Analyse

Seattle

Etude de la consommation énergétique

Sofiane Mouhab

15 mars 2021

1 - Généralités

1.1 - Problématique

Des relevés minutieux ont été effectués par vos agents en 2015 et en 2016. Cependant, ces relevés sont coûteux à obtenir, et à partir de ceux déjà réalisés, vous voulez tenter de prédire les émissions de CO2 et la consommation totale d'énergie de bâtiments pour lesquels elles n'ont pas encore été mesurées. Votre prédiction se basera sur les données déclaratives du permis d'exploitation commerciale (taille et usage des bâtiments, mention de travaux récents, date de construction..) Vous cherchez également à évaluer l'intérêt de l'"ENERGY STAR Score" pour la prédiction d'émissions, qui est fastidieux à calculer avec l'approche utilisée actuellement par votre équipe.

1.2 - Objectif

N°1 :

prédire les émissions de CO2 et la consommation totale d’énergie

N°2 :

évaluer l’intérêt de l’"ENERGY STAR Score" pour la prédiction d’émissions

N°3 :

Réaliser une courte analyse exploratoire.

N°4 :

Tester différents modèles de prédiction afin de répondre au mieux à la problématique.

1.3 - Condition de mise en oeuvre

Pour pouvoir sereinement réaliser ses quatres objectifs, il nous faut donc diverses informations qui pourrait se trouver dans notre base de données.

À nous donc, d'examiner celle-ci, de déterminer à quel point les informations sont viables, ou perfectible.

Il y a donc 3 grandes interrogations :

- A-t-on assez de données ?
- Peut-on faire des prévisions cohérente ?
- Dans quelle mesure l'energy score est important
-

Passons de suite à ce travail, en commençant par rapidement prendre connaissance des données en présence...

2 - Les données

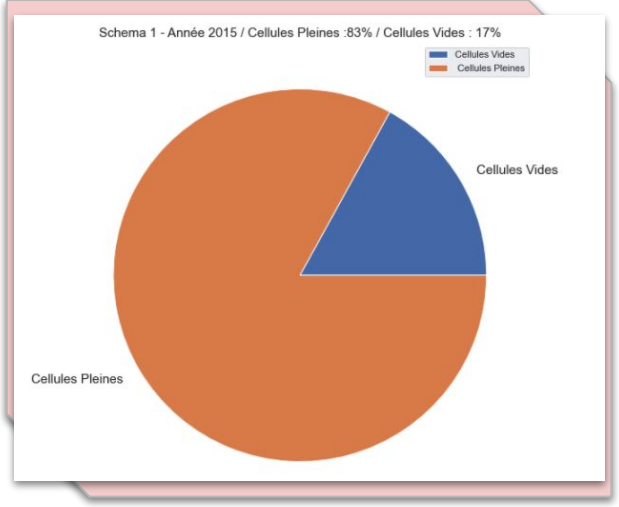
2.1 - Description

Dans ce fichier volumineux nous comptons 2 Fichiers , les relevés pour 2015 et 2016, voyons ce qu'ils contiennent :

	Lignes	Colonnes	Information Lignes	Information Colonnes
Année 2015	3340	47	Chaque lignes correspond à un batiment de la Ville de Seatle	Diverses informations telle que : - Usage - Localisation - Consommation de gaz - Consommation d'électricité - Rejet de CO2 ...
Année 2016	3376	46		

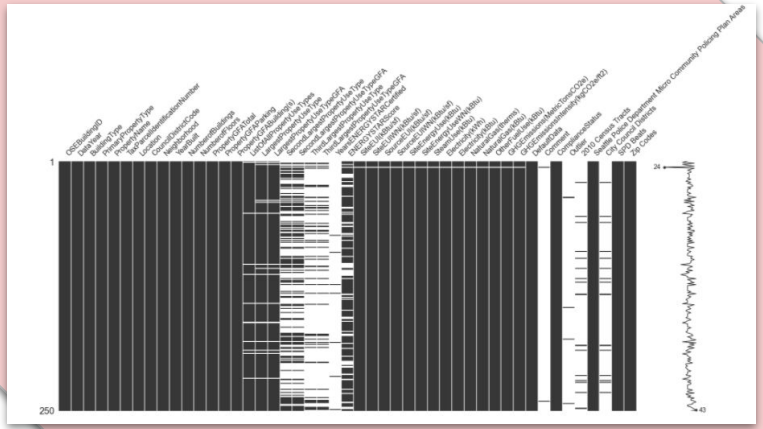
2.2 - L'année 2015

Notre premier fichier concerne le relevé de l'année 2015. On constate 17% de cellule vide. C'est donc un fichier tout à fait viable, la matrice ci-dessous est assez révélatrice de ce constat.



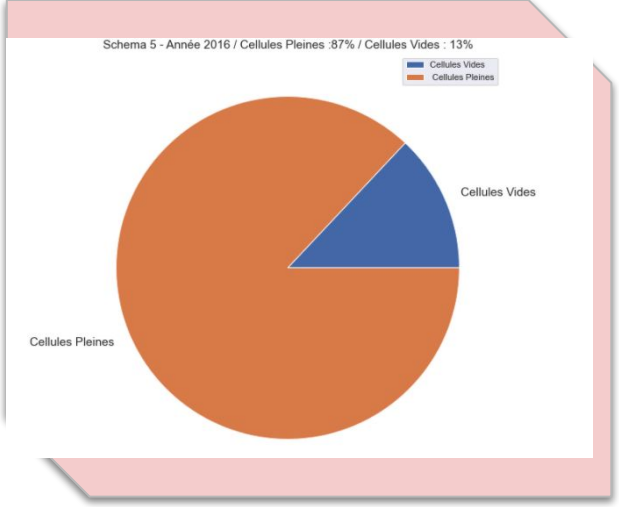
	OSEBuildingID	DataYear	BuildingType	PrimaryPropertyType	PropertyName	TaxParcelIdentificationNumber	Location	CouncilDistrictCode	Neighborhoo
0	1	2015	NonResidential	Hotel	MAYFLOWER PARK HOTEL	659000030	{'latitude': '47.61219025', 'longitude': '-122...	7	DOWNTOWI
1	2	2015	NonResidential	Hotel	PARAMOUNT HOTEL	659000220	{'latitude': '47.61310583', 'longitude': '-122...	7	DOWNTOWI
2	3	2015	NonResidential	Hotel	WESTIN HOTEL	659000475	{'latitude': '47.61334897', 'longitude': '-122...	7	DOWNTOWI

3 rows × 47 columns



2.3 - L'année 2016

Notre premier fichier concerne le relevé de l'année 2016. On constate 13% de cellule vide. C'est donc un fichier trés interessant et trés complet



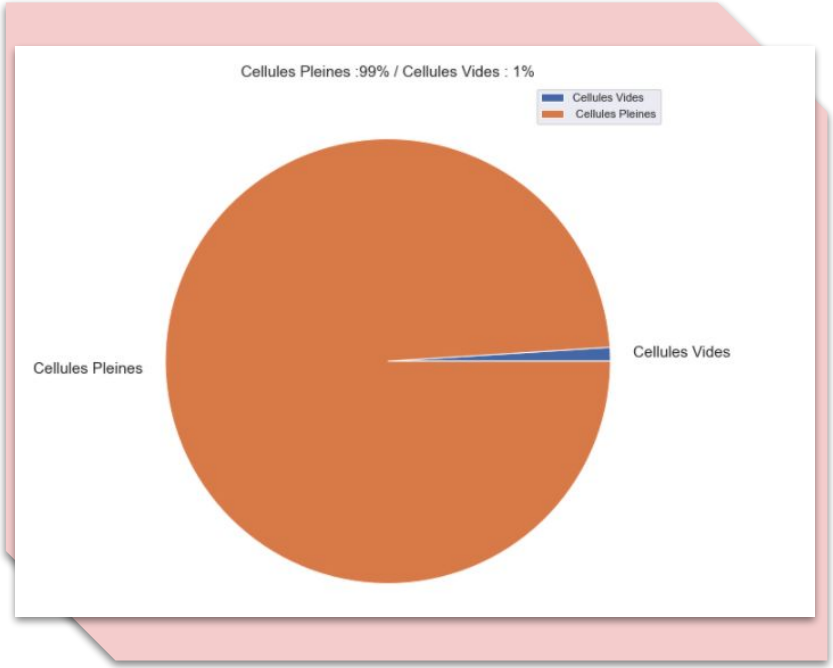
	OSEBuildingID	DataYear	BuildingType	PrimaryPropertyType	PropertyName	Address	City	State	ZipCode	TaxParcelIdentificationNumber	...
0	1	2016	NonResidential	Hotel	Mayflower park hotel	405 Olive way	Seattle	WA	98101.0	0659000030	...
1	2	2016	NonResidential	Hotel	Paramount Hotel	724 Pine street	Seattle	WA	98101.0	0659000220	...
2	3	2016	NonResidential	Hotel	5673-The Westin Seattle	1900 5th Avenue	Seattle	WA	98101.0	0659000475	...

