

Projet de Prédiction de Consommation Énergétique et d'Émissions de CO₂

Pour une Seattle neutre en carbone d'ici 2050

Présenté par : Abdourahamane LY



Sommaire

- Contexte du Projet : Vers la Neutralité Carbone
- Aperçu des Données Analysées
- Insights Clés de l'Analyse Exploratoire
- Préparation Stratégique des Données
- Comparaison des Modèles de Prédiction
- Optimisation des Modèles Sélectionnés
- Facteurs d'Influence Clés sur la Consommation et l'Émissions de CO₂
- Conclusion

Contexte du Projet : Vers la Neutralité Carbone

1

Objectif Ambitieux

Atteindre la neutralité carbone d'ici 2050, un engagement clé pour l'avenir durable de Seattle.

2

Données Fondatrices

Analyse de 3376 lignes × 46 colonnes dont 1546 bâtiments non résidentiels.

3

Mission Stratégique

Prédire la consommation énergétique et les émissions de CO₂ à partir des caractéristiques structurelles.

4

Enjeu Économique

Réduire les coûts futurs des relevés manuels, optimisant les ressources de la ville.

Aperçu des Données Analysées

Notre analyse s'appuie sur un jeu de données initial de 3376 observations et 46 variables. Après un processus de nettoyage et de préparation, le jeu de données final contient des informations sur 1083 bâtiments non résidentiels de Seattle.

Les variables du jeu de données initial se répartissent comme suit :

- 15 variables qualitatives
- 30 variables quantitatives (dont 22 réelles et 8 entières)
- 1 variable booléenne

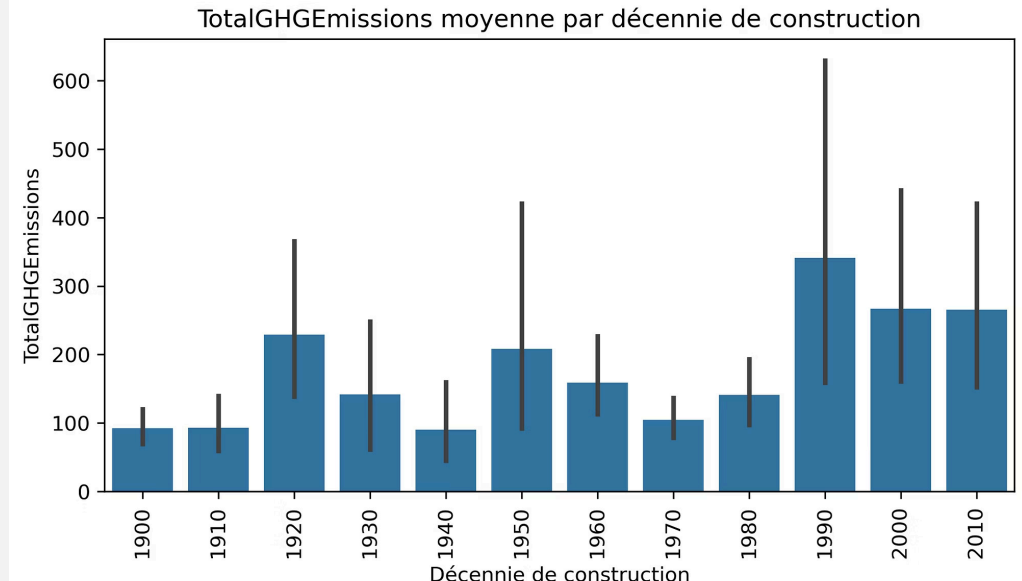
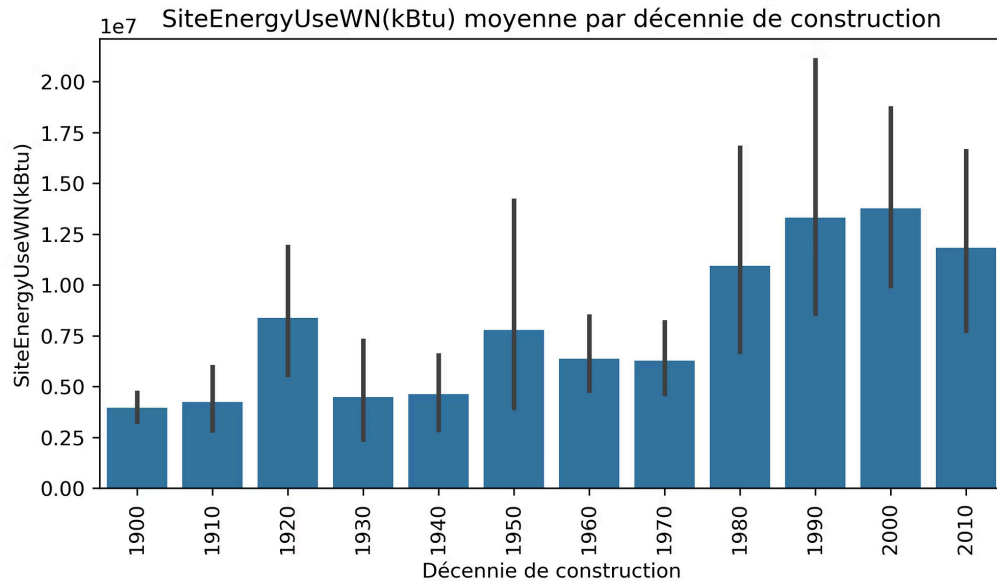
Étape de Traitement	Nb. Lignes	Nb. Colonnes
Jeu de données initial	3376	46
Exclusion des bâtiments résidentiels	1546	46
Nettoyage des données (valeurs aberrantes, dédoublonnage...)	1083	46
Suppression des colonnes redondantes et à risque de fuite	1083	9

