



# Prévision de la Consommation d'Énergie Électrique



# INTRODUCTION

## Prévision de la Consommation d'Énergie Électrique

Ce projet propose une analyse avancée de la consommation énergétique en France, en exploitant les séries temporelles pour dévoiler des tendances et modèles significatifs. En considérant les fluctuations saisonnières et les anomalies, nous visons à améliorer la prévision énergétique pour optimiser la gestion et la distribution des ressources, tout en étudiant l'impact des facteurs environnementaux.



# Cartographie du projet : De l'analyse initiale à la prédiction avancée



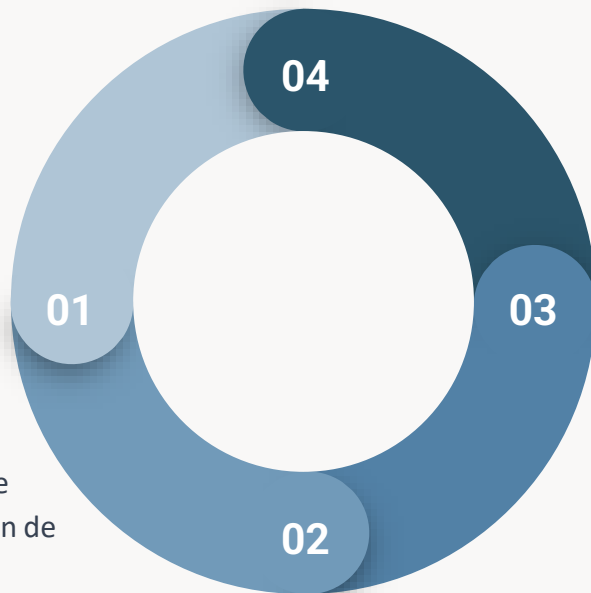
## Présentation de la Base de Données

Fondation sur les données : révéler les chiffres qui animent le récit énergétique



## Préparation et Traitement de la Base de Données

De la donnée brute à la connaissance approfondie : maîtriser la préparation de l'information



## Analyse Exploratoire des Données (AED)

L'art de l'enquête des données : décrypter l'histoire cachée derrière les chiffres



## Modélisation

des modèles prédictifs pour esquisser l'avenir de la consommation énergétique

# 01

## Présentation de la Base de Données



# LES DONNEES

## Architecture et Provenance des Données : Fondements de l'Analyse Énergétique

Date	Consumption at Normal Temperature (MW)
01/01/2014	60955
02/01/2014	67923
03/01/2014	70153
04/01/2014	66541
05/01/2014	64399

Date	Filière	Valeur (GW)
01/01/2019	Enveloppe basse 2014 - 2019	53,90419485
02/01/2019	Enveloppe basse 2014 - 2019	64,63343265
03/01/2019	Enveloppe basse 2014 - 2019	64,12361827
04/01/2019	Enveloppe basse 2014 - 2019	63,32177608
05/01/2019	Enveloppe basse 2014 - 2019	67,94584104

Date	Pic journalier consommation (MW)	Température moyenne (°C)
01/01/2012	59610	11.7
04/01/2012	78337	8.1
13/01/2012	78700	5.3
14/01/2012	74666	2.4

### Données utilisées

L' étude se concentre sur une base de données énergétique couvrant la période de 2020 à 2023, détaillant la consommation énergétique en France.

### Les variables qui seront utilisées

Les principales variables étudiées sont la consommation brute d'énergie et la température. Nous intégrons des données complémentaires pour enrichir notre analyse.

### Sources des données



# 02

## Préparation et Traitement de la Base de Données

# Les étapes du traitement des bases des données

Finalisation de la base de données pour une analyse détaillée, en établissant une cohérence et une uniformité essentielles pour l'extraction de données fiables.

01

## Harmonisation et Structuration des Données

Mise à jour du format des colonnes pour une lisibilité accrue.  
Standardisation des noms de variables pour une meilleure compréhension.  
Conversion des valeurs énergétiques en mégawatts (MW).

02

## Optimisation et Validation des Données

Filtrage de la colonne 'Filière' pour ne conserver que les données pertinentes.  
Vérification de la cohérence des données à travers les différentes sources

03

## Intégration et Unification des Bases de Données

Synchronisation des formats de date pour une intégration sans faille.  
Fusion des ensembles de données pour créer une base unique.

04

## Consolidation et Préparation pour Analyse

Alignement méthodique des données pour assurer une jointure précise.  
Constitution d'une base de données unifiée pour faciliter l'analyse approfondie.

# Table Finale après la jointure des données – 1430 lignes et 4 colonnes

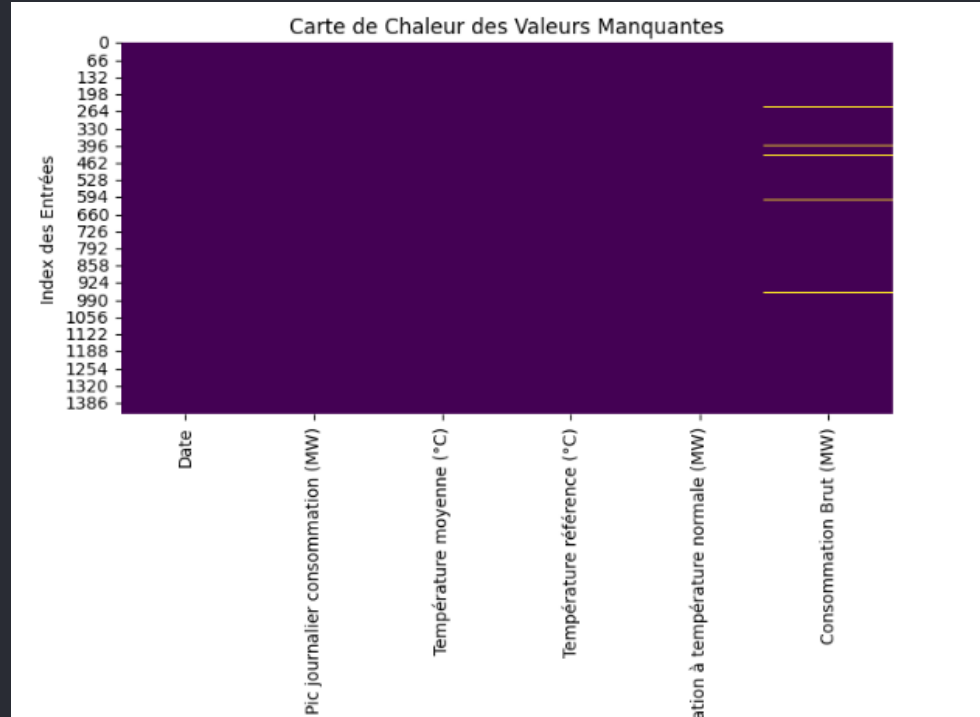
Date	Pic journalier consommation (MW)	Température moyenne (°C)	Température référence (°C)	Consommation à température normale (MW)	Consommation Brut (MW)
01/01/2020	66221	4.0	4.6	60767	63465.5
02/01/2020	72399	6.1	4.6	66879	71354.5
03/01/2020	69650	8.3	4.6	68430	70778.5
04/01/2020	67013	6.7	4.6	66059	NaN
05/01/2020	68154	5.8	4.6	64850	NaN
...	...	...	...	...	...
26/11/2023	64804	4.0	7.2	53997	56591.0
27/11/2023	71635	6.3	7.0	60207	62901.5
28/11/2023	71993	6.5	6.9	61348	63253.0
29/11/2023	72806	4.9	6.8	61397	64661.0
30/11/2023	71243	6.1	6.7	60982	63784.0