3 rows × 46 columns

2.4 - Filtre : Fusion des 2 databases

Le jeu de donnés semblent tout à fait viable. Après un nettoyage, qui consistait surtout à une mise en forme des variables pour qu'elles correspondent d'une année à l'autre (exemple des adresses….)

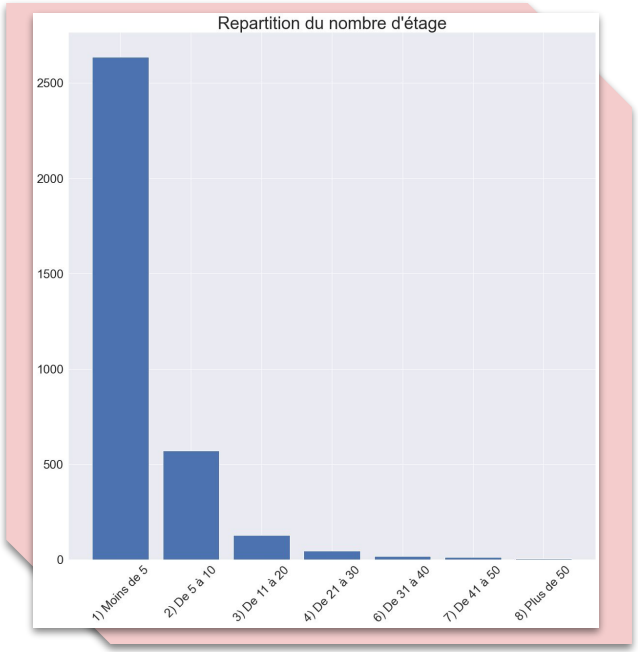
On devine ici une très bonne base de données complète et opérationnelle

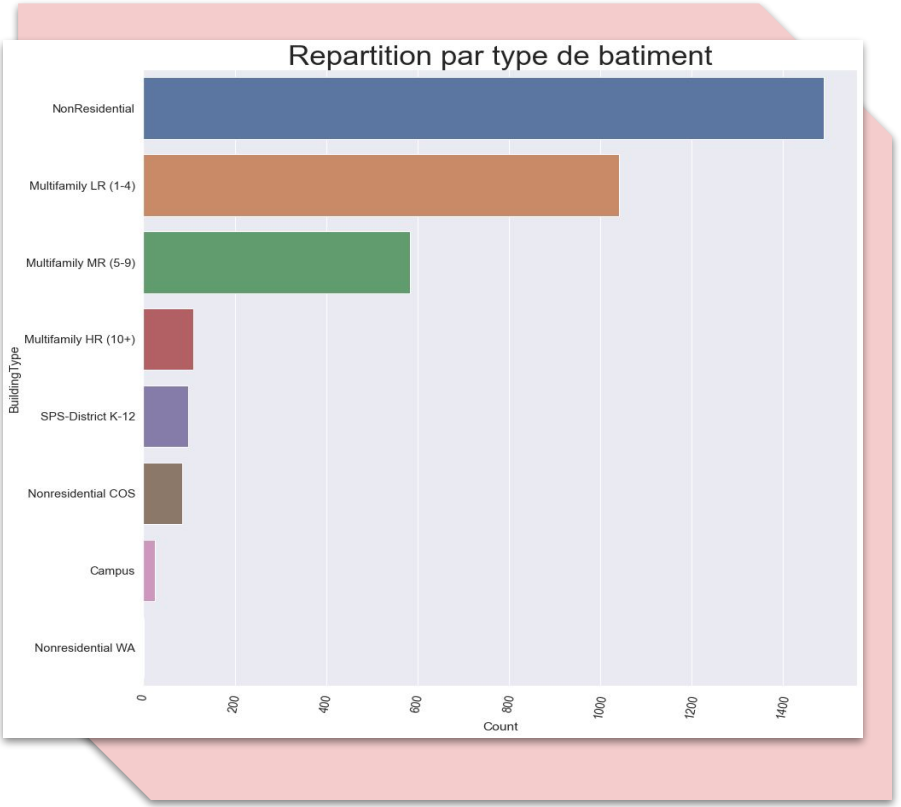


3 - Analyse

3.1 - Analyse des données - Partie 1

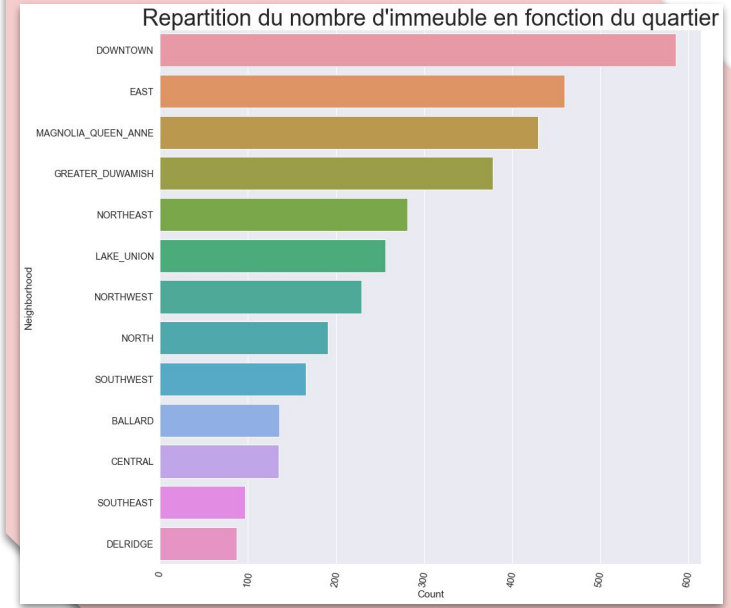
Quelques données pour mieux cerner la ville de Seattle

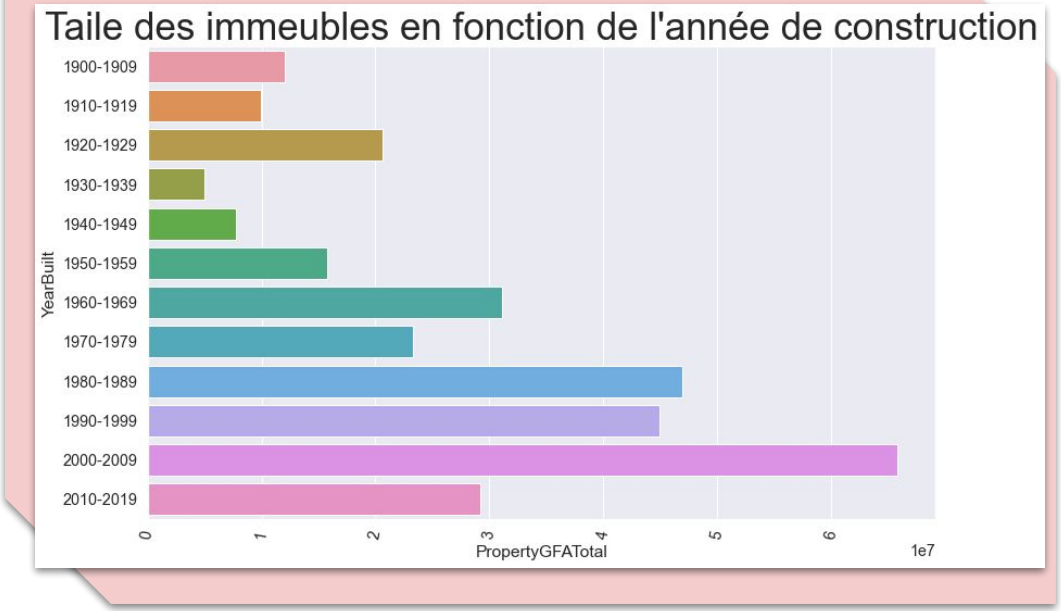




On constate une ville axée sur l'habitation mais avec une part non négligeable de bâtiment non-résidentiel. Une ville qui contrairement à l'imaginaire possèdent une très grande majorité d'immeuble à moins de 5 étages