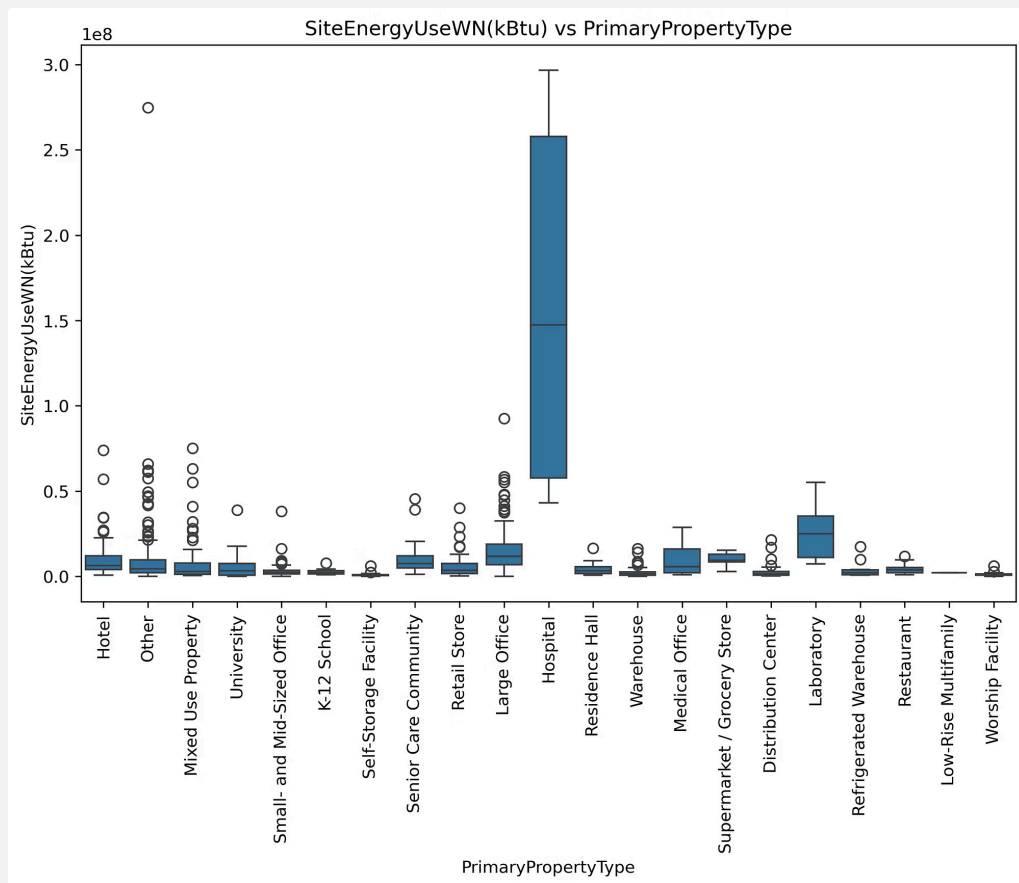
Insights Clés de l'Analyse Exploratoire



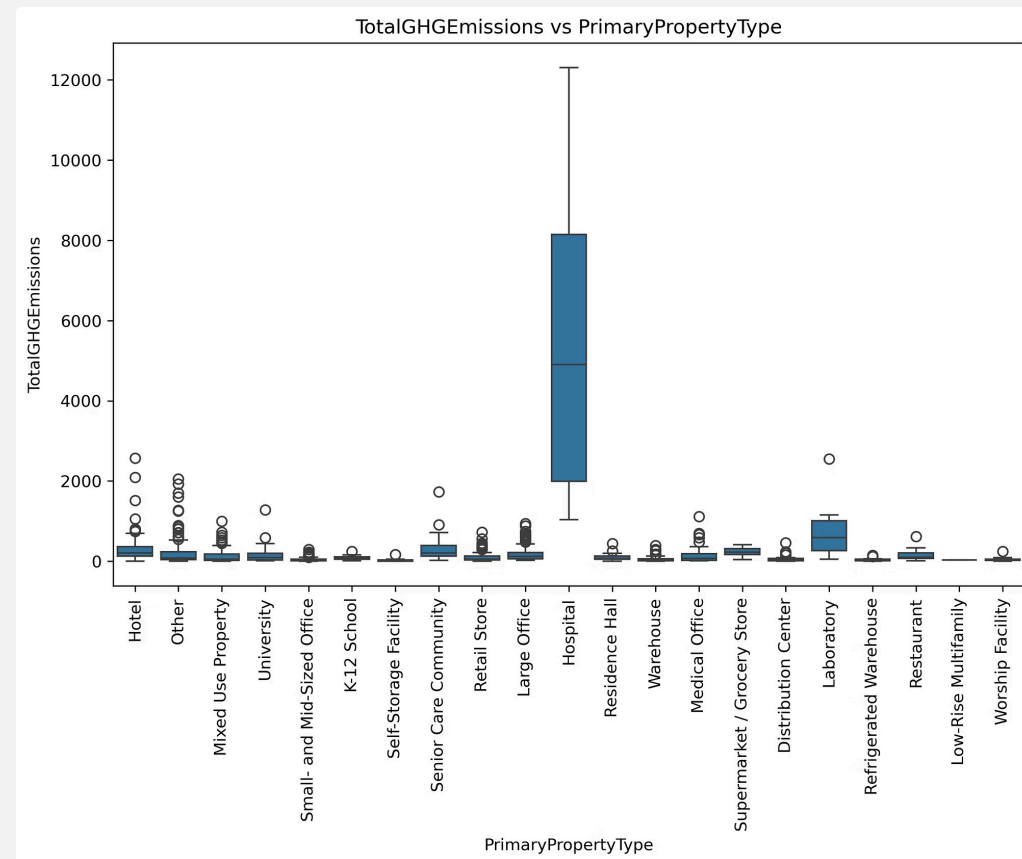
Impact de l'Ancienneté

Les bâtiments construits dans les années 1990 et 2000 affichent les plus fortes consommations énergétiques et émissions de CO₂ moyennes, tandis que ceux construits avant 1960 se montrent globalement plus sobres.





Les hôpitaux et laboratoires figurent parmi les bâtiments les plus énergivores, tandis que les lieux de culte et les entrepôts affichent une consommation bien plus modérée.

