

TotalEnergies

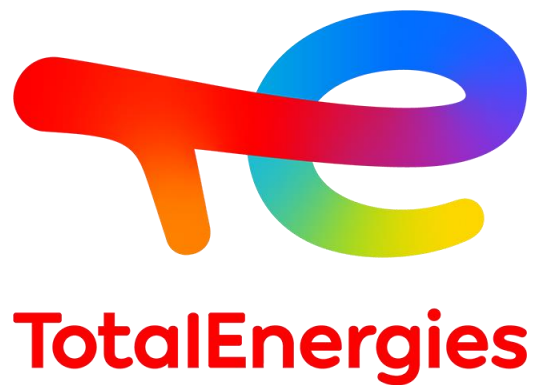
Rapport de stage

Dirigé par Eurydice Laffayrerie

Encadré par Yoan Saint-Pierre

Février 2021 – Aout 2021





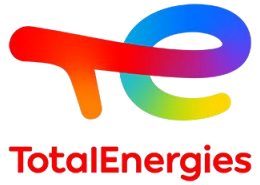
Sommaire

Introduction

I – Prédiction d'appels

II – Suivi Conso

Conclusion



Introduction

Contexte

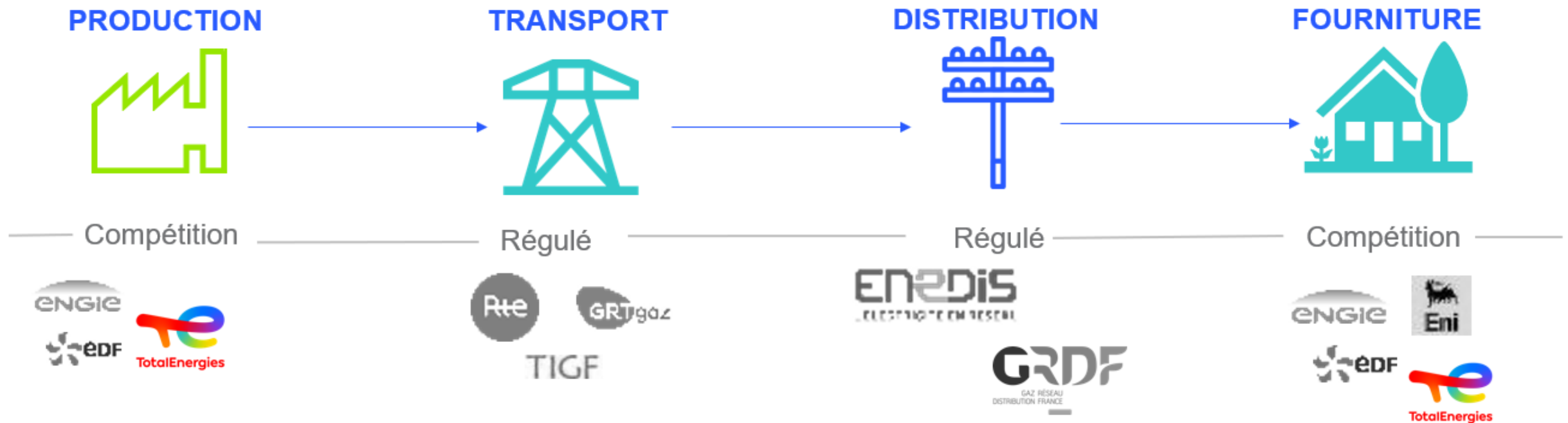
Description de l'entreprise

Objectifs et enjeux du stage

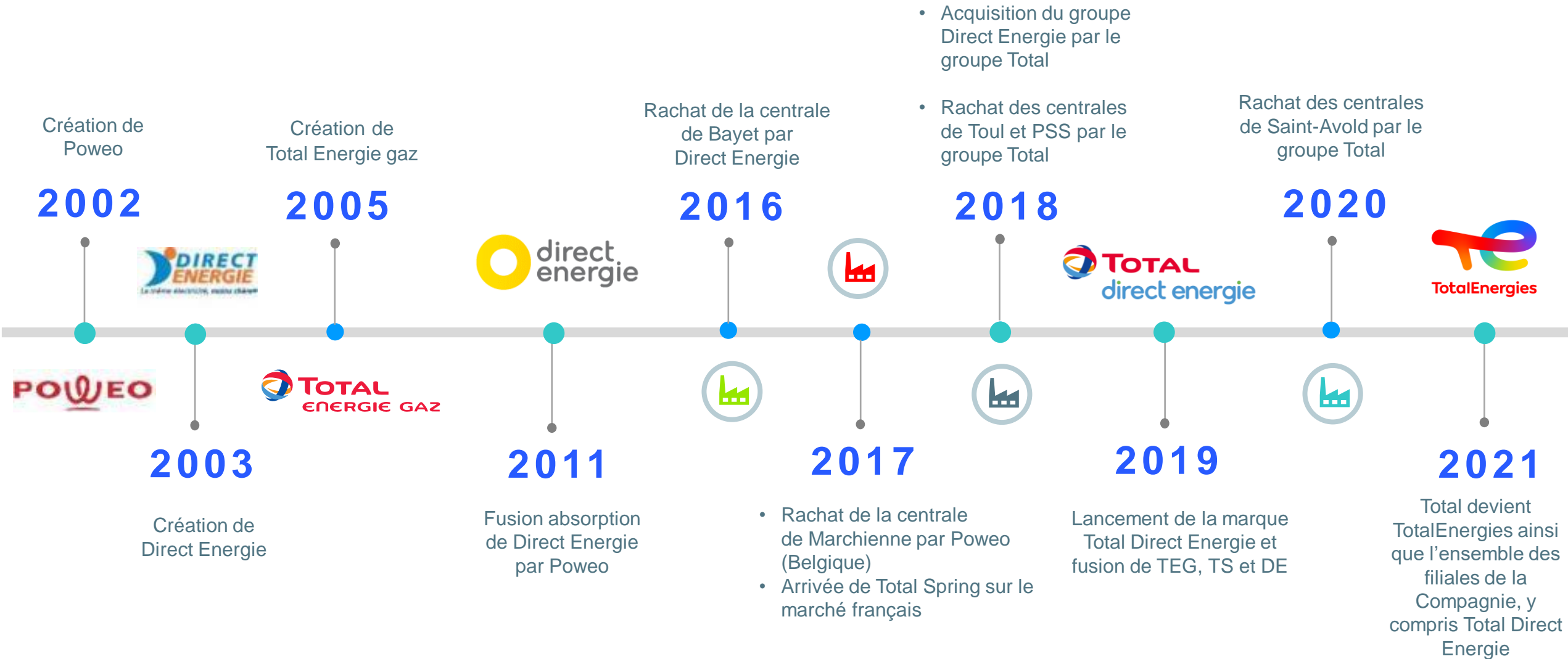
Frameworks utilisés

Fonctionnement du marché Français de l'électricité et du gaz

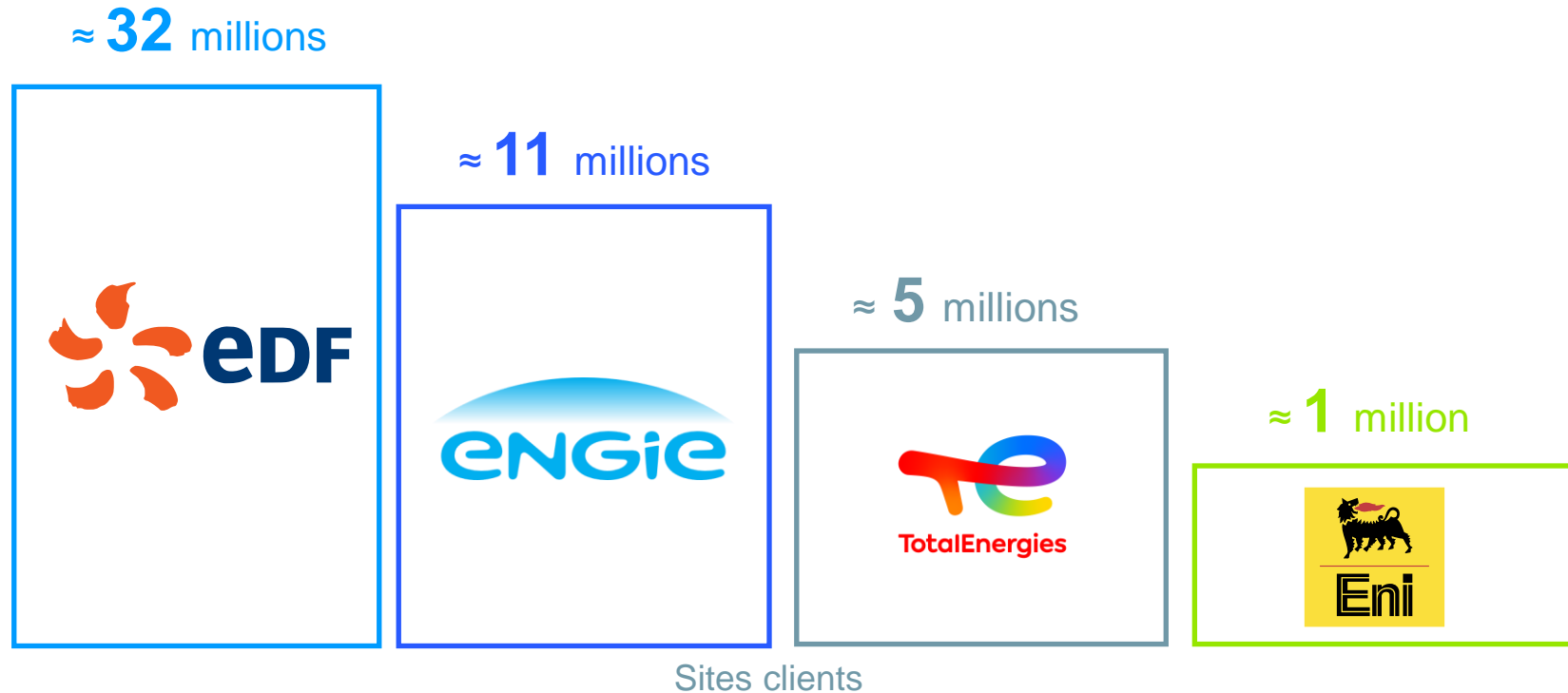
Fin du monopole d'EDF votée le 10 février 2000 -> Apparition de Tarif Réglementé de Vente (TRV)