3.2 - Analyse des données - Partie 2



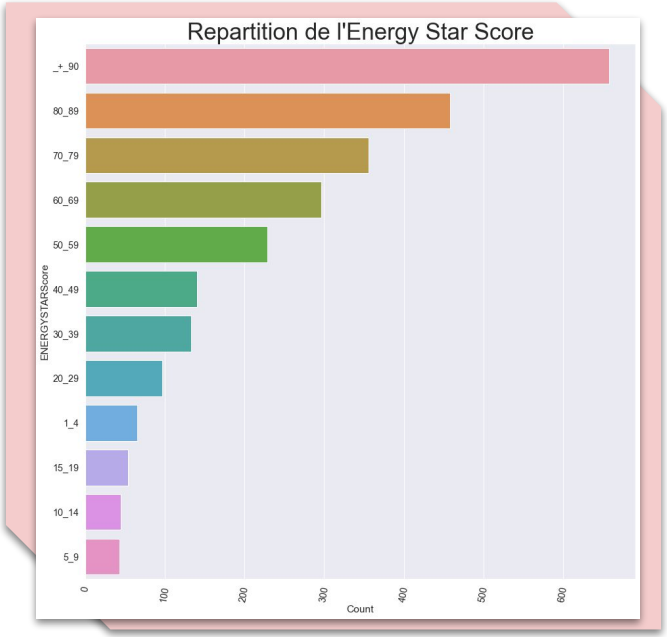


On constate que les Quartiers qui possèdent le plus d'immeuble sont :

- DownTown
- East
- Magnolia Green Anne

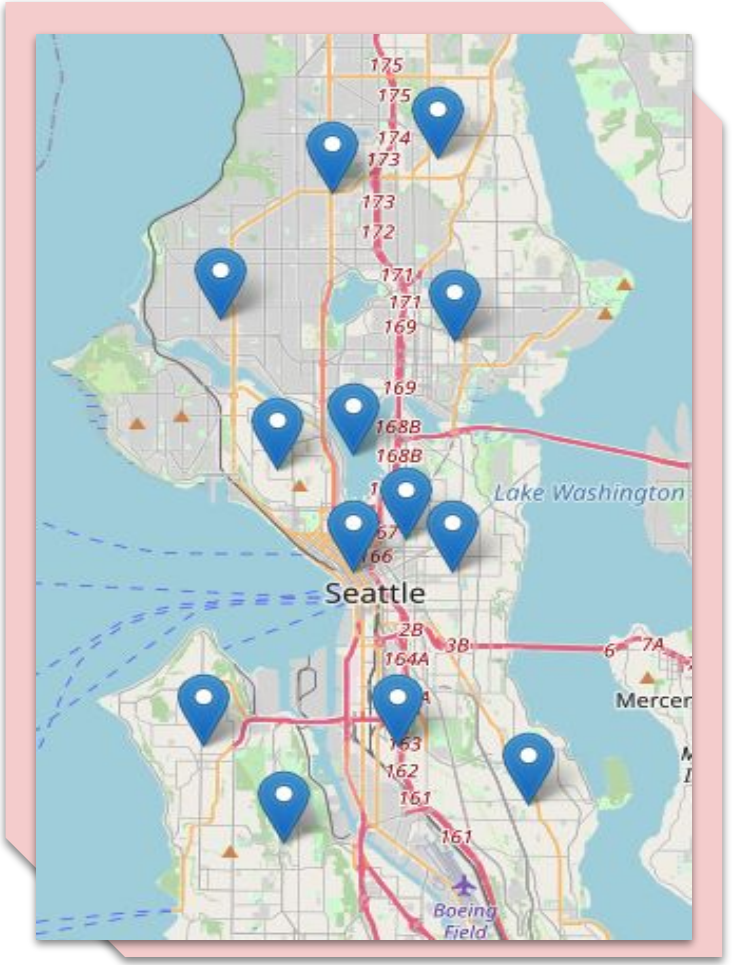
Seattle est une ville globalement récente, avec un essor particulier à partir des années 1980

3.3 - Analyse des données - Partie 3

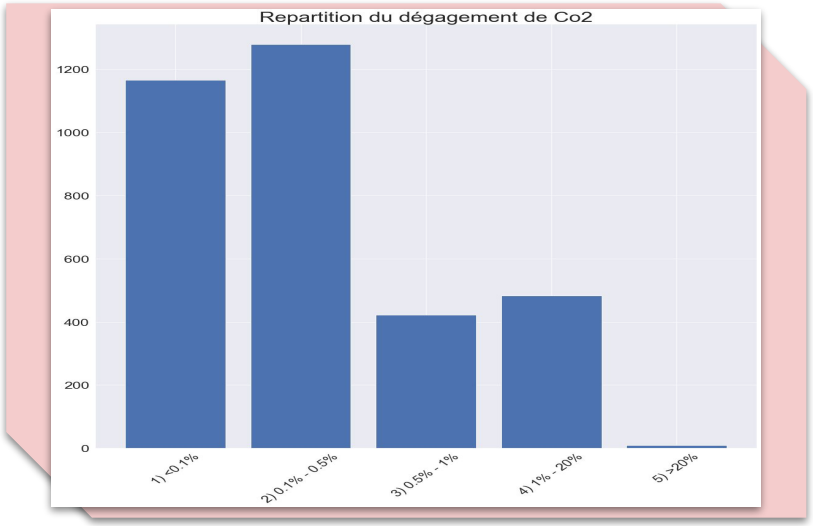


L’energy Star Score est une échelle de comparaison d’énergie (de 1 à 100), Une grande partie des immeubles sont supérieur à 70%

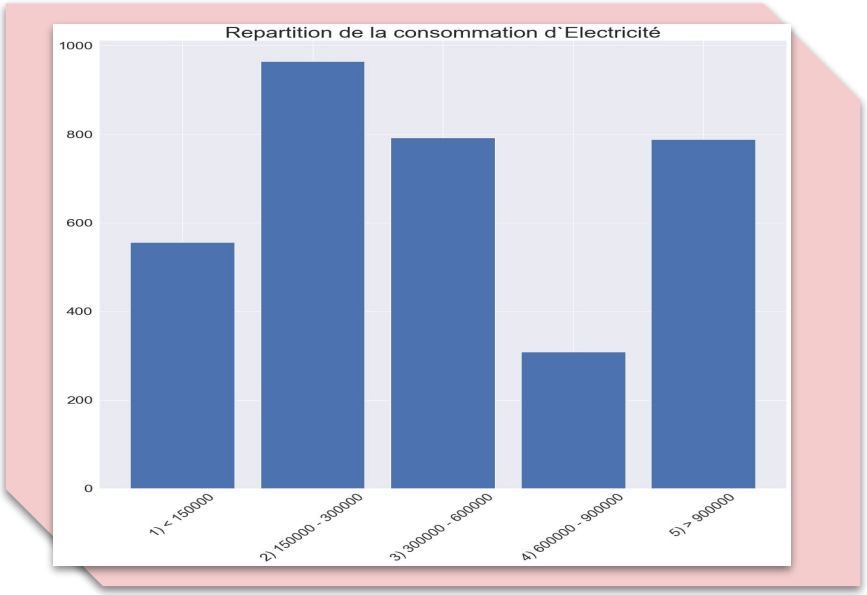
Une analyse des latitudes et longitudes données nous permet d’établir cette carte interactive. Et d’avoir une idée précise de l’emplacement des différents quartiers



