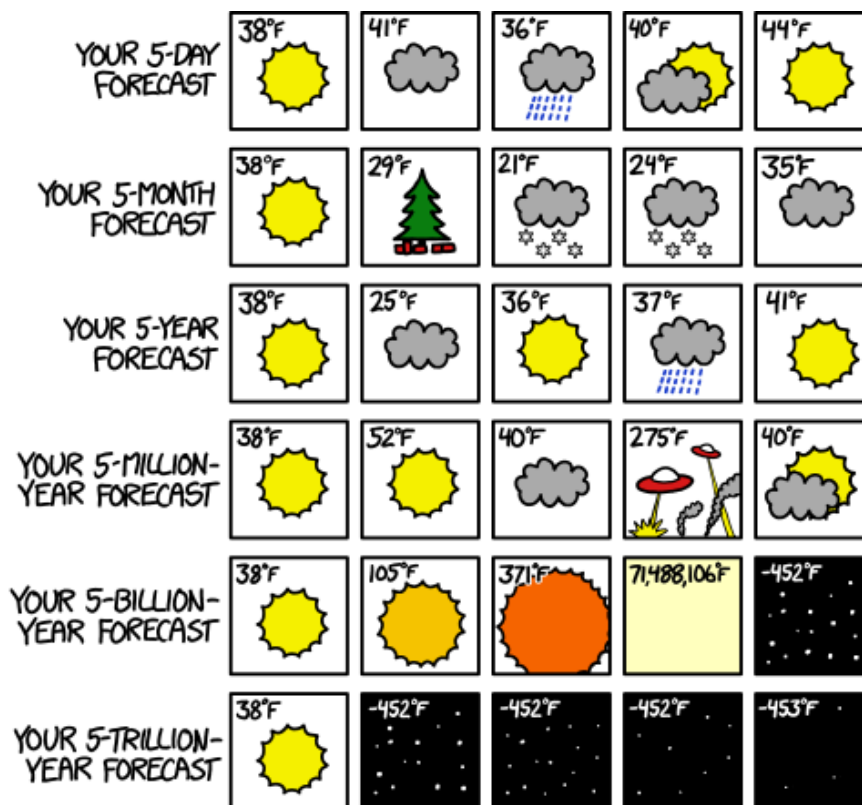




## Time series forecasting



Your 5 day forecast

Source: [xkcd](https://xkcd.com/218/)

Formatrices : Marta Rybczynska et Elise Pupier

# Objectifs du module

## Une introduction de base à l'analyse et la prévision des séries temporelles

- Comprendre les défis de l'analyse de données de séries chronologiques
- Comprendre en quoi les étapes de prédiction de séries chronologiques sont différentes des autres prédictions de machine learning
- Utiliser des modèles de séries chronologiques linéaires conventionnels pour faire des prévisions univariées
- Utiliser une bibliothèque de prévisions de séries chronologiques pour les prévisions univariées

## Démarche pédagogique

- Durée du projet : 5 jours
- Travail en autonomie, mais échangez autant que possible entre vous !
- 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 «[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.

# Étape 1 : Exploration & Analyse des séries temporelles (1 jour)

## 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.

Une série temporelle est une série dont les observations sont mesurées à différents points dans le temps. Comme il s'agit d'évaluer l'évolution d'un phénomène au cours du temps, l'ordre des observations est important.

Les séries temporelles sont présentes dans de nombreux domaines d'application. Leur étude permet de comprendre les tendances passées ou prévoir les comportements futurs.

L'objectif de cette partie est donc de proposer des outils d'analyse exploratoire permettant de tirer une information utile de ces séries temporelles.

La librairie Pandas contient des fonctions dédiées aux séries temporelles, qui simplifient énormément leur étude.

## Compétences

- Développer une approche méthodologique pour l'analyse de séries temporelles
- Utiliser Pandas pour l'étude des séries temporelles

## Consignes

- Effectuer les tâches définies dans le Jupyter notebook NB1\_Time\_Series\_Forecasts.ipynb. Ce notebook vous servira de guide pour construire votre méthodologie afin d'appréhender au mieux toute série temporelle.

