



# Projet de Prédiction du Prix de l'Électricité France- Danemark



[Cliquez ici pour accéder  
au Dashboard](#)



## Groupe 5

- **Franck Fambou**
- **Koffi Adjai**
- **Charlotte Martineau**
- **Djouhra Ould-younes**
- **Youssef Sidk**

## Table des matières

I.	Cadrage du projet et objectifs .....	2
II.	Jeux de données et enrichissement .....	2
III.	Contraintes éthiques et techniques .....	3
	Éthiques : .....	3
	Problématiques techniques : .....	3
	Limites rencontrées : .....	4
IV.	Nettoyage et transformation des données .....	5
	1. Collecte et préparation .....	5
	2. Feature Engineering .....	5
V.	Modélisations .....	5
	1. France (FR) .....	5
	2. Danemark (DK) .....	7
VI.	Évaluation et interprétabilité .....	9
	SHAP values .....	9
	• 2017-2019 Danemark LGBM : .....	10
VII.	Analyse comparative .....	10
	1. Danemark : un système éolien à forte volatilité .....	10
	2. France : un socle nucléaire sous tension .....	12
	3. Synthèse comparative France vs Danemark .....	13
	LES ANNEXES .....	14
	Quelques Graphiques sur l'EDA France .....	14
	Quelques Graphiques sur l'EDA Danemark .....	16

## I. Cadrage du projet et objectifs

Ce projet a pour but de prédire le prix de l'électricité day-ahead en Europe.

Dans le contexte de transition énergétique, la volatilité des prix de l'électricité constitue un enjeu majeur pour les opérateurs, gouvernements et consommateurs. Les fluctuations dépendent de facteurs multiples : production renouvelable, conditions météorologiques, niveau de consommation, disponibilité du réseau et coût des matières premières. La France et le Danemark représentent deux modèles énergétiques différents (nucléaire et éolien), rendant leur comparaison particulièrement pertinente. Nous pourrions ainsi observer quels sont les facteurs (features) qui comptent le plus pour la prédiction du prix.

L'objectif de ce projet est de prédire le **prix journalier de l'électricité day-ahead (€/MWh)** pour la France et le Danemark à partir de variables explicatives issues de la production, de la consommation et des conditions météorologiques, pour comparer les performances prédictives et identifier les variables influentes spécifiques à chaque pays.

Ce projet pourrait répondre à différents acteurs, et différents besoins :

- **Opérateurs énergétiques** : anticiper les prix pour optimiser leurs stratégies d'achat/vente
- **Gestionnaires de réseau** : prévoir les tensions sur le marché
- **Décideurs politiques** : comprendre les facteurs de volatilité pour orienter les politiques énergétiques
- **Consommateurs industriels** : planifier leur consommation en fonction des prix

Ce modèle prédictif robuste capable d'estimer les prix day-ahead avec une précision mesurée (comme la RMSE et la MAE qui sont des métriques adaptées aux modèles de régression)

- Dashboard interactif permettant de visualiser prédictions vs. Valeurs réelles par pays et période
- Identification des variables les plus influentes dans les variations de prix pour chaque pays

Insights sur les différences structurelles entre mix énergétiques français et danois

## II. Jeux de données et enrichissement

**Source principe** : [Open Power System Data - Time Series](#)

*Variables disponibles pour France (FR) :*

- Prix day-ahead (€/MWh) → **variable cible**
- Consommation électrique (load\_actual, load\_forecast)
- Production solaire (solar\_generation\_actual)
- Production éolienne terrestre (wind\_onshore\_generation\_actual)

*Variables disponibles pour Danemark (DK) :*

- Prix day-ahead pour DK1 et DK2 (bidding zones)
- Consommation électrique totale et par zone
- Capacités et production éolienne (onshore + offshore)
- Capacités et production solaire
- Profils de production (wind\_profile, solar\_profile)

*Caractéristiques techniques :*

- **Format** : CSV
- **Granularité temporelle** : 15 min, 30 min, 60 min (nécessite resampling journalier)
- **Période couverte** : 2015-2020 (version 2020-10-06)
- **Source** : ENTSO-E Transparency Platform

*Données complémentaires :*

Dans l'objectif d'augmenter la précision du modèle, nous avons décidé d'enrichir les données avec d'autres sources de données complémentaires.

- **Production nucléaire française** (disponibilité des centrales). Il était nécessaire de rajouter les informations nucléaires pour la France, car la production électrique française repose principalement sur le nucléaire, qui fournit autour de deux tiers de la production d'électricité ces dernières années, avec un recours plus limité au gaz et au charbon. Cela crée un profil de prix dominé par la disponibilité du parc nucléaire, la demande et quelques aléas (maintenance, vagues de froid), mais moins directement par la variabilité météo à court terme
- Données météorologiques (vitesse du vent, couverture nuageuse). Contrairement à la France, le Danemark est l'un des pays avec la plus forte part d'éolien au monde, avec près de 60% de l'électricité générée par le vent en 2023. Les prix y sont beaucoup plus sensibles aux épisodes de vent fort (prix très bas voire négatifs) et aux périodes anticycloniques sans vent (pics de prix), ce qui en fait un cas d'école pour l'impact des renouvelables variables sur le marché. Nous avons donc choisi d'enrichir les données avec des variables liées.

*Sources*: RTE/ODRE, Energinet, World Bank Commodity Prices, ENTSO-E

### III. Contraintes éthiques et techniques

**Éthiques :**

- Transparence sur l'utilisation de l'IA (conformité règles Epitech)
- Citation rigoureuse des sources de données

**Problématiques techniques :**

Nous avons rencontré quelques problématiques lors de l'enrichissement des données.

### *Données manquantes (NaN) :*

Les données sur le prix de l'électricité sont manquantes pour la France sur la période 2017/12/05 - 2020/09/30 et sur la Danemark, manquant sur l'année 2020 (période d'analyse 2017-2020 finalement). Nous n'avons pas réussi à les compléter même en passant par **l'API ENTSO-E**.

Notre première hypothèse était de tenter de compléter ces données avec un modèle de machine learning, mais il nous est apparu qu'il était impossible d'évaluer la performance du modèle sans avoir de données de comparaison. Nous avons donc décidé de faire deux modèles, un entraîné sur la période 2015-2017 et l'autre sur la période 2020-2025.

### *Résolutions temporelles hétérogènes :*

Afin d'établir un référentiel temporel commun aux deux pays, nous avons standardisé l'ensemble des données en conservant les horaires universelles (UTC) ou détrimement horaires locaux (CET).

### *Zones multiples :*

Pour le Danemark, les deux zones DK1 et DK2 ont été modélisées séparément afin de respecter la structure réelle du marché. Les données étant disponibles à un niveau local pour chaque zone de prix, un modèle dédié a été entraîné pour DK1 et un autre pour DK2. Cette approche par zones multiples améliore la précision globale : les modèles capturent mieux les spécificités de consommation, de mix de production et de volatilité propres à chaque zone que dans un schéma fusionné unique.

### *Colonnes différentes :*

Les colonnes disponibles varient fortement selon les pays, ce qui impose une modélisation différenciée. Dans le cas du Danemark, il n'existe pas de production nucléaire dans le mix électrique, et une grande partie de la génération provient de l'éolien, en particulier offshore. Les variables explicatives dominantes pour DK1/DK2 sont donc l'éolien (onshore/offshore), la consommation et la météo, alors que dans d'autres pays les colonnes liées au nucléaire ou à d'autres technologies thermiques jouent un rôle central.

### **Limites rencontrées :**

- *Variables météo* : nous avons enrichi les données avec des variables sur la production solaire/éolienne, et température.
- *Événements géopolitiques non capturés*, comme les chocs géopolitiques majeurs comme la crise du COVID-19 ou la guerre en Ukraine. Ces événements ont provoqué des ruptures fortes dans les séries de prix (confinements, effondrement puis rebond de la demande, tensions sur le gaz, plafonnements, interventions publiques) que le modèle ne peut pas anticiper à partir des seules variables techniques. Le modèle reste aveugle aux décisions politiques exceptionnelles (boucliers tarifaires, plafonds de prix, suspensions de centrales) et aux chocs exogènes globaux.
- *Granularité journalière* : Le passage à une granularité journalière simplifie les données et permet de se concentrer sur la tendance globale des prix, mais il masque complètement la dynamique intra-journalière comme les pointes du matin et du soir, les prix négatifs en milieu

de journée lors de forts épisodes solaires ou les épisodes très courts de tension ne sont plus visibles à ce niveau. Cela convient cependant avec notre objectif de prédire le prix par jour.