Histoire de la branche



Place sur le marché

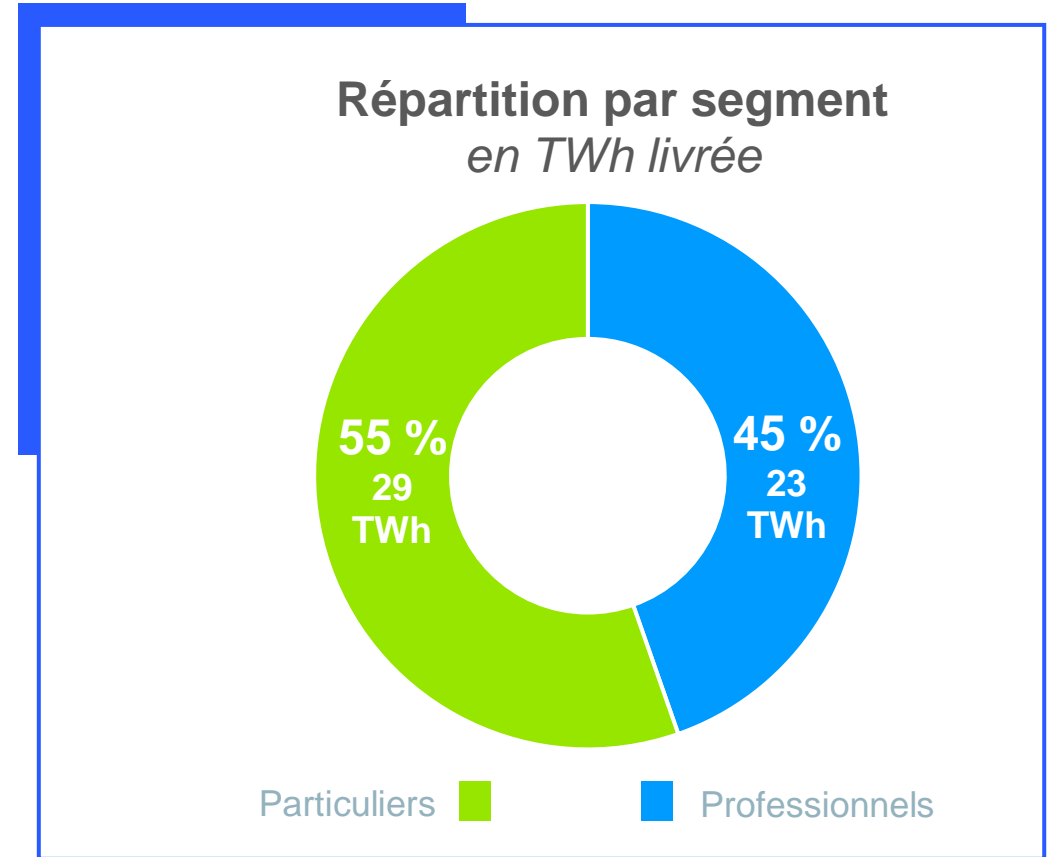
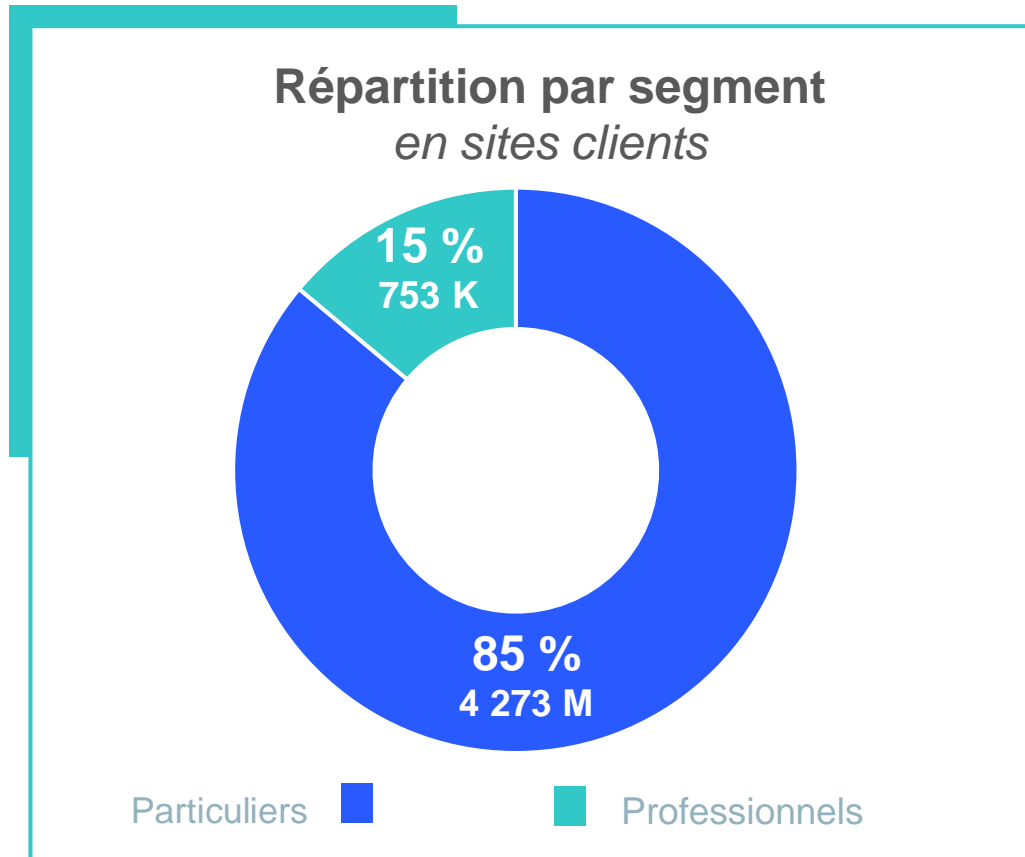


- 32 millions sites résidentiels
- 77% au tarif réglementé



- 11 millions sites résidentiels
- 40% au tarif réglementé

Proportion de clients professionnels / particuliers



Objectifs et enjeux du stage

Prédiction d'appels

Objectif :

- Prédire les appels des clients 3 mois à l'avance sur différents flux d'appels.

Enjeux :

- Permet aux chefs d'équipes de former des équipiers en prévisions d'un gros flux à venir
- Evite d'investir dans une équipe trop grande inutilement

Suivi Conso

Objectifs :

- Estimer la consommation de chauffage d'un client
- Prédire l'étiquette énergétique (DPE) du logement des clients
- Analyser l'impact de l'accompagnement de TE sur la consommation des clients

Enjeux :

- Un pas vers la transition énergétique
- Diminution de la facture
- Fidélisation des clients

Frameworks utilisés



Framework open source de calcul distribué
Utile pour le Data Processing dans le cas de
gros volumes

