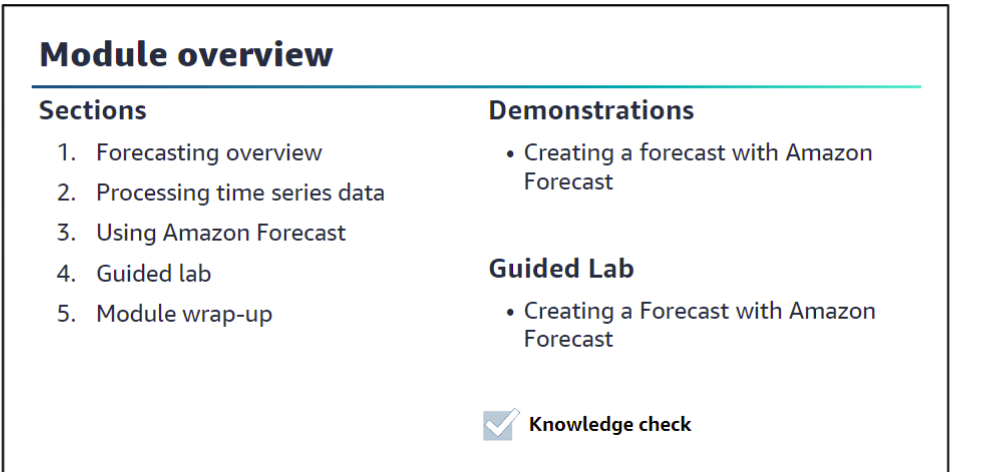
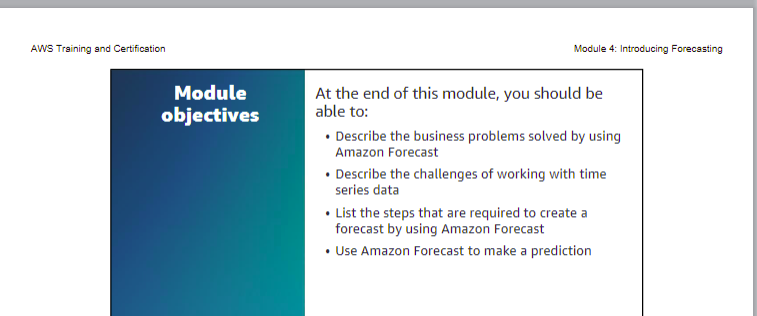
**Module 4: Introduction to Forecasting.**

****

Aperçu des prévisions  
• Traitement des données de séries chronologiques  
• Utilisation d'Amazon Forecast  
• Atelier guidé  
• Conclusion du moduleLe module inclut également un atelier pratique guidé où vous apprendrez à utiliser Amazon Forecast pour travailler avec des données de séries chronologiques[.](https://docs.aws.amazon.com/fr_fr/forecast/latest/dg/what-is-forecast.html)

Enfin, vous devrez compléter un test de connaissances qui évaluera votre compréhension des concepts clés abordés dans ce module.



Après avoir terminé ce module, vous devriez être capable de :

• Décrire les problèmes d'entreprise résolus en utilisant Amazon Forecast

• Décrire les défis liés au travail avec des données de séries chronologiques

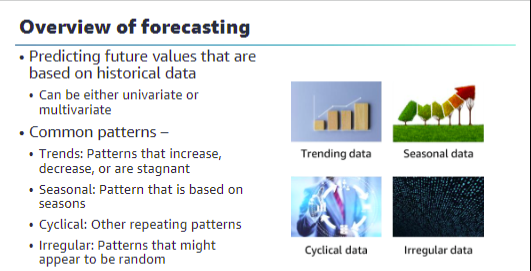
• Énumérer les étapes nécessaires pour créer une prévision en utilisant Amazon Forecast

• Utiliser Amazon Forecast pour faire une prédiction

**Section 1: Forecasting overview.(** **Aperçu des prévisions)**

**Prezentare generală a previziunilor**

*Vous commencez par un examen de ce que signifie la prévision et vous découvrez quelques cas d'utilisation de la prévision*

**

**La prévision est un domaine important de l'apprentissage automatique**. Elle est importante car de nombreuses opportunités de prédiction ***des résultats futurs sont basées******sur des données******historiques****.* Beaucoup de ces opportunités impliquent une **composante temporelle**. Bien que la composante temporelle ajoute plus d'informations, elle rend également les problèmes de séries temporelles plus difficiles à gérer que d'autres types de prédictions.*On peut considérer* ***que les données de séries******temporelles se répartissent en deux grandes catégories****. Le* ***premier type est univarié****, ce qui signifie qu'il n'a qu'une seule variable.* ***Le second type est multivarié***, *ce qui signifie qu'il a plus d'une variable.*

**En plus de ces deux catégories**, la plupart des ensembles de données de séries temporelles suivent également l'un des modèles suivants :

• **Tendance** - Un modèle qui montre les valeurs à mesure qu'elles augmentent, diminuent ou restent les mêmes au fil du temps

• **Saisonnier** - Un modèle répétitif basé sur les saisons d'une année

• **Cyclique** - Une autre forme de modèle répétitif

• **Irrégulier** - Des changements dans les données au fil du temps qui semblent aléatoires ou qui n'ont pas de modèle discernable.

**

Vous pouvez **utiliser la prévision pour une gamme de domaines**. Parmi les applications les plus courantes, on peut citer :

• **Les applications marketing**, telles que la prévision des ventes ou les projections de la demande.