3.4 - Analyse des données - Partie 4

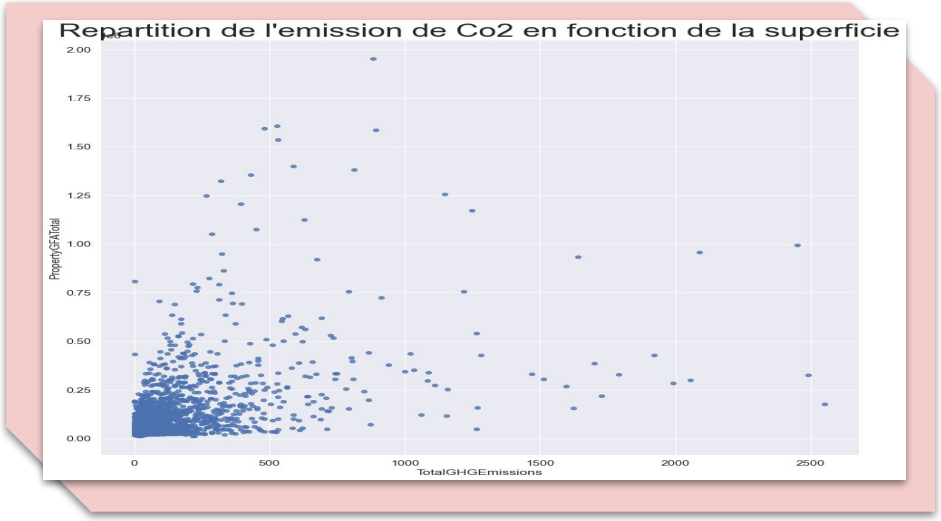


On constate qu'une large proportion d'immeuble dégage assez peu de Co2.

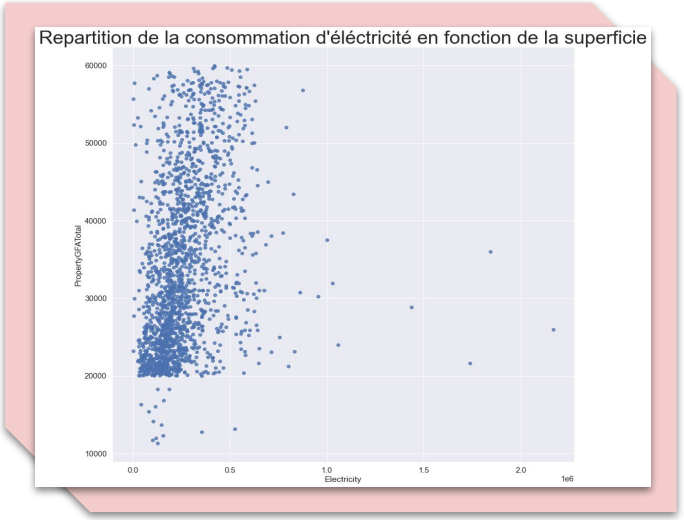


Au contraire la consommation d'électricité semble globalement équitablement réparti entre les différents immeubles

3.5 - Analyse des données - Partie 5

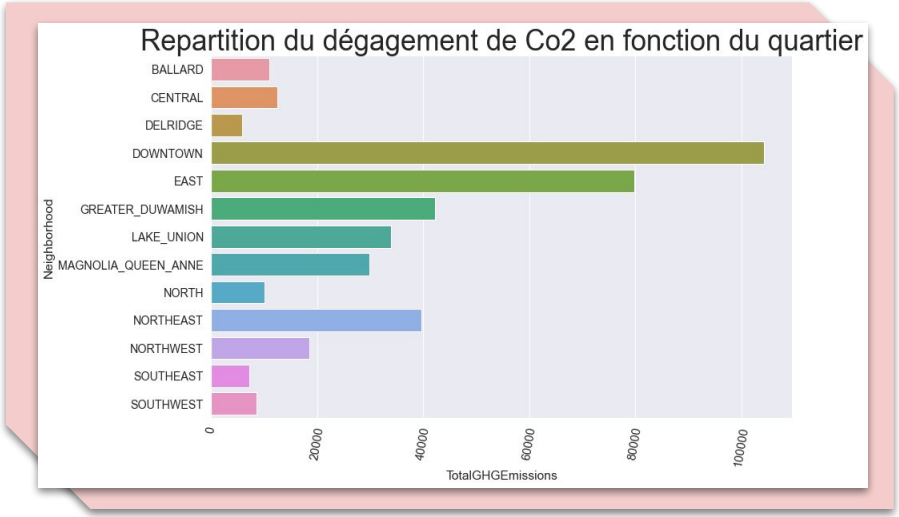


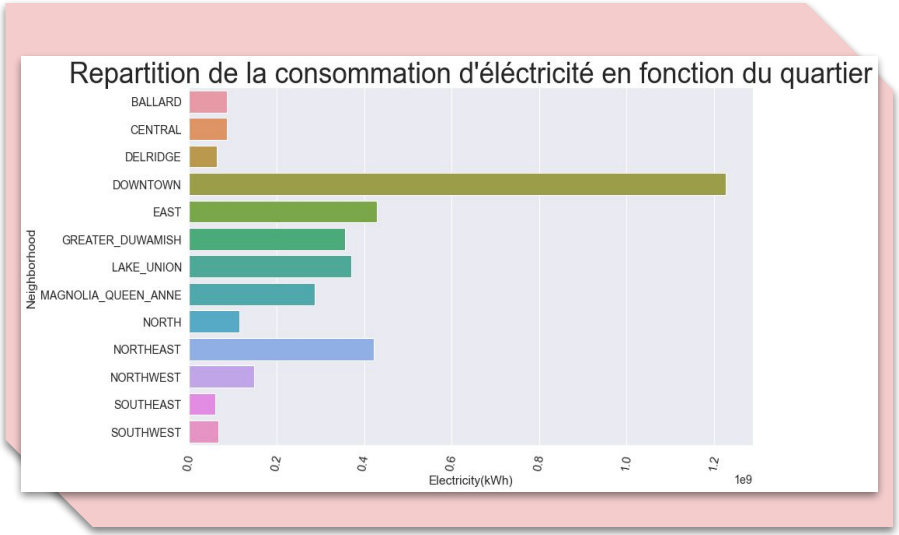
On peut par curiosité commencer avant nos prédictions à analyser quelques variables, par exemple l'émission de Co2 en fonction de la superficie



Autre hypothèse envisageable, une corrélation entre la consommation d'électricité et la superficie de l'immeuble

3.6 - Analyse des données - Partie 6





Intéressons nous aux différents quartiers de la ville, souvent révélateur de caractéristiques particulières, on observe ici le dégagement de Co2 et la consommation d'électricité

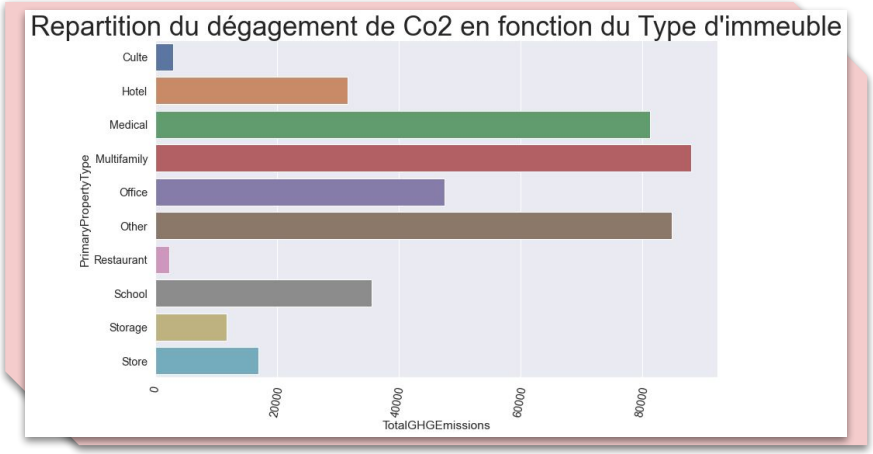
Quartier qui dégage le plus de Co2 :

