HYPER-PARAMETRES

- GridSearchCV
- CrossValidation

PIPELINE

- Preprocessor
- Param_grid

MACHINE LEARNING
_METHODOLOGIE

EVALUATION

- MAE: Mean Absolute Error
- MSE: Mean Squared Error
- MSE: Root Mean Squared Error
- MAPE: Mean Squared Error
- R2: Coefficient de détermination

PERFORMANCES

- Learning Curve
- Erreur

MACHINE LEARNING

_RESULTATS CONSOMMATION D'ENERGIE

Ensemble 1

	MAE	MSE	RMSE	MAPE	R²	DUREE
Dummy Regressor	0.928967	1.488114	0.062922	1.219883	-0.023667	0.000000
Linear Regression	0.512108	0.544238	0.035733	0.737725	0.625621	0.000998
SGDRegressor	0.512812	0.548677	0.035735	0.740727	0.622568	0.003962
Ridge	0.512136	0.544229	0.035735	0.737719	0.625628	0.000969
Lasso	0.936244	1.453823	0.064187	1.205746	-0.000079	0.000997
RandomForestRegressor	0.352690	0.286635	0.026025	0.535383	0.802825	0.524627

Ensemble 2

	MAE	MSE	RMSE	MAPE	R²	DUREE
Dummy Regressor	0.893152	1.364758	0.059346	1.168229	-0.039475	0.000973
Linear Regression	0.461018	0.392803	0.031288	0.626740	0.700819	0.000997
SGDRegressor	0.477541	0.401253	0.032841	0.633446	0.694383	0.003989
Ridge	0.461035	0.392803	0.031289	0.626740	0.700820	0.000997
Lasso	0.900144	1.317242	0.060496	1.147712	-0.003284	0.000997
RandomForestRegressor	0.225022	0.088238	0.015710	0.297048	0.932793	0.640307

Ensemble 3

	MAE	MSE	RMSE	MAPE	R²	DUREE
Dummy Regressor	0.934381	1.516331	0.061695	1.231394	-0.053944	0.000000
Linear Regression	0.492229	0.439495	0.033258	0.662944	0.694523	0.000000
SGDRegressor	0.509862	0.477834	0.034356	0.691256	0.667875	0.003989
Ridge	0.492266	0.439486	0.033261	0.662938	0.694530	0.000000
Lasso	0.936984	1.454460	0.062529	1.206010	-0.010940	0.000000
RandomForestRegressor	0.328106	0.185994	0.023116	0.431270	0.870723	0.587610

MACHINE LEARNING
_PARAMETRES IMPORTANCE ET PERFORMANCES

PARAMETRES

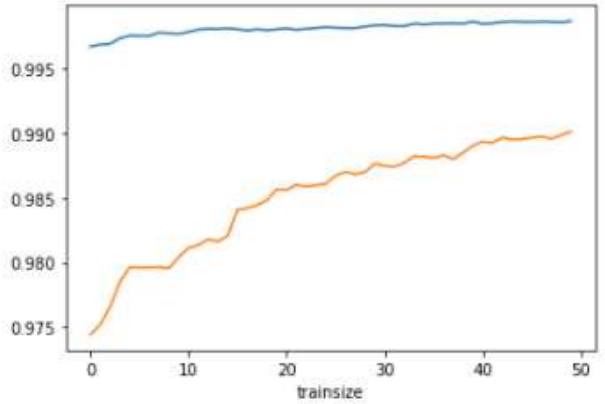
- N_estimators: 50
- Min_samples_leaf: 1
- Max_features: Auto

IMPORTANCE

PrimaryPropertyType SteamUse(kBtu)

```
array([1.12371066e-03, 5.77143927e-03, 8.46019892e-01, 1.44375280e-01,
       8.73154211e-04, 3.10458147e-04, 5.07220225e-04, 4.82931635e-04,
       5.35913188e-04])
```