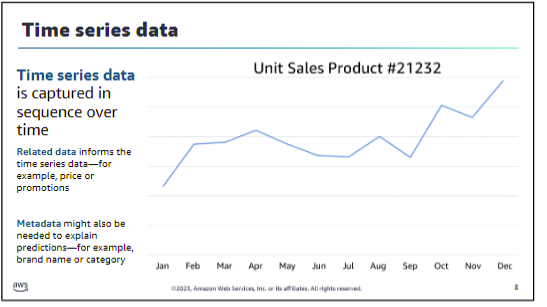
• **Les systèmes de gestion des stocks** pour anticiper les niveaux de stock requis. Souvent, ce type de prévision inclut des informations sur les délais de livraison.

• **La consommation d'énergie** pour déterminer quand et où l'énergie est nécessaire.

**• Les systèmes de prévision météorologique** pour les gouvernements et les applications commerciales telles que l'agriculture**.**

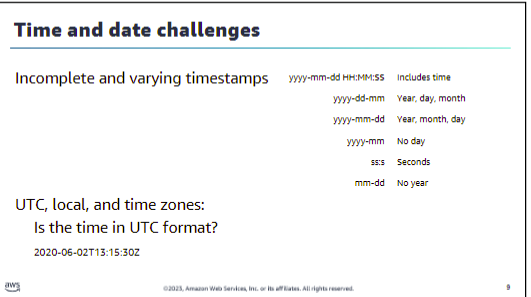
**Section 2 : Traitement des données de séries chronologiques**

Le travail avec des données de séries chronologiques présente plusieurs défis uniques, que vous allez maintenant examiner.



**Les données de séries temporelles** sont capturées dans une séquence chronologique sur une période définie. **L'introduction du** **temps** **dans un modèle d'apprentissage automatique a un impact positif car le modèle peut tirer du sens de l'évolution des points de données au fil du temps.** Les données de séries temporelles ont tendance à être corrélées, ce qui signifie qu'il existe une dépendance entre les points de données. Comme vous avez un **problème de régression** (*trecerea de la stadii mai inalte de dezvoltare la stadii mai inapoiate*)- et parce que **la régression suppose** **l'indépendance des points de données** - vous devez développer une méthode pour gérer la dépendance des données. **Le but** de cette méthode **est d'augmenter la validité des prédictions**. En plus des **données de séries temporelles**, vous **pouvez ajouter** des **données connexes pour enrichir un modèle de prévision**. Par exemple, pour une prédiction sur les ventes au détail, vous pourriez inclure des informations sur le produit vendu (comme l'identification de l'article ou le prix de vente). Ces informations s'ajoutent au nombre d'unités vendues par période.

**Le troisième type de données est les métadonnées sur l'ensemble de données.** Par exemple, pour un ensemble de données de vente au détail, vous pourriez vouloir inclure des **métadonnées** pour regrouper les résultats, comme un nom de marque. **Un autre exemple de métadonnées** pourrait être l'inclusion d'un genre pour la musique ou les vidéos.



**Plus vous avez de données, mieux c'est**. Un défi auquel vous pouvez vous attendre avec de multiples sources de données est l'horodatage des données. Vous trouverez des différences dans le format d'horodatage, ainsi que d'autres défis, tels que des données incomplètes. Dans certains cas, vous pourriez être en mesure de déduire les données manquantes. Par exemple, imaginez que vous ayez des données qui contiennent à la fois le mois et le jour, mais pas l'année. Supposons que les données semblent suivre une séquence à travers les numéros de mois dans la base de données, et se répètent après 12. Dans ce cas, vous pouvez ajouter l'année si vous savez quand les données ont commencé. Vous pouvez déduire les années futures, en fonction de l'ordre des données.

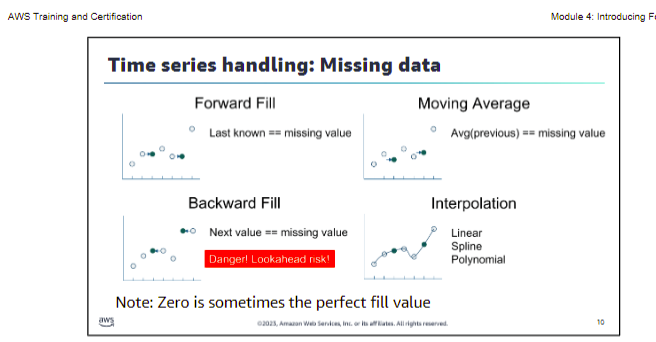
Beaucoup de données sont stockées au format Temps Universel Coordonné (UTC), mais toutes les données ne sont pas en UTC. Vous devriez vérifier si l'horodatage est en heure locale ou universelle.

Parfois, l'horodatage ne représente pas le moment que vous pensez. Par exemple, supposons que vous ayez une base de données de voitures qui ont été entretenues dans un garage. L'horodatage indique-t-il l'heure à laquelle la voiture est arrivée, a été terminée ou a été récupérée ? Ou indique-t-il quand la dernière entrée a été saisie dans le système ?

Si vous essayez de modéliser l'apport calorique horaire des patients, mais que vous n'avez que des données quotidiennes, vous devez alors ajuster votre échelle de temps cible. Il se peut que vous n'ayez pas d'horodatage présent dans vos données.

Vous pourriez avoir d'autres moyens d'extrapoler une série temporelle, en fonction des données et du domaine. Par exemple, vous pourriez avoir des mesures de longueur d'onde ou des vecteurs dans une image.

L'heure d'été est différente dans le monde entier. En raison de l'heure d'été, certaines heures se produisent deux fois par an dans leurs fuseaux horaires.