5 millions de clients + clef Conso Live



databricks

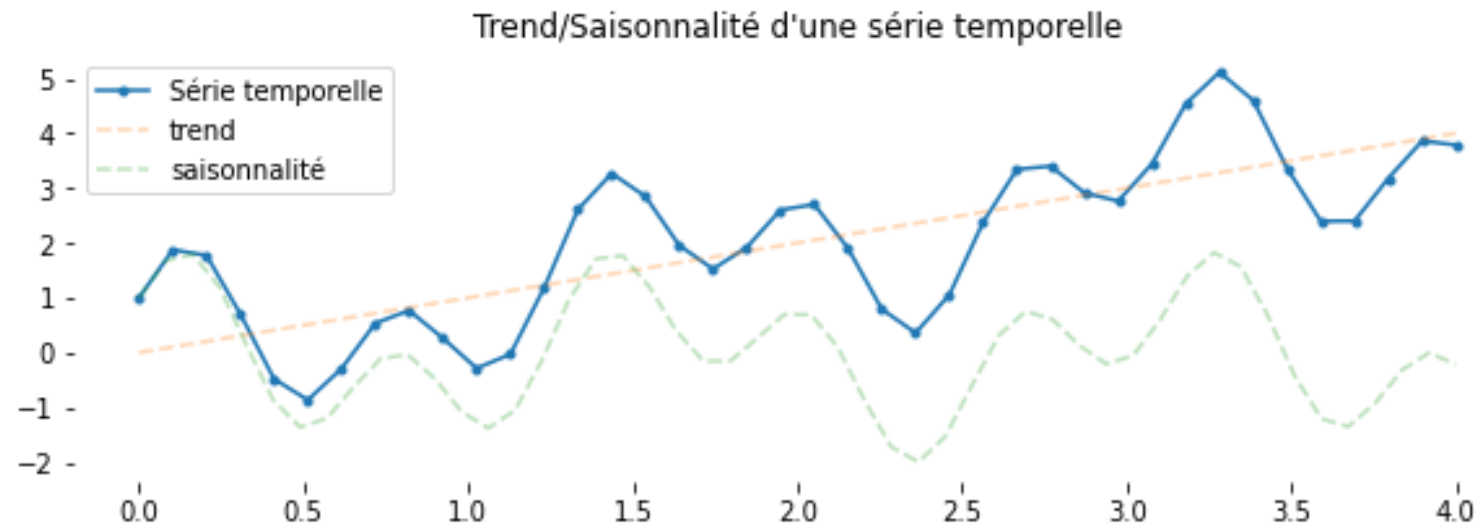
Plateforme web de partage de notebooks
Créée pour fonctionner avec Spark
(clusters de calculs)

Facilite grandement la mise en
production

I - Prédiction d'appels

- 1 – Méthodologie Générale
- 2 – Prédiction d'appels Pro
- 3 – Prédiction d'appels Vente

Vocabulaire des séries temporelles



$$y(t) = trend(t) + saison(t) + holiday(t) + err(t)$$

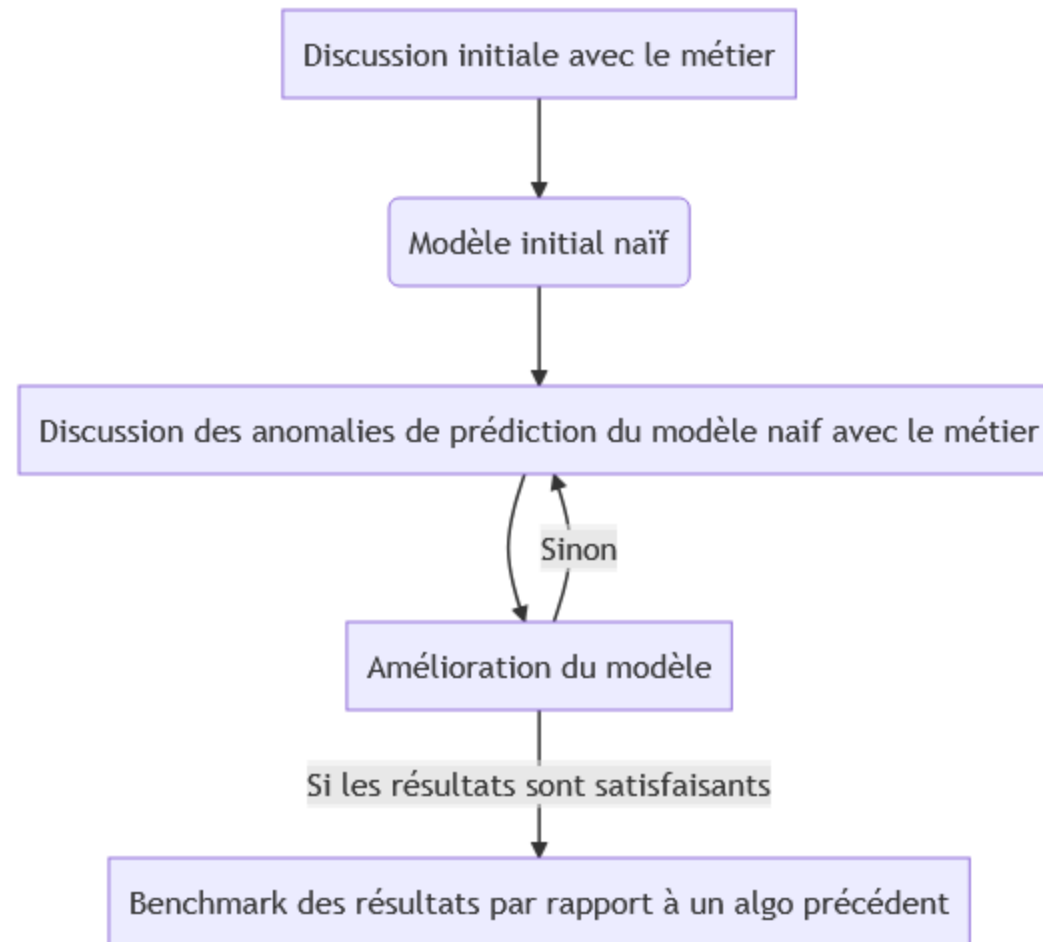


- Prend en compte des **variations de trend** (linéaire ou logistique par morceaux) de manière automatisée
- Prend en compte la saisonnalité sur plusieurs mailles (mensuelle, **annuelle**, hebdomadaire)
- Prend en compte les jours fériés/événements extraordinaires (ex: **covid**)
- Bon rapport **précision**/temps de tunage du modèle
- Robuste aux outliers et aux données manquantes
- Tunable en utilisant les connaissances métier

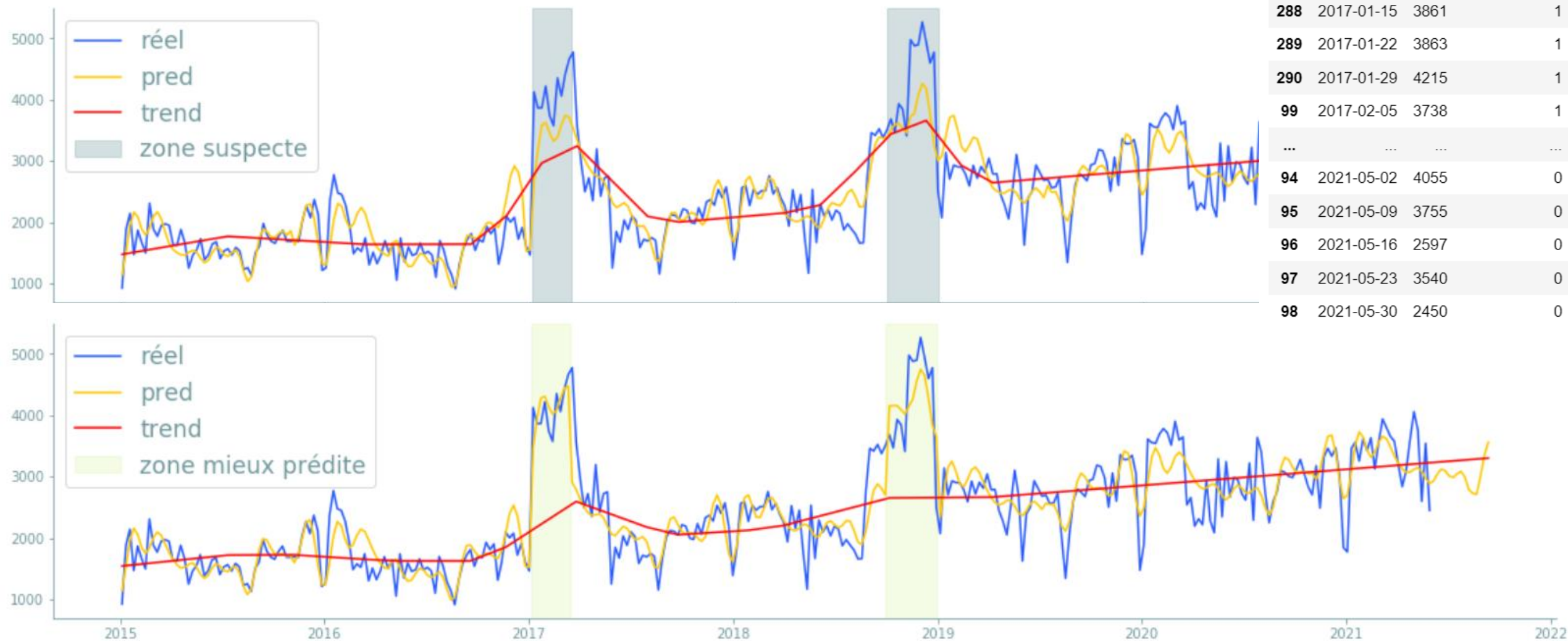


- Nécessite un **historique** de données suffisant pour voir apparaître une saisonnalité