# Gestion des Données Incomplètes

La visualisation par carte de chaleur a révélé des valeurs manquantes cruciales dans 'Consommation Brut (MW)'. Nous avons adopté l'imputation médiane pour sa robustesse aux valeurs extrêmes, assurant l'intégrité des analyses ultérieures.

La médiane choisie, 47554.5 MW, reflète la tendance centrale sans distorsion due aux anomalies, préservant la distribution originale des données pour des prédictions fiables.



	count	mean	std	min	25%	50%	75%	max
Consommation Brut (MW)	1393.0	51667.233309	10388.460078	35288.0	43868.0	47554.5	60611.5	77591.0

# Optimisation des Données pour l'analyse et pour la prédiction Prédiction Énergétique

L'ingénierie des caractéristiques est une phase critique qui affine notre jeu de données, permettant une analyse temporelle précise et une compréhension approfondie des motifs de consommation d'énergie.



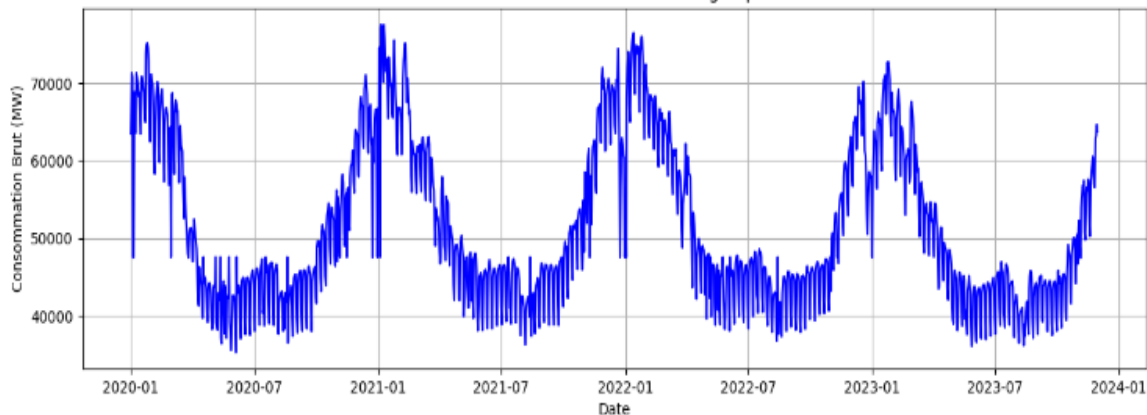
# 03



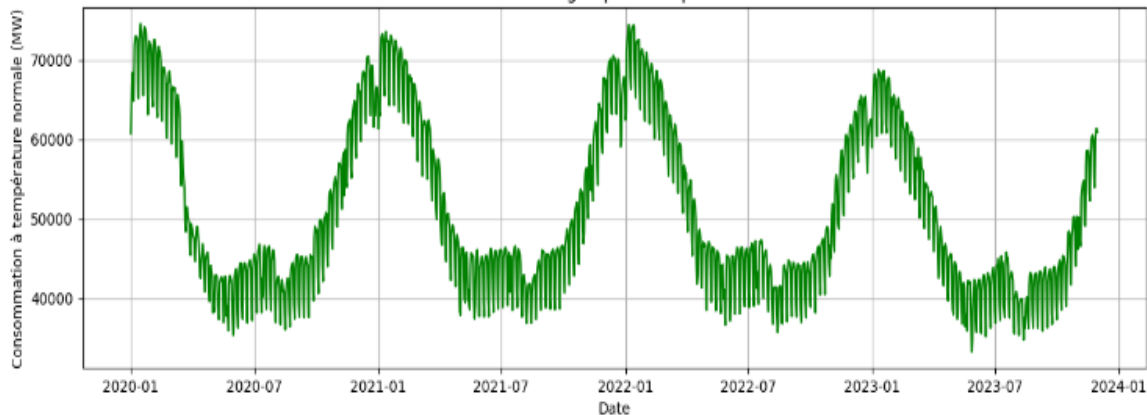
## Analyse Exploratoire des Données (AED)