Une occurrence (**ocazie) fréquente dans les problèmes de prévision** du monde réel **est l'absence de valeurs** dans les données brutes. Les valeurs manquantes rendent plus difficile pour un modèle de générer une prévision. L'exemple principal dans le commerce de détail est une situation de rupture de stock dans la prévision de la demande. Si un article est en rupture de stock, les ventes pour la journée seront nulles. Si la prévision est générée sur la base de ces valeurs de ventes nulles, la prévision sera incorrecte. Les valeurs manquantes peuvent être marquées comme manquantes pour diverses raisons. Elles peuvent survenir en raison de l'absence de transaction, ou éventuellement en raison d'erreurs de mesure. Peut-être qu'un service qui surveillait certaines données ne fonctionnait pas correctement, ou que la mesure n'a pas pu être effectuée correctement. **Les données manquantes peuvent être calculées de plusieurs façons :**

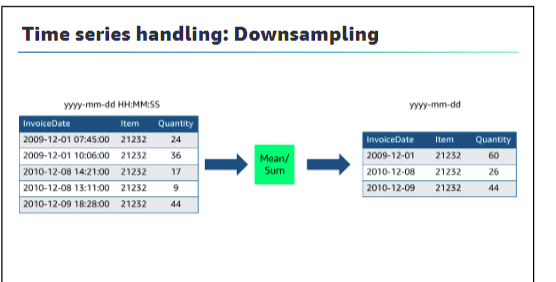
• **Remplissage vers l'avant (Forward fill)** - Utilise la dernière valeur connue pour la valeur manquante.

• **Moyenne mobile** - Utilise la moyenne des dernières valeurs connues pour calculer la valeur manquante.

• **Remplissage vers l'arrière (Backward fill)** - Utilise la prochaine valeur connue après la valeur manquante. *Attention, il est potentiellement dangereux d'utiliser le futur pour calculer le passé, ce qui est mauvais en prévision.* Cette pratique est connue sous le nom de "**lookahead**" **et devrait être évitée**.

• **Interpolation** - Utilise essentiellement une équation pour calculer la valeur manquante.

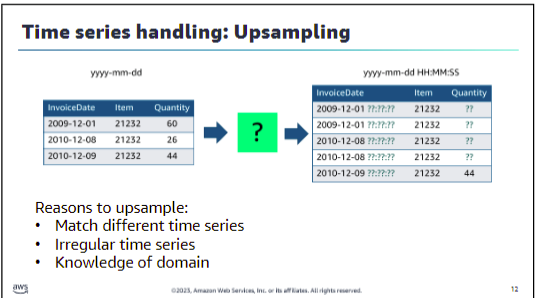
Vous avez également le choix d'utiliser un remplissage par zéro. Ce choix est souvent utilisé dans le commerce de détail, un domaine où les données de ventes manquantes ne devraient pas être calculées. Les données manquantes représentent l'absence de commandes ce jour-là. Il serait judicieux d'enquêter sur les raisons, mais vous ne voulez pas remplir la valeur manquante dans ce cas.



Vous pouvez obtenir des données à différentes fréquences. Par exemple, vous pourriez avoir des données de ventes qui incluent l'horodatage exact de l'enregistrement de la vente. Cependant, les données d'inventaire pourraient ne contenir que l'année, le mois et le jour du niveau d'inventaire. Lorsque vous avez des données qui sont à une fréquence différente des autres ensembles de données, ou qui ne sont pas compatibles avec votre question, *vous pourriez avoir besoin de* ***sous-échantillonner.*** *(****downsample****)*

***Le sous-échantillonnage*** *signifie passer d'un temps plus finement granulaire à un temps moins finement granulaire*. Cet exemple convertit un ensemble de données horaires en un ensemble de données journalières.

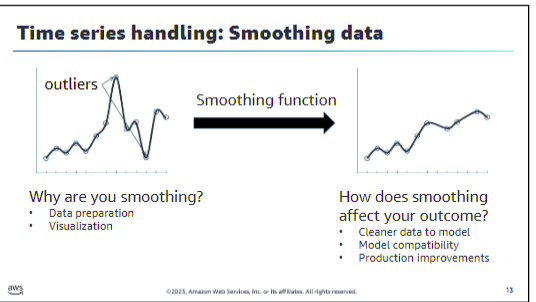
Lorsque vous ***sous-échantillonnez***, *vous devez décider comment combiner les valeurs.* Dans le cas des données de ventes, la somme des quantités est la plus logique. Si les données concernent la température, vous voudrez peut-être trouver la moyenne. Comprendre vos données vous aide à décider quelle est la meilleure ligne de conduite.



**Raisons de sur-échantillonner :(upsample)**

• Faire correspondre différentes séries temporelles  
• Séries temporelles irrégulières  
• Connaissance du domaine

*L'inverse du sous-échantillonnage* ***est* le sur-échantillonnage**. Le problème avec **le sur-échantillonnage** *est qu'il est difficile à* *réaliser dans la plupart des cas*. Supposons que vous vouliez sur-échantillonner vos données de ventes de ventes quotidiennes à ventes horaires. À moins d'avoir une autre source de données à consulter, vous ne pourriez pas passer des ventes quotidiennes aux ventes horaires. Dans certains cas, vous devez utiliser des données ou des connaissances supplémentaires. Par exemple, si vous devez faire correspondre la fréquence d'une autre série temporelle, vous pourriez avoir une série temporelle irrégulière ou des connaissances spécifiques du domaine qui pourraient aider. Dans ces cas, vous devez être prudent sur la façon dont vous convertissez les données. Pour l'exemple du commerce de détail, le mieux que vous puissiez faire est de créer une seule commande pour la journée à une heure spécifiée. Pour la température, vous pouvez copier la température quotidienne dans chacun des créneaux horaires, ou utiliser une formule pour calculer une courbe**.**

****

*Les valeurs aberrantes peuvent être un problème en science des données. Il en va de même pour les données de séries temporelles*.