- *Différences structurelles France/Danemark* : il ne s'agit pas tant d'une limite que d'une observation. Les systèmes électriques français et danois reposent sur des architectures et des mix très différents (nucléaire pilotable vs éolien dominant, interconnexions, régulation, structure des coûts). Les modèles sont donc entraînés sur des distributions de données différentes, avec des variables clés qui n'ont ni la même importance ni la même signification dans chaque pays. La comparaison France/Danemark sert à analyser des différences de comportements (volatilité, rôle du vent vs nucléaire), pas à construire un modèle généralisable à toute l'Europe. Les résultats sont donc valables dans le cadre de ces deux systèmes et de la période étudiée, mais ne garantissent pas une transférabilité à d'autres pays ou à des configurations futures très différentes.

## IV. Nettoyage et transformation des données

### 1. Collecte et préparation

- Téléchargement des datasets CSV (60 min resolution) (Data originel + Data enrichie via API ENTSO-E e)
- Extraction des variables France (FR\_) et Danemark (DK\_1\_, DK\_2\_)
- Nettoyage : suppression des lignes de valeurs manquantes, renommage des colonnes, détection outliers
- Resampling journalier (moyennes/sommes selon variables)

### 2. Feature Engineering

- *Variables temporelles* : jour, semaine, mois, saison, jours fériés
- *Lags (décalage)* : prix/consommation/production J-1, J-7, J-30
- *Moyennes mobiles* : 7j, 14j, 30j sur production et consommation
- *Ratios* : production renouvelable/consommation, solaire/éolien
- *Variables météo proxy* : solar\_profile, wind\_profile (Danemark)
- *Interactions* : production éolienne × saison, consommation × température proxy

## V. Modélisations

### 1. France (FR)

**Train/test split = split temporel** : Train (2015-2017), Test (3 mois) + train (2020-2025)

**Modèles candidats** : LightGBM & SAMRIMAX

LightGBM contrairement à SARIMAX par exemple qui est un modèle linéaire (une ligne droite + des cycles), LightGBM est un modèle à base d'arbres de décision (Gradient Boosting). Voici pourquoi c'est crucial pour notre étude sur les prix de l'électricité :

- Gestion des Effets de Seuil (Non-linéarité) :
  - Logique linéaire (SARIMAX) : **"Si la demande augmente de 1GW, le prix augmente de 10€, quel que soit le niveau actuel."**
  - Réalité physique : *Si la demande est faible, +1GW ne change rien (on utilise du nucléaire pas cher). Mais si le réseau est saturé, +1GW oblige à démarrer une centrale à gaz ou charbon très chère, faisant exploser le prix. LightGBM capture ces seuils ("Si demande > 60GW ALORS prix +++").*
- Interactions Complexes :

LightGBM détecte automatiquement des combinaisons comme : **"S'il fait froid (demande chauffage) ET qu'il n'y a pas de vent (pas d'éolien), alors le prix s'envole"**. Un modèle classique doit être programmé manuellement pour voir cette interaction.

➤ 2015-2017 :

#### SARIMAX

MAE : 4.76€/MWh RMSE : 6.93 R<sup>2</sup> : 0.65 MAPE : 0.07

#### LGBM (Base)

MAE : 0.41 RMSE: 1.21 R<sup>2</sup>: 0.995 MAPE : 0.01%

#### LGBM optimisé (GridSearch – CrossVal)

config : {'learning\_rate': 0.05, 'n\_depth': -1, 'n\_estimators': 500, 'num\_leaves': 50}

MAE: 0.16 €/MWh RMSE: 0.28 €/MWh R<sup>2</sup>: 1.00 MAPE: 0.00%

➤ 2020-2025 :

#### SARIMAX

MAE: 18.47 €/MWh RMSE: 23.08 €/MWh R<sup>2</sup>: 0.638 MAPE: 54.50 %

#### LGBM (Base)

RMSE: 2.21 MAE: 0.85 R<sup>2</sup>: 0.997 MAPE: 3.04%

#### LGBM optimisé (GridSearch – CrossVal)

Hyperparamètres optimaux : {'learning\_rate': 0.05, 'max\_depth': -1, 'n\_estimators': 200, 'num\_leaves': 100}

RMSE: 1.86 MAE: 0.61 R<sup>2</sup>: 0.998 MAPE: 2.06% Amélioration : 0.24 €/MWh

LightGBM est un gradient boosting sur arbres optimisé pour la vitesse, la gestion de grandes volumétries et de nombreuses variables. Il semble plutôt adapté à notre dataset qui contient de nombreuses données et variables.

Il gère nativement les valeurs manquantes, supporte bien les relations non linéaires et les interactions (par exemple, effet combiné “forte demande + peu de vent + prix du gaz élevé”), ce qui est crucial sur des marchés très dépendants du contexte.

LightGBM fait du gradient boosting, c'est-à-dire qu'il construit une suite d'arbres de décision, chaque nouvel arbre essayant de prédire les erreurs du modèle courant par rapport à la cible, son objectif est bien sûr de réduire au maximum la fonction coût

## 2. Danemark (DK)

### ➤ 2017-2019:

Cette étude porte sur la prédiction des prix de l'électricité ("Day-Ahead") pour les deux zones de marché du Danemark : **DK1 (Ouest - Jylland)** et **DK2 (Est - Sjælland)**. Bien que connectées, ces deux zones présentent des dynamiques distinctes : DK1 est fortement couplée au réseau allemand et très exposée à l'éolien, tandis que DK2 est historiquement plus liée au marché suédois. L'objectif est d'évaluer la robustesse des modèles prédictifs face à cette double complexité sur la période **2017-2019**.

Pour les deux zones, le LightGBM optimisé améliore nettement le modèle de base : le RMSE passe de

#### DK2

##### LGBM (Base)

RMSE :8.35 MAE :6.04  $R^2$  :39.90%

##### LGBM optimisé (GridSearch – CrossVal)

params : {'learning\_rate': 0.01,  
'max\_depth': -1, 'n\_estimators': 500,  
'num\_leaves': 31}

RMSE :7.68 MAE :5.44  $R^2$  :49.17%

##### SARIMAX

RMSE :8.51 MAE :6.36  $R^2$  :37.57%

#### DK1

##### LGBM (Base)

RMSE :9.18 MAE :6.72  $R^2$  :37.56%

##### LGBM optimisé (GridSearch – CrossVal)

params : {'learning\_rate': 0.01,  
'max\_depth': -1, 'n\_estimators': 500,  
'num\_leaves': 31}

RMSE :8.44 MAE :6.06  $R^2$  :19.39%

##### SARIMAX

RMSE :10.43 MAE :8.29  $R^2$  :19.39%

8,35 à 7,68 €/MWh pour DK2 et de 9,18 à 8,44 €/MWh pour DK1, tandis que le  $R^2$  gagne près de 10 points (jusqu'à environ 49% pour DK2 et 47% pour DK1). Le SARIMAX, lui, reste légèrement moins performant que le LightGBM optimisé, avec des RMSE autour de 8,5–10,4 €/MWh et des  $R^2$  autour de 38% pour DK2 et 19% pour DK1, ce qui confirme l'intérêt des modèles de gradient boosting pour capturer la structure non linéaire des prix, surtout dans la zone ouest plus volatile (DK1).

### ➤ 2020-2025 :

Dans cette section, l'analyse se concentre sur la modélisation des prix spot de l'électricité pour les zones DK1 (Ouest) et DK2 (Est) sur la période 2020-2025. L'objectif est d'évaluer la capacité de



différents modèles à reproduire la dynamique horaire des prix, en tenant compte à la fois de la demande, des prévisions de charge et des productions renouvelables (éolien, solaire). Les modèles retenus combinent un algorithme de gradient boosting (LightGBM) dans une version de base puis optimisée par recherche d'hyperparamètres, ainsi qu'un modèle statistique SARIMAX intégrant des variables exogènes. Les performances sont comparées sur un jeu de test tenu à l'écart de l'entraînement, à l'aide d'indicateurs standards de régression (RMSE, MAE, coefficient de détermination  $R^2$ ), afin de quantifier précisément le gain apporté par chaque approche avant d'interpréter en détail les résultats.