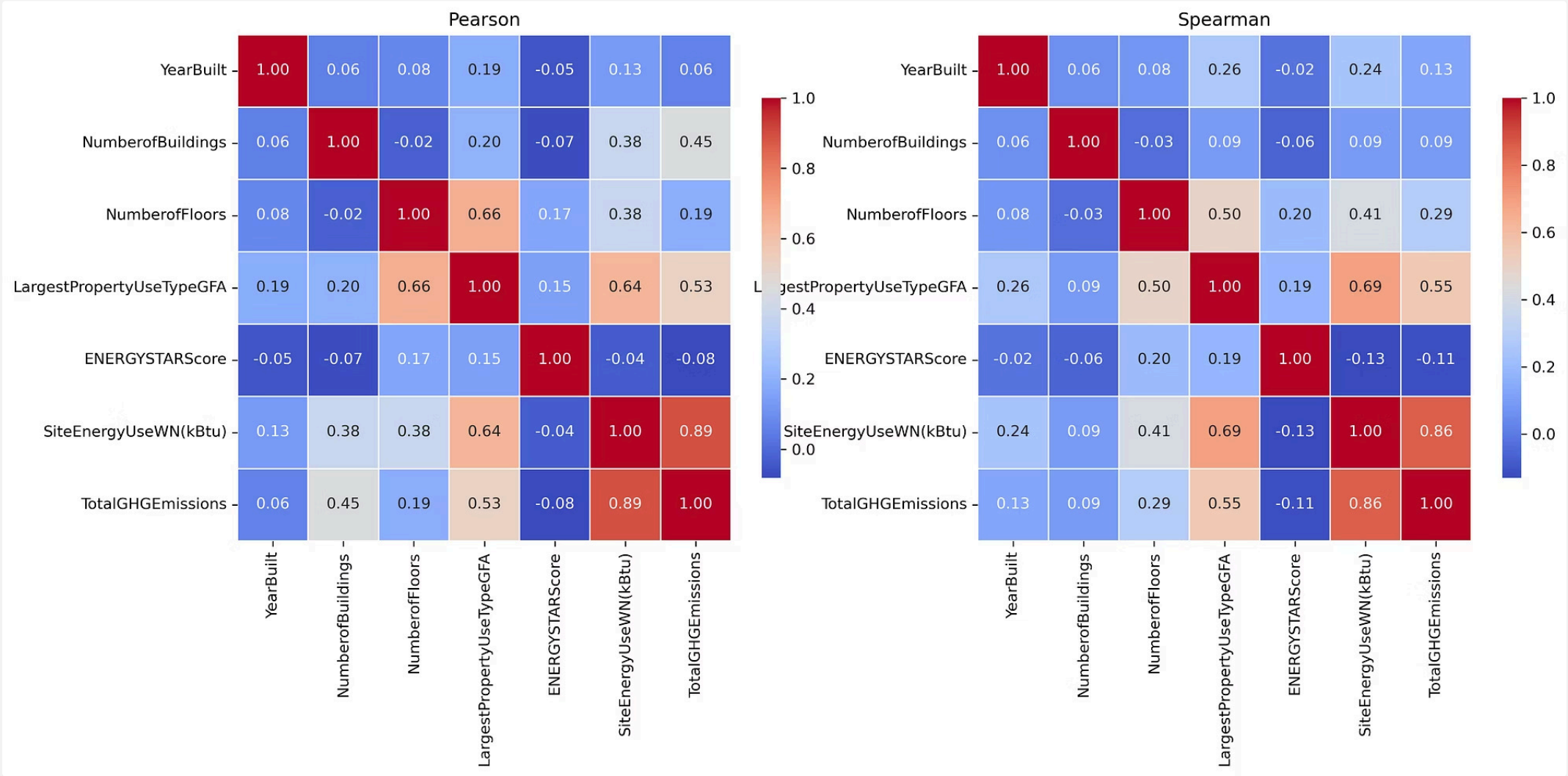


Les hôpitaux se distinguent par des émissions de gaz à effet de serre bien supérieures à tous les autres types de bâtiments, reflétant une activité intensive et une forte demande énergétique.



Corrélation Forte

Une corrélation linéaire de **0.89** et monotone de **0.86** entre consommation et émissions, confirmant leur interdépendance.