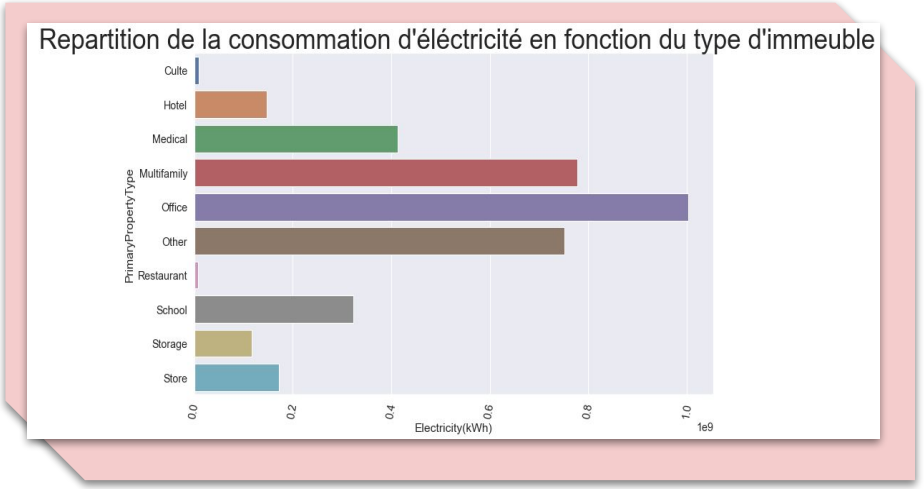
- DownTown
- East
- NorthEast

Quartier qui consomme le plus de d'electricité :

- DownTown
- East
- NorthEast

3.7 - Analyse des données - Partie 7





Les différents types d'immeuble peuvent de même apporter de nombreuses information sur les données que nous devons prévoir

Type d'immeuble qui dégage le plus de Co2 :

- Immeuble d'habitation
- Immeuble divers
- Immeuble à vocation médicale

Type d'immeuble qui consomme le plus d'électricité:

- Immeuble à vocation professionnel
- Immeuble d'habitation
- Immeuble divers

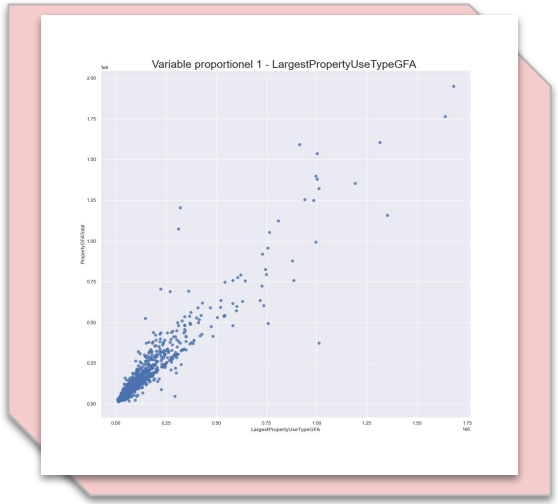
4 - Prédiction

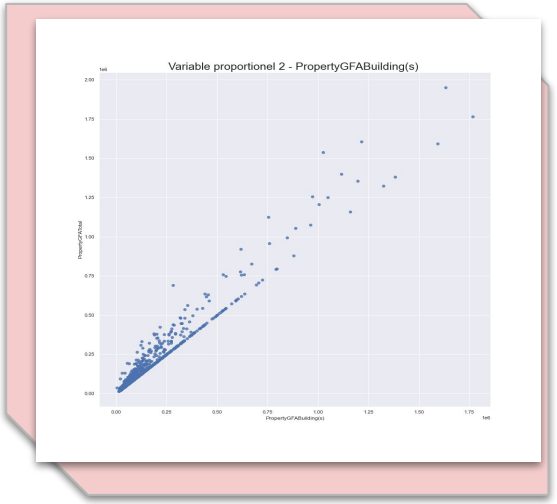
4.1. Variables choisis

Nous avons sélectionné pour notre étude les variables suivantes :

- Superficie
- Energy Star Score
- Latitude
- Longitude
- Quartier
- Type d'immeuble

Ces choix ont été faits après éliminations des variables non-fixes, mais aussi des données proportionnelles





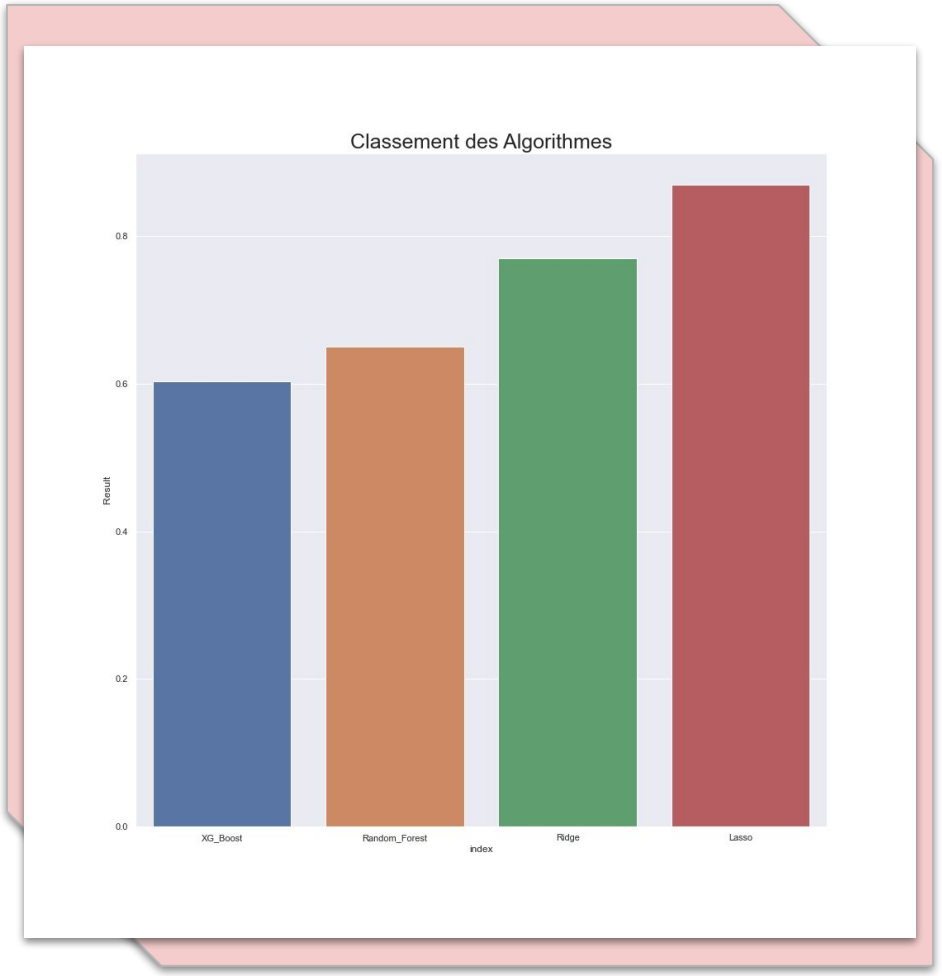
4.2.1 Co2 - Prédictions basiques

RMSE :