Si vous examinez des données de ventes et que vous voyez une commande avec une quantité inhabituellement élevée d'articles, vous pourriez ne pas vouloir inclure cette commande dans vos calculs de prévision. La taille de la commande pourrait ne jamais se répéter. La suppression de ces valeurs aberrantes et anomalies est connue sous le nom de lissage.( ***smoothing***)

*Le lissage (****smoothing****) de vos données peut vous aider à gérer les valeurs aberrantes et autres anomalies*.

**Vous pourriez envisager le lissage (*smoothing*) pour les raisons suivantes :**

• *Préparation des données* - Suppression des valeurs erronées et des valeurs aberrantes  
• *Visualisation* - Réduction du bruit dans un graphique.

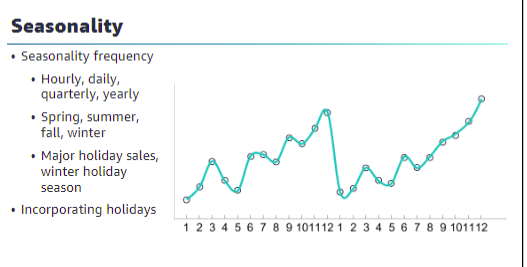
Comprenez pourquoi vous lissez les données et l'impact que cela pourrait avoir. Vous pourriez vouloir que le résultat soit une réduction du bruit et la création d'un meilleur modèle.

**Cependant, il est tout aussi important de considérer ces questions :**

*Votre lissage pourrait-il compromettre le modèle ?*

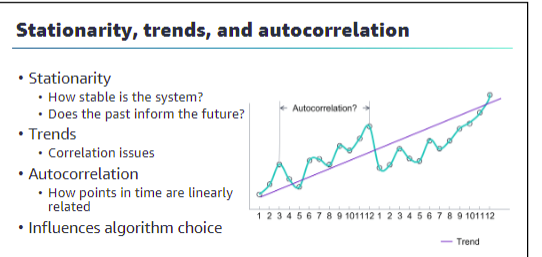
*Le modèle s'attend-il à des données bruitées ?*

*Pouvez-vous également lisser les données en production ?*



**La saisonnalité** dans les données est toute sorte d'observation répétitive où la fréquence de l'observation est stable. Par exemple, dans les ventes, on observe généralement des ventes plus élevées à la fin d'un trimestre, et au cours du quatrième trimestre. Le commerce de détail grand public présente ce schéma de manière encore plus prononcée au quatrième trimestre. Sachez que les données peuvent avoir plusieurs types de saisonnalité dans le même ensemble de données.

Souvent, vous pourriez vouloir incorporer des informations de saisonnalité dans votre prévision. Les fêtes locales sont un bon exemple pour les ventes.



**Il est important de connaître le niveau de stabilité d'un système**. **Le niveau de stabilité**, ***ou stationnarité***, *peut vous indiquer dans quelle mesure vous devez vous attendre à ce que le comportement passé du système informe son comportement futur*. **Un système avec une faible stabilité n'est pas bon pour prédire l'avenir.**

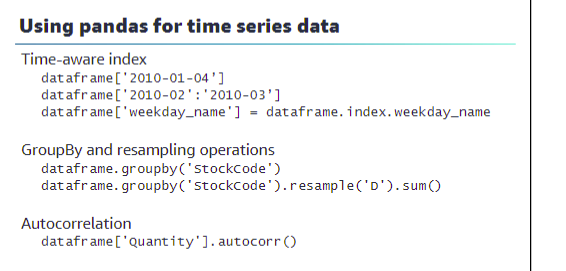
Souvent, vous voudrez déterminer la tendance d'une série temporelle. Cependant, ajuster la série pour la tendance peut rendre difficile la comparaison de la série avec une autre série qui a également été ajustée pour la tendance. Les tendances peuvent dominer les valeurs de la série, ce qui peut vous amener à surestimer la corrélation entre les deux séries. Ce phénomène a été montré dans le sujet précédent.

L'autocorrélation est l'un des problèmes spécifiques auxquels vous êtes confronté avec les données de séries temporelles. Comme vous l'avez vu dans d'autres problèmes d'apprentissage automatique, **l'objectif de la construction d'un modèle ML est de séparer le signal du bruit. L'autocorrélation est une forme de bruit car les observations séparées ne sont pas indépendantes les unes des autres.** Une série temporelle avec autocorrélation pourrait surestimer la précision du modèle produit. Certains des algorithmes que vous verrez dans ce module peuvent aider à corriger l'autocorrélation.

**Les corrélations ne signifient pas causalité**.

**Soyez prudent** ***lorsque vous interprétez vos propres données***, et **soyez prudent** ***avec les corrélations*** - vous ne voulez pas agir sur des corrélations qui n'ont pas de signification dans le monde réel. À titre d'expérience, disons que vous générez deux ensembles de données de séries temporelles aléatoires de nombres entre 0 et 1. Vous constaterez qu'ils ont une faible corrélation. ***Cependant, si*** ***vous introduisez la même pente dans les deux ensembles de données, vous verrez une forte corrélation.***

Ces facteurs, ainsi que la saisonnalité, peuvent influencer le modèle que vous sélectionnez pour produire votre prévision. Certains algorithmes gèrent la saisonnalité et l'autocorrélation, mais d'autres non.