Influence des Catégories de Bâtiments

Analyse de l'impact des différentes catégories de bâtiments sur la consommation et les émissions.

```
=====
TotalGHGEmissions vs PrimaryPropertyType → effet significatif (F=79.36, p=0.000000)
TotalGHGEmissions vs LargestPropertyUseType → effet significatif (F=30.26, p=0.000000)
TotalGHGEmissions vs SecondLargestPropertyUseType → pas d'effet significatif (F=0.38, p=0.999769)
TotalGHGEmissions vs ThirdLargestPropertyUseType → effet significatif (F=1.73, p=0.025557)
=====
```

Les émissions de GES sont fortement liées au type principal de bâtiment. L'usage principal joue un rôle significatif, tandis que les usages secondaires et tertiaires ont un impact marginal.

```
=====
SiteEnergyUseWN(kBtu) vs PrimaryPropertyType → effet significatif (F=47.95, p=0.000000)
SiteEnergyUseWN(kBtu) vs LargestPropertyUseType → effet significatif (F=22.75, p=0.000000)
SiteEnergyUseWN(kBtu) vs SecondLargestPropertyUseType → pas d'effet significatif (F=0.50, p=0.995088)
SiteEnergyUseWN(kBtu) vs ThirdLargestPropertyUseType → effet significatif (F=1.60, p=0.047657)
=====
```

La consommation énergétique varie également fortement selon le type principal de bâtiment. L'usage principal influence significativement la consommation, alors que les usages secondaires et tertiaires n'ont qu'une influence marginale.

Préparation Stratégique des Données

Une préparation méticuleuse des données a été effectuée pour garantir la robustesse et la performance de nos modèles.

- ① • **Élimination des Redondances** : Nous avons procédé à la suppression des variables redondantes afin d'optimiser le jeu de données.
- **Séparation Train/Test** : Le jeu de données a été divisé en ensembles d'entraînement (80%) et de test (20%), avec une stratification basée sur le **PrimaryPropertyType** pour assurer une représentativité équilibrée.
- **Pipeline de Transformation** :
 - Application de **RobustScaler** pour la normalisation des variables numériques.
 - Utilisation de **LabelEncoder** pour le traitement des variables catégorielles.
- **Évaluation de l'Impact d'ENERGY STAR** : Le score **ENERGY STAR** a été inclus dans une première phase de modélisation, puis exclu dans une seconde pour évaluer son influence sur la précision des prédictions.



Comparaison des Modèles de Prédiction

Évaluation comparative des performances pour la consommation énergétique et les émissions de CO₂ avec la variable
ENERGY STAR SCORE

Émissions de CO₂

Modèle	R ²	RMSE	MAE
Dummy	-0.0002	853.37	216.23
Linear	0.43	643.70	192.98
SVR	-0.013	858.72	162.67
RandomFores t	0.64	513.61	128.65
XGBoost	0.86	323.97	125.96
GradientBoos ting	0.59	544.56	133.28

Modèle Final : XGBoost (R² = 0.86)

Consommation d'Énergie

Modèle	R ²	RMSE (kBtu)	MAE (kBtu)
Dummy	-3.43	22.8M	8.95M
Linear	0.59	14.7M	5.41M
SVR	-0.05	23.4M	7.19M
RandomForest	0.52	15.9M	4.94M
XGBoost	0.21	20.3M	5.91M
GradientBoosti ng	0.30	19M	5.83M

Modèle Final : Linear (R² = 0.59)

Comparaison des Modèles de Prédiction

Évaluation comparative des performances pour la consommation énergétique et les émissions de CO₂ sans la variable
ENERGY STAR SCORE

Émissions de CO₂

Modèle	R ²	RMSE	MAE
Dummy	-0.0002	853.37	216.23
Linear	0.43	644.40	180.69
SVR	-0.013	858.59	162.63
RandomForest	0.84	345.51	107.12
XGBoost	0.95	194.55	103.48
GradientBoosting	0.94	219.3	93.73

Modèle Final : XGBoost (R² = 0.95)

Consommation d'Énergie

Modèle	R ²	RMSE (kBtu)	MAE (kBtu)
Dummy	-3.43	22.82M	8.95M
Linear	0.58	14.83M	5.37M
SVR	-0.05	23.41M	7.19M
RandomForest	0.82	9.63M	4.40M
XGBoost	0.69	12.69M	4.77M
GradientBoosting	0.75	11.36M	4.77M

