## ligne horizontale

Projet Analyse d’une série temporelle de type Cryptomonnaie

22.05.2023

**─**

BETAM Ludovic EPI DIGITAL SCHOOL 4EME IA&DS

Encadrant :

Madame Saoussen BEN JABRA

**Sommaire :**

[**Chapitre 1 : Contexte General**](#_au51mny0sx6) **1**

[Cas d’usage](#_4p7xi5bvhxdr) 2

[L'importance du Machine Learning](#_f6vlqm1mr560) 4

[Deep Learning](#_ic9qh6sb0z2n) 5

[**Chapitre 2 : État de l’art**](#_32m4toq6e6c5) **6**

[Preprocessing/Cleaning](#_qt1quw13k7q9) 6

[Décomposition de la série temporelle](#_74x7rooaasb9) 8

[● Machine Learning](#_60sfk8jv8g30) 9

[Modèle ARIMA](#_cylml5layx78) 10

[● Deep Learning](#_5aaiefti9ewx) 12

[Modèle RNN](#_4ydc8m3zab1l) 12

[Facebook Prophet](#_vcyu0cl38vhe) 14

[**Chapitre 3 : Résultats**](#_ax7honhzc1am) **15**

[Modèle ARIMA](#_q94mo1b4gwwy) 15

[Modèle RNN](#_jnav0kfawhle) 16

[Prophet](#_ha0uy6tpg9i2) 19

[**Conclusion**](#_tirvradkszdu) **24**

[Bibliographie :](#_fkmfw5ge514f) 25

# 

# 

# 

# Chapitre 1 : Contexte General

Notre projet porte sur la prédiction du cours d'une crypto-monnaie en utilisant des techniques de machine learning et de deep learning. Nous collectons les données historiques du cours de la crypto-monnaie et les préparons en les divisant en ensembles d'entraînement, de validation et de test. Ensuite, nous construisons un modèle basé sur des algorithmes de machine learning tels que les réseaux de neurones récurrents (RNN) . Nous entraînons le modèle sur les données d'entraînement en ajustant les paramètres pour minimiser la fonction de perte, généralement l'erreur quadratique moyenne (RMSE). Nous évaluons ensuite la performance du modèle en utilisant différentes métriques telles que le RMSE, le coefficient de détermination (R^2) ou la précision. Une fois le modèle entraîné, nous utilisons les données de test pour évaluer sa capacité à prédire avec précision les variations du cours de la cryptomonnaie. Notre objectif est de développer un modèle robuste et précis qui peut être utilisé pour prendre des décisions éclairées dans le trading ou l'investissement en cryptomonnaie.

# **Objectifs**

**Notre objectif est de développer un modèle analytique qui peut être utilisé pour prendre des décisions éclairées dans le trading ou l'investissement en cryptomonnaie.**

## 

## **Cas d’usage**

Une crypto monnaie est une forme de monnaie numérique basée sur des principes de cryptographie. Contrairement aux monnaies traditionnelles émises, une cryptomonnaie est une forme de monnaie numérique basée sur des principes de cryptographie. Contrairement aux monnaies traditionnelles émises par les gouvernements, les cryptomonnaies sont décentralisées et utilisent des technologies de registres distribués, tels que la blockchain, pour enregistrer et vérifier les transactions.

# 

Les crypto-monnaies sont généralement créées à travers un processus appelé "minage", où des ordinateurs résolvent des problèmes mathématiques complexes pour valider les transactions et sécuriser le réseau. Une fois qu'une transaction est validée, elle est ajoutée à la blockchain, qui est un registre public et immuable contenant l'historique de toutes les transactions effectuées.

Les crypto-monnaies offrent plusieurs caractéristiques distinctives. Elles sont souvent conçues pour être sécurisées, transparentes, pseudonymes (les transactions sont associées à des adresses plutôt qu'à des identités réelles) et rapides pour les transferts de fonds. De plus, certaines cryptomonnaies permettent également l'exécution de contrats intelligents, qui sont des programmes auto-exécutables avec des conditions prédéfinies.

Le Bitcoin, créé en 2009, est la première et la plus connue des crypto-monnaies. Depuis lors, de nombreuses autres crypto-monnaies ont émergé, telles que Ethereum, Ripple.

Une série temporelle est une séquence de données observées à des intervalles de temps réguliers, et son analyse permet d'identifier des motifs, des tendances et des relations cachées. Dans le domaine des investissements, l'analyse des séries temporelles revêt une importance particulière pour les prévisions et les prises de décision éclairées.

Dans ce projet de fin d'année en Data Analyse, nous nous concentrons spécifiquement sur l'analyse des cours de crypto-monnaie, qui sont des séries temporelles représentant l'évolution des prix des différentes cryptomonnaies au fil du temps. Comprendre les fluctuations des cours de crypto-monnaies est crucial pour les investisseurs et les traders qui cherchent à maximiser leurs rendements et à minimiser les risques. En utilisant des techniques d'analyse des séries temporelles, nous pouvons extraire des informations significatives des données historiques et effectuer des prédictions sur les tendances futures.

**Ce rapport présente les différentes étapes de mon projet, depuis la collecte des données jusqu'à l'analyse exploratoire et la modélisation de modèles prédictifs du cours d’une crypto monnaie le Bnb\*.**

Bnb est l'abréviation de Binance Coin, une crypto monnaie qui a été lancée en 2017 par la plateforme d'échange de cryptomonnaies Binance. Le Binance Coin est un jeton utilisé principalement sur la plateforme Binance, qui est l'une des plus grandes et des plus populaires plateformes d'échange de cryptomonnaies au monde. Le Bnb est utilisé pour payer les frais de transaction sur la plateforme Binance, bénéficier de réductions sur les frais de transaction, participer aux ventes de jetons et à d'autres fonctionnalités de l'écosystème Binance. Le Binance Coin repose sur la blockchain Binance Chain et est basé sur le standard de jeton ERC-20 d'Ethereum. Le Bnb a gagné en popularité en raison de la popularité de la plateforme Binance et de son utilisation étendue au sein de l'écosystème Binance.

Binance est une plateforme d'échange de cryptomonnaies fondée en 2017 par Changpeng Zhao. Elle est aujourd'hui l'une des plus grandes et des plus populaires plateformes d'échange de cryptomonnaies au monde. Binance permet aux utilisateurs d'acheter, de vendre et de trader une large gamme de cryptomonnaies, offrant ainsi un accès facile au marché des cryptomonnaies.

# 

## **L'importance du Machine Learning**

Le machine learning est une branche de l'intelligence artificielle qui permet aux ordinateurs d'apprendre à partir des données sans être explicitement programmés. Dans l'analyse de séries temporelles, le machine learning offre des capacités avancées pour modéliser, prévoir et comprendre les motifs complexes et les comportements changeants des données temporelles.

L'importance du machine learning dans l'analyse de séries temporelles réside dans sa capacité à capturer les relations non linéaires, les effets de retard, les tendances saisonnières et d'autres facteurs dynamiques qui caractérisent souvent ces données. Les algorithmes de machine learning peuvent extraire des informations à partir de variables passées, détecter les schémas cachés et généraliser ces connaissances pour effectuer des prédictions sur de nouvelles données.

En utilisant des techniques de machine learning telles que les réseaux de neurones, les modèles ARIMA (AutoRegressive Integrated Moving Average), En utilisant des techniques de machine learning telles que les réseaux de neurones, les modèles ARIMA (AutoRegressive Integrated Moving Average),, il est possible d'améliorer significativement la précision des prédictions des séries temporelles. Ces algorithmes peuvent également aider à l'identification des facteurs les plus influents et à la détection des anomalies ou des comportements inhabituels.

Dans le cadre de notre projet de fin d'année en Data Analyse, nous utiliserons des techniques de machine learning pour modéliser et prédire les cours des crypto-monnaies. En entraînant nos modèles sur des données historiques, nous chercherons à exploiter les schémas et les tendances passés pour effectuer des prédictions sur les prix futurs des crypto-monnaies.

## **Deep Learning**

**Définition** : Le deep learning, ou apprentissage profond, est une branche de l'intelligence artificielle qui utilise des réseaux de neurones artificiels profonds pour apprendre à partir de données. Inspiré par le fonctionnement du cerveau humain, le deep learning permet de modéliser et d'apprendre des représentations complexes et hiérarchiques des données. En utilisant des architectures de réseaux de neurones profonds avec de nombreuses couches interconnectées, le deep learning est capable d'extraire automatiquement des caractéristiques pertinentes à partir des données brutes.

Ainsi l'utilisation du deep learning pour les prédictions de séries temporelles offre une puissance de modélisation supérieure, une capacité à apprendre automatiquement les caractéristiques pertinentes, une adaptabilité à différents types de séries temporelles et une possibilité de construction de modèles end-to-end. Cela permet d'obtenir des prédictions plus précises et une meilleure compréhension des dynamiques temporelles.