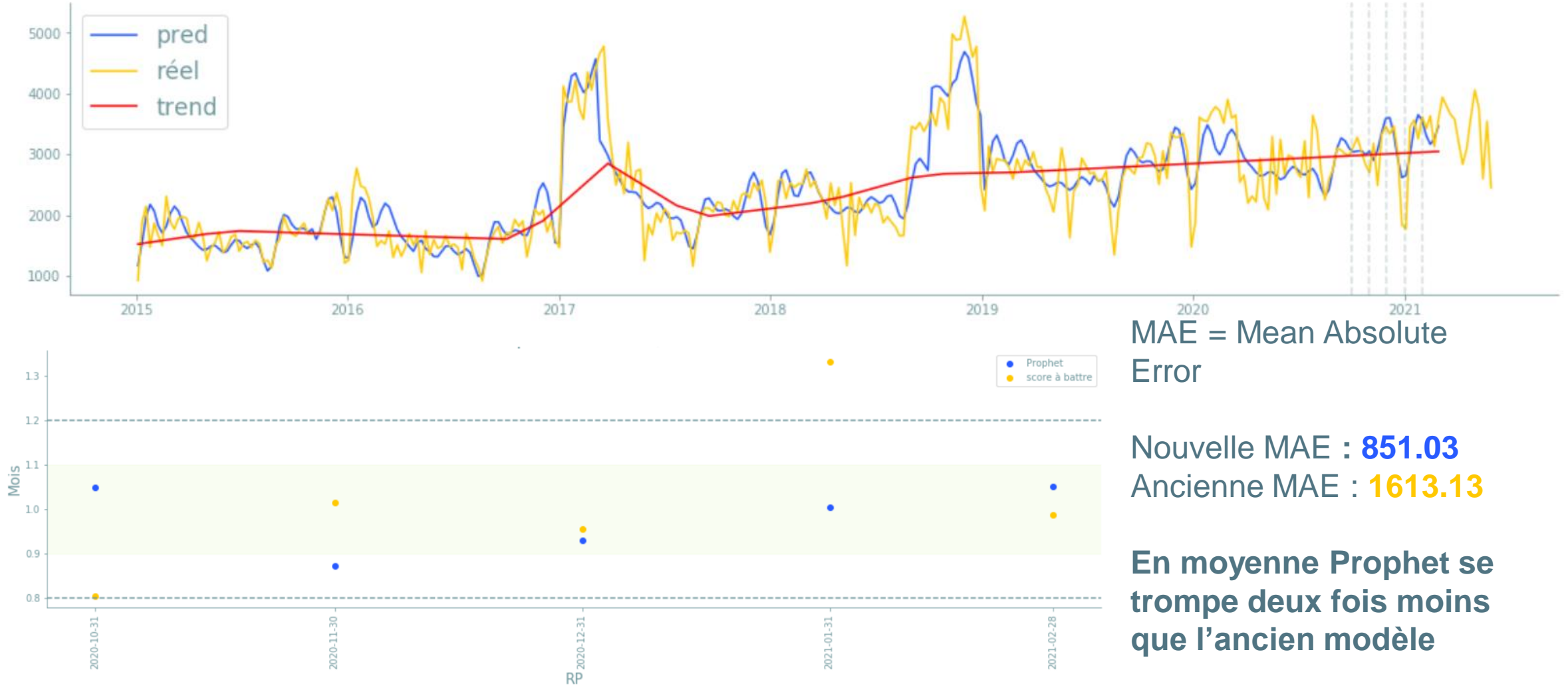
Boucle de travail



Prédiction d'appels – Service Client PRO



Résultats - Pro



MAE = Mean Absolute Error

Nouvelle MAE : **851.03**

Ancienne MAE : **1613.13**

En moyenne Prophet se trompe deux fois moins que l'ancien modèle

>> [Mise en production du notebook](#)

Prédiction d'appels – Service Client – Ventes - particulier

Suite à un changement d'architecture des flux, seuls les flux de **3099** et de **Click To Call (CTC)** ont un historique potentiellement exploitable.

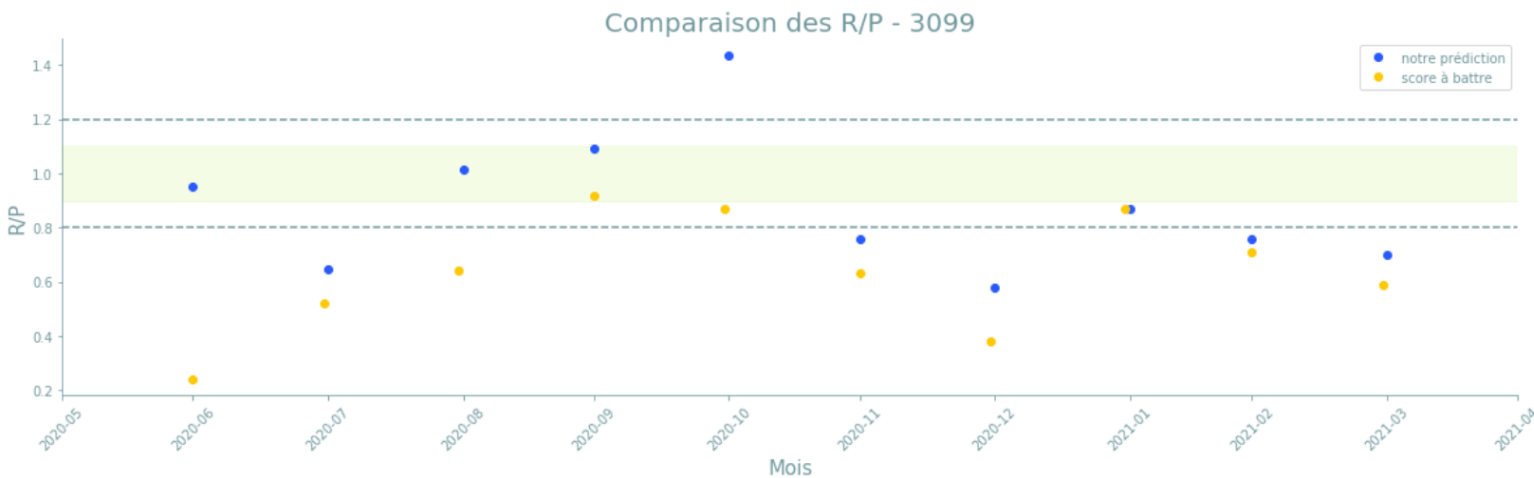
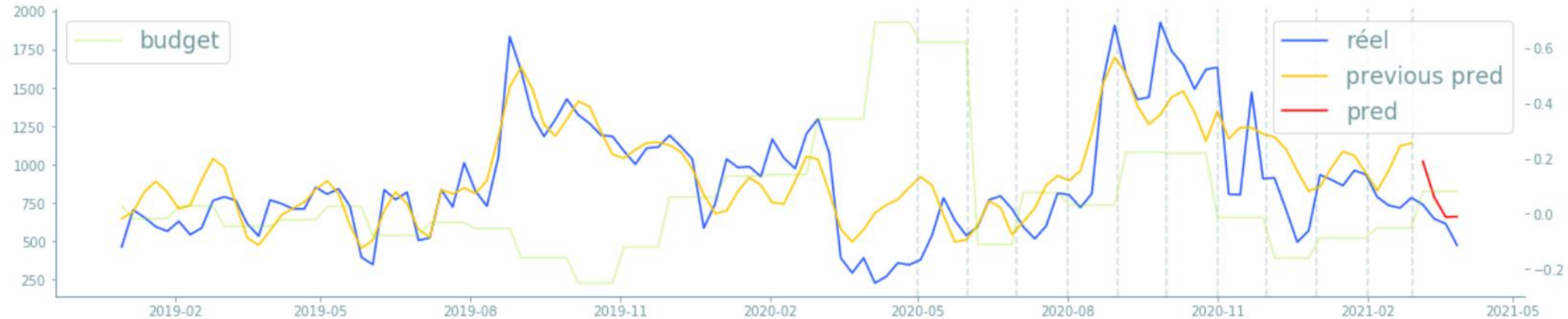
Modèle précédent : Winter-Holt (fonction built-in de Excel) manuellement lancé en début de chaque mois.

