



## Cas d'étude Data Scientist #1

2022



En premier lieu, merci de prendre le temps de réaliser ce cas d'étude : nous vous en sommes reconnaissants. Rappelez-vous qu'il n'existe jamais une seule solution à ce type de problème, et ne visez donc pas la perfection la plus totale. Finalement, exprimez votre créativité et sentez-vous libre d'appliquer votre méthode personnelle.

## Contexte

Dans le cadre de la mission proposée, le rôle du Data Scientist est d'assurer le développement de l'activité de vente d'électricité sur les marchés court terme de l'électricité. La vente de la production d'un portefeuille de centrales Boralex est réalisée sur les marchés dits physiques, au lendemain ou Day-Ahead et intrajournalier ou Intraday.

Les ordres de vente devant être passés plusieurs heures en avance (les marchés clôturant avant la livraison), ils doivent être calculés sur la base de prévisions de production. Ces prévisions sont nécessairement imparfaites et ne correspondent jamais exactement à la production injectée par les actifs sous gestion.

- Lorsque la puissance vendue sur les marchés est supérieure à la production, Boralex doit racheter a posteriori l'électricité manquante au gestionnaire de réseau, à un prix *généralement* supérieur aux prix de marché,
- Lorsque la puissance vendue est inférieure à la production, le gestionnaire de réseau rachète l'électricité injectée en surplus à un prix *généralement* inférieur aux prix de marché.

Ce processus dit « des écarts » est utilisé par le gestionnaire de réseau pour assurer l'équilibre offre-demande sur le territoire national.

Dans cet exercice, on considérera que les ordres de vente passés sur le marché dit Day-Ahead (le seul marché court terme considéré dans le cadre de cette étude) sont égaux aux prévisions de production.

La valorisation de référence de la production est égale à la valeur de l'électricité vendue au prix Day-Ahead, c'est-à-dire le chiffre d'affaires généré par des prévisions de production parfaites / des ventes exactement égales à la puissance injectée. La valorisation nette est égale à la valorisation réelle de l'électricité, soit le produit des ordres passés sur le marché Day-Ahead par le prix Day-Ahead (on considérera que chaque ordre est accepté et converti exactement en vente réalisée), plus la somme des achats ou des reventes selon le processus des écarts. Le coût des écarts est, en €/MWh injecté, le différentiel entre la valorisation de référence et la valorisation nette.

## Tâches

L'objectif du cas d'étude est de calculer deux variables en fonction du *leadtime* :

- La nMAE (*normalized Mean Absolute Error*) de la prévision de production, soit la MAE de la prévision divisée par la capacité installée du portefeuille (93,7 MW), en %.
- Le coût des écarts tel que défini ci-avant, en €/MWh.

Une liste des tâches à effectuer est fournie ci-après :

- Effectuer les opérations de formatage nécessaires (agrégation des données de production et de prévision de production fournies par actif),
- Jointure des différentes sources de données,
- Calculer, par *leadtime*, les deux métriques, nMAE et coût des écarts.
- Tracer des graphiques de ces métriques, et les analyser brièvement sur la base des données fournies.

Il n'est à ce stade pas nécessaire de disposer d'une connaissance exhaustive des marchés de l'électricité et les informations fournies devraient vous être suffisantes pour réaliser le cas d'étude. Dans le cas contraire, n'hésitez pas à nous contacter.

## Outils

Afin de réaliser ce cas d'étude, vous pouvez utiliser le langage de programmation de votre choix. En interne, nous utilisons principalement Python.

Des données de production, de prévisions de production, de prix Day-Ahead, de prix des écarts sont communiqués dans des fichiers .csv distincts. Les données de prévision de production sont fournies selon différents *leadtime*, le différentiel entre l'heure de calcul de la prévision et l'heure d'injection.

De plus :

- Nous attendons de vous que vous ne passiez que quelques heures sur le sujet, au maximum une demi-journée,
- Utilisez les outils et les modules de votre choix,
- Votre code doit pouvoir être mis à l'échelle pour des ensembles de données plus conséquents et pour un nombre plus important d'actifs,
- Veuillez nous envoyer le cas d'étude sous forme de repository git ou au format .zip,
- Veuillez ne pas commit les données dans le repository git,
- Nous rappelons que les données sont confidentielles.

### **Livrables**

Le livrable finale est un repository git contenant :

- Un *readme* expliquant la manière dont vous avez résolu le problème. Quelles ont été vos hypothèses ? Comment améliorerez-vous le pipeline si vous disposiez de plus de temps ?,
- Une description rapide de comment faire fonctionner le programme. Autrement, vous pouvez nous communiquer les sorties du code,
- Votre code.

Une nouvelle fois, n'hésitez pas à nous contacter pour toute question complémentaire.

Dans l'attente de votre cas d'étude !

L'équipe Marchés de Boralex