# Décryptage des Tendances Énergétiques :

Tendance de la Consommation Énergétique Brut



Consommation Énergétique à Température Normale



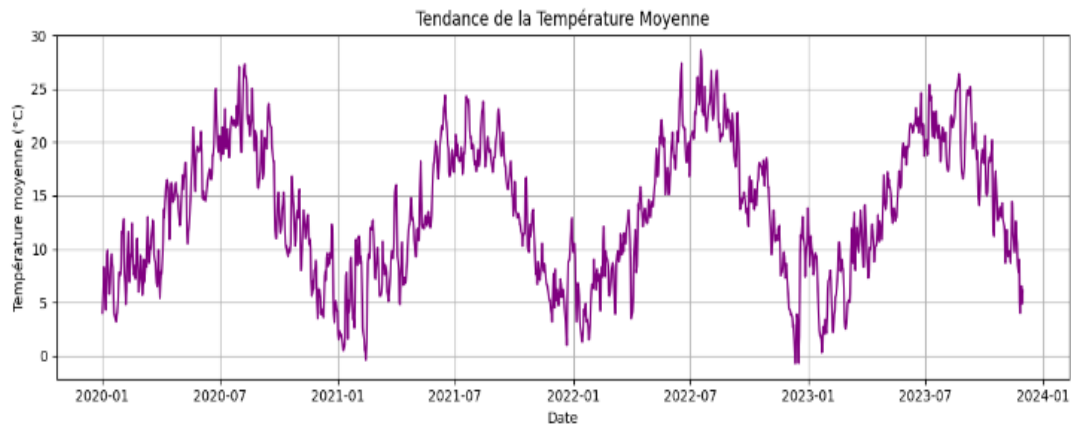
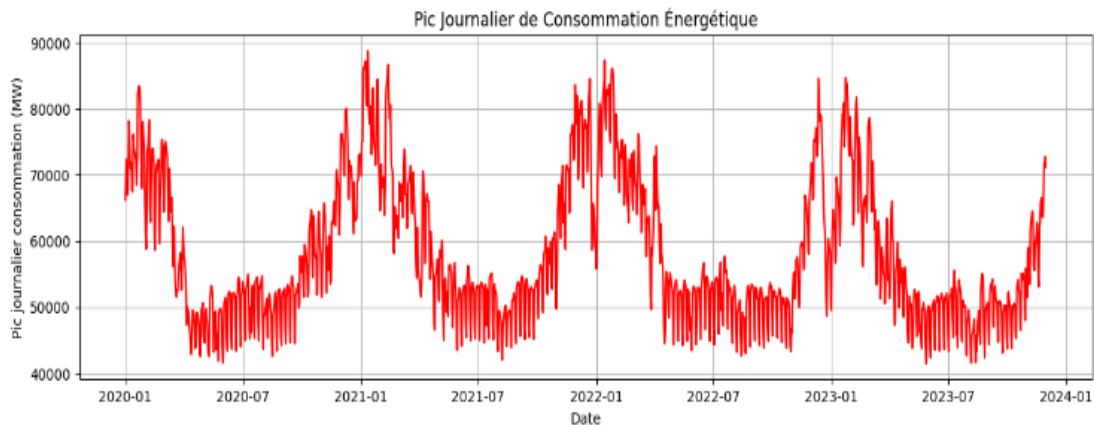
## Fluctuations Saisonnnières:

Le graphique en bleu révèle les variations saisonnières de la consommation brute d'énergie, soulignant les pics durant les mois d'hiver et d'été, périodes de forte demande énergétique.

## Impact Climatique:

La courbe verte met en évidence l'ajustement de la consommation d'énergie en fonction des températures normales, dévoilant l'influence directe du climat sur la demande en énergie.

# Décryptage des Tendances Énergétiques :



## Pics de Demande Énergétique:

Le graphique rouge illustre les points culminants de consommation énergétique, révélant les périodes de forte demande et leur corrélation avec les conditions météorologiques ou les événements spécifiques.

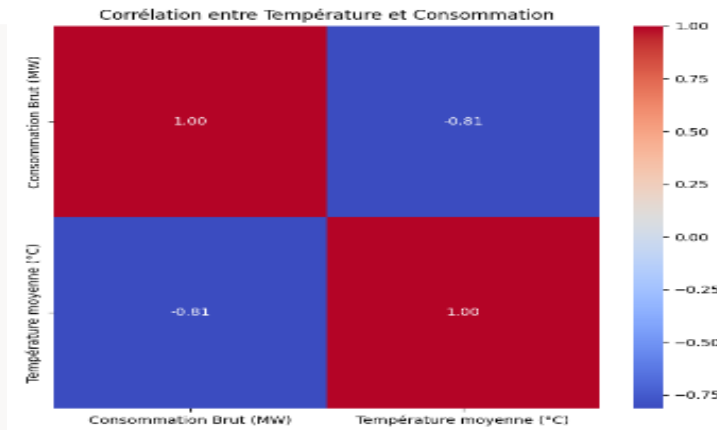
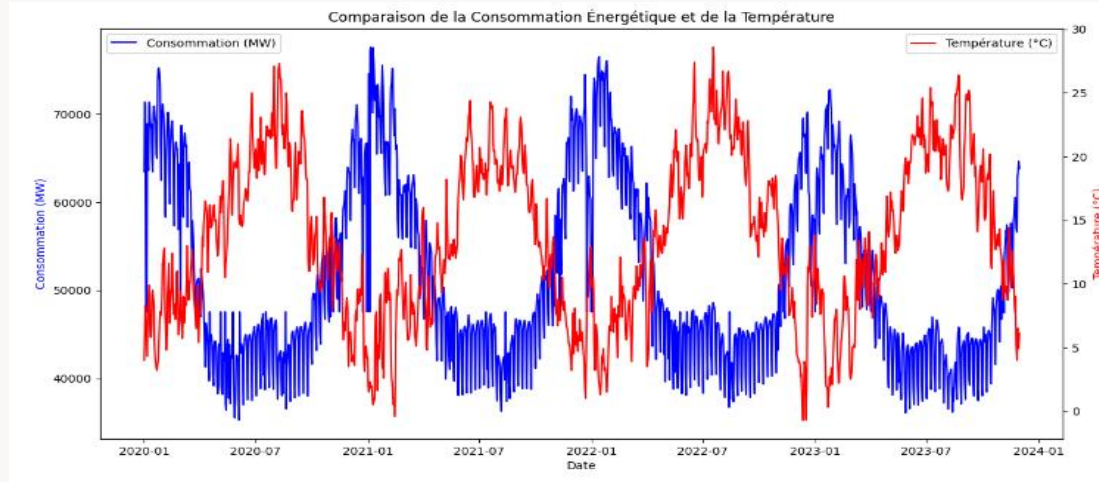
## Contexte Climatique:

La courbe violette décrit l'évolution de la température moyenne, mettant en perspective les besoins énergétiques face aux changements climatiques.