Comme on travaille sur les ventes, il est pertinent d'introduire le budget mensuel prévisionnel comme variable exogène dans notre modèle

Objectifs du projet :

- Automatiser la prédiction
- Améliorer la prédiction

Résultats - 3099



MAE = Mean Absolute Error

Ancienne MAE = 2477.80

Nouvelle MAE = 1064.6

En moyenne **on se trompe 2 fois moins** que dans la précédente prédiction

>> [Mise en production du notebook](#)

Résultats - CTC



Série temporelle difficile à prédire car :

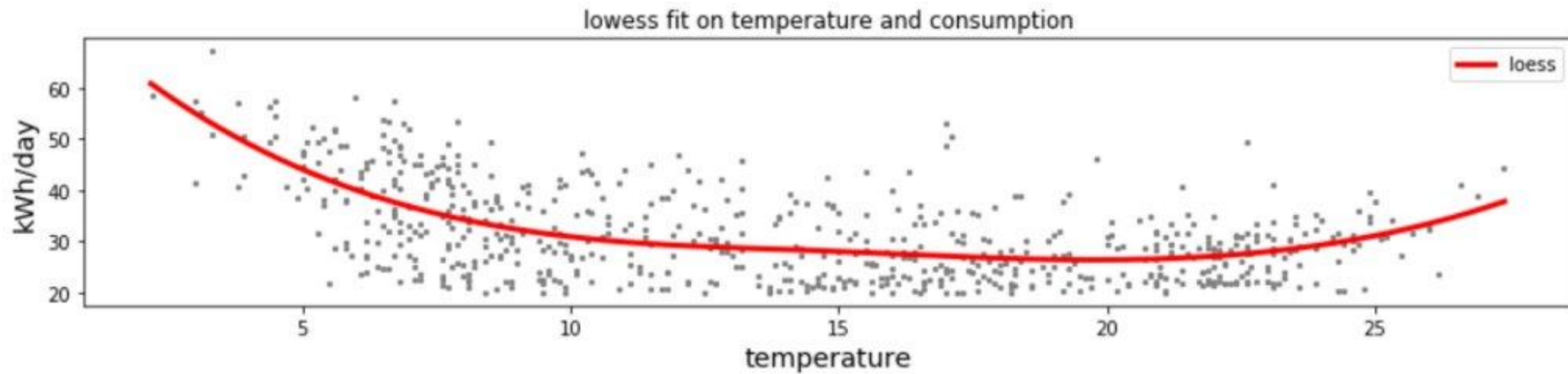
- peu de saisonnalité
- peu corrélée au budget
- pics imprévisibles dans le futur du point de vue de notre modèle

II – Suivi Conso

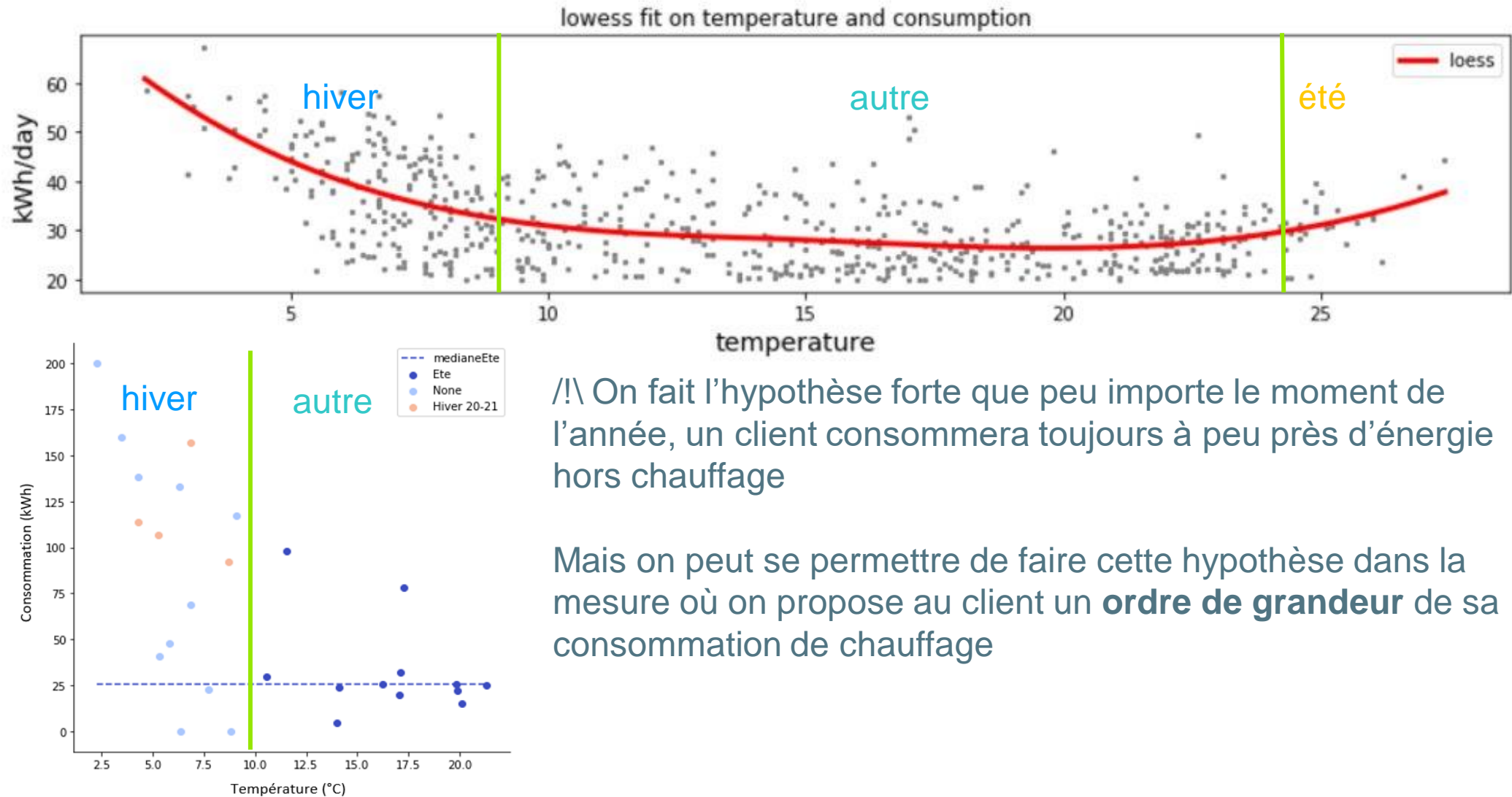
- 1 – Bilan hiver
- 2 – Analyse d'impact
- 3 – Prédiction d'étiquettes DPE

Bilan hiver - Enjeux

- Combien d'électricité/de gaz ai-je utilisé pour me chauffer ?
- Est-ce que les autres consomment autant pour se chauffer ?



Bilan hiver - Algo

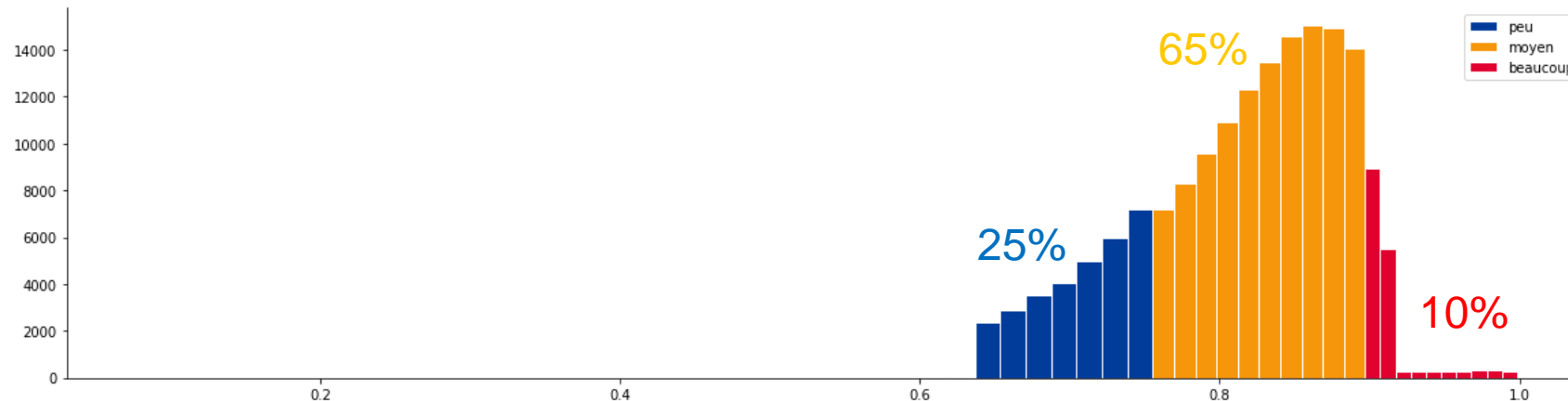
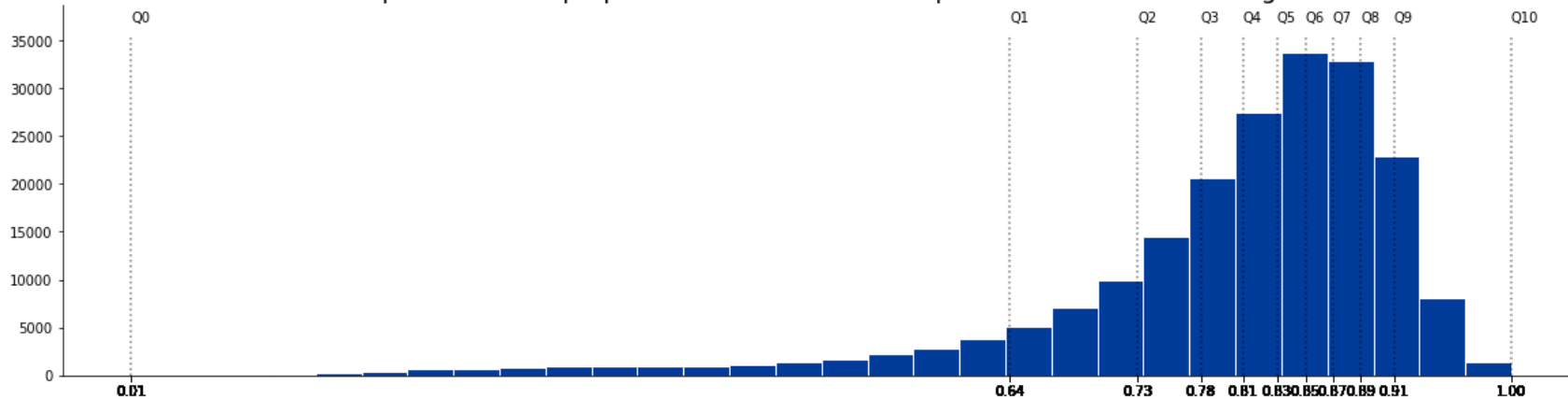