## DK2

### LGBM (Base)

RMSE :16.00 MAE :8.73  $R^2$  :91.32%

### LGBM optimisé (GridSearch – CrossVal)

Meilleurs hyperparamètres :  
{ 'learning\_rate': 0.1, 'max\_depth': 20,  
'n\_estimators': 300, 'num\_leaves': 50 }

RMSE :15.8 MAE :8.76  $R^2$  :91.5%

### SARIMAX(2 derniers mois)

RMSE :39.35 MAE :31.14  $R^2$  :26.8%

## DK1

### LGBM (Base)

RMSE :15.58 MAE :8.34  $R^2$  :91.3%

### LGBM optimisé (GridSearch – CrossVal)

Meilleurs hyperparamètres :  
{ 'learning\_rate': 0.05, 'max\_depth': 10,  
'n\_estimators': 300, 'num\_leaves': 70 }

RMSE :15.53 MAE :8.26  $R^2$  :91%

### SARIMAX

RMSE :35.36 MAE :26.74  $R^2$  :19%

Cette synthèse compare clairement les performances des modèles entre DK1 (droite) et DK2 (gauche) sur 2020-2025.

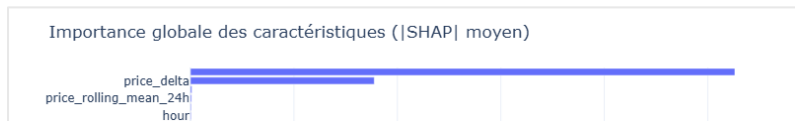
Dans les deux zones, le LightGBM de base atteint déjà un bon niveau de précision avec des RMSE autour de 16 €/MWh pour DK2 et 15,6 €/MWh pour DK1, et des  $R^2$  proches de 0,91, ce qui montre que la dynamique des prix est bien capturée. L'optimisation par GridSearch n'apporte qu'un gain marginal (RMSE à 15,8 €/MWh pour DK2 et 15,5 €/MWh pour DK1), signe que le modèle de base était déjà bien calibré sur ces séries.

À l'inverse, le SARIMAX sur les deux derniers mois reste nettement moins performant : les RMSE montent autour de 39 €/MWh pour DK2 et 35 €/MWh pour DK1, avec des  $R^2$  autour de 0,26–0,19 seulement. Cela confirme que, dans ce contexte de forte non-linéarité et de nombreux effets exogènes (vent, solaire, charge), les modèles de type gradient boosting surperforment largement une approche purement statistique ARIMA/SARIMAX

## VI. Évaluation et interprétabilité

### SHAP values

- 2015\_2017 France LGBM:



La variable `price_delta` ressort comme de loin la plus influente, ce qui signifie que

les variations récentes de prix (hausse/baisse entre deux pas de temps) expliquent une grande partie des décisions du modèle sur cette période plus « calme » que la crise 2020-2025. La moyenne glissante sur 24 h (`price_rolling_mean_24h`) et les variables temporelles comme l'heure jouent un rôle secondaire, surtout pour capter la saisonnalité intra-journalière et lisser le bruit, mais leur poids reste nettement inférieur à celui de la dynamique de court terme matérialisée par `price_delta`.

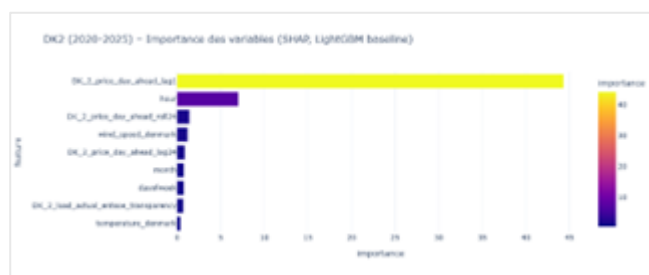
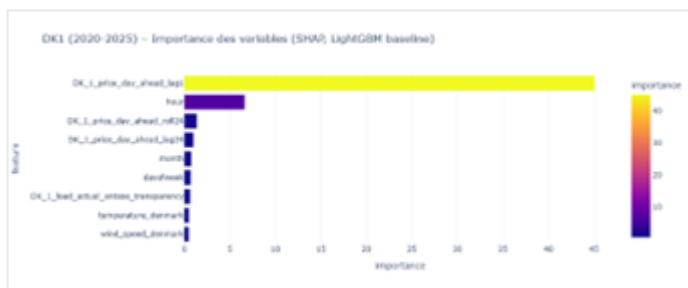
- 2020\_2025 France LGBM :



Sur cette période, `price_lag_1h` écrase toutes les autres variables : le prix à une heure donnée s'explique surtout par le prix une heure plus tôt, signe d'une forte inertie

du marché français. Les features dérivées comme `price_delta` ou les moyennes glissantes (par exemple `price_rolling_min_6h`) n'apportent qu'un signal additionnel de court terme, beaucoup moins influent que cette mémoire immédiate des prix.

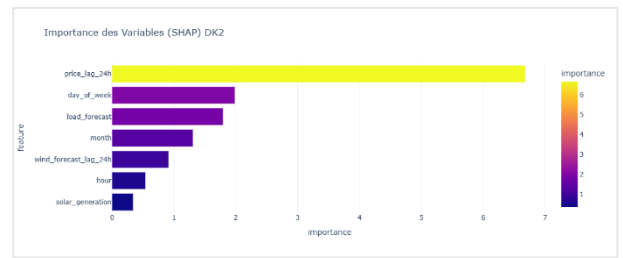
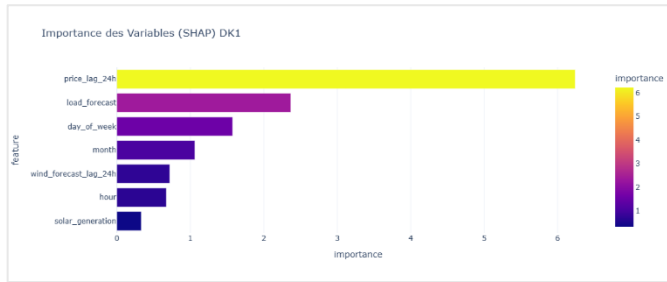
- 2020\_2025 Danemark LGBM :



Ces deux graphiques SHAP montrent que, sur 2020-2025, les modèles LightGBM de DK1 et DK2 ont une structure explicative très similaire. La variable dominante est le prix day-ahead décalé d'une heure, ce qui traduit une forte autocorrélation de court terme.

Les variables de calendrier (heure, mois, jour de semaine), les moyennes glissantes, le vent, puis la charge et la température n'apportent qu'un effet complémentaire, beaucoup plus faible, ce qui indique que l'information déjà contenue dans la série de prix reste déterminante.

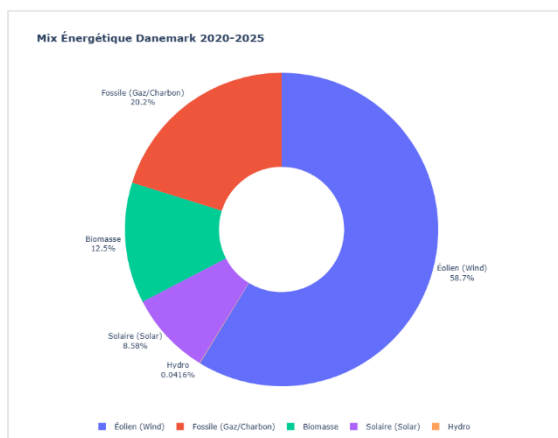
- 2017-2019 Danemark LGBM :



Dans les deux cas, le prix de la veille à la même heure (price\_lag\_24h) domine nettement, ce qui montre que les patterns journaliers structurent fortement le marché danois. La prévision de charge et les variables de calendrier (jour, mois), puis le vent, l'heure et le solaire jouent un rôle plus secondaire, essentiellement pour affiner la prévision autour de ce signal principal.