Modèle Final : RandomForest (R² = 0.82)

Optimisation des Modèles Sélectionnés

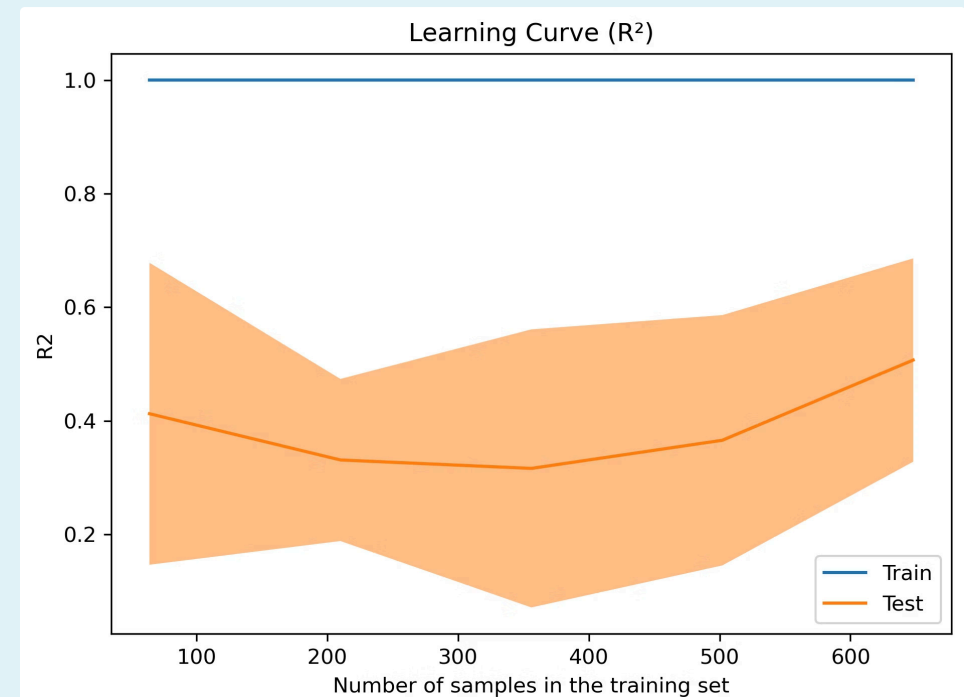
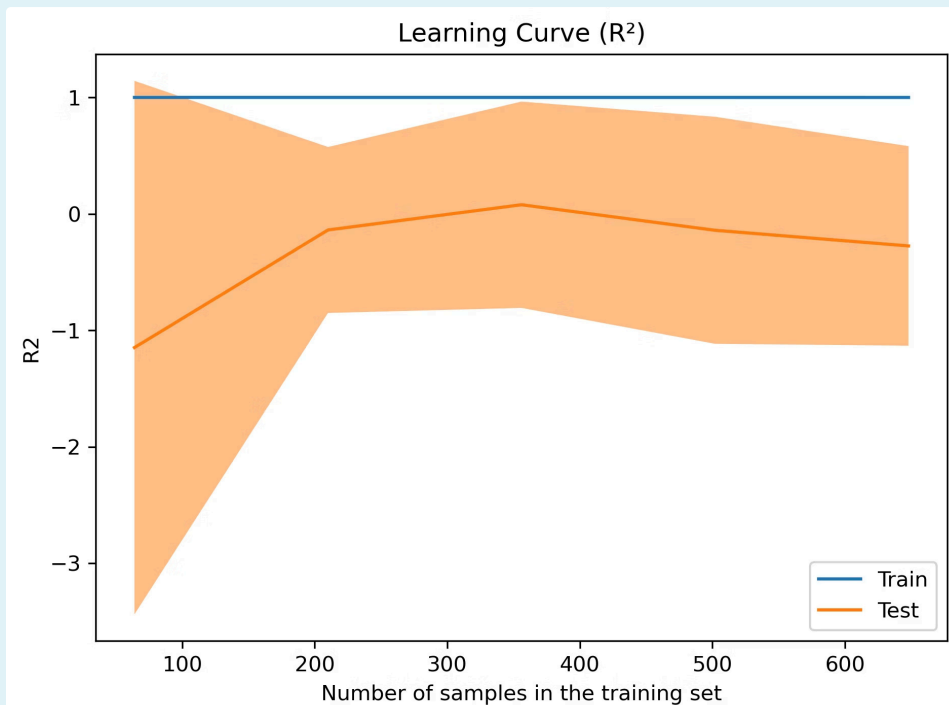
XGBoost (Émissions CO₂)