## 

## 

## 

## 

## 

## 

## 

## 

## 

# Chapitre 2 : État de l’art

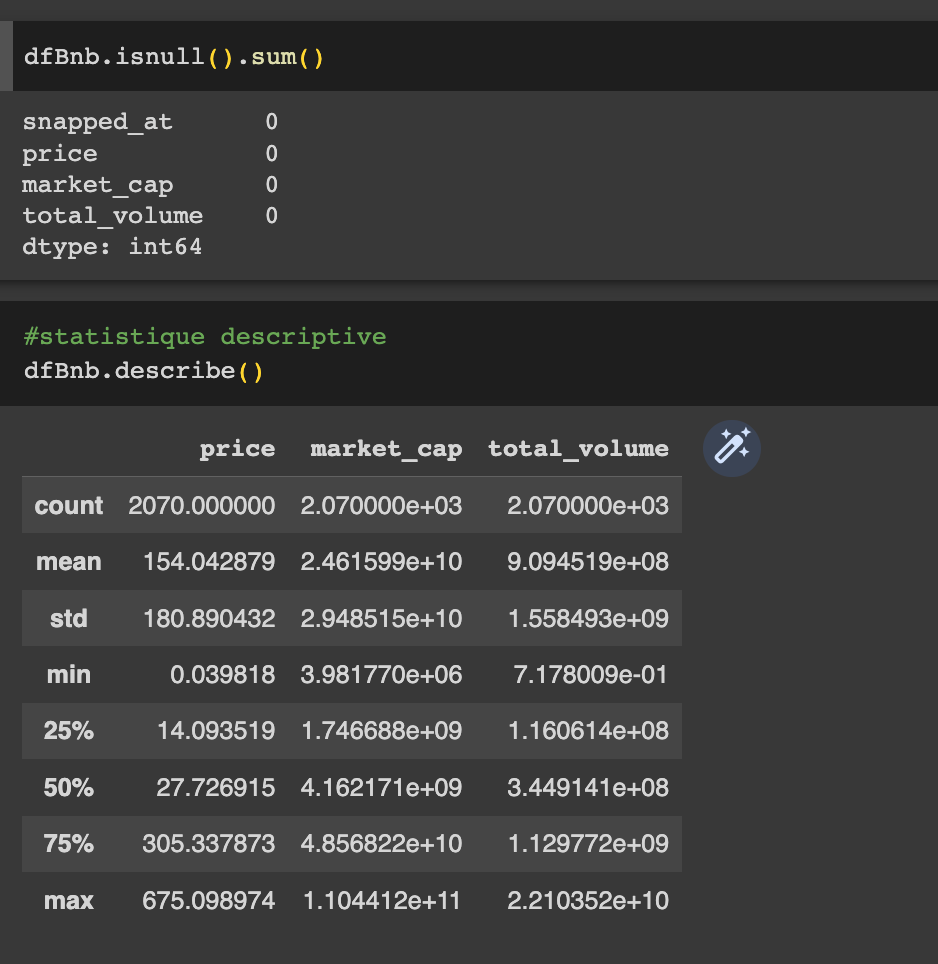
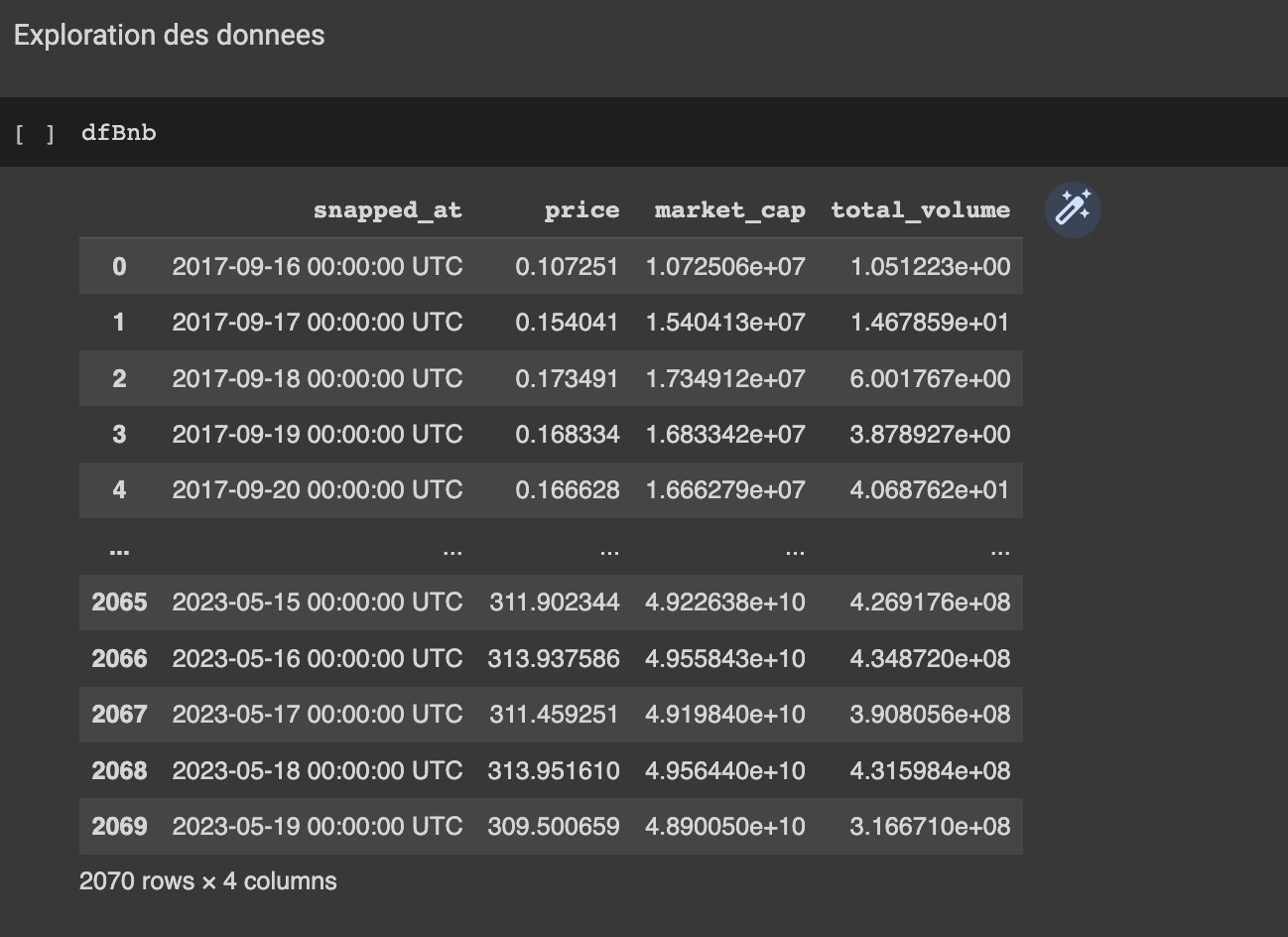
## **Preprocessing/Cleaning**

La collecte de données de haute qualité et leur préparation soigneuse sont fondamentales pour mener à bien une analyse de séries temporelles des cours de cryptomonnaie. La fiabilité et la pertinence des résultats obtenus dépendent en grande partie de la qualité des données utilisées. Les données ainsi traitées dans notre rapport nous viennent du site :<https://www.coingecko.com/fr> qui est une plateforme d'échange et d'analyse du marché de la cryptomonnaie.

Pour notre projet nous allons nous concentrer sur la Crypto Monnaies BNB.

Les données relatives aux différentes monnaies sont accessibles gratuitement et peuvent être télécharger au format .csv qui est l'un des formats le plus pratique à l'analyse de données.

Ainsi, après l'importation de nos données et les bibliothèques nécessaires, on passe à l'exploration et la visualisation de celles-ci..



Ainsi, notre datasets est composés de 4 colonnes :

**Snappet-at** : qui fait référence à la date exacte du relevé des valeurs du jour.

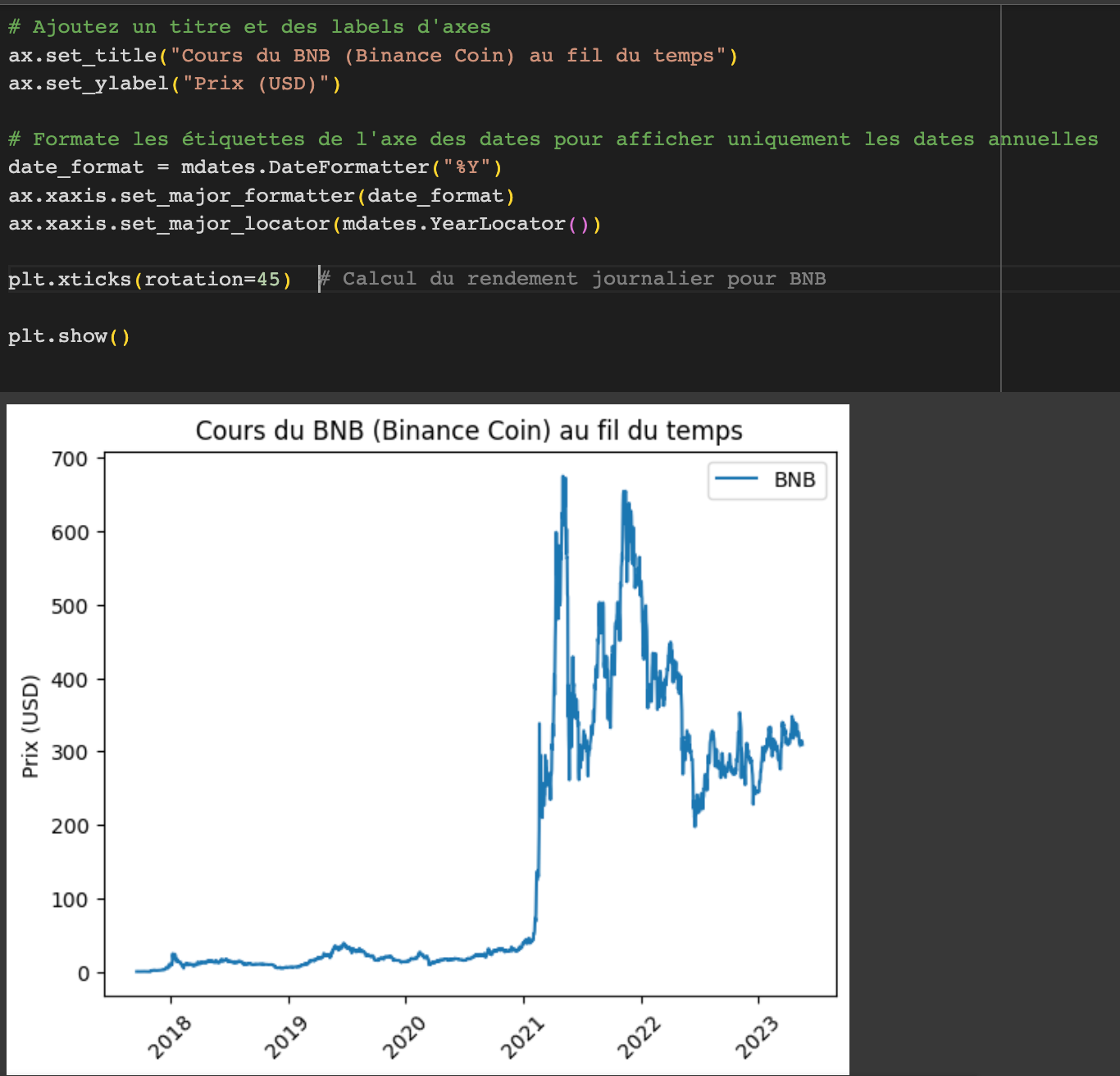
**Price** : la valeur du jour de la monnaie

**Market Cap** : qui fait référence à la *capitalisation boursière\** de la monnaie

**Volume** : Il représente la quantité d'unités de la crypto-monnaie qui ont été échangées.

*\*La capitalisation boursière d'une crypto monnaie est la valeur totale de toutes les unités de cette crypto monnaie en circulation. C'est une mesure de la valeur globale d'une crypto-monnaie sur le marché. La capitalisation boursière est généralement calculée en multipliant le prix d'une unité de la cryptomonnaie par le nombre total d'unités en circulation.*

Après avoir vérifié que notre jeu de données ne contient aucune valeurs nulles on va visualiser la courbe afin d'observer son évolution du cours du temps:



*Ainsi on observe une augmentation qu'on peut qualifier d'exponentielle aux alentours de 2021.*

**Remarque** : De nombreux facteurs Socio-économiques influencent la valeurs sur le marché d’une monnaie, dans le cas des crypto-monnaies l’augmentation de leurs valeurs en 2021 est probablement dû à la Pandémie de la Covid-19 et de la perte de confiance de certains investisseurs dans le secteur bancaire dit centralisé.