***La bibliothèque pandas*** *a été développée en pensant à l'analyse de données financières. En tant que telle, elle est efficace pour gérer les données de séries temporelles.*

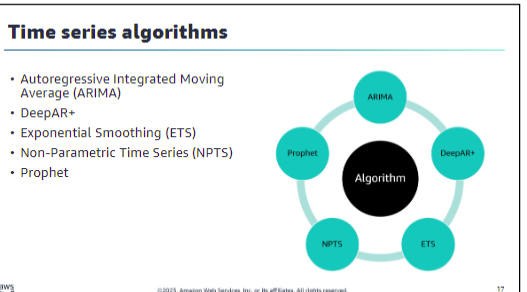
Vous pouvez *définir l'index de votre DataFrame pandas comme étant un datetime, ce qui vous permet d'utiliser la date et l'heure pour sélectionner vos données*. Vous pouvez utiliser des plages qui contiennent des dates partielles. Vous pouvez également extraire des parties de dates, telles que l'année, le mois, le nom du jour de la semaine, et plus encore.

Pour les tâches de regroupement et de rééchantillonnage, pandas dispose de fonctions intégrées pour faire les deux.

Enfin, pandas peut vous donner des informations sur l'autocorrélation.

For more information about pandas and time series data, see “Time Series/Date Functionality” in the pandas documentation at

https://pandas.pydata.org/pandas-docs/stable/user\_guide/timeseries.html



*L'une des tâches dans la construction d'une application de prévision est de choisir un algorithme approprié. Le type d'ensemble de données que vous utilisez et les caractéristiques de cet ensemble de données devraient déterminer votre choix d'algorithme.*

**Amazon Forecast prend en charge ces cinq algorithmes** :

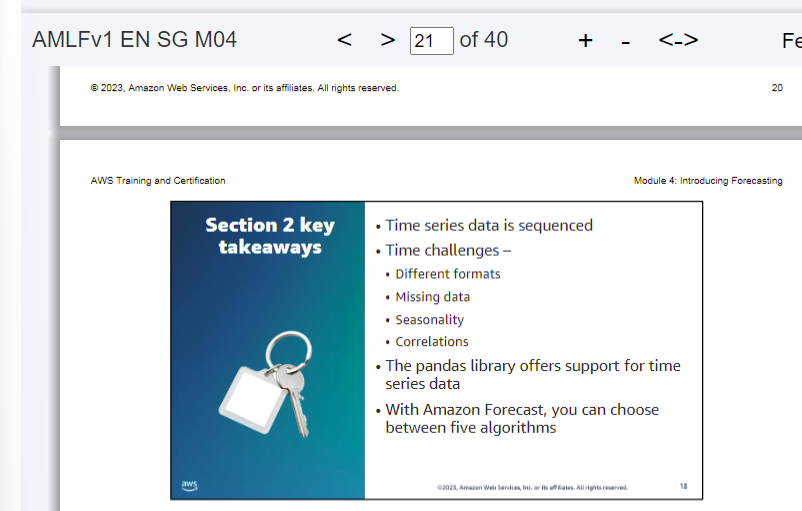
• **Moyenne mobile intégrée autorégressive (ARIMA)** - Cet **algorithme** ***supprime les autocorrélations***, qui pourraient influencer le modèle d'observations.

• **DeepAR+** - Un **algorithme** ***d'apprentissage supervisé pour la prévision de séries temporelles unidimensionnelles.*** Il utilise un réseau de neurones récurrent pour entraîner un modèle sur plusieurs séries temporelles.

• **Lissage exponentiel (ETS)** : Cet ***algorithme est utile pour les ensembles de données avec saisonnalité.*** Il utilise une moyenne pondérée pour toutes les observations. Les poids sont diminués au fil du temps.

• **Séries temporelles non paramétriques (NPTS)** - ***Les prédictions sont basées sur l'échantillonnage d'observations passées***. Des versions spécialisées sont disponibles pour les ensembles de données saisonniers et climatologiques.

• **Prophet** - Un modèle de série temporelle bayésien. Il est utile pour les ensembles de données qui s'étendent sur une longue période, ont des données manquantes ou de grands écarts



*Quelques points clés à retenir de cette section du module incluent* :

• **Les données de séries temporelles** sont des données séquencées qui incluent un élément temporel, ce qui les rend différentes des ensembles de données réguliers

• Certains des défis liés au temps comprennent :

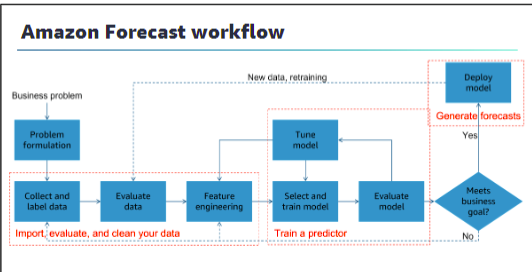
* *La gestion de différents formats de temps*
* *La gestion des données manquantes par* **sous**-**échantillonnage**, **sur-échantillonnage** et **lissage(*smoothing)***
* La gestion de la **saisonnalité**, comme les jours de la semaine et les **cycles** annuels
* L'évitement des mauvaises corrélations

• La bibliothèque pandas offre un support pour les données de séries temporelles grâce à des fonctions qui traitent du temps

• Avec Amazon Forecast, vous pouvez choisir entre cinq algorithmes :

* ARIMA
* DeepAR+
* ETS
* NPTS
* Prophet

**Section 3: Using Amazon Forecast.**

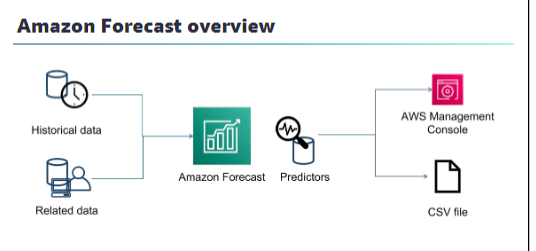


*Lorsque vous générez des prévisions, vous pouvez appliquer le pipeline de développement d'apprentissage automatique que vous utilisez tout au long de ce cours.*

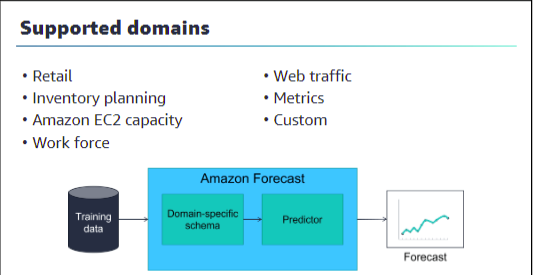
• *Importer vos données* - Vous devez importer autant de données que vous en avez - à la fois des données historiques et des données connexes. Vous devriez effectuer une évaluation de base et une ingénierie des caractéristiques avant d'utiliser les données pour entraîner un modèle.