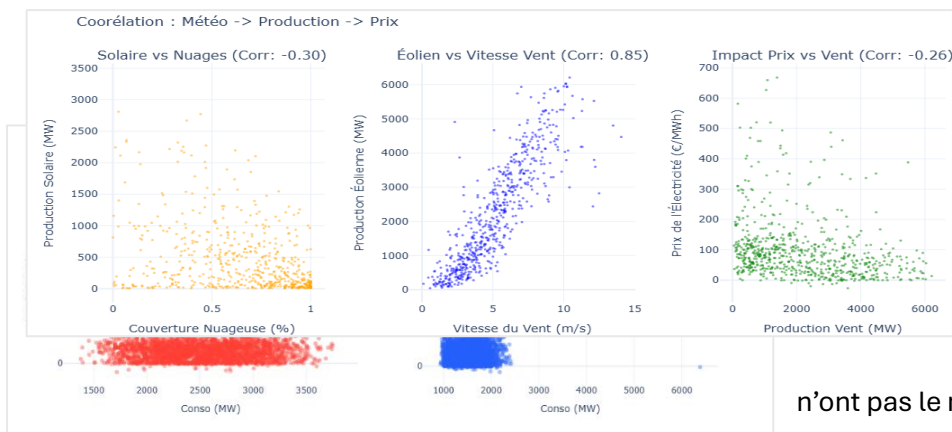
## VII. Analyse comparative

### 1. Danemark : un système éolien à forte volatilité



Sur la période 2020–2025, le mix électrique danois est dominé par l'éolien, qui représente près de 60 % de la production, devant la biomasse, le fossile et le solaire.

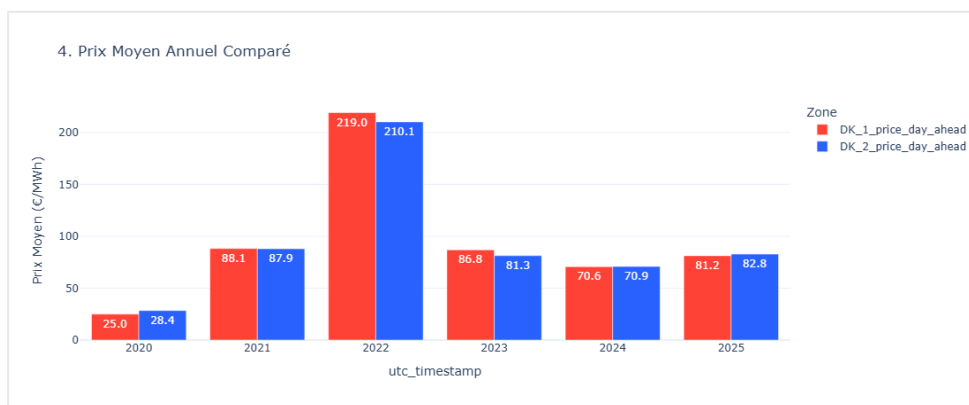
La corrélation très forte observée entre vitesse du vent et production éolienne (corrélation 0,85 sur le nuage de points « Éolien vs Vitesse de vent ») confirme que la météo, et en particulier le vent, pilote directement l'offre d'électricité au Danemark.



La distinction entre DK1 (Ouest) et DK2 (Est) met également en évidence le rôle des contraintes réseau. Les deux zones n'ont pas le même profil de prix, ce qui

reflète des capacités d'interconnexion et des profils de consommation différents.

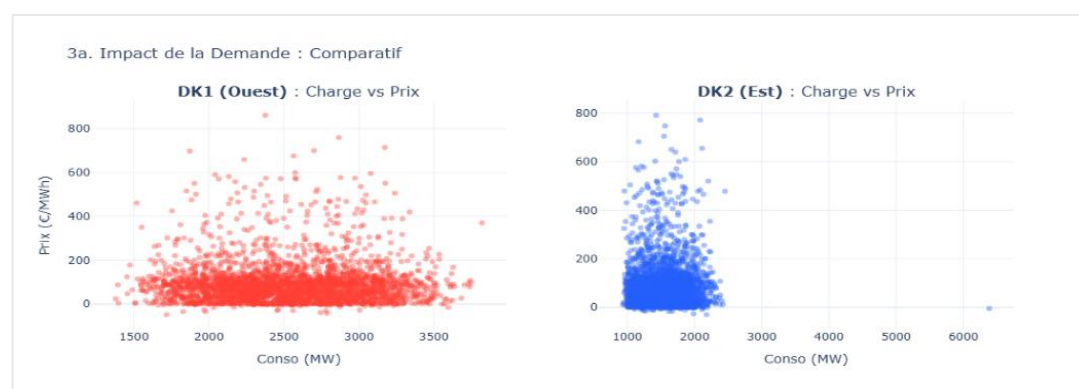
En revanche, la structure des prix et les facteurs qui les expliquent différent : DK1 est plus exposée aux épisodes de vent fort et de surproduction, tandis que DK2, plus urbaine et



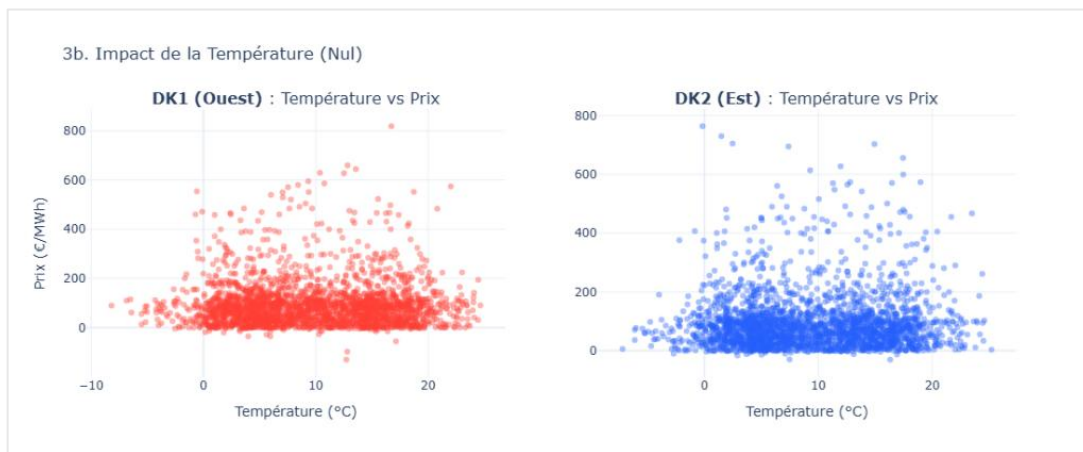
interconnectée, est davantage influencée par la consommation et la température, comme le montre la Matrice de Facteurs d'Influence (DK1/DK2)



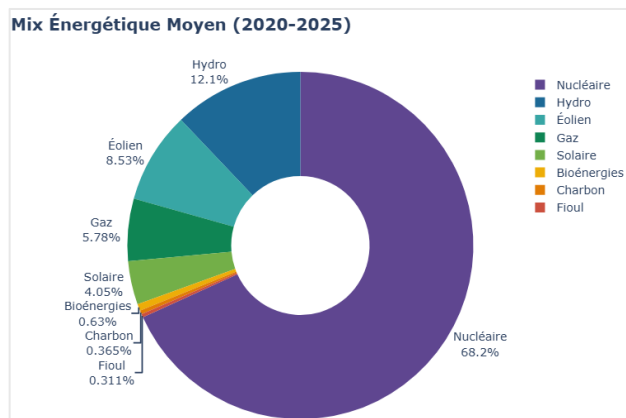
Lors des tempêtes, le surplus d'énergie de l'Ouest ne peut pas toujours être totalement évacué vers l'Est ou les pays voisins, ce qui provoque un effondrement local du prix en DK1.



En DK2, zone plus urbaine et interconnectée, la Température apparaît plus corrélée à la consommation, et les corrélations entre prix, consommation et vent sont plus modérées, signe que la demande locale et les échanges avec l'extérieur jouent un rôle plus important.