## 

## 

## 

## 

## 

## 

## 

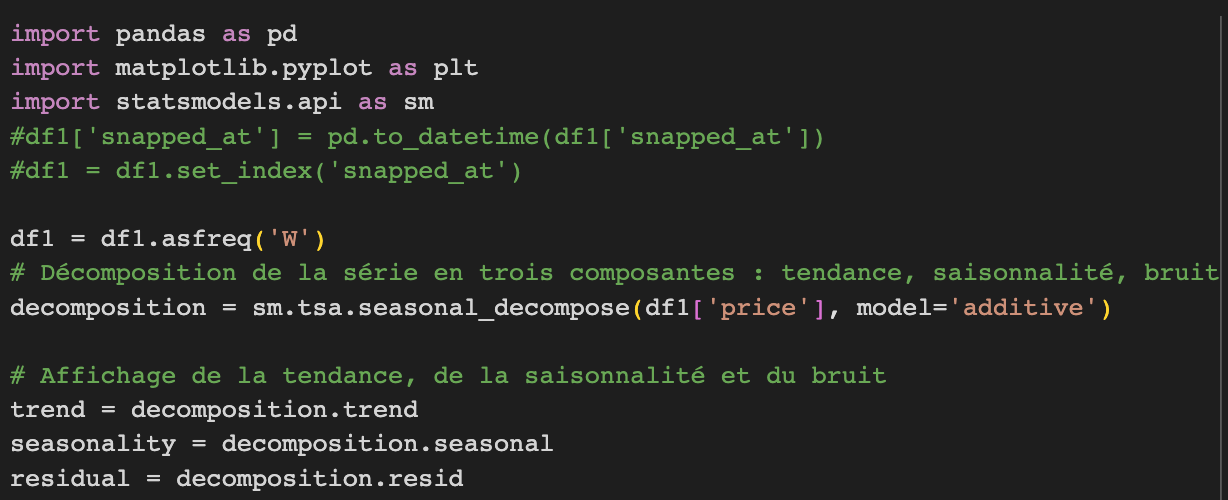
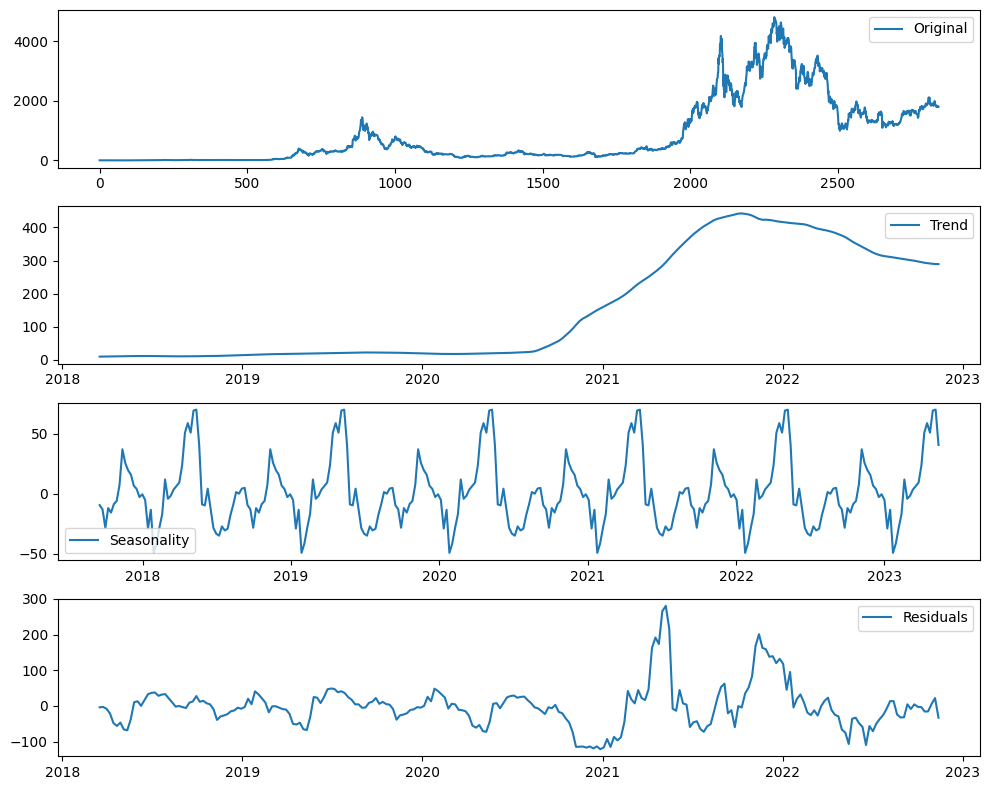
## 

## 

## 

## **Décomposition de la série temporelle**

*Décomposons le cours du BnB en fonction de la Tendance, la saisonnalité et du bruit blanc* :

*Sur l'image ci-dessus on observe bien une tendance à la hausse au début de l'année 2021 avant de diminuer légèrement à partir de 2022.*

*L'observation de la saisonnalité montre bien un motif périodique qui se répète tous les ans. De plus, on voit bien que l'amplitude reste relativement similaire.*

### **Machine Learning**

## **Modèle ARIMA**

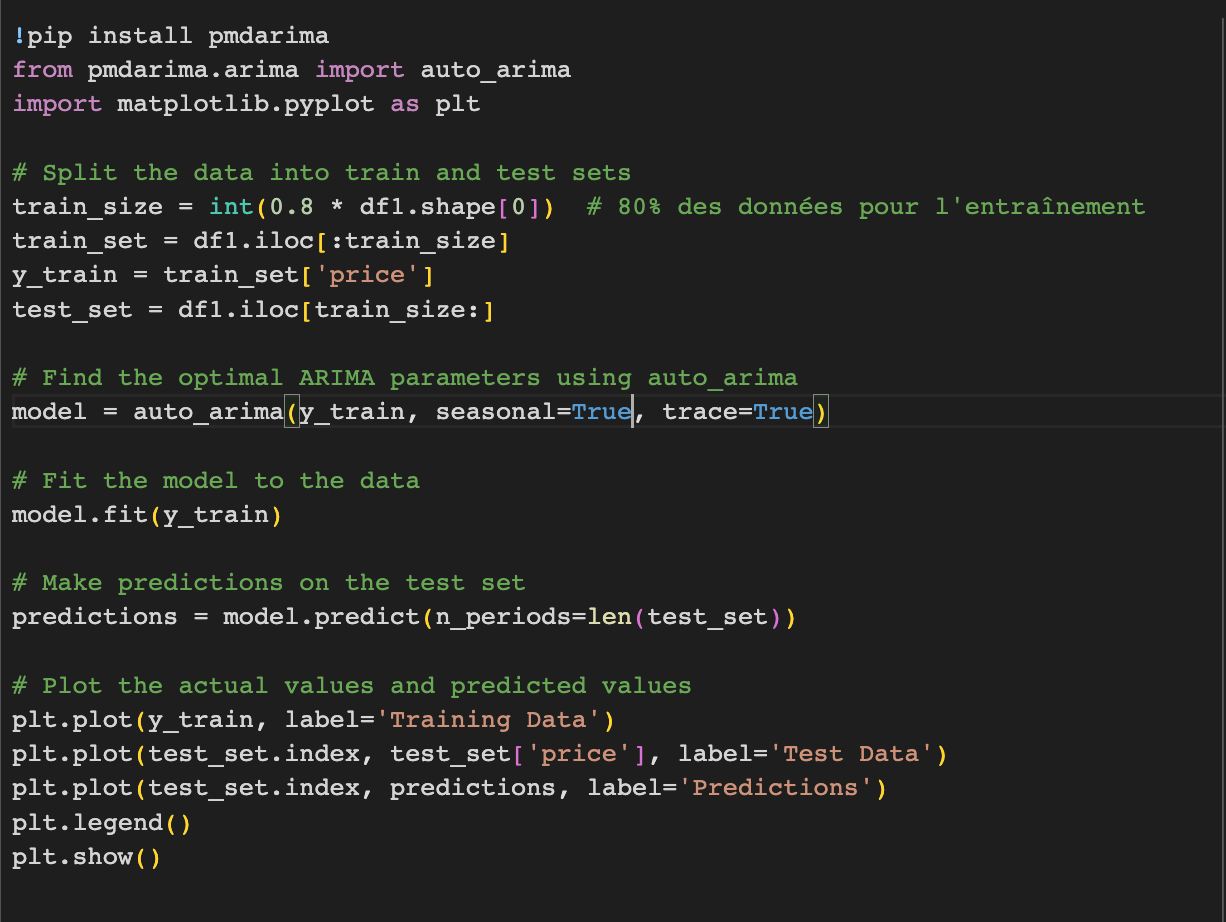
**Le modèle ARIMA** (AutoRegressive Integrated Moving Average) est un modèle statistique largement utilisé pour l'analyse et la prévision des séries temporelles. Il permet de modéliser et de capturer les tendances, les motifs saisonniers et les comportements autorégressifs dans les données temporelles.

Le modèle ARIMA est composé de trois composantes principales : **l'autorégression** (AR), l'intégration (I) et la moyenne mobile (MA):

La composante **autorégressive** (AR) capture la dépendance linéaire entre les observations passées et l'observation actuelle. Elle utilise les valeurs précédentes de la série temporelle pour prédire la valeur actuelle. L'ordre de l'autorégression (représenté par p) indique combien de valeurs précédentes sont utilisées pour la prédiction.

La composante **d'intégration** est utilisée pour rendre la série temporelle stationnaire, c'est-à-dire en éliminant les tendances et les variations non stationnaires. Cela peut impliquer la différenciation de la série temporelle, c'est-à-dire prendre la différence entre les observations successives jusqu'à ce que la série devienne stationnaire. L'ordre d'intégration (représenté par d)

Enfin, **la moyenne mobile** (MA) modélise les erreurs résiduelles entre les observations réelles et les prédictions basées sur les valeurs passées. Elle utilise une moyenne pondérée des erreurs précédentes pour prédire la valeur actuelle. L'ordre de la moyenne mobile (représenté par q) indique le nombre d'erreurs passées utilisées pour la prédiction

**Implementation**

### **Deep Learning**

## **Modèle RNN**

Un modèle RNN (Récurrent Neural Network) est un type de modèle d'apprentissage profond utilisé pour traiter des données séquentielles. Contrairement aux réseaux de neurones traditionnels, les RNNs peuvent maintenir une mémoire interne ou un état caché qui leur permet de capturer les dépendances à long terme dans les séquences de données.

Le modèle RNN est constitué de cellules récurrentes qui se répètent à chaque étape de la séquence. Chaque cellule prend en entrée l'état caché précédent et l'entrée actuelle, puis génère une sortie et met à jour son état caché. Cette sortie est ensuite utilisée comme **l'entrée de la cellule à l'étape suivante**. Cela permet au modèle de prendre en compte les informations précédentes lors de la prédiction de la prochaine valeur dans la séquence.

L'architecture du modèle RNN comprend des cellules récurrentes qui se répètent à chaque étape de la séquence. Chaque cellule reçoit en entrée une donnée de la séquence et une mémoire interne appelée "état caché". La cellule combine ces informations pour générer une sortie et mettre à jour son état caché.