• *Entraîner un prédicteur* - Pour entraîner un prédicteur, vous devez choisir un algorithme. Si vous n'êtes pas sûr de l'algorithme le mieux adapté à vos données, vous pouvez laisser Amazon Forecast choisir en sélectionnant AutoML comme algorithme. Vous devez également sélectionner un domaine pour vos données, mais si vous n'êtes pas sûr du domaine le plus approprié, vous pouvez sélectionner un domaine personnalisé. Les domaines ont des types de données spécifiques qu'ils requièrent. Pour plus d'informations, consultez la section Domaines de jeux de données prédéfinis et types de jeux de données dans la documentation d'Amazon Forecast.

• *Générer des prévisions* - Dès que vous avez un modèle entraîné, vous pouvez l'utiliser pour faire une prévision en utilisant un groupe de jeux de données d'entrée. Après avoir généré une prévision, vous pouvez interroger la prévision, ou vous pouvez l'exporter vers un compartiment Amazon Simple Storage Service (Amazon S3). Vous avez également la possibilité de chiffrer les données de la prévision avant de les exporter.



Le processus global pour travailler avec Amazon Forecast consiste à importer des données historiques et connexes. Amazon Forecast inspecte les données, identifie les données clés et sélectionne un algorithme approprié. Il utilise l'algorithme pour entraîner et optimiser un modèle personnalisé et produire un prédicteur. Vous créez des prévisions en appliquant le prédicteur à votre ensemble de données. Ensuite, vous pouvez soit récupérer ces prévisions dans la console AWS, soit les exporter sous forme de fichiers délimités par des virgules. Vous pouvez également utiliser une **interface de programmation** **d'application (API**) et des **commandes d'interface en ligne de** **commande (CLI)** pour créer et récupérer des prévisions.



Lorsque vous travaillez avec Amazon Forecast, vous devez sélectionner le domaine approprié.

Vous pouvez choisir parmi la liste suivante :

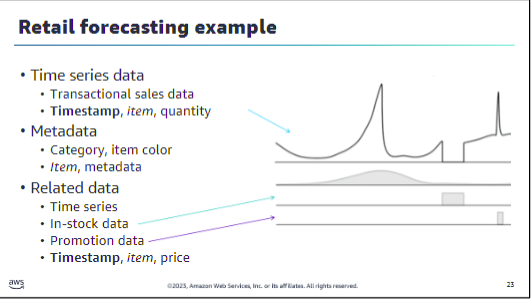
• *Vente au détail* - Demande de produits  
• *Planification des stocks* - Besoins en matières premières  
• *Capacité EC2* - Demande de capacité pour Amazon Elastic Compute Cloud (Amazon EC2)  
• *Main-d'œuvre* - Projections de charge de travail  
• *Trafic Web* - Trafic projeté vers un ou plusieurs sites Web  
• *Métriques* - Projection de métriques telles que les revenus, les ventes ou les flux de trésorerie  
• *Personnalisé* - Projections pour un domaine que vous ne pouvez pas associer à l'un des domaines précédents

En sélectionnant un domaine, vous améliorez l'efficacité du prédicteur. Chaque domaine a des types de données spécifiques que vous fournissez lors de la construction du prédicteur. Par exemple, le domaine Vente au détail attend des identifiants d'articles, un horodatage pour l'observation, le nombre de ventes pour cet article et l'horodatage spécifié

**For more information** on the supported domains, see the topic “Predefined Dataset Domains and Dataset Types” at

https://docs.aws.amazon.com/forecast/latest/dg/howitworks-

domains-ds-types.html.



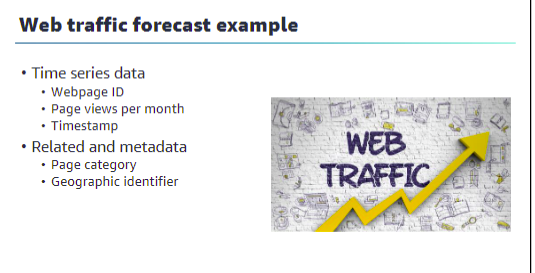
*L'exemple suivant montre les données dont vous avez besoin pour une prévision de la demande dans le commerce de détail.*

**Pour la série temporelle, vous devez fournir** :

• **Horodatage (timestamp)-** Le moment où la transaction a eu lieu, idéalement au format UTC  
• **Article(item)** - L'identifiant de l'article  
• **Quantité** - Combien d'articles ont été vendus

Les métadonnées de l'article peuvent inclure, par exemple, la catégorie ou la couleur de l'article. Le lien de retour vers les données de la série temporelle n'inclut que l'identifiant de l'article, car les métadonnées de l'article ne changent généralement pas.

Le prix de vente ou d'autres données promotionnelles sont des exemples de données connexes qui peuvent créer une prévision plus utile. Pour lier cela à l'article, vous devez inclure l'horodatage et l'identifiant de l'article ID.