Le modèle final sélectionné a un **R² de 0.95** avec les hyperparamètres par défaut

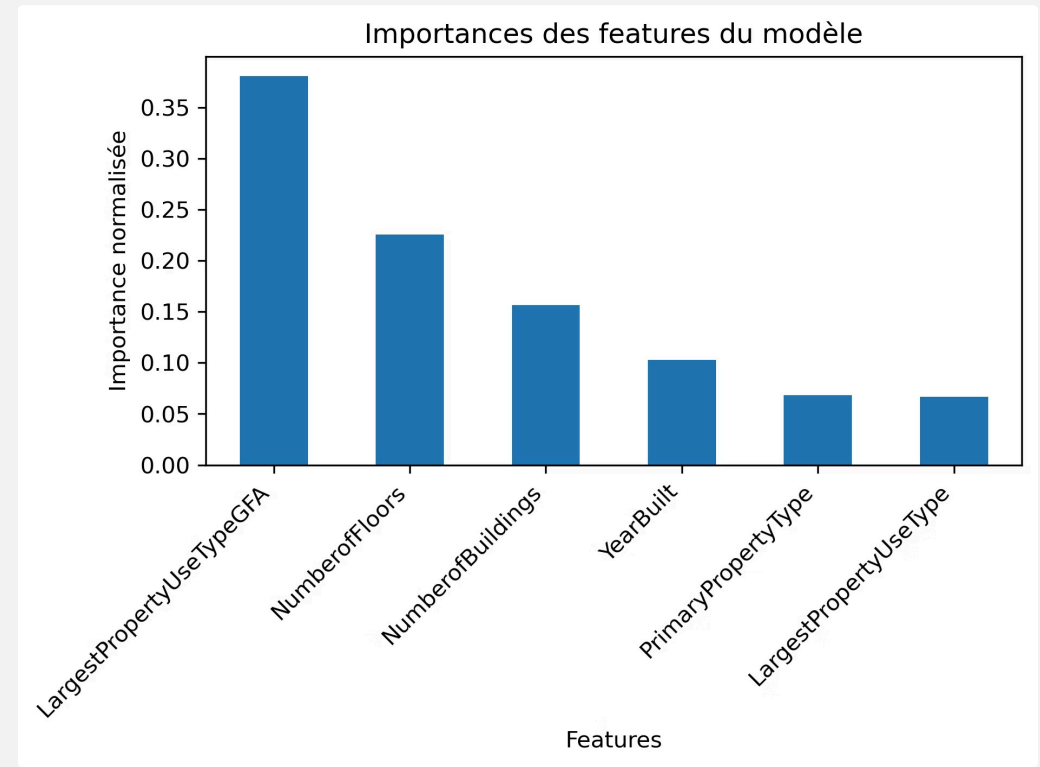
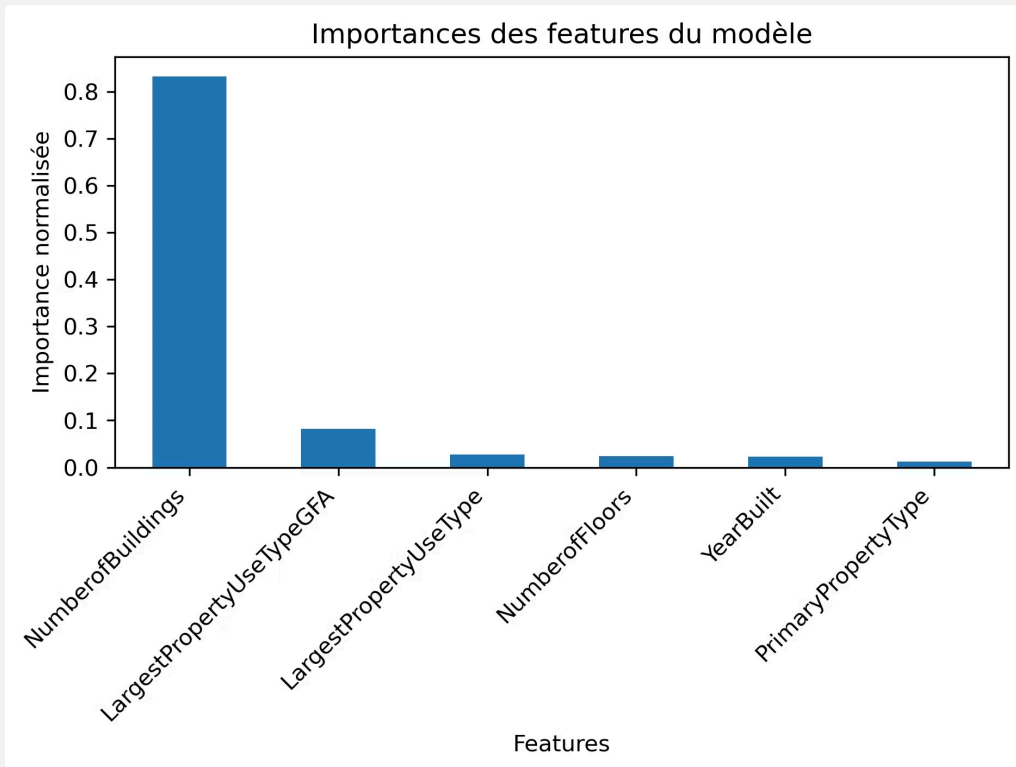
RandomForest (Consommation Energie)