/!\ On fait l'hypothèse forte que peu importe le moment de l'année, un client consommera toujours à peu près d'énergie hors chauffage

Mais on peut se permettre de faire cette hypothèse dans la mesure où on propose au client un **ordre de grandeur** de sa consommation de chauffage

Bilan hiver - Résultats

Répartition de la proportion de consommation qu'on attribue à du chauffage



On envoie ensuite un mail aux clients restants du type : « Cet hiver, on a estimé que tu avais consommé **85%** de ton électricité pour te chauffer, c'est **normal** comparé aux autres clients de ta région »

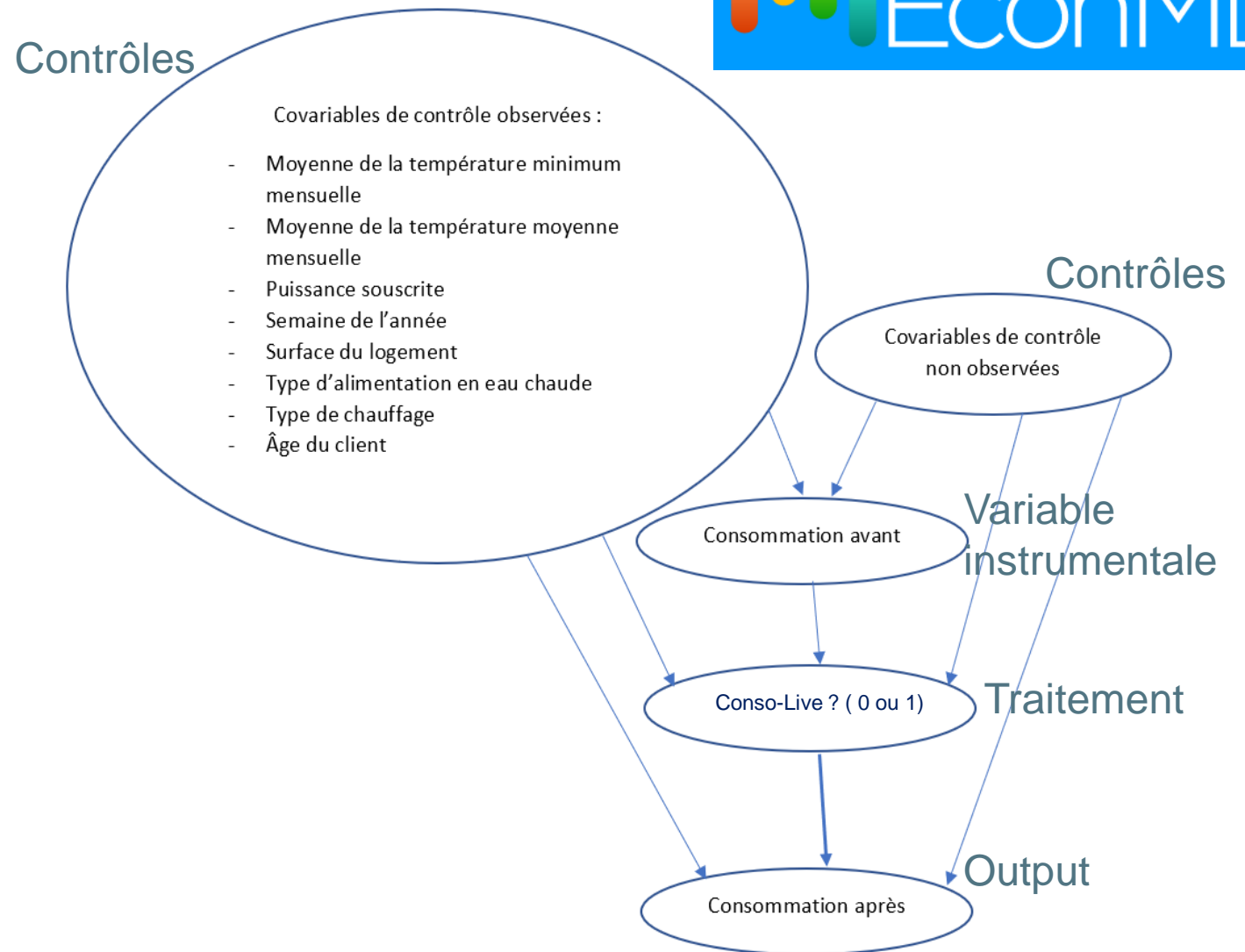
Analyse d'impact

Intérêts :

- Vérifier l'efficacité d'un service proposé
- Utiliser cette analyse comme argument lors de démarchage ou de campagnes publicitaires
- Voir même dans le meilleur des cas, acquérir un label officiel de réduction d'énergie

Objectif : explorer et découvrir une librairie d'économétrie pour proposer une méthodologie d'analyse d'impact

On étudie l'impact de la clef **Conso-Live** sur la consommation d'électricité de nos clients car c'est un service en lequel on a confiance qu'il ait un réel impact



Analyse d'impact résultats

Uncertainty of Mean Point Estimate :

- Mean point : -0.441 kWh/an
- Pvalue : 0.25
- CI_lower_mean : -1.072
- CI_upper_mean : 0.189

Ce sont des résultats décevants qu'on peut expliquer par :

- Une pose de problème un peu trop ambitieuse (impact sur 1 an de la réduction > impact le mois suivant l'installation de conso-live)
- Un faible nombre de clients (à cause de la restriction de 2 ans d'historique)
- Un modèle trop pauvre en features de contrôle

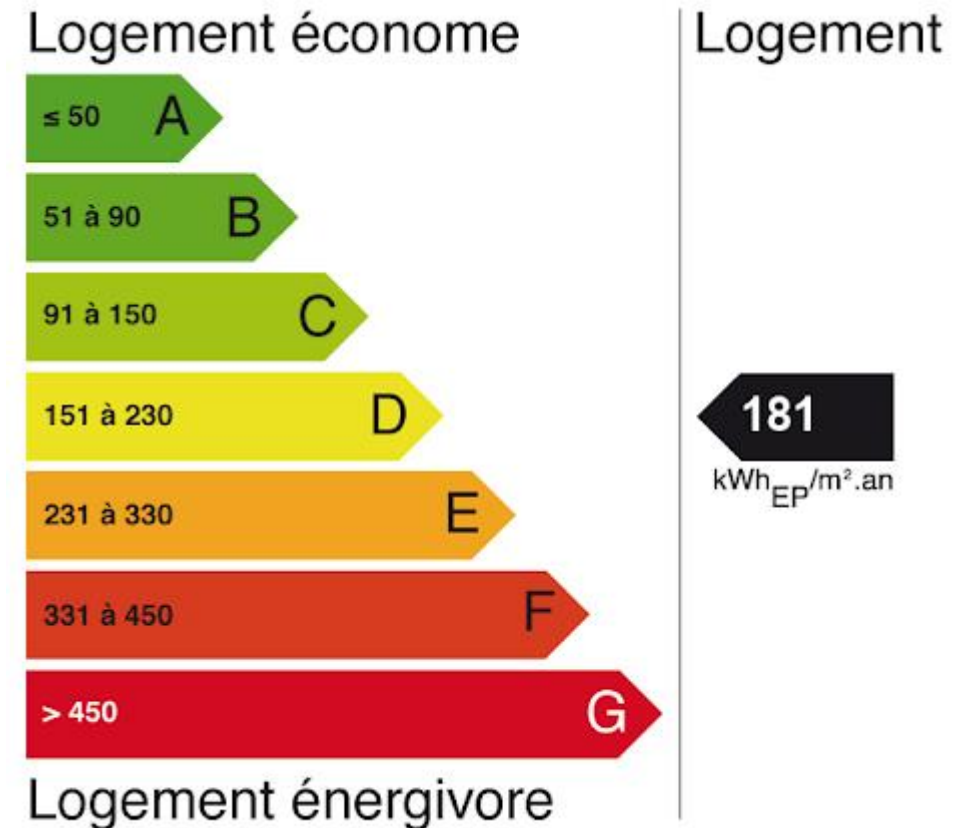
Prédiction d'étiquettes DPE

Objectifs :

- Quels sont les foyers thermosensibles ?
- Pour quelle raison ? Logement Energivore ?