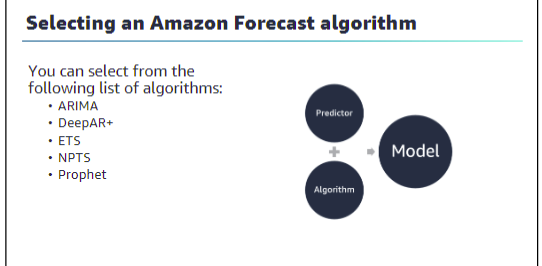


L'exemple suivant montre le type de données dont vous avez besoin pour une prévision de trafic web.

Pour la série temporelle, vous devez fournir :  
• Identifiant de la page web  
• Nombre de vues de la page par mois  
• Horodatage

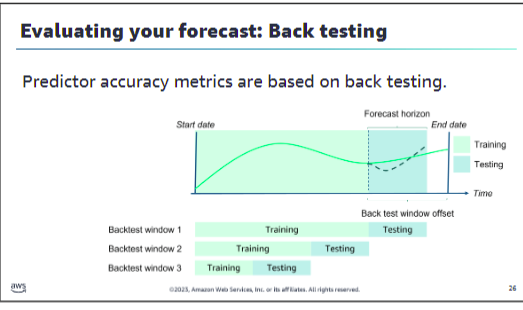
Les données connexes qui créent une prévision plus utile comprennent :  
• Catégorie de la page (comme navigation ou catégorie de contenu)  
• Identifiant géographique pour le client web.

Vous pourriez également avoir besoin des métadonnées suivantes :  
• Région  
• Informations sur les promotions de vente



Les prédicteurs Amazon Forecast utilisent un algorithme pour entraîner un modèle. Ils utilisent ensuite le modèle pour faire une prévision en utilisant un groupe de jeux de données d'entrée. Pour vous aider à démarrer, Amazon Forecast fournit les algorithmes prédéfinis suivants :• ARIMA  
• DeepAR+  
• ETS  
• NPTS  
• ProphetVous pouvez également utiliser la fonction AutoML, qui essaie tous les algorithmes pour voir lequel est le meilleur pour prédire les données.

For more information about Amazon Forecast algorithms, see Choosing an “Amazon Forecast Algorithms” in the AWS documentation at https://docs.aws.amazon.com/forecast/latest/dg/aws-forecast-choosing-recipes.html.



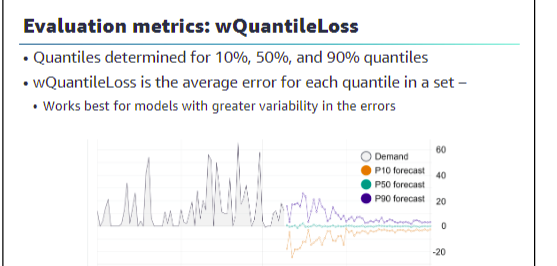
Lorsque vous préparez des données pour l'entraînement en apprentissage automatique, vous réservez généralement des données à utiliser pour valider et évaluer le modèle. Les données que vous réservez sont habituellement un échantillon aléatoire de vos données disponibles. Avec les données de séries temporelles, vous devez traiter vos données différemment en raison d'une corrélation avec le temps.

Lorsque vous importez vos données, Amazon Forecast les divise en ensembles de données d'entraînement et de test, comme le montre le diagramme. Les données d'entraînement sont utilisées pour entraîner le modèle, qui est ensuite testé par rapport aux données qui ont été mises de côté.

Vous pouvez spécifier plusieurs fenêtres de rétro-test, qui diviseront les données plusieurs fois, entraîneront le modèle et utiliseront des métriques pour déterminer quel modèle donne les meilleurs résultats. La fenêtre de rétro-test par défaut est de 1.

Vous pouvez modifier la façon dont Amazon Forecast divise les données en définissant le paramètre BackTestWindowOffset lors de la création du prédicteur. Si vous ne définissez pas cette valeur, les algorithmes utilisent des valeurs par défaut.

Après avoir entraîné un modèle, vous devrez mesurer sa précision, ce que vous apprendrez ensuite.



*Quantiles déterminés pour les quantiles 10%, 50% et 90%  
• wQuantileLoss est l'erreur moyenne pour chaque quantile dans un ensemble  
• Fonctionne mieux pour les modèles avec une plus grande variabilité dans les erreurs*

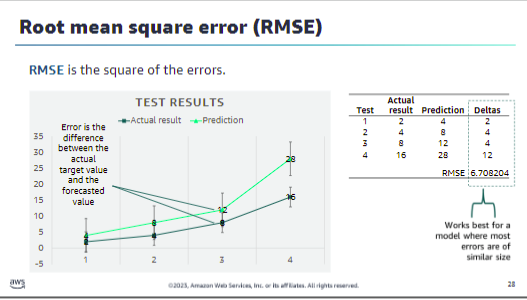
La première métrique d'évaluation d'Amazon Forecast est la perte quantile pondérée (wQuantileLoss). Lorsqu'Amazon Forecast crée une prévision, il fournit des prédictions probabilistes à trois quantiles distincts - 10 pour cent, 50 pour cent et 90 pour cent. Ces quantiles de prédiction vous montrent le niveau d'incertitude associé à chaque prévision.

Un quantile P10 prédit que, 10 pour cent du temps, la valeur réelle sera inférieure à la valeur prédite. Par exemple, supposons que vous êtes un détaillant. Vous voulez prévoir la demande de produits pour des gants d'hiver qui se vendent bien uniquement pendant l'automne et l'hiver. Disons que vous n'avez pas suffisamment d'espace de stockage et que le coût du capital investi est élevé, ou que le prix d'être surstocké en gants d'hiver vous préoccupe. Alors, vous pourriez utiliser le quantile P10 pour commander un nombre relativement faible de gants d'hiver. Vous savez que la prévision P10 surestime la demande de vos gants d'hiver seulement 10 pour cent du temps, donc vous serez en rupture de stock de vos gants d'hiver 90 pour cent du temps.