Ces analyses graphiques éclairent sur les besoins en énergie et aident à ajuster la gestion des ressources pour une efficacité maximale.



# Synthèse des Indicateurs de Consommation et Température

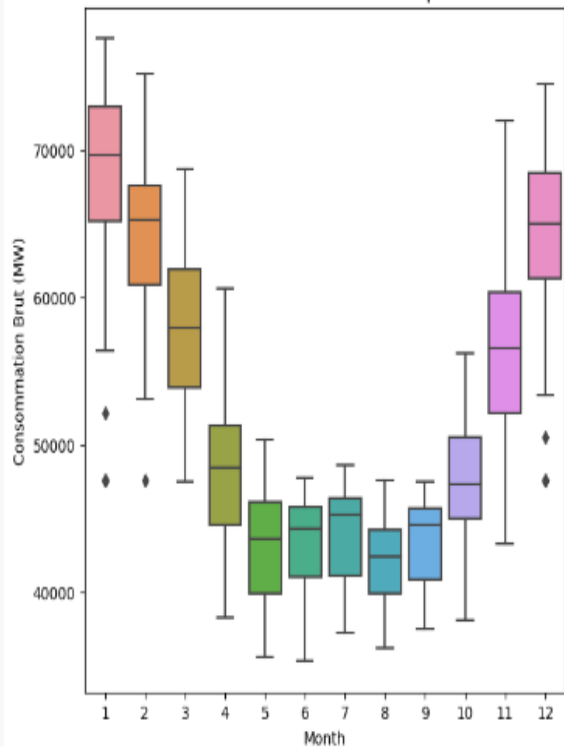


## Description explicative

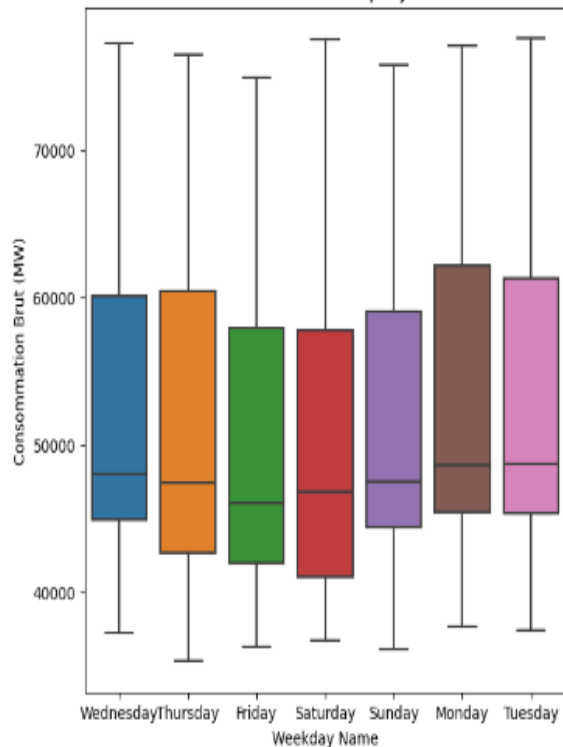
La visualisation révèle un lien clair entre la consommation d'énergie et la température : plus il fait froid, plus la demande énergétique pour le chauffage augmente. La corrélation négative significative démontrée par la matrice renseigne sur l'impact direct du climat sur la consommation énergétique.

# Analyse Saisonnière & Hebdomadaire de la Consommation Énergétique

Distribution de la Consommation par Mois



Distribution de la Consommation par Jour de la Semaine



## VARIABLES RETENUES

Cette visualisation dévoile les tendances de consommation énergétique mensuelles et hebdomadaires, mettant en évidence des consommations hivernales élevées contre des basses estivales et une stabilité journalière, ce qui est essentiel pour optimiser la planification énergétique

# 04



## Modélisation

# Affinement et Prédiction dans la Modélisation Énergétique

## Validation Statistique & Stationnarité

Avant de plonger dans la modélisation, une évaluation approfondie de la stationnarité de la série temporelle est effectuée, utilisant des outils visuels et des tests statistiques pour garantir l'adéquation des modèles

## Inspection de l'Auto-Corrélation

Des fonctions d'auto-corrélation et de corrélation partielle sont analysées pour identifier les liens temporels internes, orientant la sélection des paramètres AR et MA des modèles ARIMA ou SARIMA