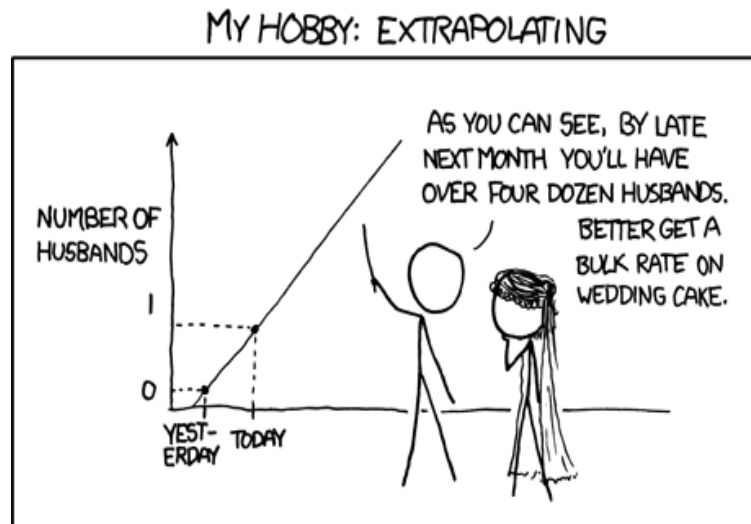
## Livrables

- Le Notebook NB1\_Time\_Series\_Forecasts.ipynb entièrement complété.

## Ressources

- <https://www.youtube.com/watch?v=FsroWpkUuYI&list=PLjwX9KFWtvNnOc4HtsvaDf1XYG3O5bv5s>
- [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/timeseries.html](https://pandas.pydata.org/pandas-docs/stable/user_guide/timeseries.html)
- <https://jakevdp.github.io/PythonDataScienceHandbook/03.11-working-with-time-series.html>
- <https://www.kaggle.com/code/prashant111/complete-guide-on-time-series-analysis-in-python>
- [https://perso.math.univ-toulouse.fr/jydauxoi/files/2017/04/poly\\_eleves.pdf](https://perso.math.univ-toulouse.fr/jydauxoi/files/2017/04/poly_eleves.pdf)

## Étape 2 : Prédiction des séries temporelles avec ARIMA & Co (3 jours)



### 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 NB2\_Time\_Series\_Forecasts.ipynb. Ce notebook vous guidera à travers le processus d'utilisation des modèles de base ARIMA sur les données de séries temporelles pour faire des prévisions.

### Livrables

- Le notebook NB2\_Time\_Series\_Forecasts.ipynb entièrement complété.
- 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 !).
- 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.

## Ressources

- AR, MA, ARIMA :

<https://www.youtube.com/watch?v=Mc6sBAUdDP4&list=PLjwX9KFWtvNnOc4HtsvaDf1XYG3O5bv5s&index=6>

[https://www.youtube.com/watch?v=zNLG8tsA\\_Go&list=PLjwX9KFWtvNnOc4HtsvaDf1XYG3O5bv5s&index=6](https://www.youtube.com/watch?v=zNLG8tsA_Go&list=PLjwX9KFWtvNnOc4HtsvaDf1XYG3O5bv5s&index=6)

<https://www.youtube.com/watch?v=dXND1OEBABI&list=PLjwX9KFWtvNnOc4HtsvaDf1XYG3O5bv5s&index=7>

- ACF & PACF :

<https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/>

- Métriques :

<https://towardsdatascience.com/time-series-forecast-error-metrics-you-should-know-cc88b8c67f27>

- Cross-validation :

<https://stats.stackexchange.com/questions/14099/using-k-fold-cross-validation-for-time-series-model-selection>

<https://robjhyndman.com/hyndsight/tscv/>

- *Forecasting: Principles and Practice* livre (Rob Hyndman) :

<https://otexts.com/fpp2/>

<https://otexts.com/fpp2/non-seasonal-arima.html>

<https://otexts.com/fpp2/transformations.html>

<https://otexts.com/fpp2/seasonal-arima.html>

- *Documentation de Pmdarima*:

<http://alkaline-ml.com/pmdarima/>

## Étape 3 : Demandez au Prophet ! (1 jour)

### Objectifs de l'activité

- Comprendre et utiliser Facebook Prophet pour effectuer des prévisions univariées.

### Compétences

- Apprendre à maîtriser un modèle tiers et comparer ses performances aux modèles précédents.

### Consignes

- Effectuer les tâches définies dans le Jupyter notebook NB3\_Time\_Series\_Forecasts.ipynb.

### Livrables

- Le notebook NB3\_Time\_Series\_Forecasts.ipynb entièrement complété.

### Ressources

- <https://peerj.com/preprints/3190/>
- <https://www.youtube.com/watch?v=2XFro0nIHQM&list=PLjwX9KFWtvNnOc4HtsvaDf1XYG3O5bv5s&index=10>
- <https://www.youtube.com/watch?v=95-HMzxsgHY>
- <https://www.youtube.com/watch?v=pOYAXv15r3A>