Un quantile P50 prédit que 50 pour cent du temps, la valeur réelle sera inférieure à la valeur prédite. En continuant l'exemple des gants d'hiver, disons que vous savez qu'il y aura une demande modérée pour les gants, et que vous n'êtes pas préoccupé par le surstockage. Alors, vous pourriez choisir d'utiliser le quantile P50 pour commander des gants.

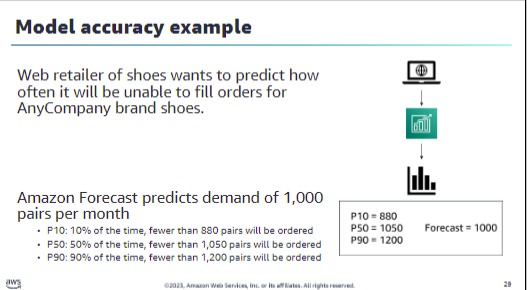
Un quantile P90 prédit que 90 pour cent du temps, la valeur réelle sera inférieure à la valeur prédite. Supposons que vous déterminiez que le fait d'être en rupture de stock de gants entraînera d'importantes pertes de revenus - par exemple, le coût de ne pas vendre de gants est extrêmement élevé ou le coût du capital investi est faible. Dans ce cas, vous pourriez choisir d'utiliser le quantile P90 pour commander des gants.

Amazon Forecast calcule également la perte (erreur) associée à chaque quantile. La perte quantile pondérée (wQuantileLoss) calcule à quel point la prévision d'un certain quantile s'écarte de la demande réelle dans les deux sens. Des métriques wQuantileLoss plus basses signifient que les prévisions du modèle sont plus fiables.



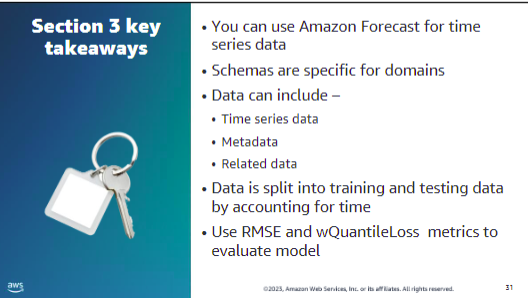
L'erreur quadratique moyenne (RMSE - Root Mean Square Error) est une autre méthode pour évaluer la fiabilité de vos prévisions. Comme wQuantileLoss, RMSE calcule l'écart entre les valeurs prévues et les données de test réelles.

La RMSE trouve la différence entre la valeur cible réelle dans l'ensemble de données et la valeur prévue pour cette période, puis elle élève au carré ces différences. L'exemple montre comment calculer la RMSE. La valeur RMSE représente l'écart-type des erreurs de prédiction. Ce test est bon pour la validité des prévisions lorsque les erreurs sont principalement de la même taille (c'est-à-dire qu'il n'y a pas beaucoup de valeurs aberrantes). Des métriques RMSE plus basses indiquent que les prévisions du modèle sont plus fiables.



Cet exemple montre comment un détaillant en ligne pourrait utiliser les métriques de précision pour évaluer une prévision. Le détaillant souhaite prédire la demande pour les ventes d'une marque particulière de chaussures. Il saisit les registres de ventes de cette marque dans Amazon Forecast pour créer un prédicteur.

Le prédicteur fournit une demande prévue de 1 000 paires avec les valeurs P10, P50 et P90. Les valeurs wQuantileLoss indiquent que 10 pour cent du temps (P10), moins de 880 paires seront vendues. Ensuite, 50 pour cent du temps (P50), moins de 1 050 paires seront vendues. Enfin, 90 pour cent du temps (P90), moins de 1 200 paires seront vendues. Le détaillant peut alors utiliser ces valeurs pour déterminer quel niveau de stock conserver. La détermination est basée sur leur évaluation du risque de ne pas pouvoir satisfaire les commandes, ou d'avoir un excès de stock.



*Quelques points clés à retenir de cette section du module incluent :*

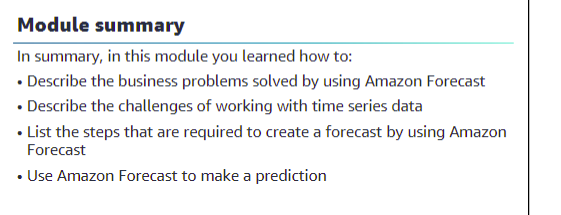
• Vous pouvez utiliser Amazon Forecast pour entraîner et utiliser un modèle pour les données de séries temporelles

• Il existe des schémas spécifiques définis pour des domaines tels que la vente au détail et la planification de la capacité EC2, ou vous pouvez utiliser un schéma personnalisé

• Vous devez fournir au minimum les données de séries temporelles, mais vous pouvez également fournir des métadonnées et des données connexes pour ajouter plus d'informations au modèle

• Comme pour la plupart des problèmes d'apprentissage automatique supervisé, vos données sont divisées en données d'entraînement et de test, mais cette division prend en compte l'élément temporel

• Utilisez les métriques RMSE et wQuantileLoss pour évaluer l'efficacité du modèle



En résumé, dans ce module, vous avez appris à :

• Décrire les problèmes commerciaux résolus en utilisant Amazon Forecast  
• Décrire les défis liés au travail avec des données de séries temporelles  
• Énumérer les étapes nécessaires pour créer une prévision en utilisant Amazon Forecast  
• Utiliser Amazon Forecast pour faire une prédiction  