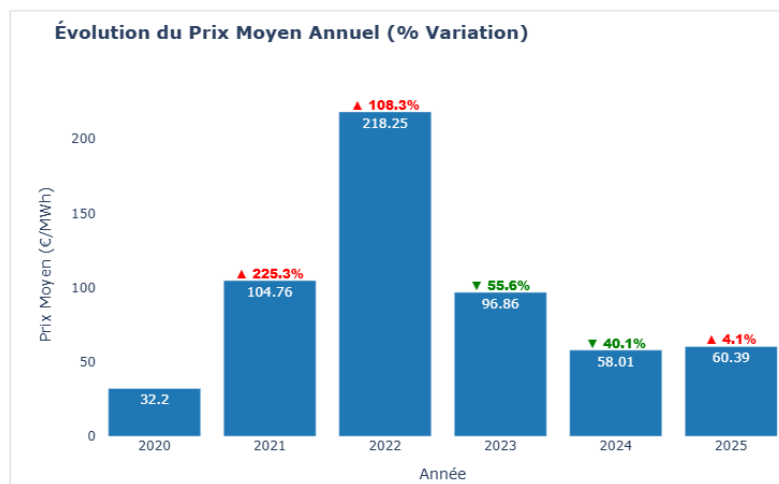


## 2. France : un socle nucléaire sous tension



En France, sur 2020–2025, le mix électrique reste largement dominé par le nucléaire, qui représente environ 68 % de la production, loin devant le gaz, l'hydro, le solaire et l'éolien (voir graphique « Mix Énergétique France 2020–2025 »). Ce socle pilotable offre une base de production relativement stable et moins directement dépendante des conditions météo instantanées que le système danois.

La distribution et la chronologie des prix day-ahead français montre parfois des prix marqués, notamment en 2022 (flambée des prix liée à la guerre en Ukraine).



Il semble y avoir une relation non linéaire entre charge et prix. Pour des niveaux de consommation faibles à modérés, le prix reste relativement stable : le parc nucléaire et les moyens de base suffisent à couvrir la demande. Au-delà de certains seuils de charge, les points s'envolent presque verticalement, signe que l'appel à des moyens plus chers (gaz, imports) provoque une forte hausse du prix marginal.

En France, la météo agit surtout via la demande (froid hivernal, canicules) et la disponibilité des moyens (hydro, contraintes de refroidissement des centrales), plus que via une production éolienne dominante. La corrélation entre température et prix reste donc plus

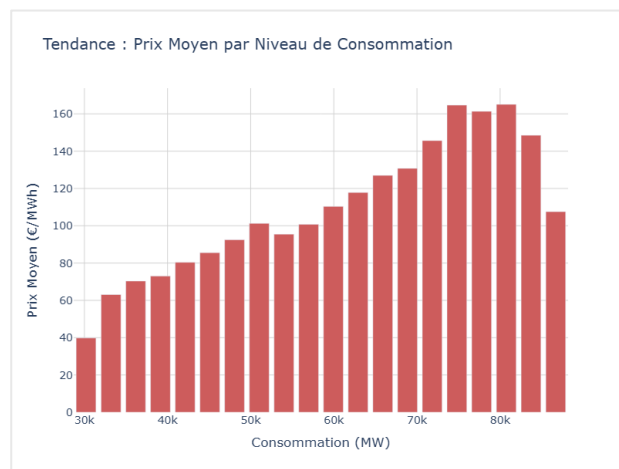
visible qu'au Danemark, comme illustré dans les graphiques « DK1 (Ouest) : Température vs Prix » et « DK2 (Est) : Température vs Prix », même si l'augmentation de la part éolienne et solaire commence à modifier cette relation classique.

### 3. Synthèse comparative France vs Danemark

La France repose sur un parc nucléaire pilotable, qui limite les prix négatifs mais expose le système à des pics élevés lorsque le gaz devient cher ou que les indisponibilités nucléaires se multiplient, tandis que le Danemark s'appuie sur un mix très éolien, dans lequel la météo pilote directement le prix et génère de nombreux épisodes de prix très bas ou négatifs lors de surproduction.

Les moteurs principaux de la volatilité ne sont pas les mêmes : en France, les chocs viennent surtout des combustibles (gaz), de la disponibilité nucléaire et des périodes de forte demande, comme on le voit sur la « Chronologie des Prix France (Moyenne Hebdo) » et la « Tendence : Prix Moyen par Niveau de Consommation ». Au Danemark, ils sont principalement liés à la variabilité du vent, aux épisodes de surproduction locale et aux limites d'export entre DK1, DK2 et les pays interconnectés, illustrés par « Distribution des Prix – Danemark (Histogramme + Boxplot) », « Éolien vs Vitesse de vent » et « Impact Prix vs Vent ».

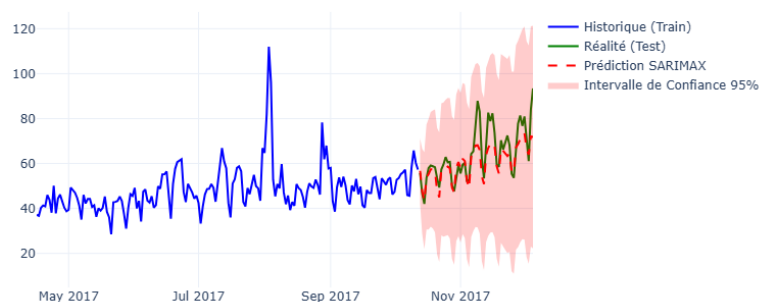
Les modèles prédictifs reflètent ces différences : les modèles français s'articulent davantage autour de la charge, du parc pilotable et du contexte gaz/nucléaire, tandis que les modèles danois doivent intégrer finement la production éolienne, les profils de vent et la granularité zonale DK1/DK2 pour capturer la dynamique de prix. Enfin, cette comparaison illustre qu'une augmentation de la part de renouvelables variables rend le système plus décarboné mais aussi plus volatil, ce qu'on observe déjà nettement dans les graphiques de prix et de mix énergétique du Danemark par rapport à la France.



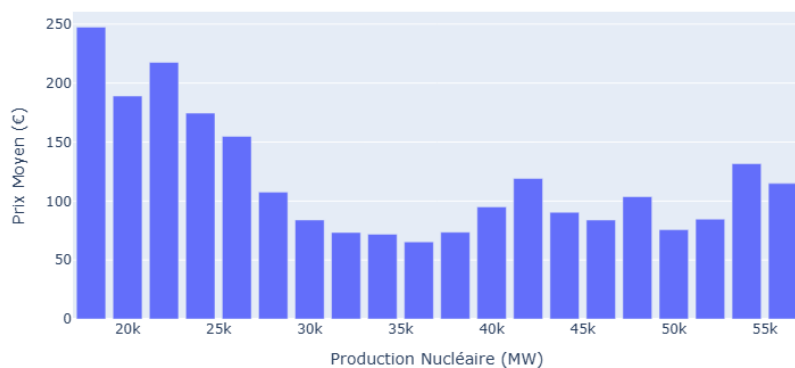
## LES ANNEXES

### Quelques Graphiques sur l'EDA France

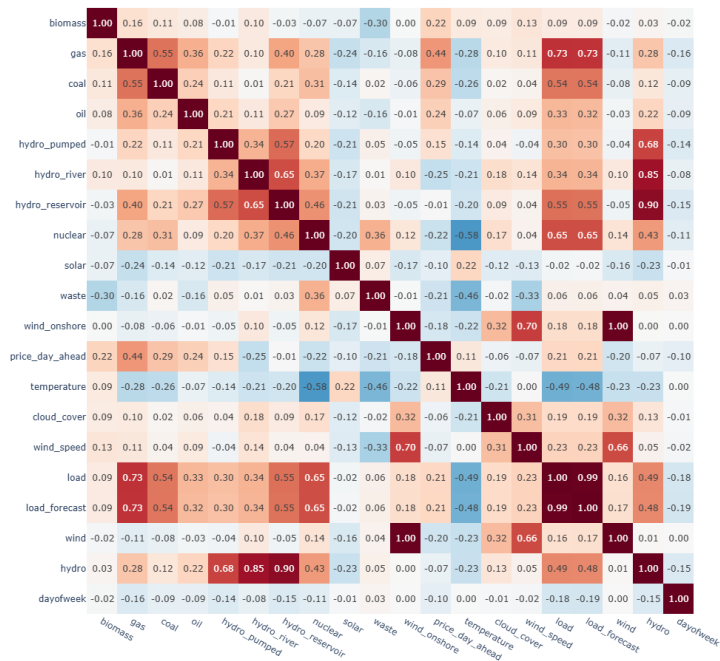
SARIMAX (MAE: 4.76 €)