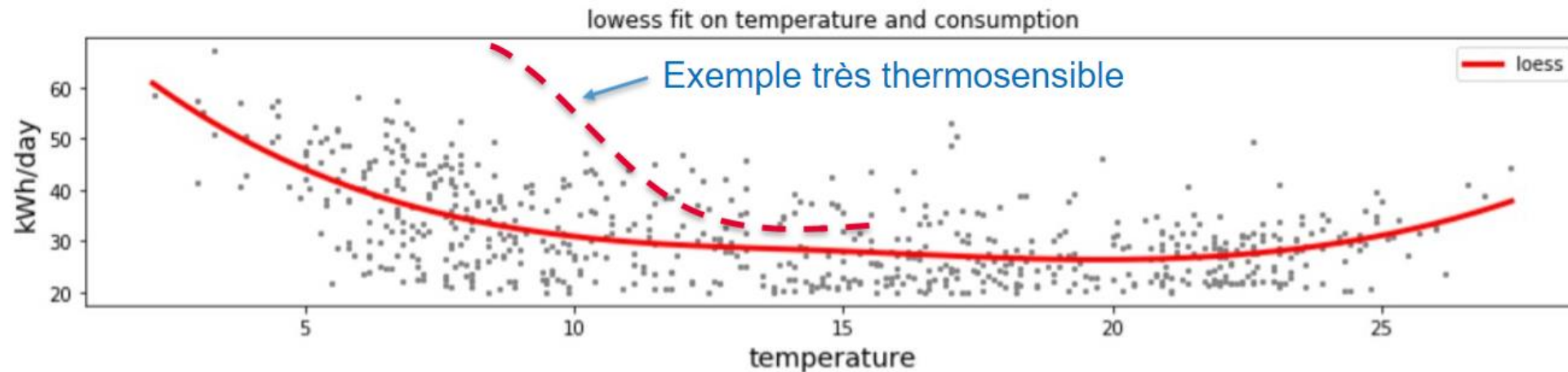
Enjeux :

- Proposer des travaux de rénovations financés par l'Etat aux clients qui en ont besoin



Prédiction d'étiquettes DPE

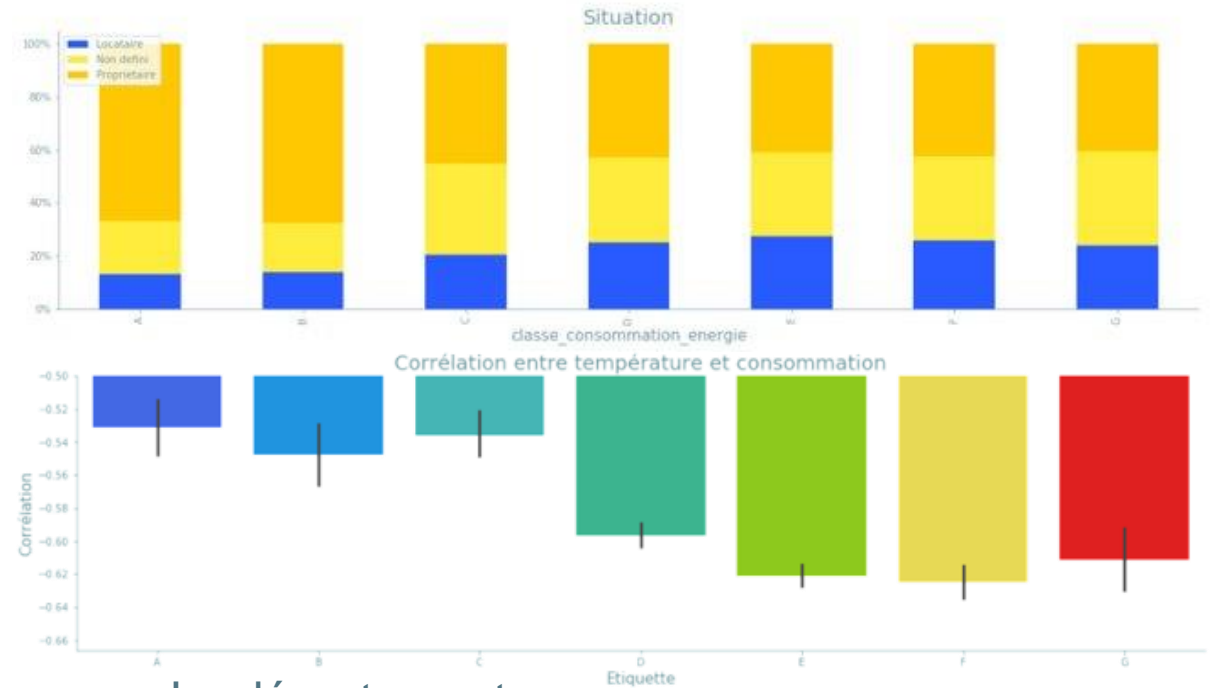
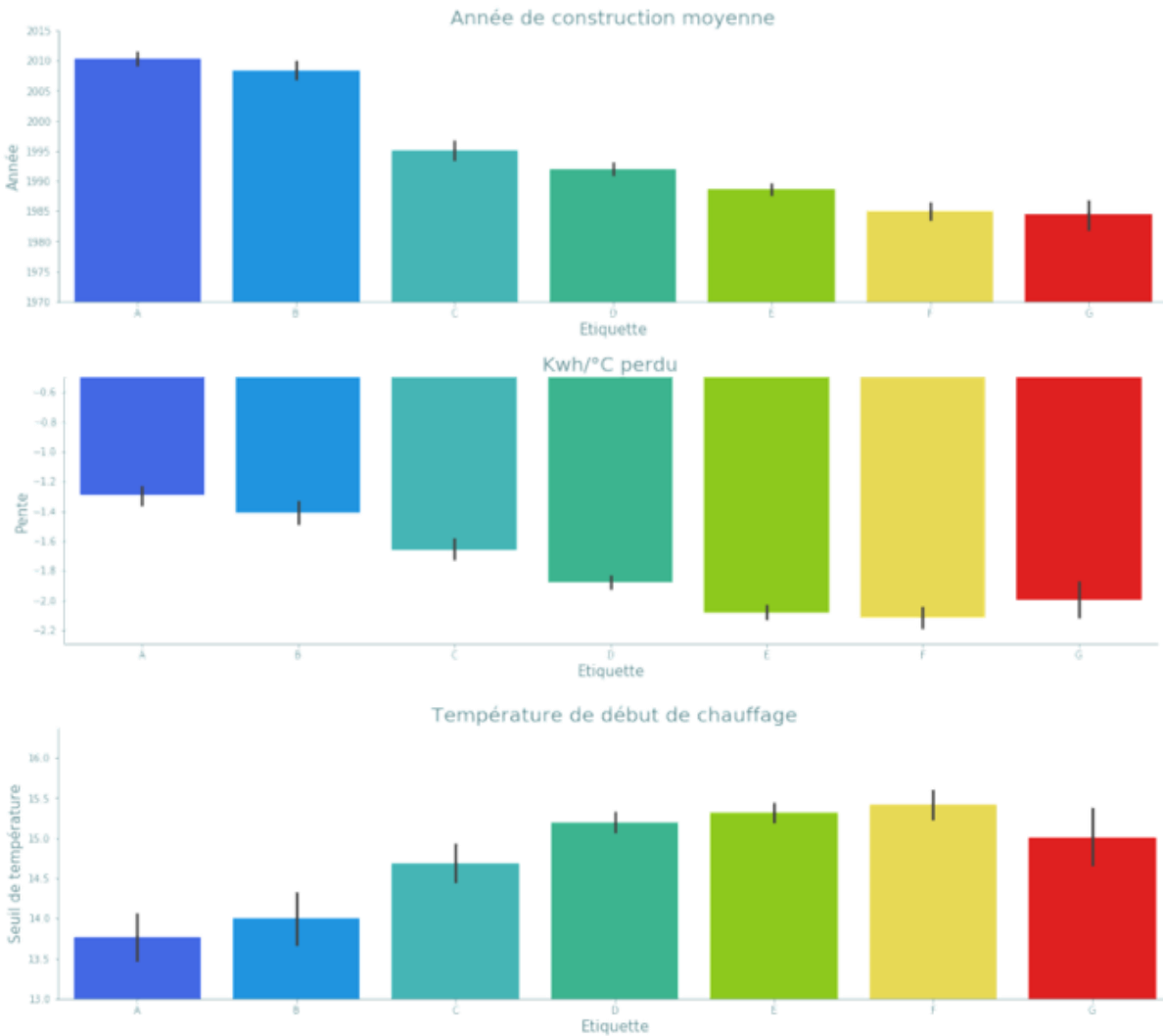
En étudiant la relation entre la consommation et la **température** de nos clients, on repère des clients bien plus thermosensibles que d'autres.