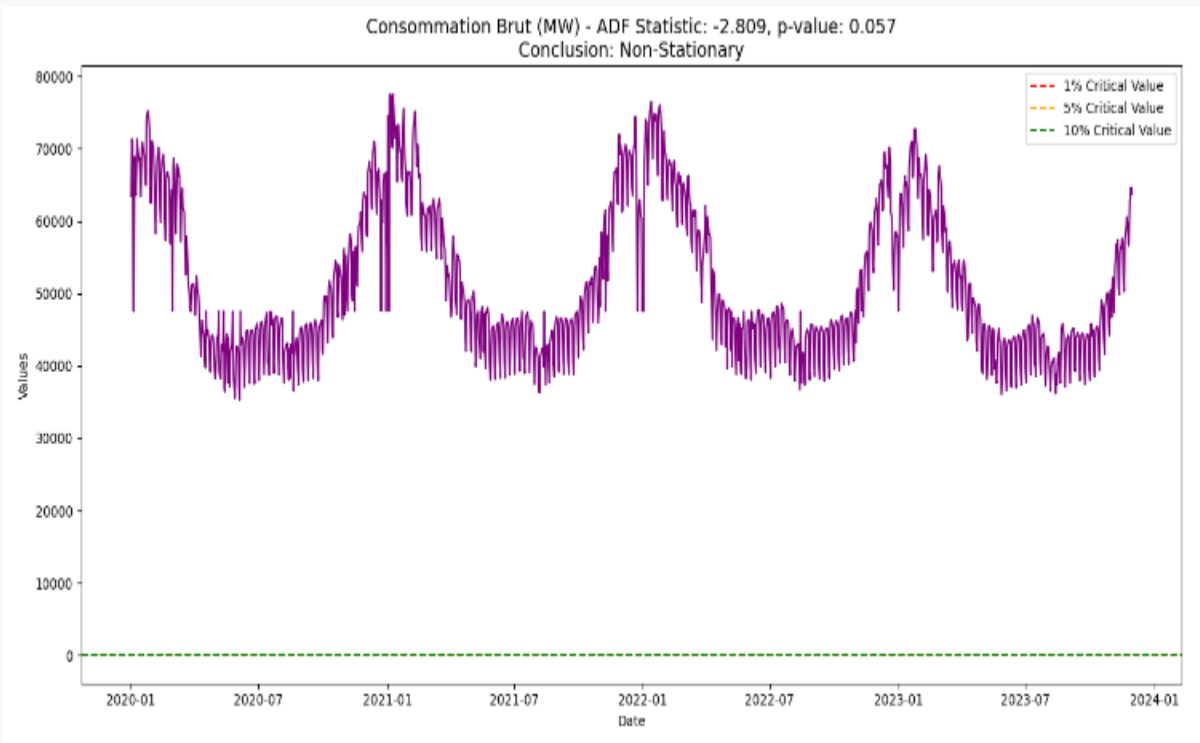
## Sélection et Test de Modèles

Une gamme de modèles, des linéaires aux avancés comme le boosting par gradient ou les LSTM, sont testés et évalués via des métriques de performance pour sélectionner le meilleur prédicteur

## Validation Croisée & Ajustement Fin

Les modèles potentiels sont soumis à une validation croisée rigoureuse pour assurer leur fiabilité, avec une adaptation minutieuse aux données historiques pour une prédiction précise et applicable

# Stabilité Temporelle de la Consommation Énergétique

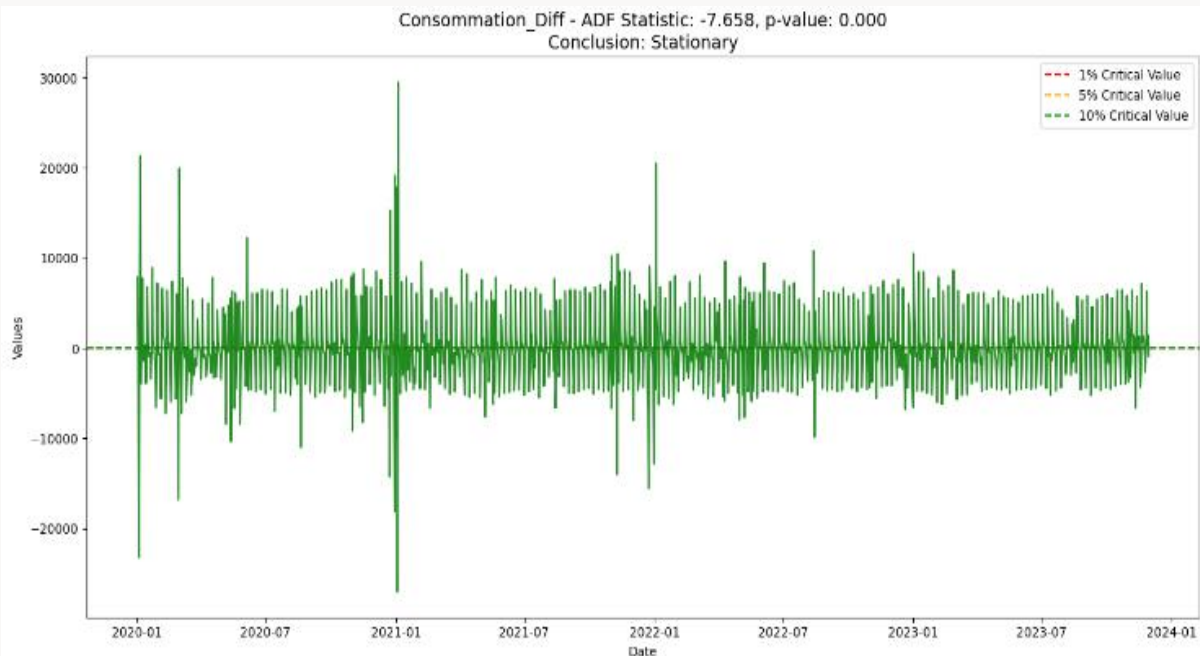


## Validation Statistique & Stationnarité

Le Test de Dickey-Fuller augmenté révèle une non-stationnarité de la consommation énergétique brute (valeur-p de 0.057). L'hypothèse nulle ( $H_0$ ) suggérant la présence d'une racine unitaire n'est pas rejetée, indiquant la nécessité d'appliquer des transformations, telles que la différenciation saisonnière, pour atteindre la stationnarité requise pour une modélisation prédictive efficace



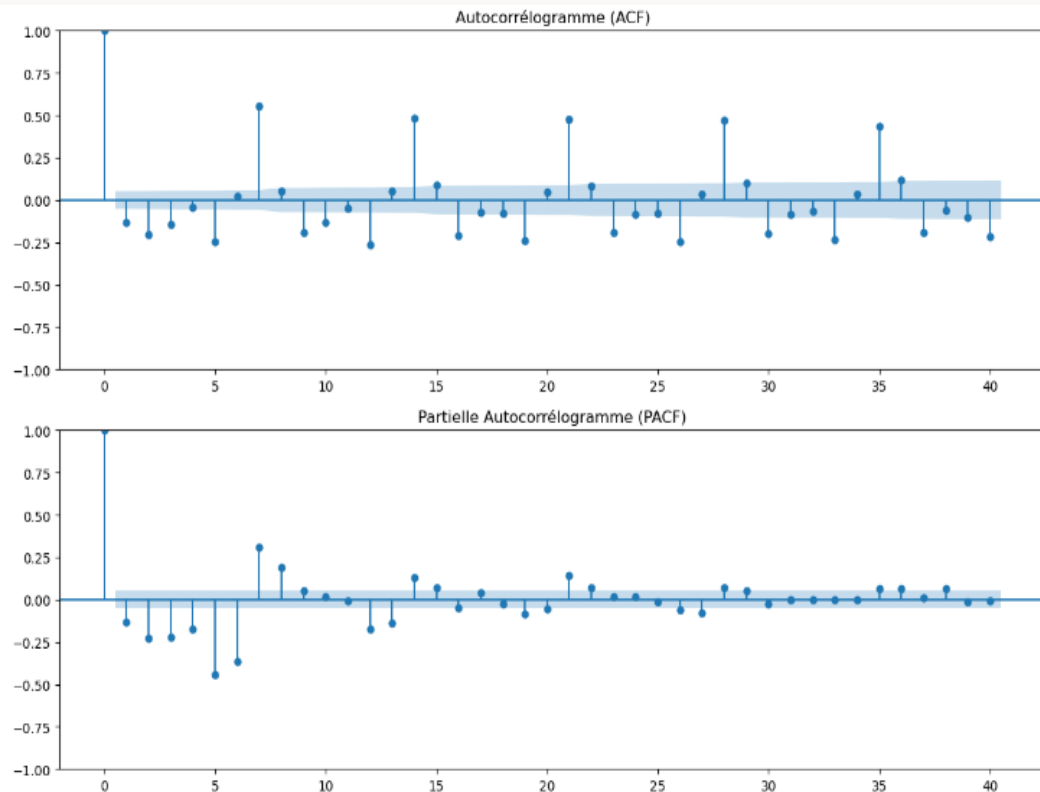
# Confirmation de Stationnarité Post-Différenciation