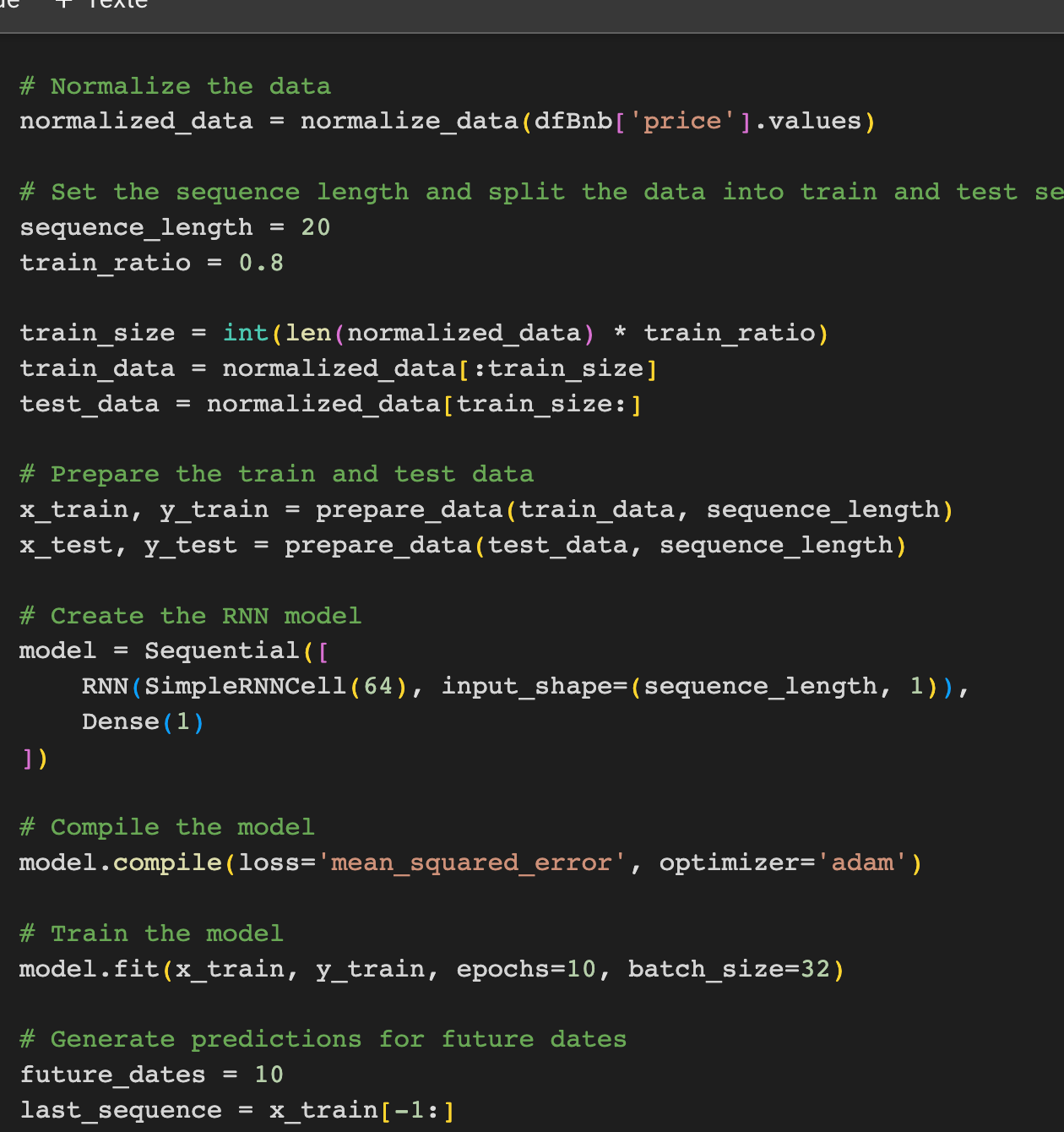
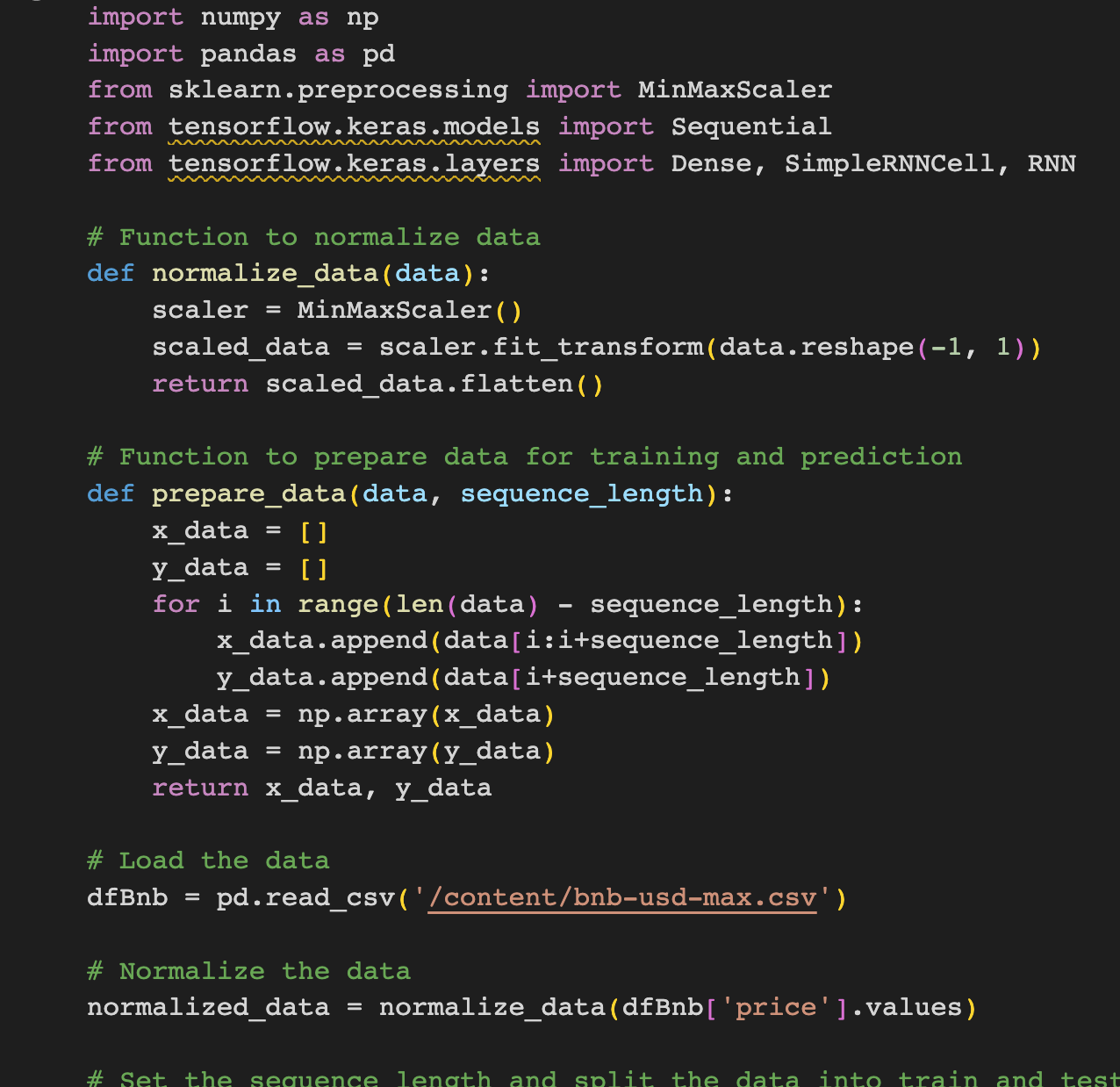
Cela permet au modèle de comprendre le contexte et de prendre des décisions basées sur les informations précédentes.

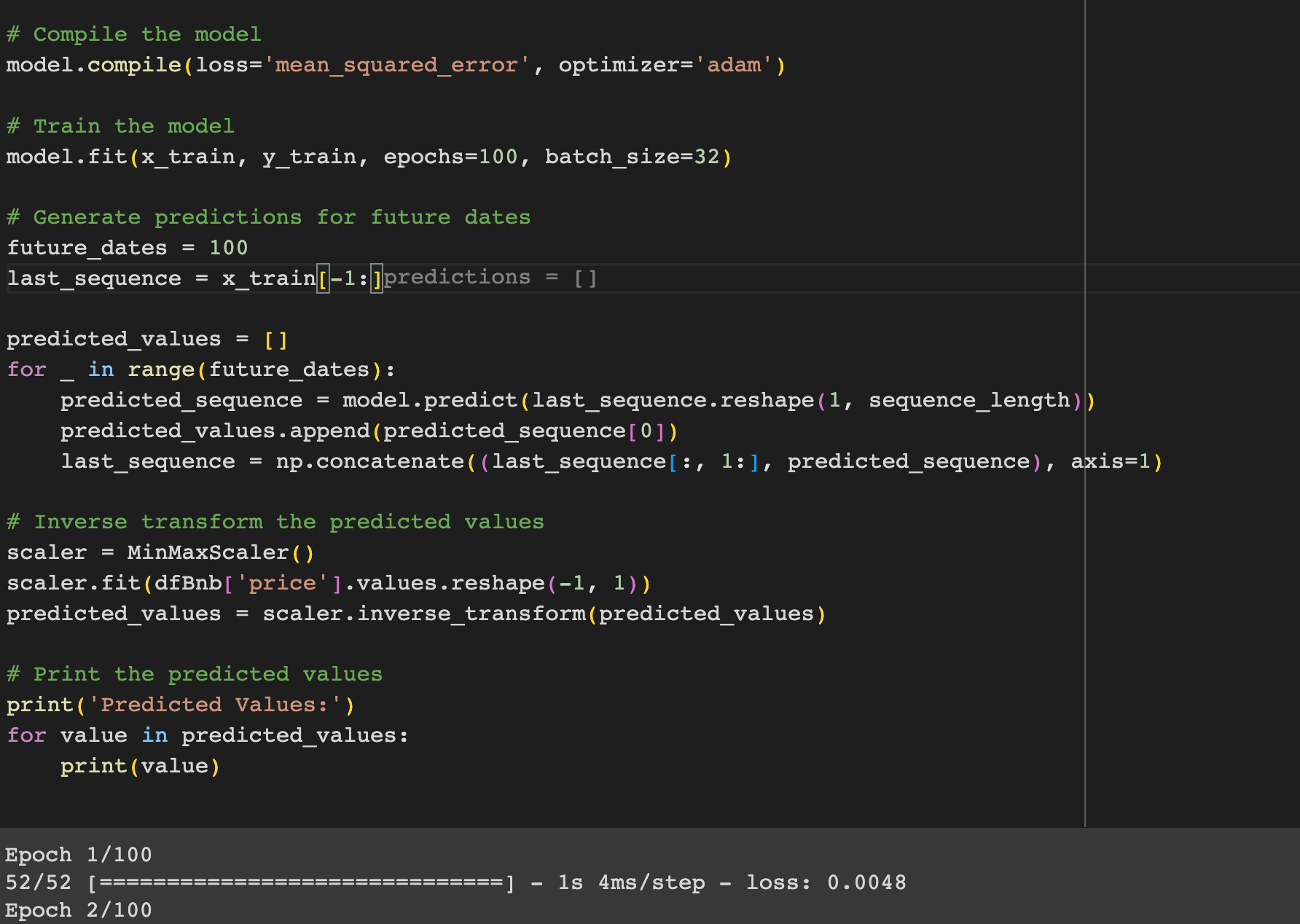
**Implementation :**

Pour implémenter un modèle RNN on utilise les bibliothèques *Keras*\* et *Tensor flow*\*

\*TensorFlow est une bibliothèque populaire et puissante d'apprentissage automatique et d'apprentissage en profondeur (deep learning). Elle a été développée par Google Brain et offre une interface conviviale pour construire, entraîner et déployer des modèles d'apprentissage automatique.

\*Keras est une interface de haut niveau pour construire des modèles d'apprentissage automatique. Il est intégré à TensorFlow et permet de créer des modèles de manière simple et intuitive





## **Facebook Prophet**

Prophet est une librairie open source (R et Python) de prévision des données de séries temporelles basée sur un modèle additif. Cette librairie permet d’analyser aisément des séries temporelles.