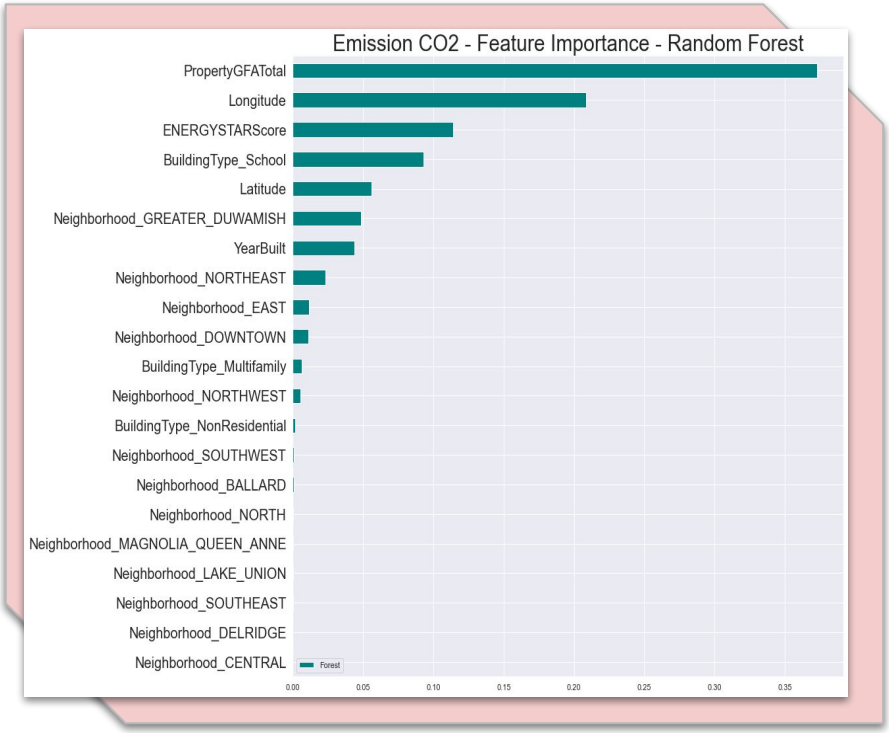
- XGBoost : 323.803
- Random Forest : 363.05
- Lasso : 426.47
- Ridge : 425.28

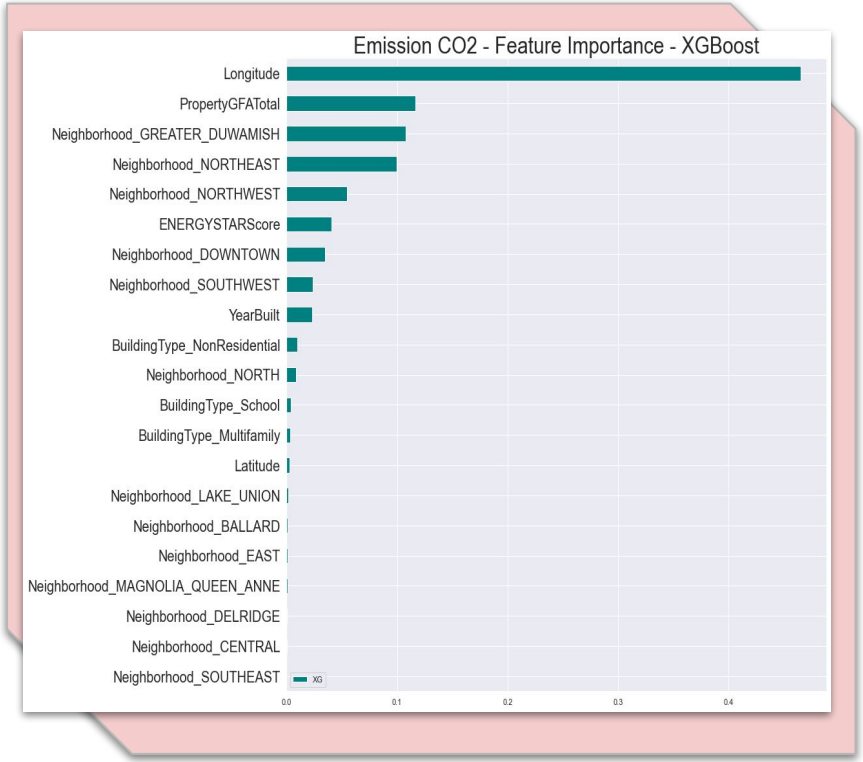
Meilleur Algorithme :

- **XGBoost**
- **Random Forest**



4.2.2 Variables importantes





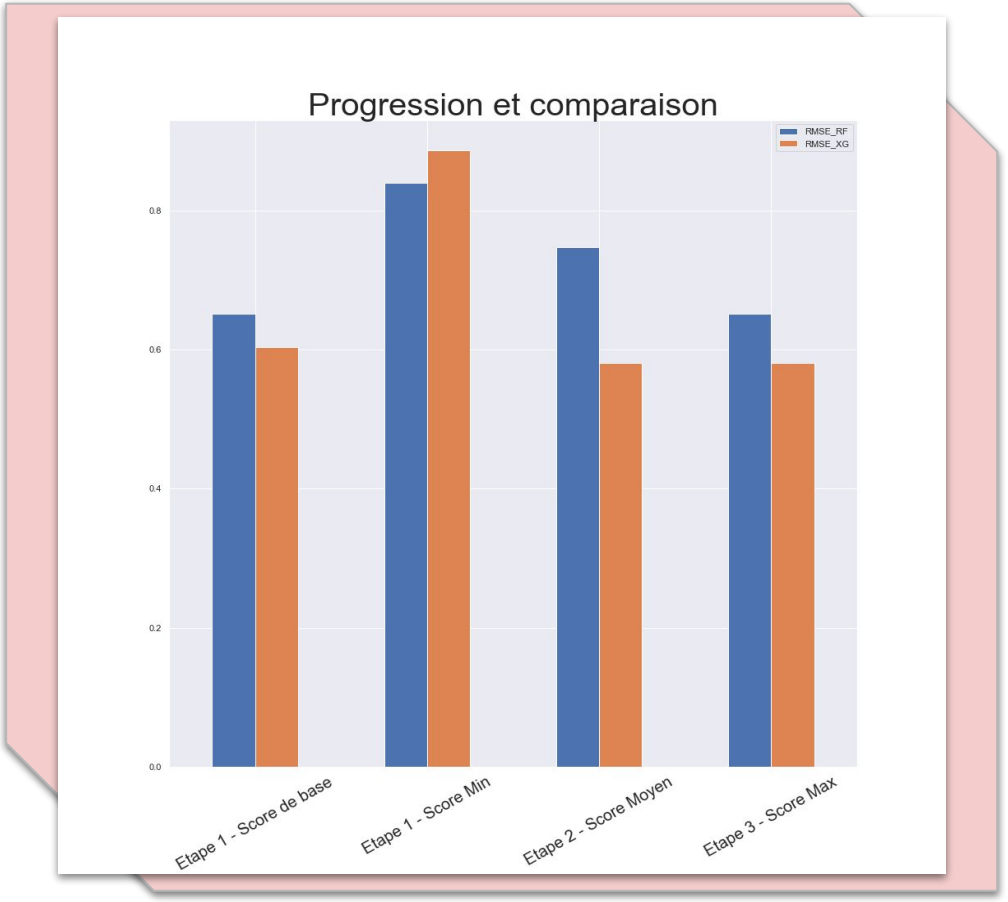
Nous pouvons à ce stade pour chacun de nos algorithmes définir les variables essentielles par ordre croissant

4.2.3 Evolution

Les différents processus d'hyper-paramétrisation et de prises en compte de l'importance des variables nous permettent d'affiner la prévision.