#### Prix Moyen Vs production nucléaire



#### Heatmap de Corrélation



#### Top 10 Corrélations Positives:

```
gas 0.443047
coal 0.288048
oil 0.240060
biomass 0.223176
load 0.209751
load_forecast 0.208779
hydro_pumped 0.153662
temperature 0.113519
hydro_reservoir -0.006784
cloud_cover -0.057756
Name: price_day_ahead, dtype: float64
```

#### Top 10 Corrélations Négatives:

```
cloud_cover -0.057756
hydro -0.069788
wind_speed -0.073151
dayofweek -0.095181
solar -0.098045
wind_onshore -0.181319
wind -0.198403
waste -0.210292
nuclear -0.223527
hydro_river -0.245726
Name: price_day_ahead, dtype: float64
```

Figure 1 : Prix vs Charge (2015-2017)

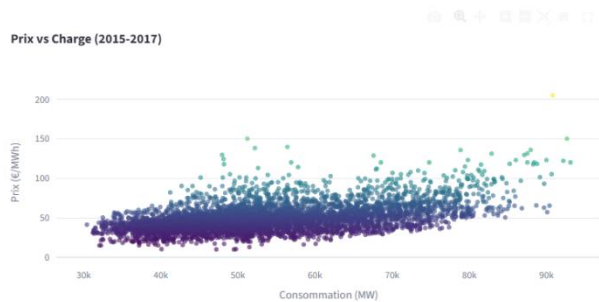
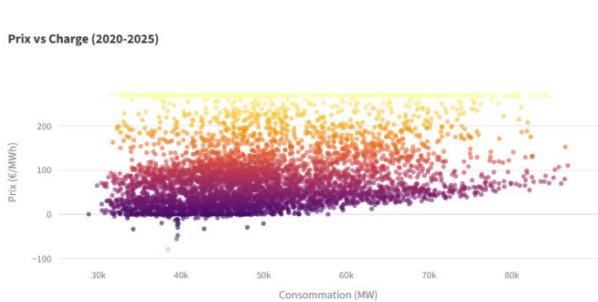
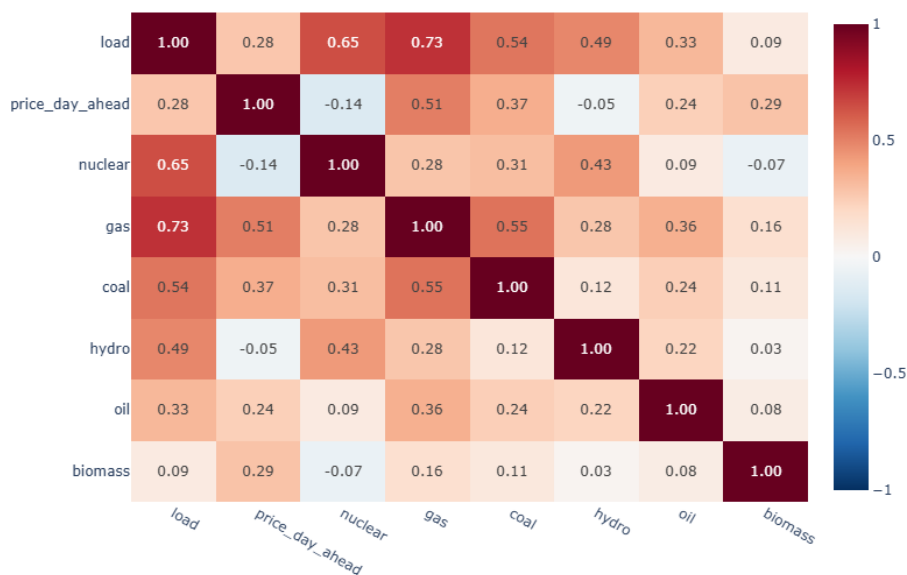


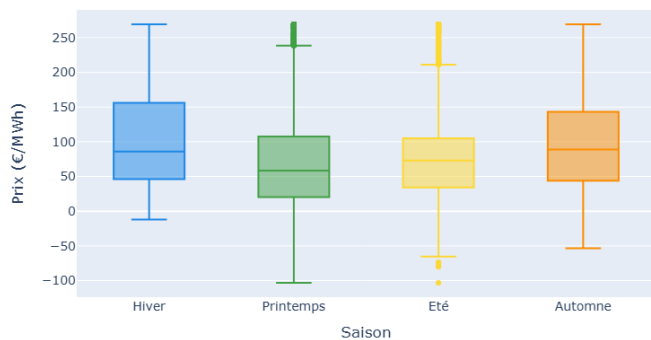
Figure 2 : Prix vs Charge (2020-2025)



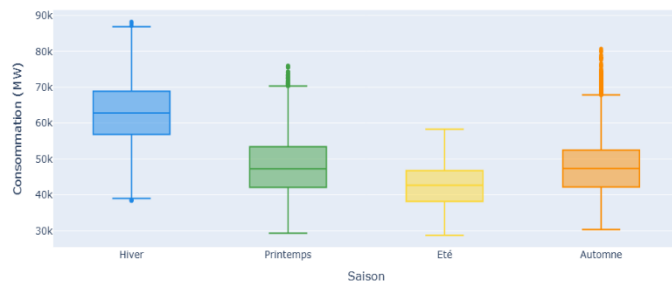
### Corrélation : Consommation, Production et Prix



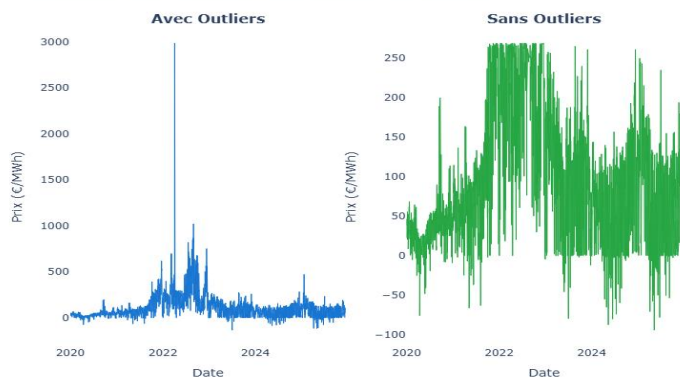
### Distribution des Prix par Saison



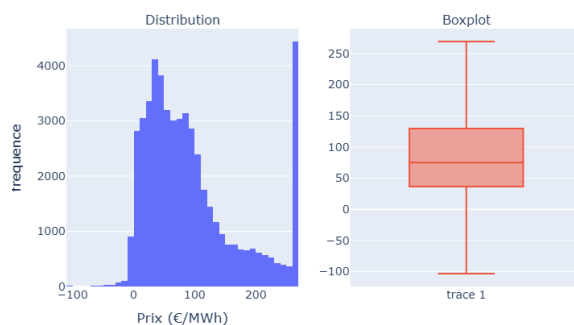
### Distribution des Consommations par Saison