Facebook prophet a pour but de résoudre les problématiques suivantes :

* La difficulté de créer des modèles de prévision fiables : Car cette discipline requiert une expérience particulière
* La rigidité et le manque de robustesse des techniques de prévisions automatiques

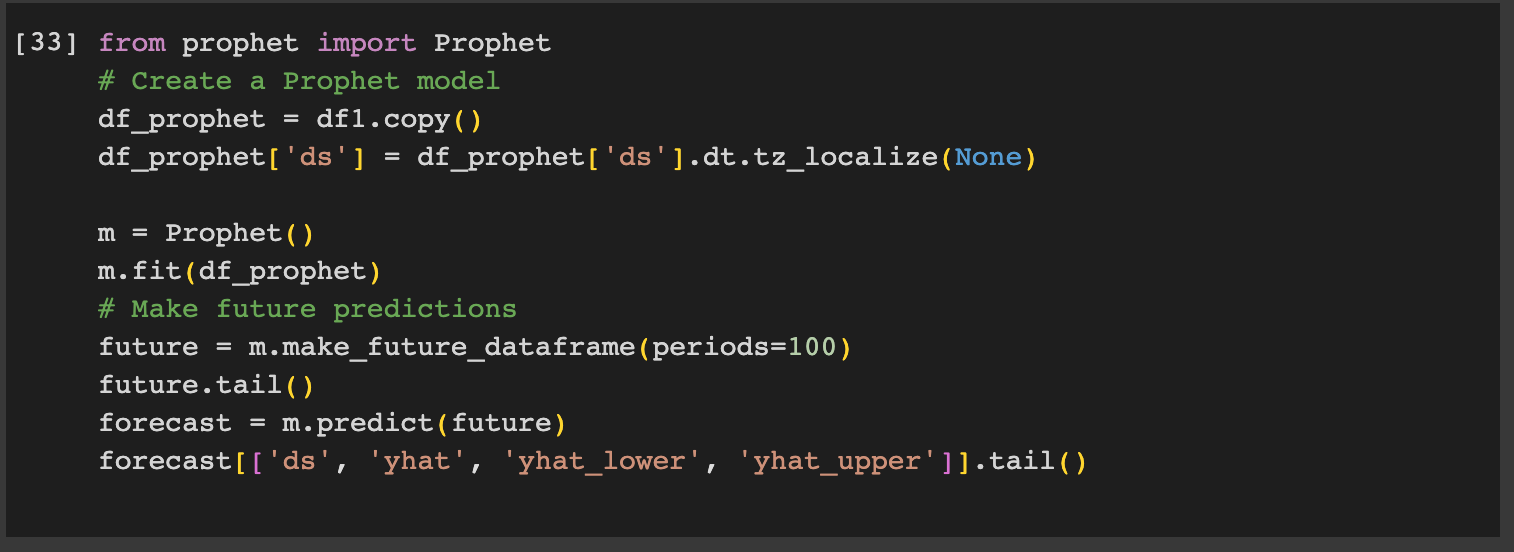
**Implémentation**

Pour utiliser Prophet pour la prévision, il faut d’abord définir et configurer un objet **Prophet()**, puis l’ajuster sur l’ensemble de données avec la fonction **fit()** et en lui transmettant les données.

L’objet **Prophet()** prend plusieurs arguments (saisonnalité, holidays…).

Par défaut, le modèle cherchera automatiquement les meilleurs paramètres pour le modèle.

La fonction **fit()** prend en entrée un DataFrame de séries temporelles. Le DataFrame doit avoir un format spécifique. La première colonne doit porter le nom **« ds »** et contenir les dates et heures. La deuxième colonne doit porter le nom **« y »** et contenir les observations.

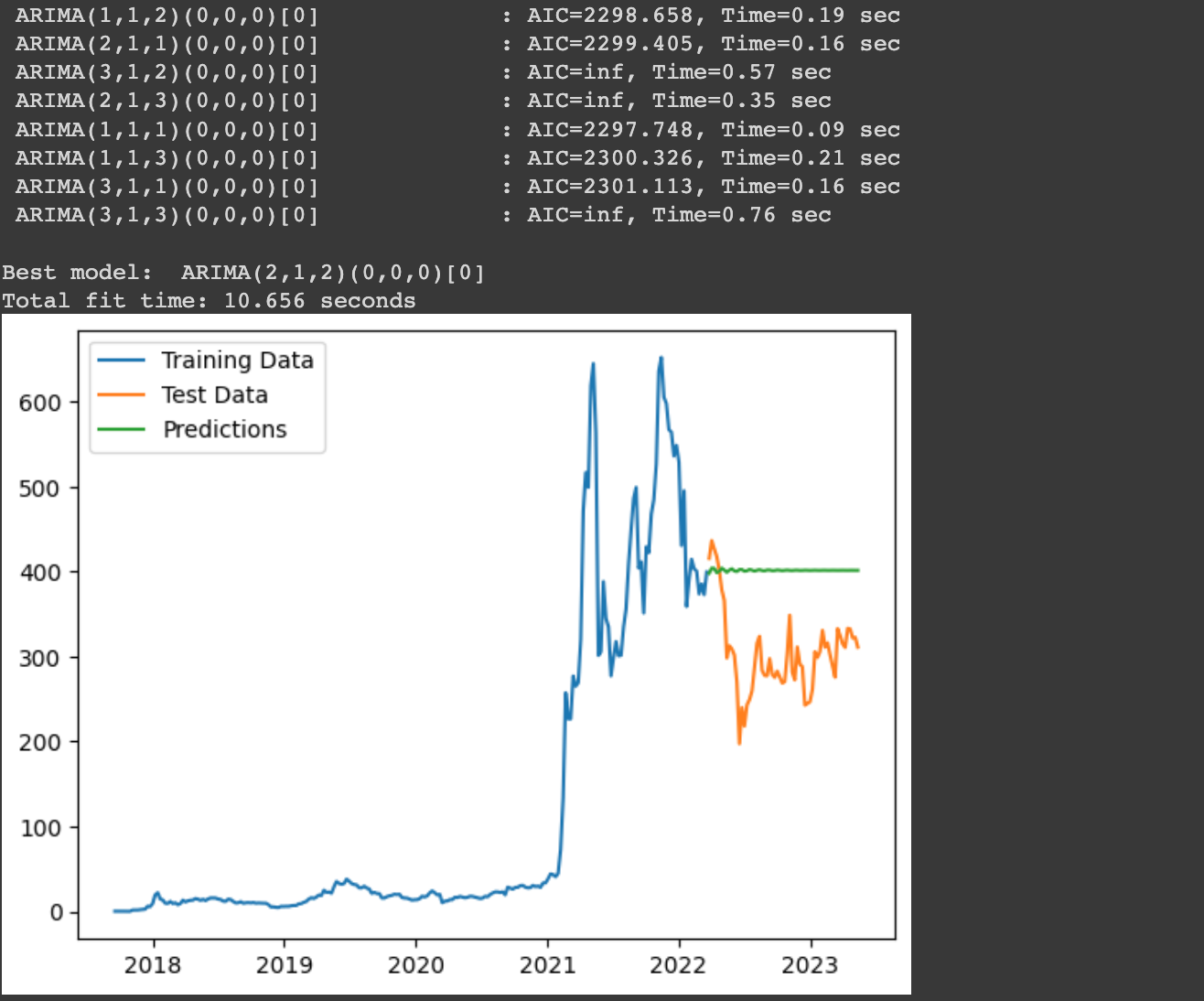


*Remarque : periods = 365 signifie ici que notre prédiction sera faite sur les 365 prochains jours.*

# Chapitre 3 : Résultats

## **Modèle ARIMA**

Ainsi après l'implémentation du modèle ARIMA, le meilleur modèle trouvé pour le cours du Bnb est un modèle (2, 1, 2) cependant on observant la courbe prédictive on se rend compte de l'imprécision de la prédiction. Cela peut s'expliquer car le  **critère de sélection** est la minimisation de **L'AIC**.

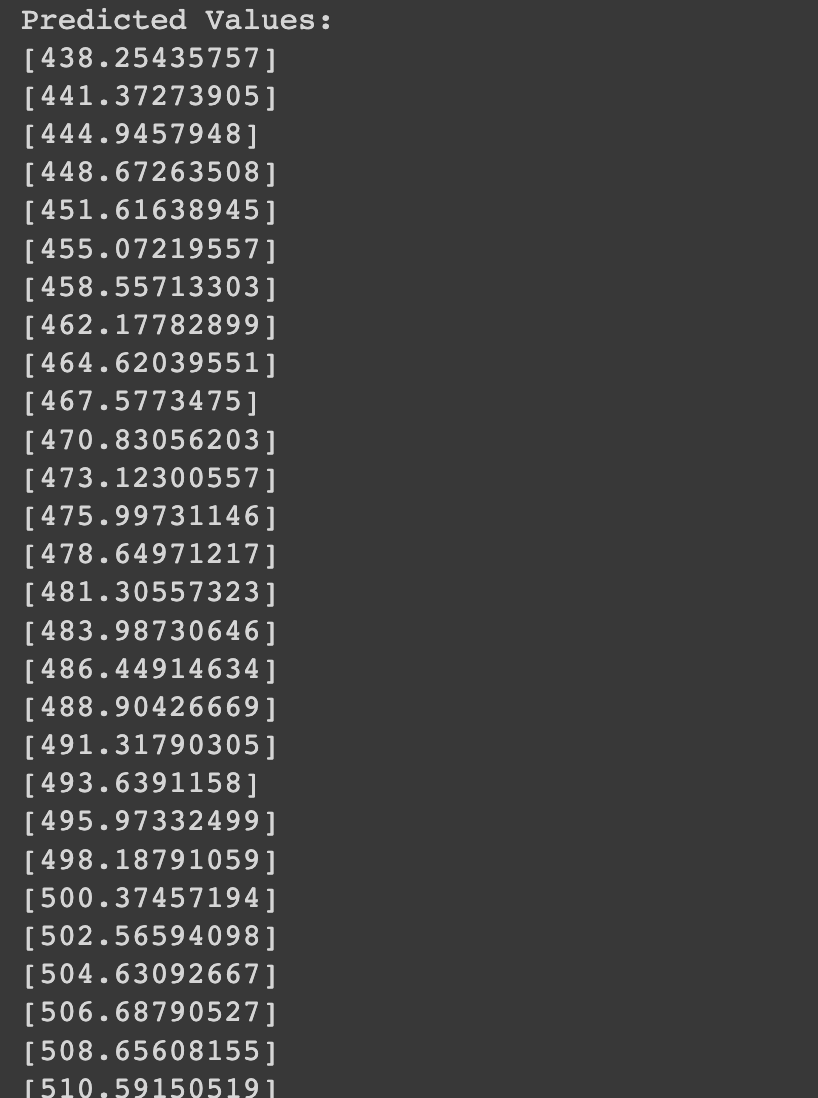


\*AIC (Akaike's Information Criterion) est un critère d'information développé par Hirotugu Akaike. Il est utilisé dans la sélection de modèles statistiques pour estimer la qualité de l'ajustement d'un modèle aux données et pour comparer différents modèles entre eux.