Meilleur Algorithme :

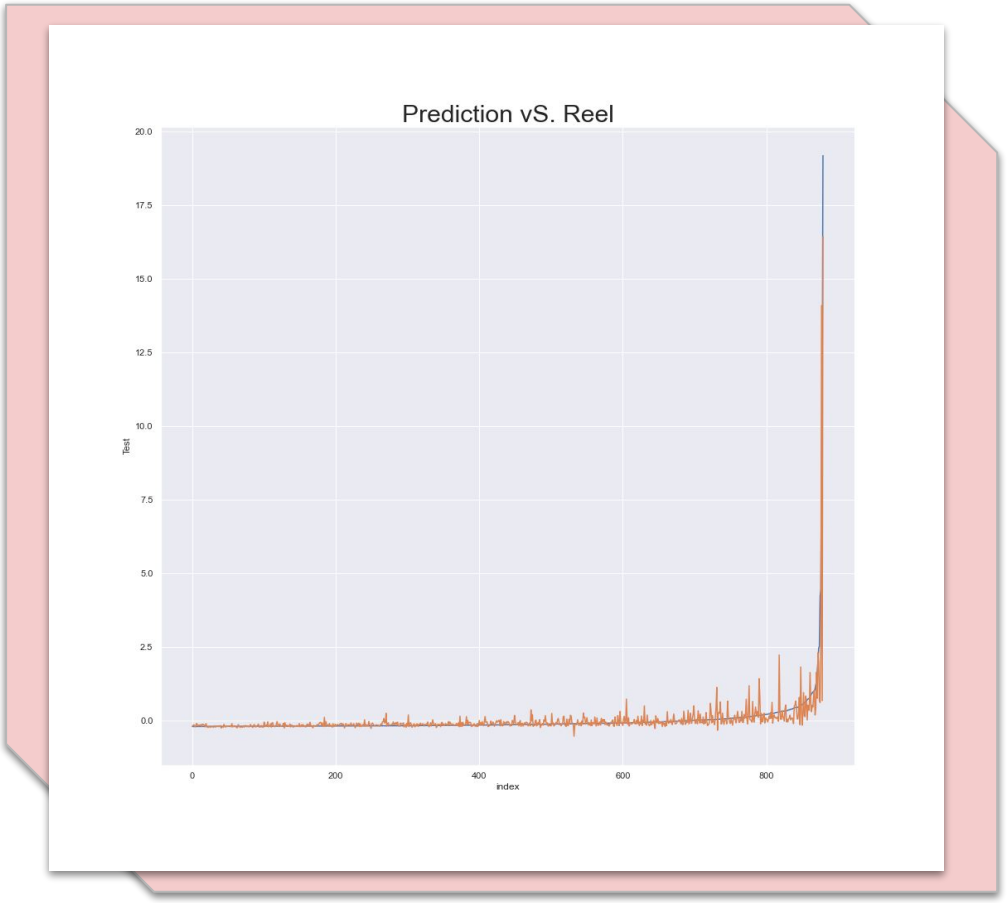
- **XGBoost**



4.2.4 Visualisation

On peut voir ici la courbe qui nous permet de comparer les résultats réels et ceux obtenus.

On constate aisément que cela semble fonctionner

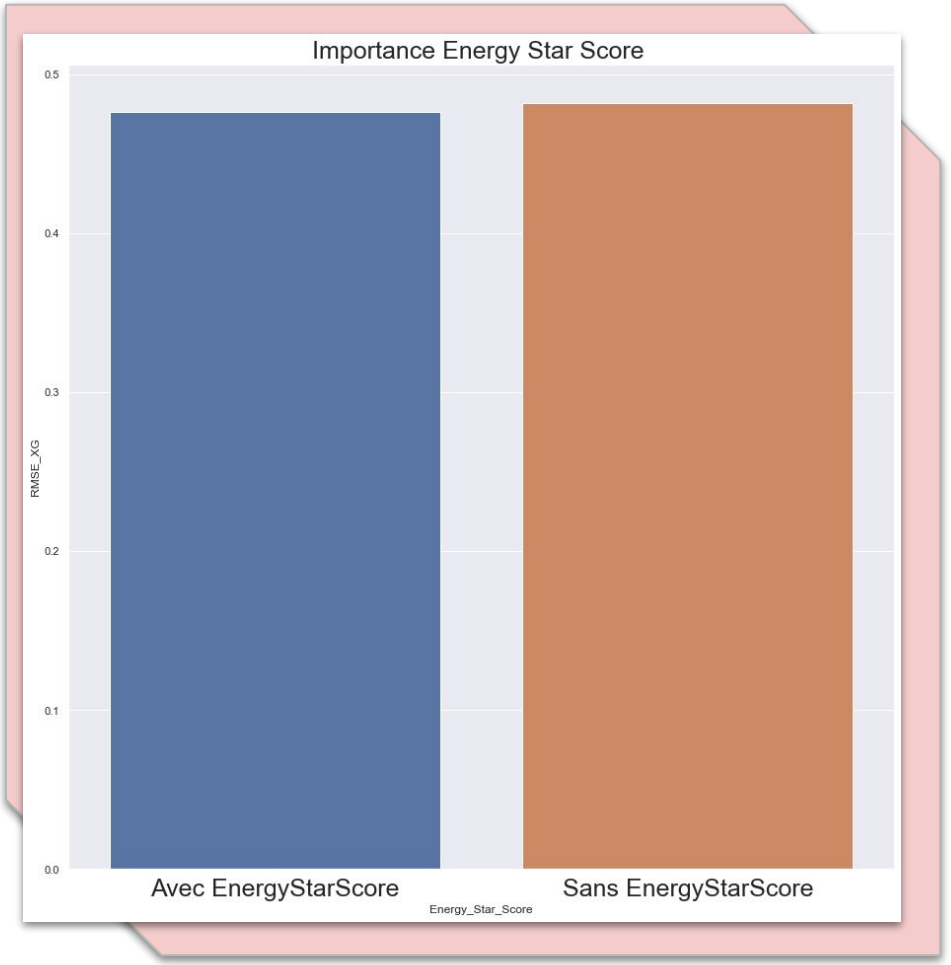


4.2.5 Importance EnergyStarScore

L’energyStarScore et son importance peuvent être étudiées à ce stade.

La complexité du procédé nous amène à un questionnement sur son utilité.

En ce qui concerne l’émission de Co2, l’energyStarScore n’influence pas de manière significative notre prédiction



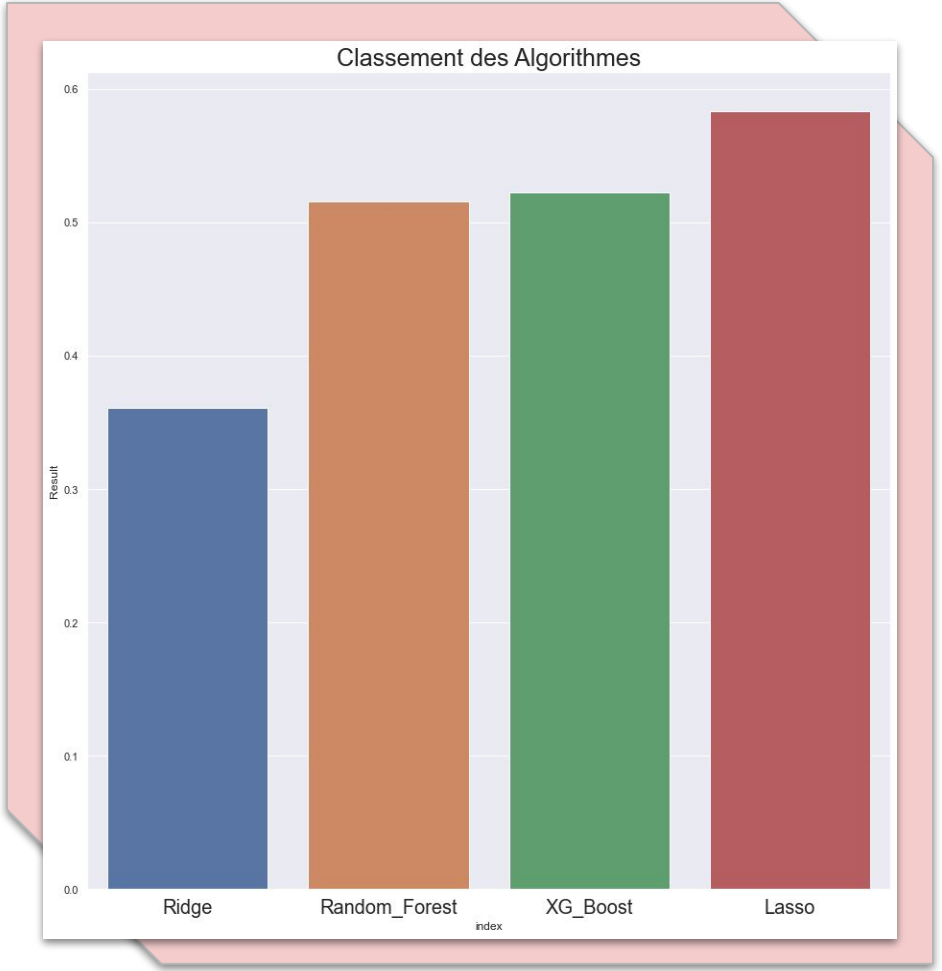
4.3 - Electricité - Prédiction basique

RMSE :