## Validation Statistique & Stationnarité

La série temporelle ajustée démontre désormais une stationnarité affirmée, avec une valeur-p de 0.000 au test ADF. Les variations centrées autour de zéro valident l'absence de tendance, préparant le terrain pour des modèles prédictifs précis

# Analyse de l'Auto-Corrélation Post-Différenciation



## Inspection de l'Auto-Corrélation

L'analyse ACF et PACF montre des autocorrélations significatives indiquant la nécessité d'un modèle ARIMA. L'outil `auto_arima` guidera la sélection des termes AR et MA adéquats

# Optimisation du Modèle Prédicatif: Entraînement, Validation et Test

**Train** 70% du  
jeu de données

Du **01-01-2020** Au **31-12-2022**

## Entraînement avec Auto ARIMA

Le modèle `auto_arma` a été appliqué sur **70%** des données pour déterminer les **paramètres optimaux**, en vue d'une adéquation parfaite avec les tendances observées



**Test** 30% du jeu  
de données

Du **01-01-2023** Au **30-11-2023**

## Validation des Paramètres du Modèle

Les paramètres sélectionnés sont testés sur le même **ensemble d'entraînement** pour affiner le modèle et s'assurer de sa précision avant la phase de test..



## Évaluation des Performances

Le modèle est ensuite évalué sur les 30% de données de test, où les métriques RMSE et R2 mesurent l'efficacité des prévisions par rapport aux données réelles

# Synthèse du Modèle ARIMA pour la Prédiction Énergétique

## Modèle ARIMA (5,0,5)

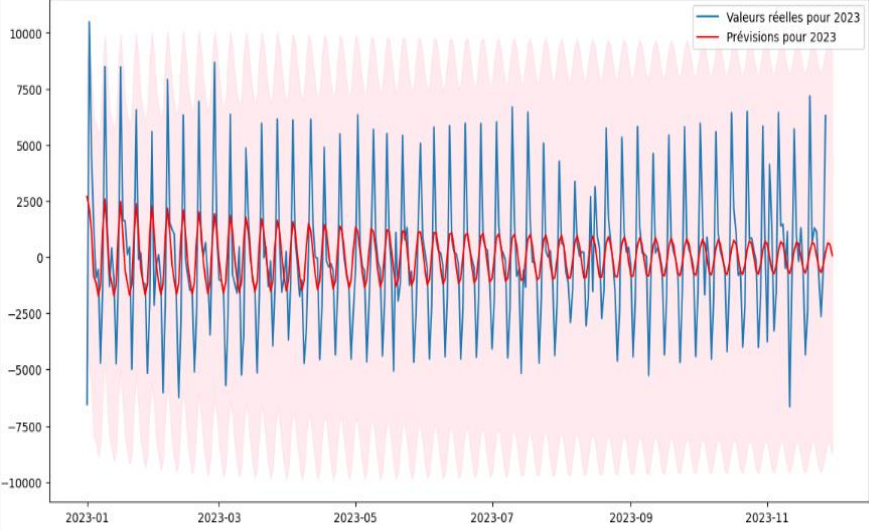
### SARIMAX Results

Dep. Variable:	Consommation_Diff	No. Observations:	1088			
Model:	SARIMAX(5, 0, 5)	Log Likelihood	-10132.348			
Date:	Thu, 11 Jan 2024	AIC	20286.696			
Time:	22:25:17	BIC	20341.548			
Sample:	01-09-2020	HQIC	20307.463			
	- 12-31-2022					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.4238	0.032	13.067	0.000	0.360	0.487
ar.L2	-1.1062	0.029	-38.122	0.000	-1.163	-1.049
ar.L3	0.2350	0.047	4.993	0.000	0.143	0.327
ar.L4	-0.6703	0.029	-22.942	0.000	-0.728	-0.613
ar.L5	-0.3694	0.033	-11.343	0.000	-0.433	-0.306
ma.L1	-0.9731	0.043	-22.703	0.000	-1.057	-0.889
ma.L2	1.3790	0.052	26.759	0.000	1.278	1.480
ma.L3	-0.9488	0.069	-13.764	0.000	-1.084	-0.814
ma.L4	0.9596	0.048	19.952	0.000	0.865	1.054
ma.L5	-0.2559	0.042	-6.105	0.000	-0.338	-0.174
sigma2	1.003e+07	3.95e-09	2.54e+15	0.000	1e+07	1e+07
Ljung-Box (L1) (Q):		0.05	Jarque-Bera (JB):			3587.28
Prob(Q):		0.83	Prob(JB):			0.00
Heteroskedasticity (H):		0.45	Skew:			-0.33
Prob(H) (two-sided):		0.00	Kurtosis:			11.90

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_5 y_{t-5} + \theta_1 \varepsilon_{t-1} + \dots + \theta_5 \varepsilon_{t-5} + \varepsilon_t$$