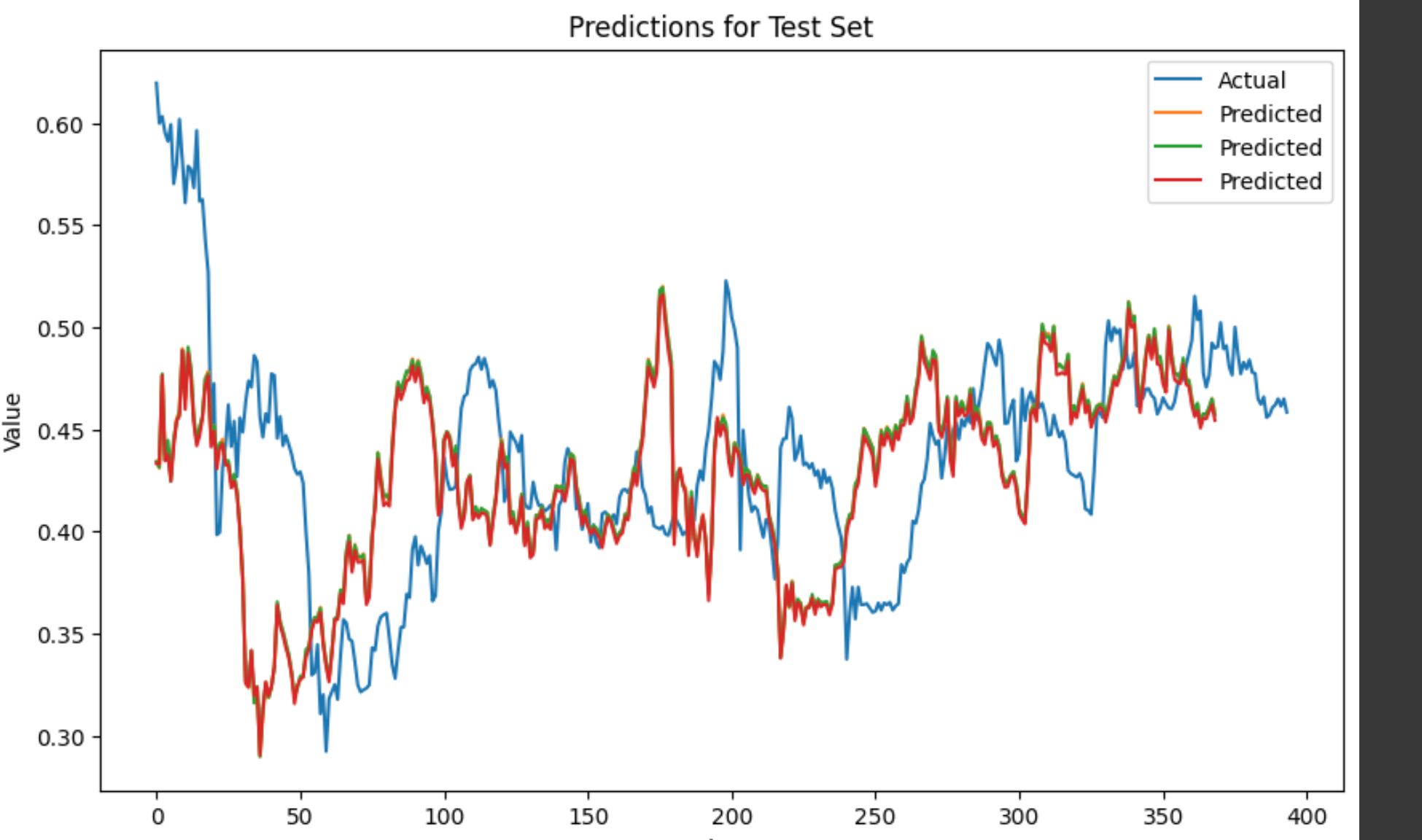
L'AIC est basé sur le principe de minimisation de l'information de Kullback-Leibler, qui mesure la divergence entre la distribution vraie des données et la distribution estimée par le modèle. L'objectif de l'AIC est de trouver un équilibre entre un bon ajustement aux données et la complexité du modèle.

## **Modèle RNN**

Notre modèle nous a permis d’afficher une liste des valeurs prédites sur les 100 prochains jours :

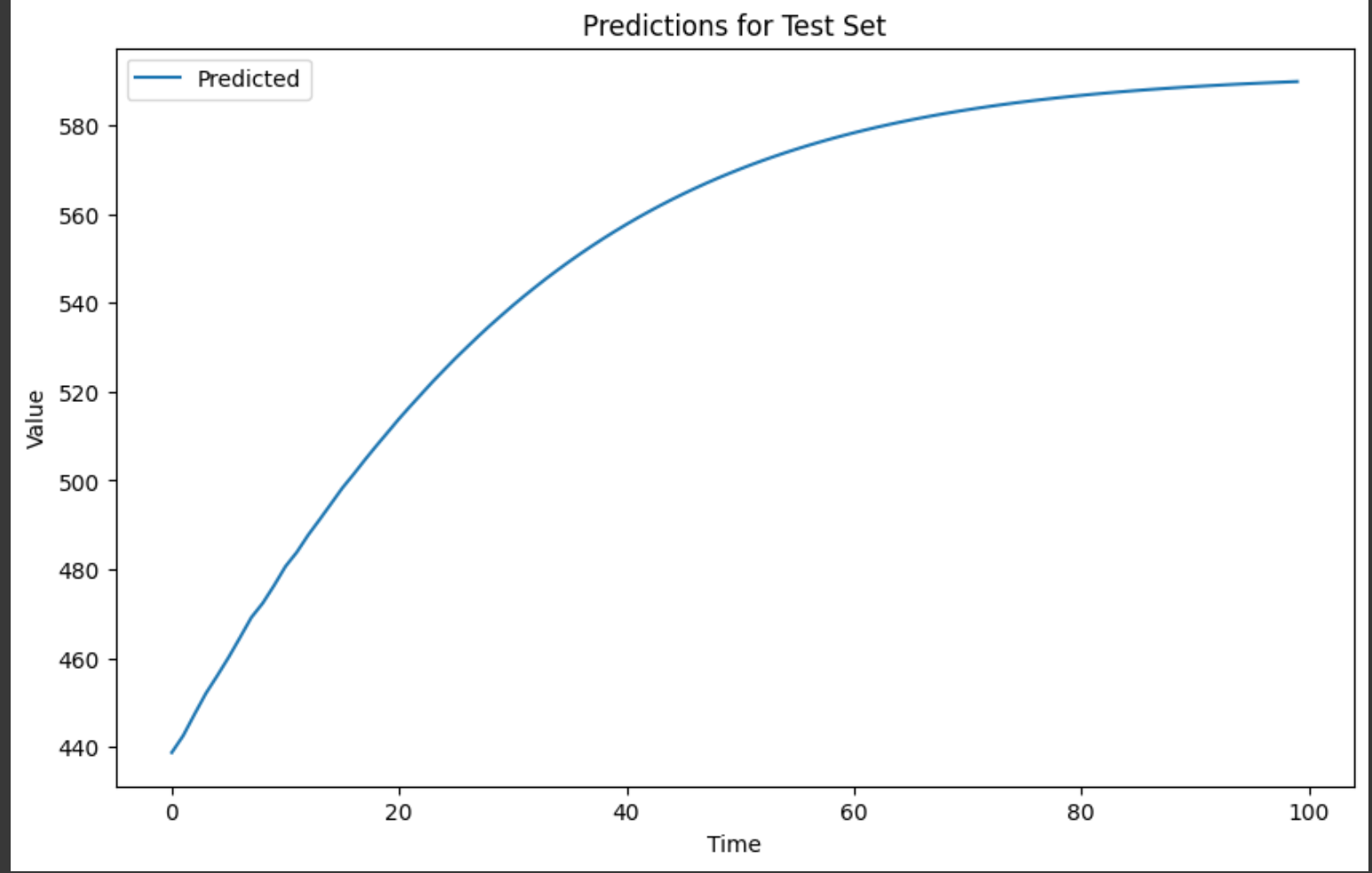


**On peut également afficher la courbe sur les données Test de notre Dataset**



*L’axe des X est en nombre de jours.*

***la courbe des valeurs prédites est la suivante :***

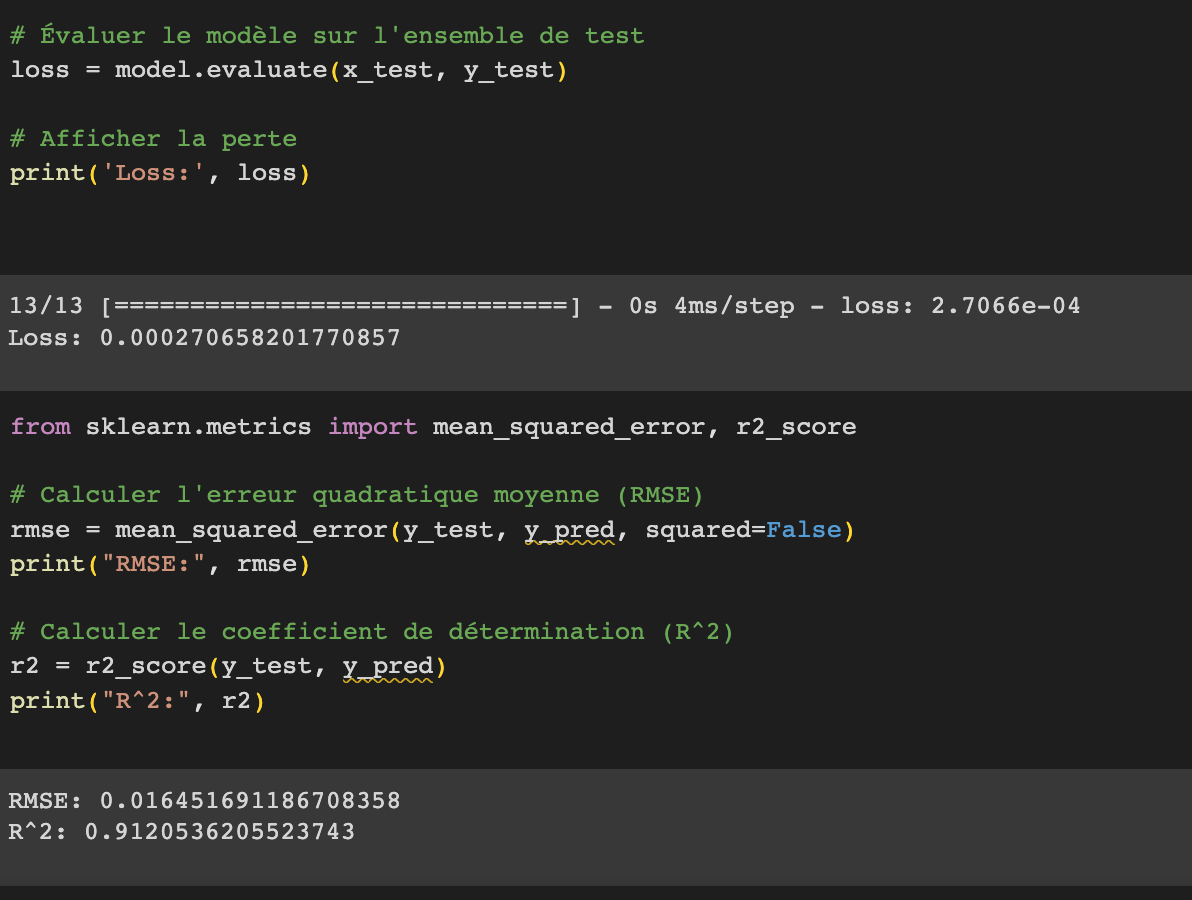
**

On observe ainsi une tendance à la hausse de notre cours du Bnb au cours des 100 prochains jours.

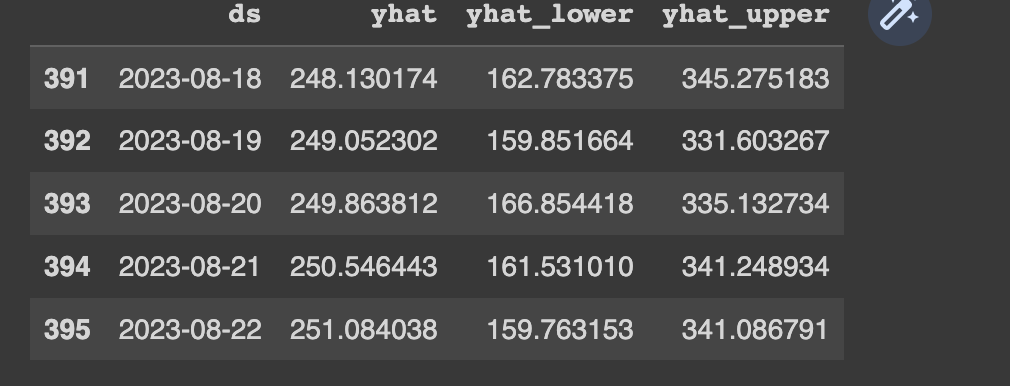
**Métriques d'évaluation :**

**Le RMSE** (Root Mean Squared Error) est une mesure de l'erreur moyenne entre les valeurs prédites et les vraies valeurs dans un problème de régression. Il est calculé en prenant la racine carrée de la moyenne des carrés des écarts entre les prédictions et les vraies valeurs. Le RMSE permet d'évaluer à quel point les prédictions sont proches des vraies valeurs, et il est souvent utilisé pour mesurer la performance des modèles de régression. Plus le RMSE est proche de zéro, meilleure est la performance du modèle.