### Évolution Temporelle du Prix avec et sans Outliers

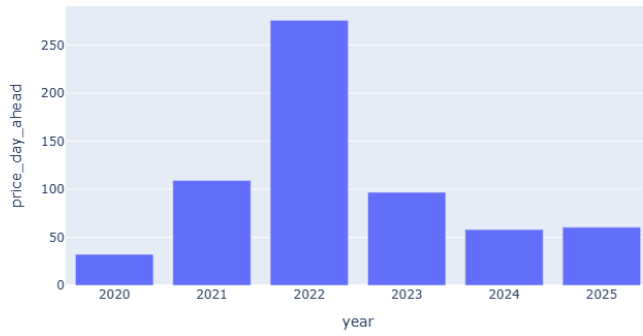


### Distribution des Prix

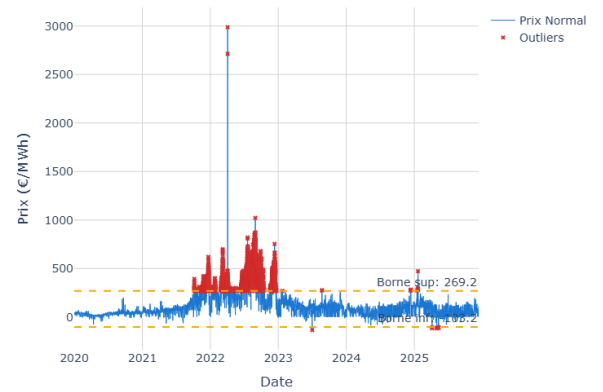




Prix Moyen Annuel

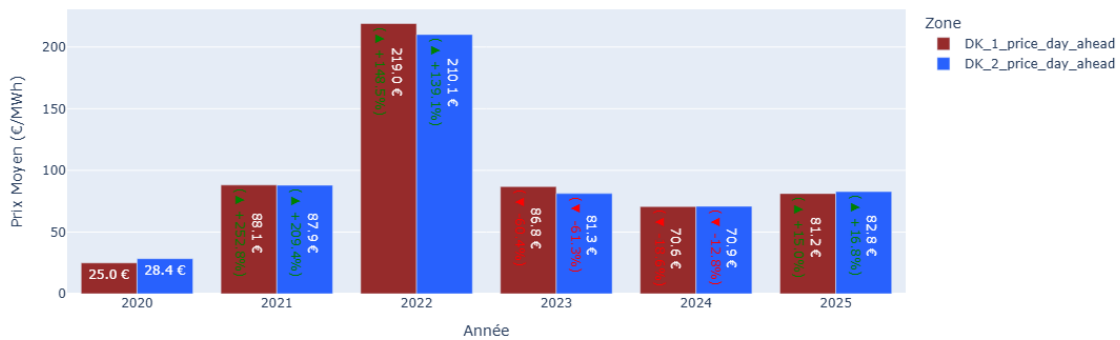


Détection des Outliers (Méthode IQR)

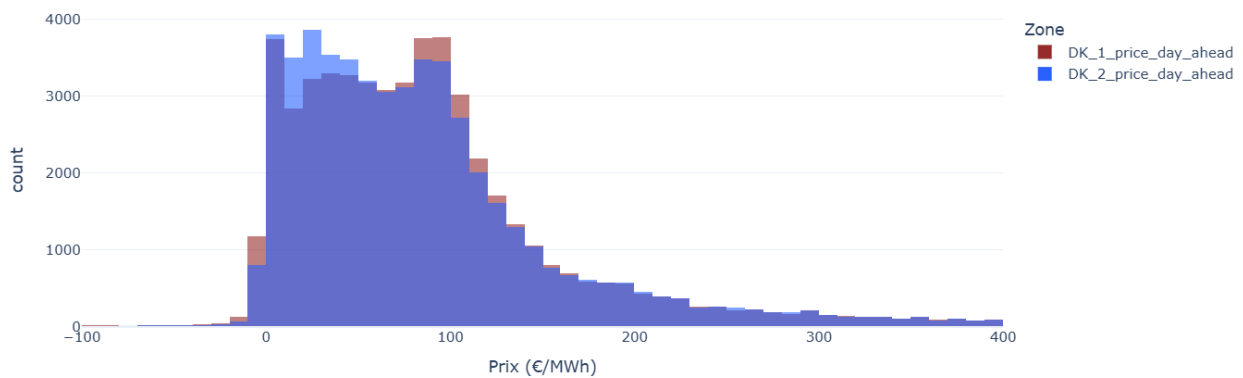


## Quelques Graphiques sur l'EDA Danemark

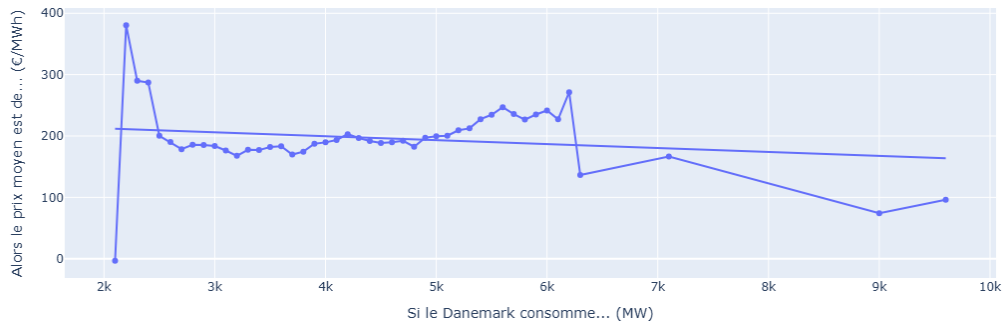
Prix Moyen Annuel et Évolution



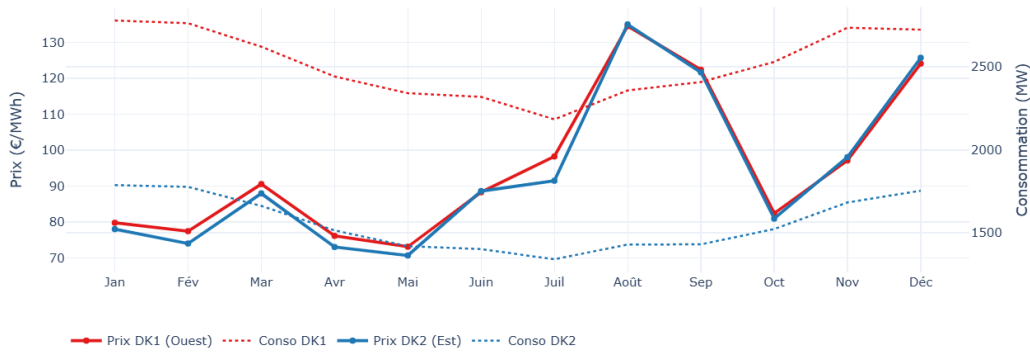
Distribution des Prix : DK1 (Volatil) vs DK2 (Stable)



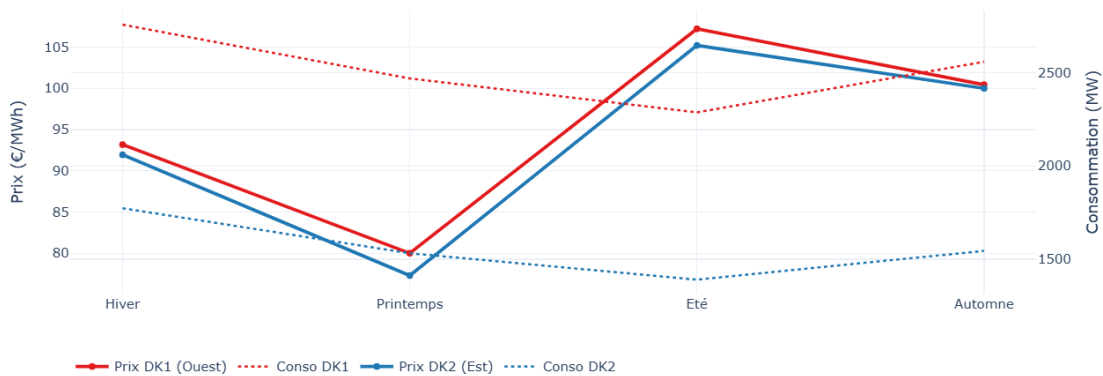
Tendance : Prix Moyen selon le Niveau de Consommation (Danemark)



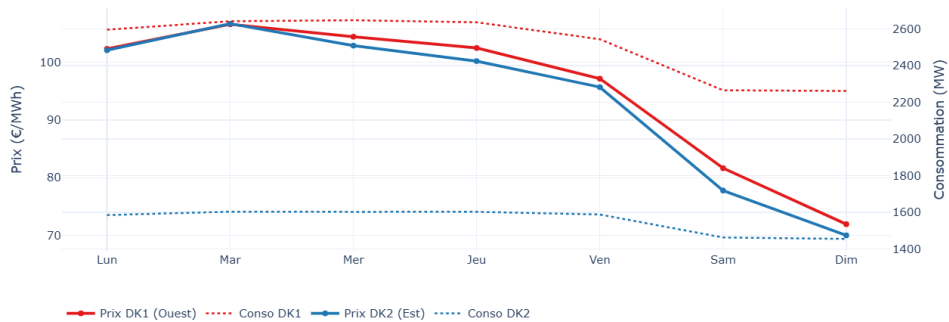
Saisonnalité Mensuelle Comparée : Prix vs Consommation (DK1 vs DK2)