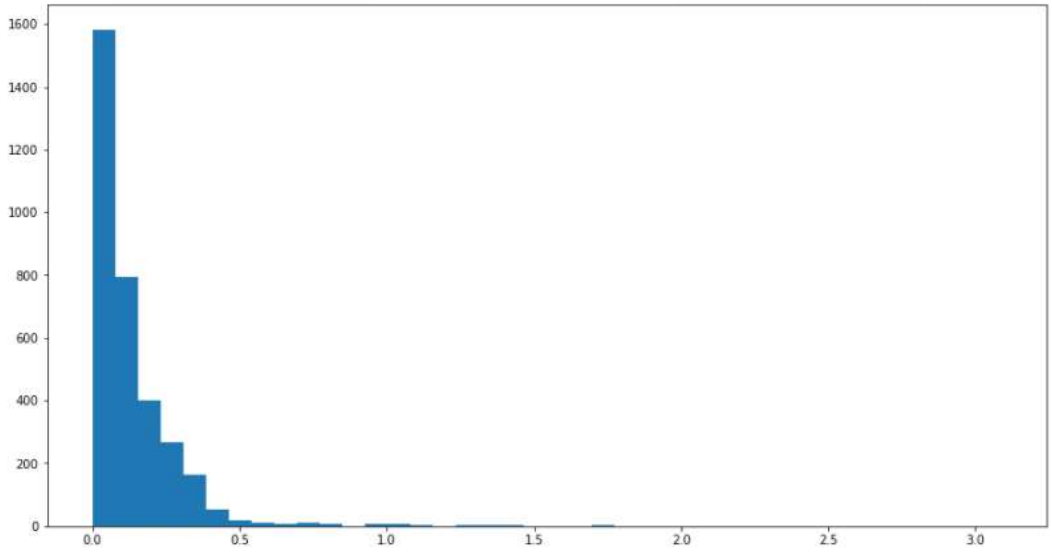
LEARNING CURVE



MACHINE LEARNING

_PARAMETRES IMPORTANCE ET PERFORMANCES

ERREURS



MACHINE LEARNING

_RESULTATS EMISSIONS CO2

Ensemble 1

	MAE	MSE	RMSE	MAPE	R²	DUREE
Dummy Regressor	1.188372	2.023436	0.699583	1.422475	-0.010366	0.000000
Linear Regression	0.838829	1.026882	0.474704	1.013352	0.487245	0.000998
SGDRegressor	0.839063	1.026091	0.474272	1.012962	0.487640	0.001994
Ridge	0.838828	1.026751	0.474779	1.013287	0.487310	0.000000
Lasso	1.186346	2.015722	0.693107	1.419761	-0.006515	0.000997
RandomForestRegressor	0.584263	0.912952	0.358739	0.955485	0.544134	0.491191

Ensemble 2

	MAE	MSE	RMSE	MAPE	R²	DUREE
Dummy Regressor	1.180964	2.034793	0.610180	1.426462	-0.009585	0.000969
Linear Regression	0.797839	0.924072	0.407038	0.961287	0.541511	0.000997
SGDRegressor	0.841509	1.025159	0.431619	1.012501	0.491356	0.002018
Ridge	0.797862	0.923961	0.407100	0.961229	0.541566	0.000996
Lasso	1.177488	2.025277	0.602280	1.423122	-0.004864	0.000000
RandomForestRegressor	1.509241	3.805450	0.639751	1.950756	-0.888117	0.460796

Ensemble 3

	MAE	MSE	RMSE	MAPE	R²	DUREE
Dummy Regressor	1.215044	2.142622	0.666864	1.463770	-0.008312	0.000000
Linear Regression	0.858142	1.046983	0.459631	1.023222	0.507293	0.000997
SGDRegressor	0.894024	1.133318	0.482982	1.064574	0.466664	0.001994
Ridge	0.858180	1.046890	0.459719	1.023176	0.507337	0.000996
Lasso	1.215079	2.143012	0.667151	1.463903	-0.008495	0.000997
RandomForestRegressor	0.334789	0.239305	0.132968	0.489188	0.887384	0.407123

Seattle

MACHINE LEARNING
_PARAMETRES ET IMPORTANCE

PARAMETRES

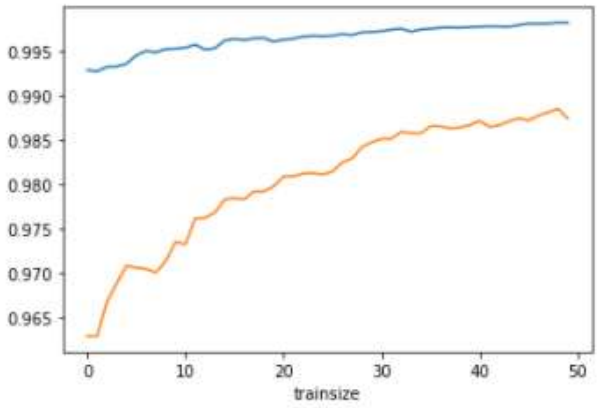
- N_estimators: 50
- Min_samples_leaf: 1
- Max_features: Auto

IMPORTANCE

Electricity(kBtu) NaturalGas(kBtu)

```
array([1.39503843e-03, 3.17391291e-02, 2.40115416e-01, 7.24726517e-01,
       1.50462574e-03, 5.19273456e-04])
```

LEARNING CURVE



Seattle

MACHINE LEARNING

_RESULTATS EMISSIONS CO2 AVEC ENERGY STAR SCORE

Ensemble 2

	MAE	MSE	RMSE	MAPE	R^2	DUREE
Dummy Regressor	1.294789	2.367264	0.661170	1.538592	-0.003716	0.001004
Linear Regression	0.793695	0.925523	0.395275	0.962041	0.607580	0.001995
SGDRegressor	0.785645	0.920958	0.378614	0.959666	0.609515	0.005984
Ridge	0.793867	0.925142	0.395466	0.961843	0.607741	0.001967
Lasso	1.294260	2.364533	0.658060	1.537704	-0.002558	0.000997
RandomForestRegressor	0.512676	0.403805	0.263383	0.635456	0.828787	0.560067

CONCLUSION
_PREVISIONS SUR LE CONSOMMATION

- **Dataset: Grand nombre de batiments**
- **Modèle: Bonnes performances (93% de variance expliqués)**
- **Learning Curve: Données suffisantes pour la prédiction**
- **Prédiction: Fiable**

CONCLUSION

_PREVISIONS SUR L'EMISSION DE CO2

- **Dataset: Grand nombre de NaN si on intègre l'Energy Star Score**
- **Modèle: Bonnes performances (88% de variance expliqués)**
- **Prédiction: Moins fiable, surtout si on intègre l'Energy Star Score qui diminue la precision du modèle**
- **Energy Star: Obsolète ?**