- XGBoost : 1488182.42
- Random Forest : 1630959.69
- Lasso : 1079048.44
- Ridge : 1084621.98

Meilleur Algorithme :

- **Random Forest**
- **Ridge**

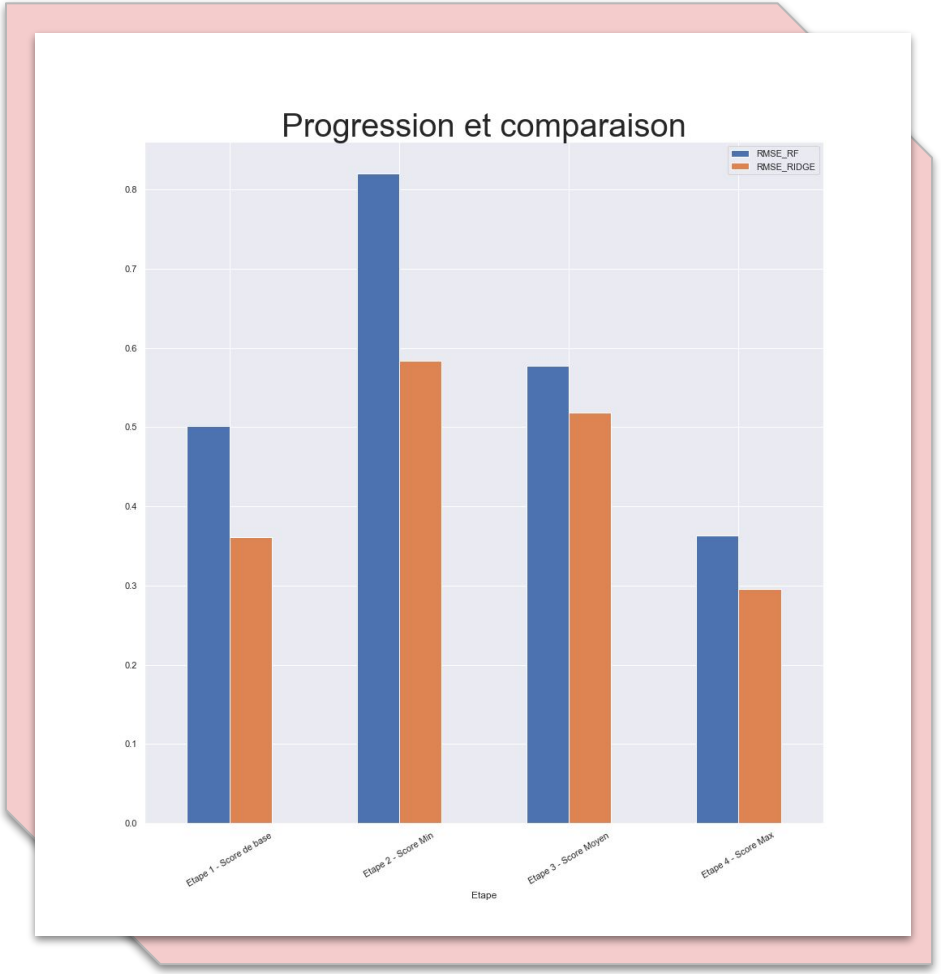


4.3.1 Evolution

On applique désormais quelques techniques pour affiner les prédictions, finalement, un algorithme se dégage

Meilleur Algorithme :

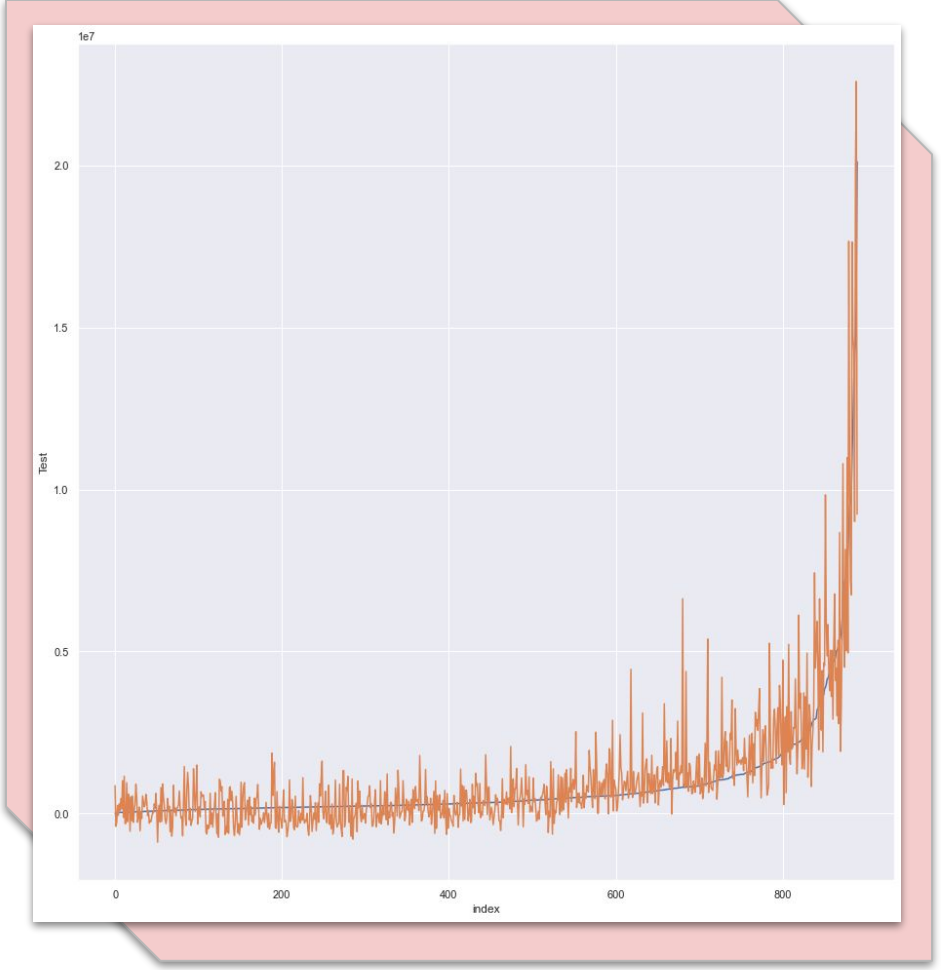
- **RIDGE**



4.3.2 Visualisation

On peut voir ici la courbe qui nous permet de comparer les résultats réels et ceux obtenus.

On constate aisément que cela semble fonctionner



4.3.3 Importance EnergyStarScore

L’energyStarScore et son importance peuvent être étudiées à ce stade.

La complexité du procédé nous amène à un questionnement sur son utilité.

En ce qui concerne l’émission de Co2, l’energyStarScore n’influence pas de manière significative notre prédiction

5 - Conclusion

5.1 - Préambule

Cette étude nous a amené a testé plusieurs algorithmes pour tenter au mieux de prévoir 2 variables importante pour la ville de Seattle.

Après étude attentive nous sommes arrivés à des estimations satisfaisantes pour 4 hypothèses :

1. Libération de CO2 en prenant en compte l'Énergie Score
2. Libération de CO2 sans l'Énergie Score
3. Consommation d'électricité d en prenant en compte l'Énergie Score
4. Consommation d'électricité sans l'Énergie Score

Voici nos conclusions

5.2 - Conclusion

	Meilleur algorithme	RMSE avec Energy Score	RMSE sans Energy Score	Energy Score nécessaire ?
Emission de CO2	XGB Regressor	270	271	NON
Consommation d'électricité	Ridge Régression	888837	913697	NON

Voici nos conclusion néanmoins un travail plus approfondi peut nous permettre d’affiner et surtout l’améliorer nos résultats