## Prévision du Modèle ARIMA (5,0,5)

Comparaison des Prévisions et des Valeurs Réelles pour 2023



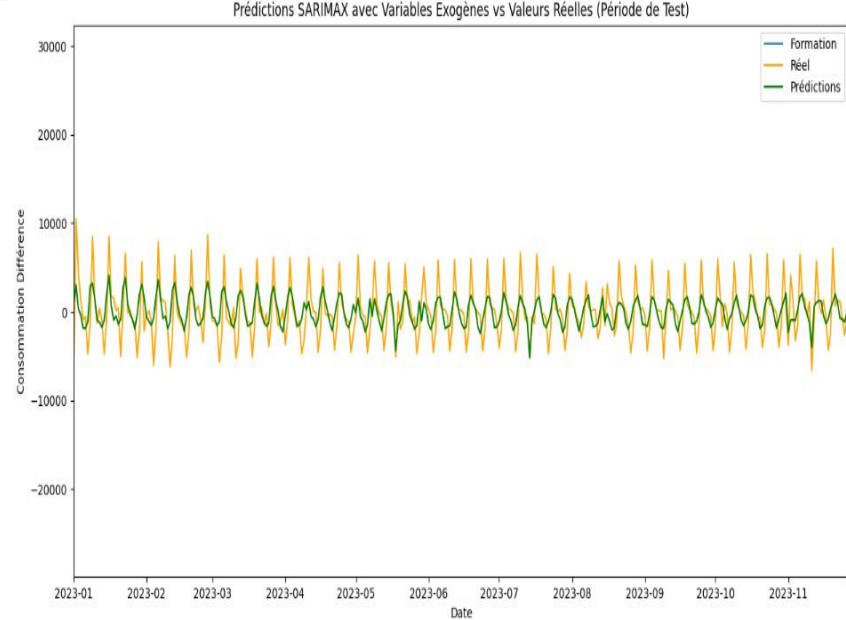
Le modèle **ARIMA(5,0,5)** révèle la complexité et les interdépendances de la consommation énergétique, avec des résultats encourageants mais perfectibles, suggérant l'exploration de modèles supplémentaires pour améliorer la précision des prévisions. Il affiche un **RMSE de 2659.52** et un **R<sup>2</sup> de 0.2791**, révélant des écarts entre prévisions et réalité. La poursuite de l'analyse avec le modèle SARIMAX, intégrant des variables exogènes, vise à affiner ces prédictions.

# Synthèse du Modèle SARIMAX pour la Prédiction Énergétique

## Modèle SARIMAX (6,05) x (2, 1, 1, 12)

SARIMAX Results						
=====						
Dep. Variable:	Consommation_Diff		No. Observations:		1088	
Model:	SARIMAX(6, 0, 5)x(2, 1, [1], 12)		Log Likelihood		-9875.211	
Date:	Thu, 11 Jan 2024		AIC		19784.421	
Time:	23:14:19		BIC		19868.618	
Sample:	01-09-2020		HQIC		19816.352	
	- 12-31-2022					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]
Température moyenne (°C)	-35.5111	13.840	-2.566	0.010	-62.638	-8.384
IsHoliday	-2731.6427	599.627	-4.556	0.000	-3906.890	-1556.395
ar.L1	-0.7063	0.056	-12.512	0.000	-0.817	-0.596
ar.L2	-0.7693	0.028	-27.951	0.000	-0.823	-0.715
ar.L3	-0.9157	0.055	-16.771	0.000	-1.023	-0.809
ar.L4	-0.6069	0.030	-20.270	0.000	-0.666	-0.548
ar.L5	-1.0257	0.037	-28.101	0.000	-1.097	-0.954
ar.L6	-0.5595	0.031	-18.113	0.000	-0.620	-0.499
ma.L1	0.2364	0.071	3.349	0.001	0.098	0.375
ma.L2	0.5345	0.057	9.394	0.000	0.423	0.646
ma.L3	0.5474	0.081	6.769	0.000	0.389	0.706
ma.L4	0.2783	0.057	4.900	0.000	0.167	0.390
ma.L5	0.7268	0.055	13.292	0.000	0.620	0.834
ar.S.L12	0.0277	0.056	0.493	0.622	-0.082	0.138
ar.S.L24	0.0105	0.070	0.151	0.880	-0.126	0.147
ma.S.L12	-0.9582	0.030	-31.467	0.000	-1.018	-0.898
sigma2	1.397e+07	0.102	1.37e+08	0.000	1.4e+07	1.4e+07
=====						
Ljung-Box (L1) (Q):	0.17	Jarque-Bera (JB):		3263.23		
Prob(Q):	0.68	Prob(JB):		0.00		
Heteroskedasticity (H):	0.35	Skew:		-0.23		
Prob(H) (two-sided):	0.00	Kurtosis:		11.64		
=====						

## Modèle SARIMAX (6,05) x (2, 1, 1, 12)



$$Consommation\_Diff_t = c + \sum_{i=1}^6 \phi_i Consommation\_Diff_{t-i} + \sum_{i=1}^5 \theta_i \varepsilon_{t-i} + \varepsilon_t + \text{saisonnalité} + \text{variables exo.}$$

Le modèle SARIMAX ajusté avec des variables exogènes telles que la température et les jours fériés réduit le **RMSE à 2533.81** et augmente le **R<sup>2</sup> à 0.3456**, montrant un modèle un peu plus précis. Les prochaines étapes incluent l'exploration de modèles de machine learning avancés tels que XGBoost et LSTM pour une précision accrue



# Modèles d'Apprentissage Automatique – Optimisation des Prévisions par Ingénierie des Caractéristiques

L'ajout de variables retardées et de fenêtres glissantes affine la précision des modèles prédictifs en capturant les tendances et la saisonnalité inhérentes à la consommation énergétique

