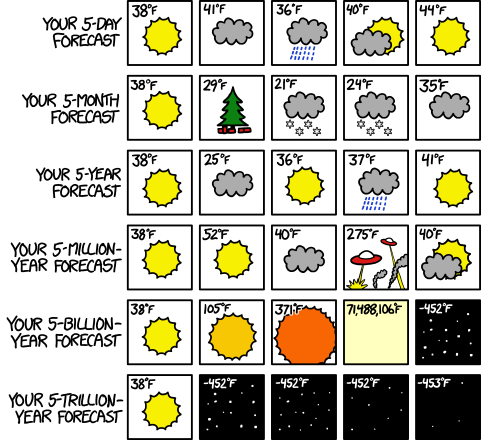


**Time series forecasting**



Your 5 day forecast

*Source:* [*xkcd*](https://xkcd.com/1606/)

**Objectifs du module**

**A basic introduction to time series analysis and forecasting**

* Understanding the challenges of time series data analysis
* Understand how time series prediction steps are different from other machine learning predictions
* Use conventional linear time series models to make univariate forecasts
* Use a time series forecasting library for univariate forecasts

**Démarche pédagogique**

* Durée du projet : 3 jours
* Work alone but exchange as much as possible between you!
* Produire vos propres scripts et mémos individuels pour terminer le projet

**Contexte**

Dans le monde numérique actuel, nous collectons des quantités incroyables de données. Tout ce que nous créons dans le monde numérique est horodaté: de nos interactions, clics, achats en ligne, à nos moniteurs de santé comme les «fitbits», à notre position GPS à partir de nos téléphones, à notre «selfie» le plus récent... Plus les capteurs deviennent bon marché et plus nous les intégrons dans nos vies. Nous collectons ainsi de nombreux points de données pour construire des séries chronologiques de notre vie et de notre société. Si certaines données que vous avez actuellement ne sont PAS une série chronologique, elles le seront très certainement dans un avenir proche.

Par extension, les tentatives de prédictions sur l'avenir se multiplient, aussi qualifiées de prévisions. Ce podcast «[Les prévisions sont toujours fausses (mais nous en avons besoin quand même)](https://www2.monash.edu/impact/podcasts/thought-capital/forecasts-are-always-wrong-but-we-need-them-anyway/)» explique à quoi ressemblent les prévisions modernes. Dans la prévision, la terminologie est particulièrement fournie et un temps certain peut être requis afin de comprendre l’ensemble des méthodes utilisées, leurs spécificités utiles pour différents types de données. Le but de ce module est de vous faire utiliser python pour faire des prévisions robustes et de commencer à appréhender la terminologie utilisée dans la prévision des séries temporelles.

**Etape 1 (1 jour) - L’analyse de séries temporelles à but de prédiction**

**Objectifs de l’activité**

* Découvrir ou redécouvrir les notions de base derrière l’analyse de séries temporelles.
* Initier sa propre boîte à outils d’analyse de données de type série temporelle.

**Compétences**

* Développer une approche méthodologique lors de l’analyse de séries temporelles
* Utiliser des librairies classiques à ce titre
* De la théorie, de la théorie et… du code!

**Consignes**

* Effectuer les tâches définies dans le jupyter notebook 2023\_Campus\_ML6-NB1.ipynb. Ce NB vous servira de guide pour construire votre méthodologie afin d’appréhender au mieux toute série temporelle.
* Réaliser un logigramme des étapes à effectuer afin d’appréhender au mieux toute série temporelle sur laquelle s’appuyer pour faire des prévisions.

**Livrables**

* Le NB complété et déposé sur l’espace github prévu à cet effet:
  + NB 2023\_Campus\_ML6-NB2.ipynb renommé avec vos Nom et Prénom mentionné → 2023\_NB1\_FrançoisRobert.ipynb.

***Ressources (implement with your own!)***

* [*https://www.kaggle.com/code/prashant111/complete-guide-on-time-series-analysis-in-python*](https://www.kaggle.com/code/prashant111/complete-guide-on-time-series-analysis-in-python)
* *…*

Si vous avez fini, un peu de préparation pour demain : [Une petite vidéo!](https://www.youtube.com/watch?v=ZoJ2OctrFLA)

**Etape 2 (2 jours) - ARIMA et compagnie**

**Objectifs de l’activité**

* Comprendre en quoi les étapes de prévision des séries temporelles sont différentes des autres prédictions d'apprentissage automatique.
* Comprendre et utiliser un modèle classique de séries temporelles linéaires, et ses déclinaisons, pour faire des prévisions univariées.

**Compétences**

* Définir une méthodologie typique utilisée lors de la réalisation de prévisions de séries temporelles.
* Utiliser la librairie pmdarima…du code, du code, du code!

**Consignes**

* Effectuer les tâches définies dans le jupyter notebook 2023\_Campus\_ML6-NB2.ipynb. Ce NB vous guidera à travers le processus d'utilisation des modèles de base ARIMA sur les données de séries temporelles pour faire des prévisions.

**Livrables**

* Un mémo qui explique (dans vos propres mots) comment le fractionnement des trains / tests et la validation croisée des données de séries chronologiques doivent être effectués.
* Un logigramme des étapes à effectuer afin d’appréhender au mieux toute série temporelle sur laquelle s’appuyer pour faire des prévisions (Choose your tool!).
* NB 2023\_Campus\_ML6-NB2.ipynb fini et déposé sur l’espace github crée à cet effet avec vos Nom et Prénom mentionné tels que 2023\_NB2\_FrançoisRobert.ipynb.

***Ressources***

* *Former les fractionnements des tests et la validation croisée dans les séries chronologiques :*

[*https://stats.stackexchange.com/questions/14099/using-k-fold-cross-validation-for-time-series-model-selection*](https://stats.stackexchange.com/questions/14099/using-k-fold-cross-validation-for-time-series-model-selection)[*https://robjhyndman.com/hyndsight/tscv/*](https://robjhyndman.com/hyndsight/tscv/)

* *Forecasting: Principles and Practice livre (Rob Hyndman) :*[*https://otexts.com/fpp2/*](https://otexts.com/fpp2/)[*https://otexts.com/fpp2/non-seasonal-arima.html*](https://otexts.com/fpp2/non-seasonal-arima.html)[*https://otexts.com/fpp2/transformations.html*](https://otexts.com/fpp2/transformations.html)
* *Documentation de Pmdarima:*[*http://alkaline-ml.com/pmdarima/*](http://alkaline-ml.com/pmdarima/)