Le **coefficient de détermination**, souvent noté R^2, est une mesure de la proportion de la variance des vraies valeurs qui est expliquée par le modèle. Il indique à quel point les prédictions du modèle sont cohérentes avec les vraies valeurs. Le coefficient de détermination peut prendre des valeurs entre 0 et 1, où une valeur de 1 indique une correspondance parfaite entre les prédictions et les vraies valeurs, et une valeur proche de 0 indique une faible adéquation du modèle aux données.



## **Prophet**

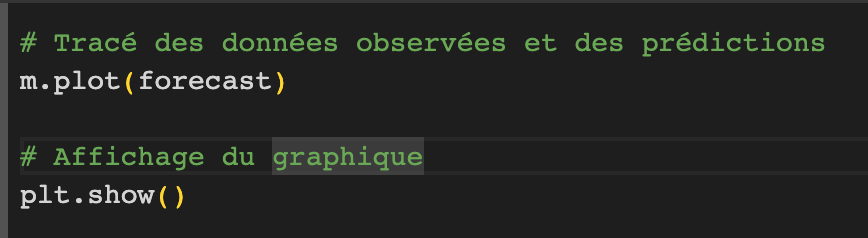


On remarque que l'output est une table contenant les dates des prédictions ainsi que la valeur du cours prédit yhat on remarque également deux autres colonnes yhat\_lower et yhat\_upper qui correspondent à un intervalle de confiance applicable sur notre valeur.

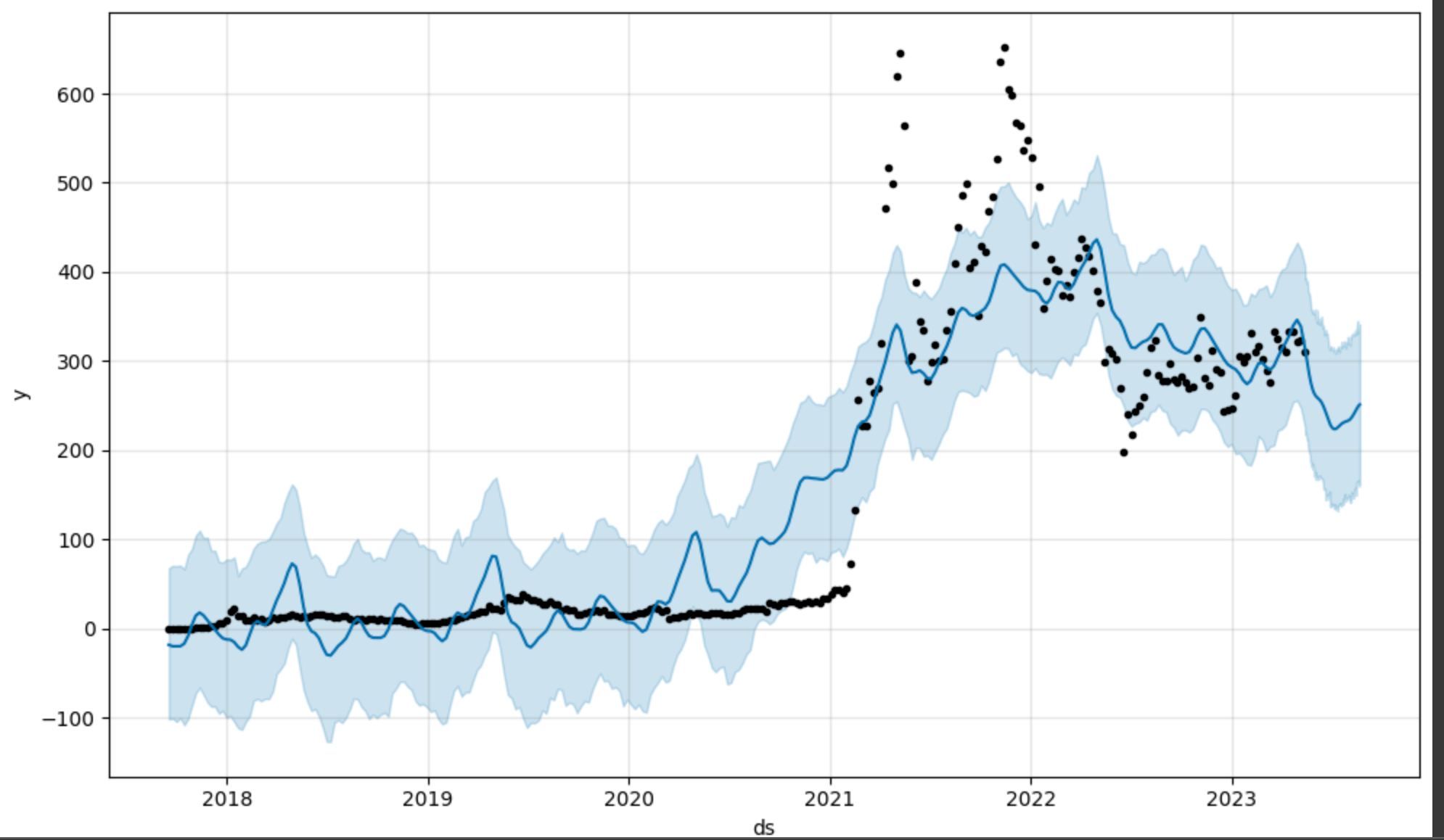
Visualisation de la Prédiction :

Prophet nous permet également de visualiser le cours de la valeur prédite.

Grâce aux commandes :



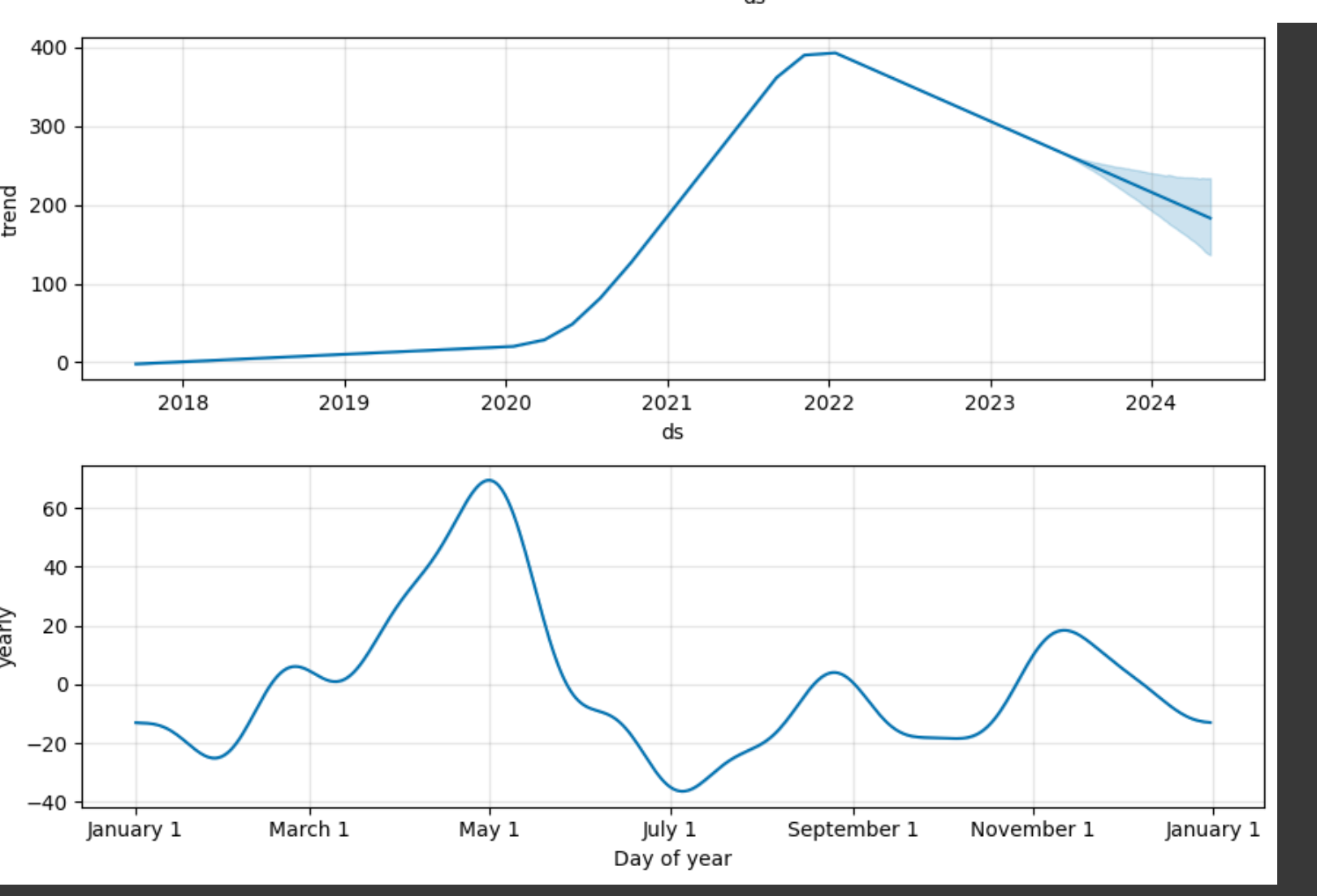
La courbe obtenue montre une vue d'ensemble des valeurs prédites et de la valeurs réelles du cours :