Mais quand il s'agit de savoir **pourquoi** ils consomment plus en hiver, on a besoin de plus d'information car on pourrait avoir affaire à :

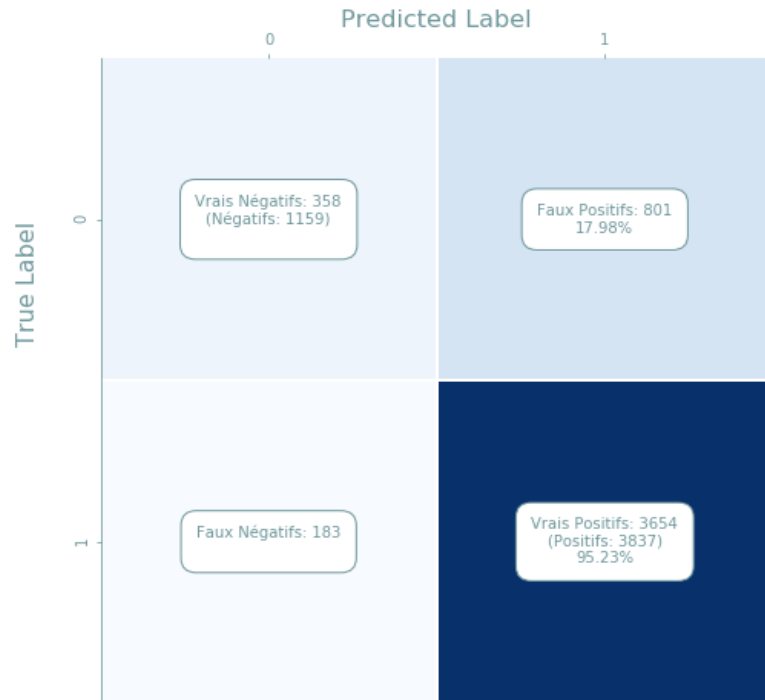
- Logement énergivore
- Habitudes de vies différentes (degré de confort, mauvaises pratiques de chauffage, ...)
- Appareils supplémentaires (piscine chauffée, sauna, ...)

Prédiction d'étiquettes DPE – description des données



- Le département
 - Le type de chauffage
 - Le type de chauffe eau
- modèle de type LightGBM (une version allégée en calculs de XGBoost)
classification binaire (bonne étiquette/mauvaise étiquette) supervisée

Prédiction d'étiquettes DPE – résultats



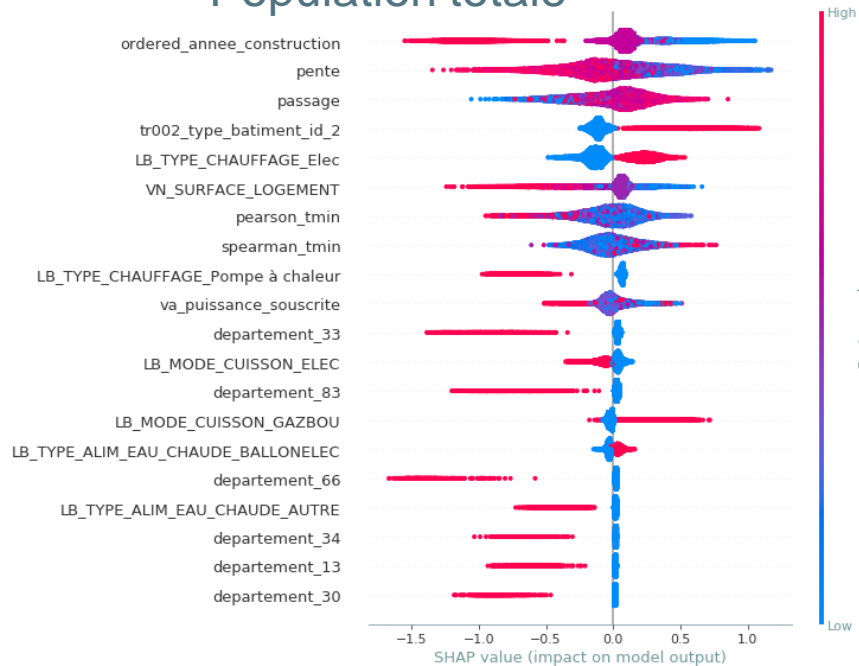
Rappel = 95.23%

18% des négatifs ont été classés
comme faux positifs

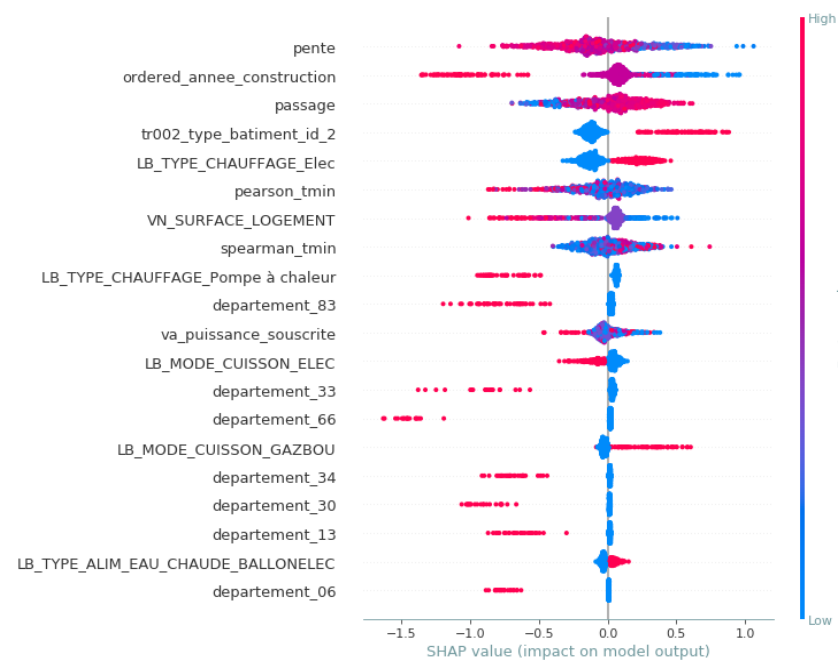
Prédiction d'étiquettes DPE – Etude des faux positifs

Faux Positifs

Population totale

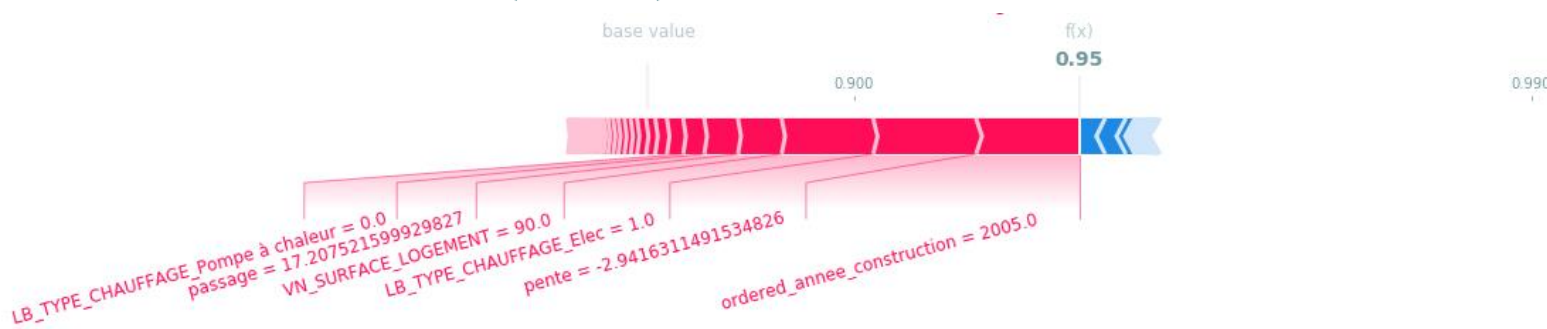


Faux Positifs



Le modèle se trompe souvent quand :

- La pente de leur nuage de points est trop forte
- La température de début de chauffage est élevée
- La taille du logement est assez importante
- Vieille année de construction

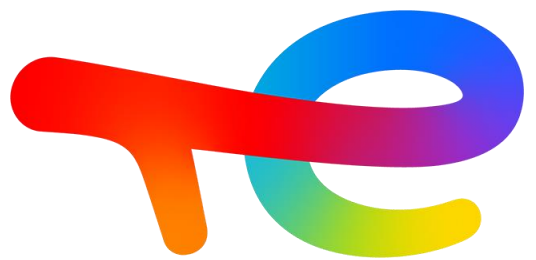


Le projet a particulièrement intéressé l'équipe de Marketing

Conclusion

Stage très enrichissant :

- J'ai pu mener plusieurs projet du début de la phase de développement jusqu'à la phase de production (prédictions d'appel) dont certains avec une certaine autonomie (Prédiction d'appels - Vente)
- J'ai travaillé autour de beaucoup de projets variés nécessitant d'apprendre de nouvelles compétences ou parfois de faire preuve de créativité
- J'ai pu découvrir le travail en entreprise (réunions régulières, projets inter-service, travail en équipe, séances de brainstorming, sprints de production et partages de connaissance hebdomadaire)



TotalEnergies

Remerciements