- **n_estimators** : 430. **max_features** : sqrt
- **min_samples_split** : 2. **bootstrap** : False
- **min_samples_leaf** : 1. **max_depth** : 19

R² de 0.84



Facteurs d'Influence Clés sur les Émissions CO₂ et la Consommation d'Energie

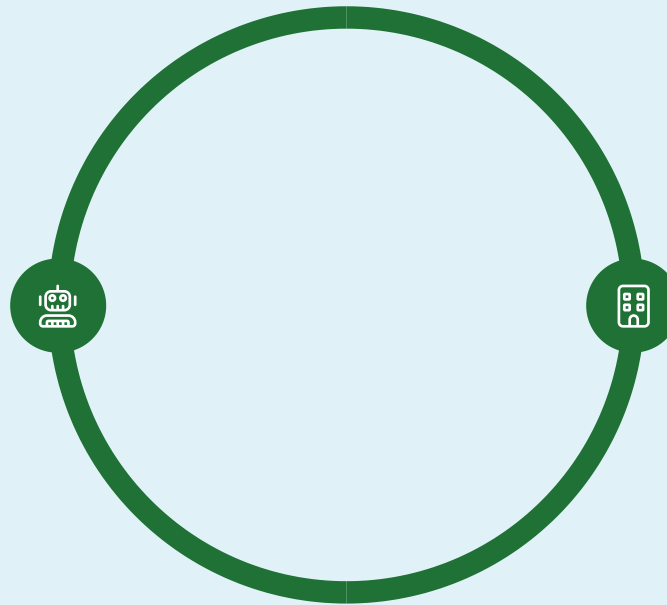


Conclusion

Nos analyses techniques ont permis de formuler des recommandations stratégiques concrètes et d'identifier des pistes d'action pour la ville de Seattle.

Amélioration Continue des Modèles

Pour optimiser la précision et la robustesse de nos modèles prédictifs, l'acquisition et l'intégration de données supplémentaires s'avèrent cruciales.



La Superficie : Un Levier Stratégique Majeur

La superficie totale (GFA) est confirmée comme le facteur le plus influent, offrant un levier essentiel pour guider les initiatives d'optimisation énergétique et de réduction des émissions.

Merci et Avez-vous des questions ?

N'hésitez pas à nous poser toutes vos questions.