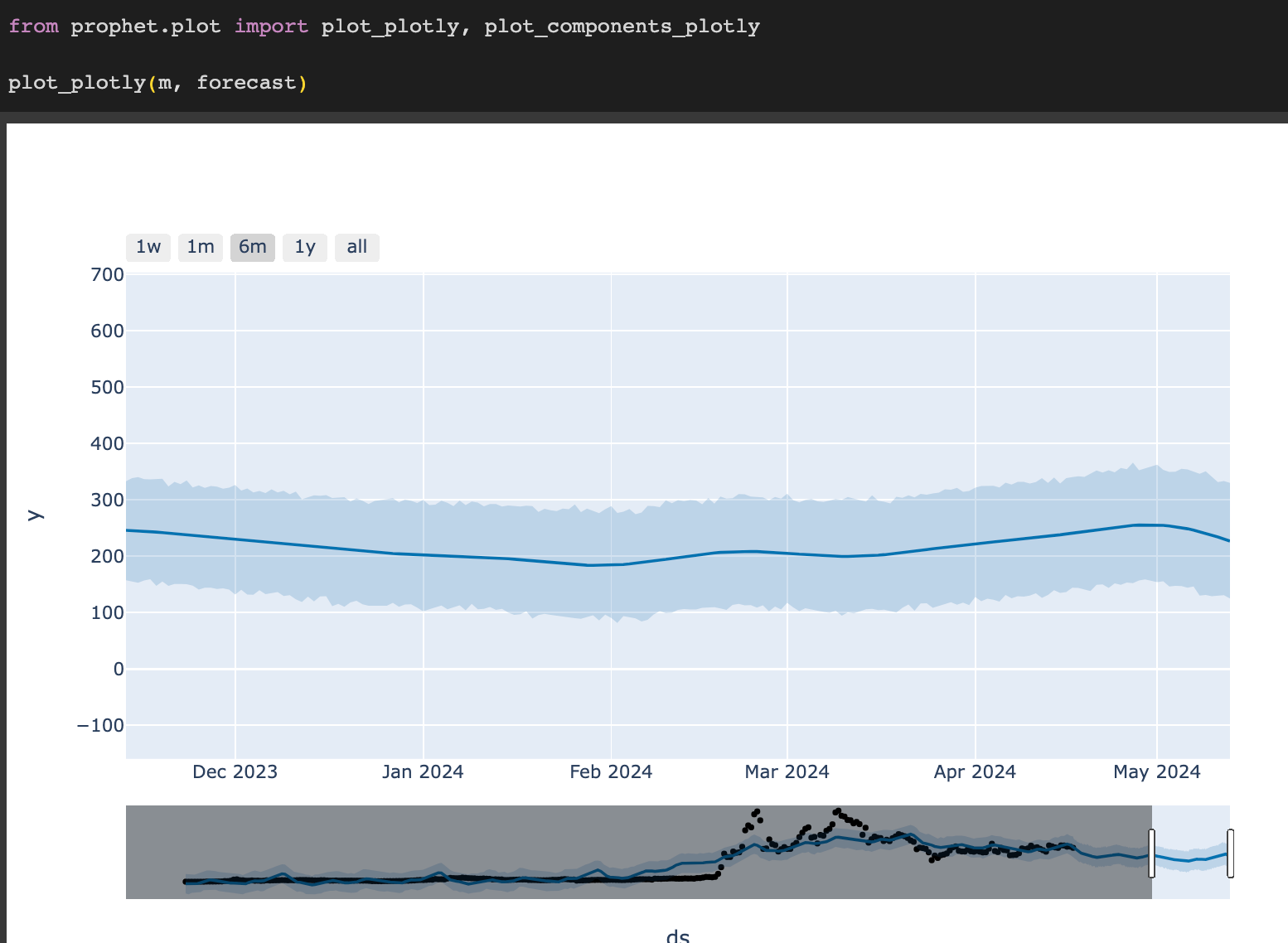
*Les points noirs correspondent aux valeurs réelles.*

**Les commandes** :fig1 = m.plot(forecast)

fig2 = m.plot\_components(forecast) *permettent de se focaliser sur la partie prédictive de notre cours :*

**

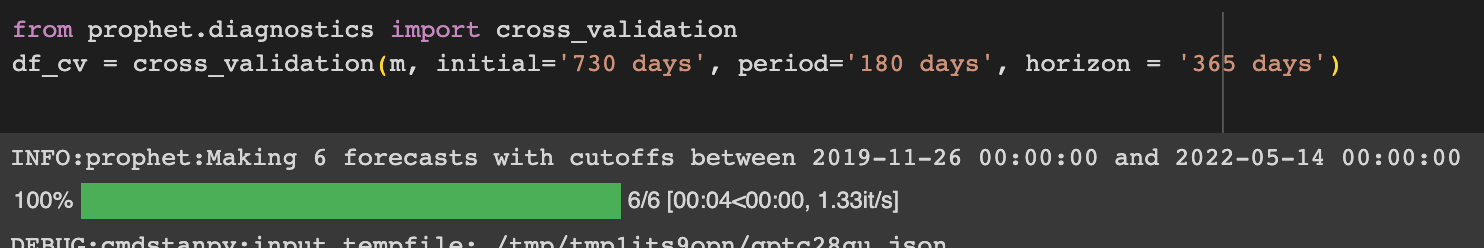
On observe une tendance et un motif saisonnier périodique similaire à notre décomposition. Cependant, on remarque que les valeurs prédites sont plus cohérentes et plus précises qu'avec le modèle ARIMA implémenter plus haut.

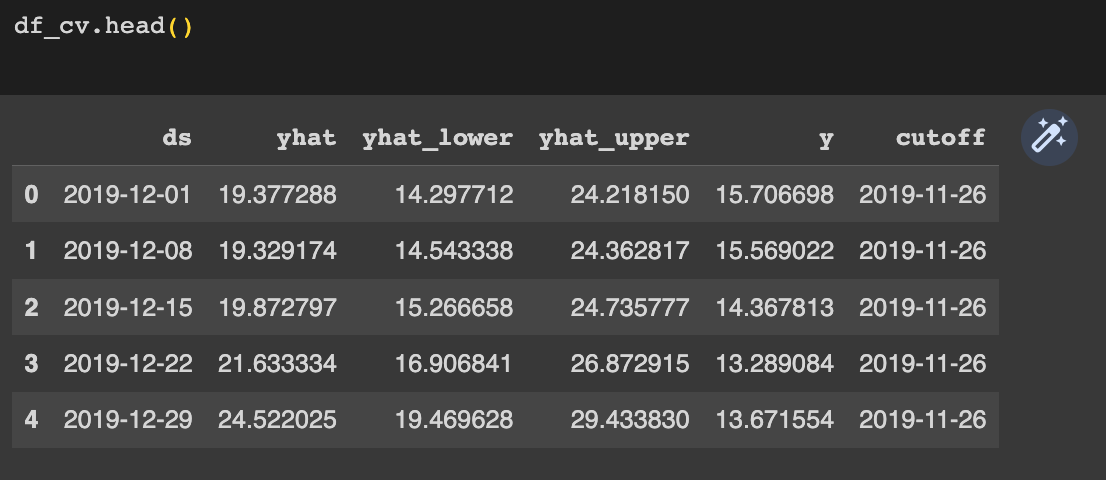


*En passant la souris sur la courbe bleu ça affiche la valeur du jour relative à la date.*

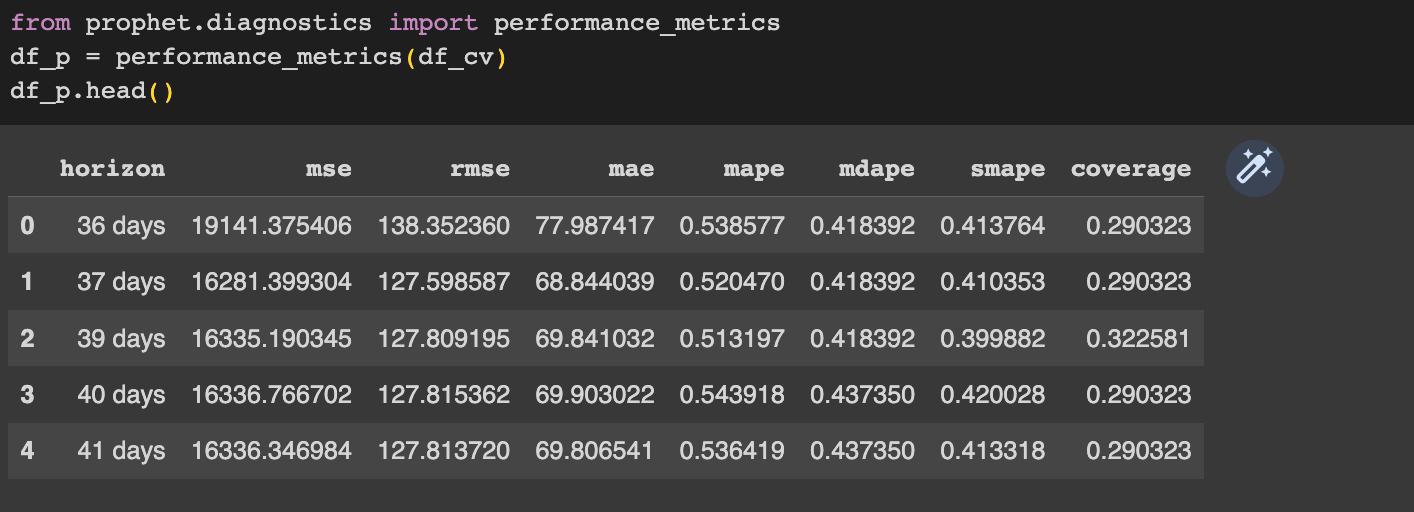
**Critère d'évaluation**

**De même que pour le modèle RNN on peut afficher les métriques d'évaluation**

****

**On obtient alors un dataFrame des mesure de Cross-Validation **

**et avec les commandes :**

****

**On a les métriques d'évaluation relative à différentes plages temporelles**

**En conclusion, le modèle Prophet est un outil puissant pour la prédiction de séries temporelles. Il offre plusieurs avantages, notamment sa facilité d'utilisation, sa capacité à gérer des données manquantes et des valeurs aberrantes, et sa flexibilité pour modéliser différents types de tendances et de saisons. Prophet fournit également des fonctionnalités de diagnostic et d'évaluation de la qualité des prédictions.**

**Cependant, il est important de noter que Prophet est conçu pour des séries temporelles avec des schémas saisonniers réguliers et des tendances non linéaires.**

# Conclusion

En conclusion, notre projet d'analyse du cours du Bnb et de prédiction de ses valeurs a démontré l'importance du machine learning et du deep learning dans ce domaine. Grâce à ces techniques, nous avons pu exploiter les données historiques du cours du Bnb pour effectuer des prédictions et anticiper les mouvements futurs du marché.

Le machine learning nous a permis de construire des modèles basés sur des algorithmes traditionnels tels que ARIMA, qui ont fourni des prédictions en se basant sur des motifs et des tendances passées dans les données. Cependant, nous avons également constaté que les modèles ARIMA ont des limites, notamment en ce qui concerne la capture de relations non linéaires et de motifs complexes.

C'est là que le deep learning a pris toute son importance. En utilisant des architectures de réseaux de neurones profonds, tels que les réseaux de neurones convolutifs (CNN) ou récurrents (RNN), nous avons pu capturer des relations complexes et non linéaires dans les séries temporelles du Bnb. Cela nous a permis d'obtenir des prédictions plus précises en prenant en compte les dépendances à différentes échelles de temps.

Cependant, il est essentiel de souligner les limites de ces approches. La prédiction des valeurs du Bnb reste un défi en raison de la nature volatile et imprévisible des marchés financiers. Les résultats obtenus ne garantissent pas nécessairement des performances futures et doivent être interprétés avec prudence. De plus, la qualité des prédictions dépend de la qualité des données disponibles et de la capacité du modèle à capturer toutes les variations et les facteurs pertinents.

Pour résumer, notre projet a démontré comment le machine learning et le deep learning peuvent être des outils puissants pour la prédiction des valeurs du Bnb. Ces techniques offrent des avantages en termes de capacité à capturer des relations complexes, à apprendre automatiquement les caractéristiques pertinentes et à fournir des prédictions précises. Cependant, il est important de prendre en compte les limites inhérentes à ces approches et d'adopter une approche prudente dans l'interprétation des résultats obtenus.

## **Bibliographie :**

[**https://facebook.github.io/prophet/docs/diagnostics.html#cross-validation**](https://facebook.github.io/prophet/docs/diagnostics.html#cross-validation)

[**https://www.coingecko.com/fr**](https://www.coingecko.com/fr)

[**https://fr.blog.businessdecision.com/technologies/outils-langage-data/**](https://fr.blog.businessdecision.com/technologies/outils-langage-data/)

[**https://machinelearningmastery.com/**](https://machinelearningmastery.com/)