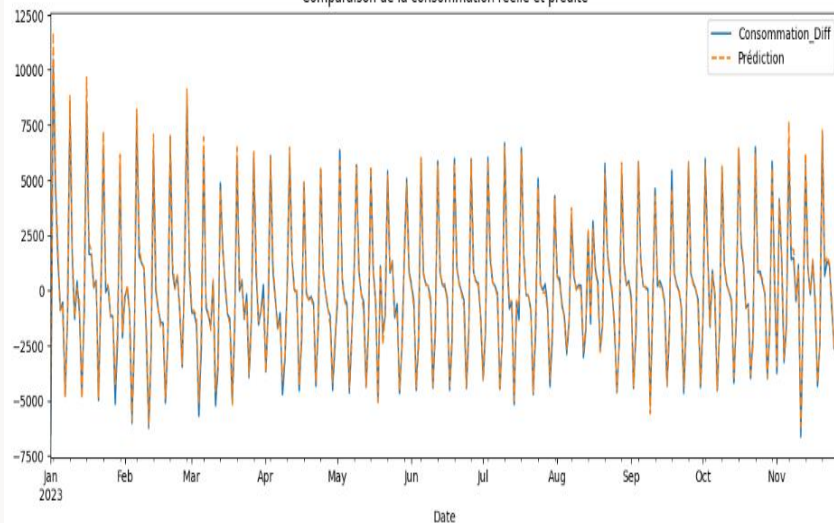


# Précision Avancée avec XGBoost en Prédiction Énergétique"

## Prédiction XGBOOST

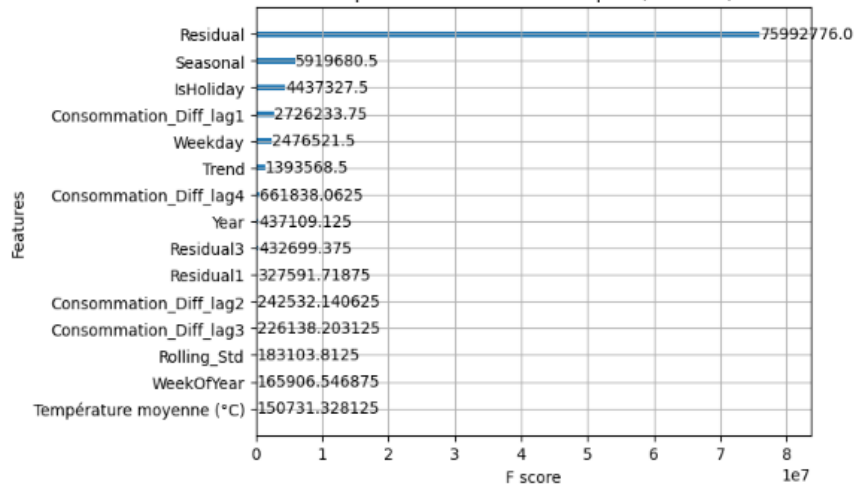
RMSE du meilleur modèle: 292.0922031654923

Comparaison de la consommation réelle et prédite



## Les variables importantes

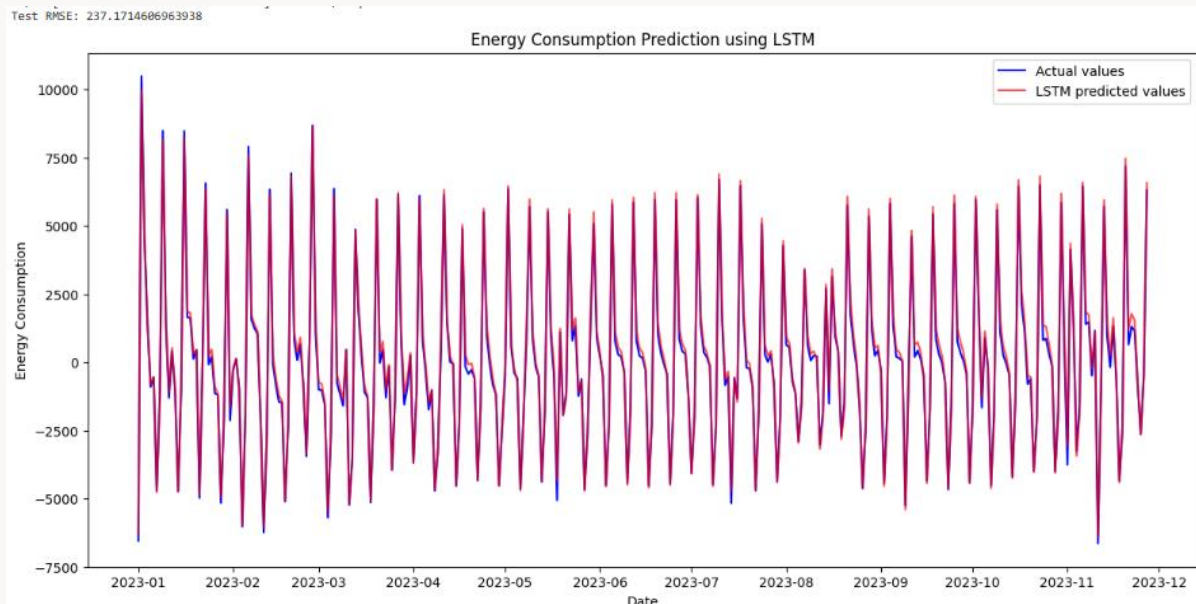
Importance des Caractéristiques (XGBoost)



Le modèle **XGBoost** a nettement amélioré la précision des prévisions avec un **RMSE de 292**. Les composantes saisonnières et résiduelles ainsi que les jours fériés ressortent comme des indicateurs clés, optimisant la prédiction de la consommation énergétique

# Optimisation Prédicative avec LSTM pour la Consommation Énergétique

## Prévision Modèle LSTM



L'analyse avec LSTM confirme son efficacité avec un RMSE de 237, offrant des prédictions de consommation énergétique fiables. Ces résultats prometteurs renforcent notre stratégie d'intégration de techniques avancées d'apprentissage automatique pour affiner la gestion énergétique

# Conclusion

## Prévision de la Consommation d'Énergie Électrique

Notre exploration des modèles prédictifs a amélioré la précision, avec XGBoost et LSTM atteignant respectivement des RMSE de 292 et MSE de 290. L'adoption d'une surveillance en continu des données énergétiques promet une gestion réactive et adaptative. L'avenir se tournera vers l'approfondissement des méthodes d'apprentissage automatique pour une gestion énergétique optimisée et écoresponsable.