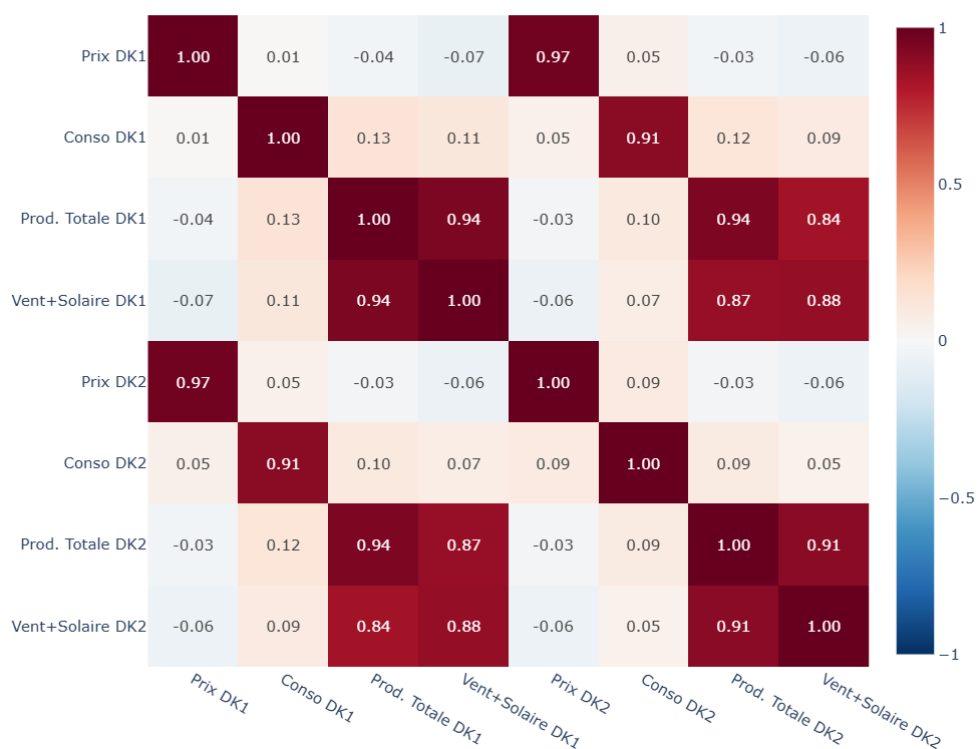
Saisonnalité Climatique : Prix vs Conso (Impact Hivernal)



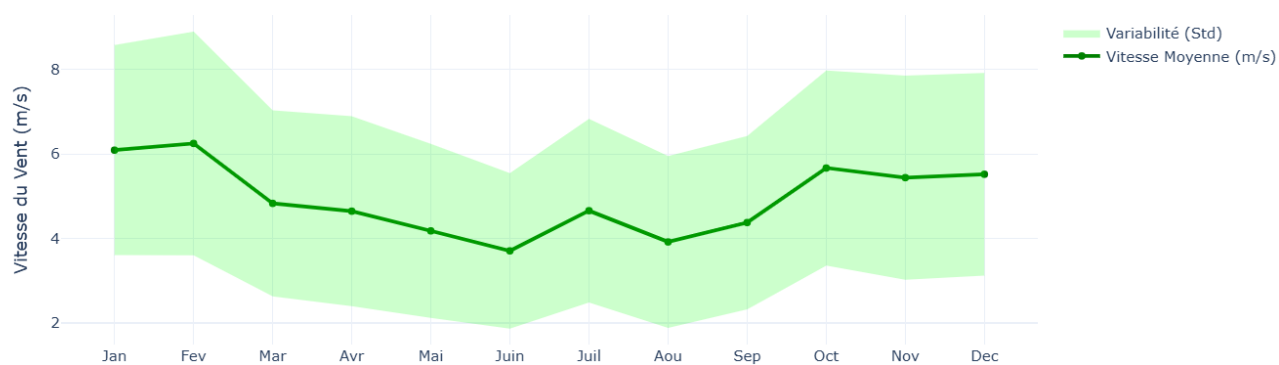
Profil Hebdomadaire : L'effet Week-end



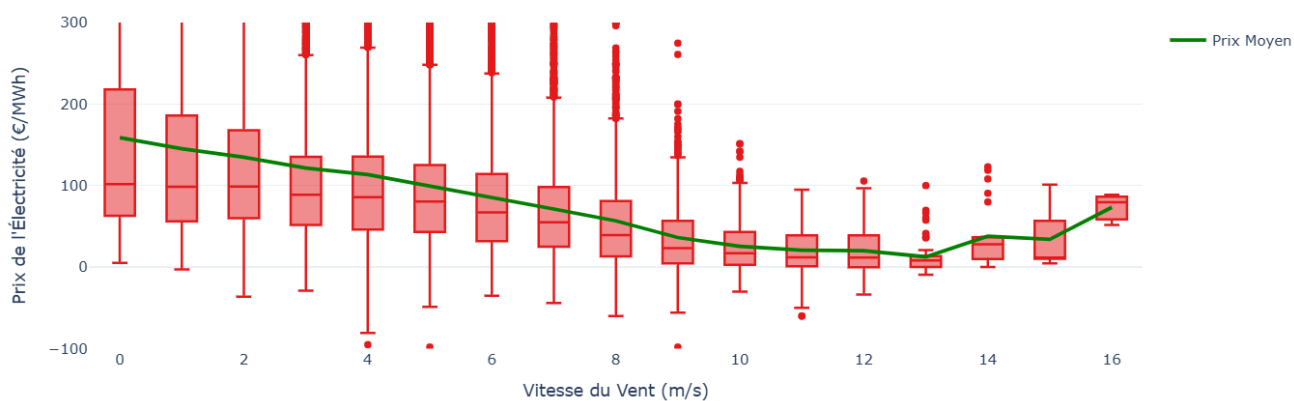
## Matrice de Corrélation : Prix, Conso & Production



## Cycle Annuel du Vent (Moteur du Mix Danois)

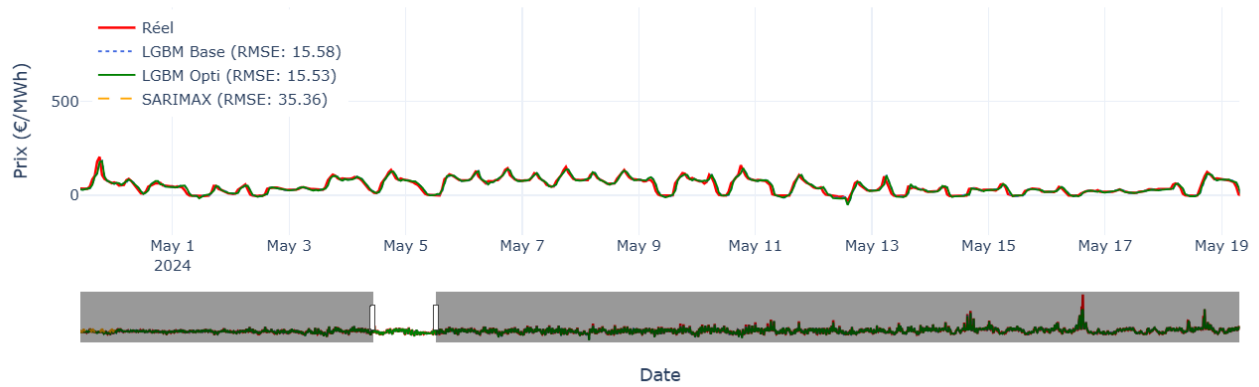


## Analyse de Sensibilité : Relation Vent vs Prix de Marché



# LGBM Base vs LGBM Optimisé vs SARIMAX

DK1



# LGBM Base vs LGBM Optimisé vs SARIMAX

DK2

