

# Exercice Data Scientist

Bienvenue sur le marché de la tondeuse !!!

Notre client, Mower4u, souhaiterait mettre de nouveaux produits sur le marché et aurait besoin de votre aide pour l'assister dans sa prise de décision.

Son principal besoin consiste à **estimer l'attractivité** de chaque produit sur le marché. De cette attractivité dépend les futures parts de marché des produits étudiés.

Pour cela, il nous a fournit une capture du marché actuel des tondeuses en vente. Les données à disposition sont les suivantes:

ID	ID du produit
Capacity	La capacité de la tondeuse (en Litres)
Failure_rate	Taux de panne
Price	Prix de la tondeuse
Prod_cost	Coût de production
Product_type	1, 2 ou 3 selon le marché
Quality	Indice de qualité
Attractiveness	Coefficient d'attractivité (entre 0 et 1, 0: pas attractif, 1: très attractif). Correspond à la cible de notre étude
Market_share	La part de marché associée au produit (correspond à une renormalisation de la colonne attractiveness)

Les tondeuses sont réparties en trois types (product\_type), soient trois micro-marchés distincts que vous pouvez supposer indépendants.

Votre mission consiste donc à **concevoir un modèle pour prédire l'attractivité de nouveaux produits** en fonction des différentes caractéristiques présentées.

Vous disposez donc de deux tables

- mower\_market\_snapshot.csv: Table contenant une capture du marché actuel
- submission\_set.csv: Table contenant les nouveaux produits que l'on veut mettre en vente, et pour lesquels il n'y a pas d'information sur l'attractivité.

Le candidat devra fournir le code généré pour construire le modèle de prédiction ainsi qu'un fichier (nom\_prenom\_attractiveness.csv) contenant les prédictions d'attractivité pour chaque nouveau produit, au format suivant:

```
id;attractiveness
21;0
32;0
72;0
etc..
```

La fonction de coût qui sera appliquée pour évaluer les résultats est le **Root Mean Square Logarithmic Error**.

$$RMSLE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\log(\tilde{y}_i + 1) - \log(y_i + 1))^2}$$

Avec:

- $y_i$  La valeur réelle d'attractivité pour le produit i
- $\tilde{y}_i$  Votre prédiction d'attractivité pour le produit i

**NB :** Il s'agit de données fictives, générées pour l'occasion, la dynamique du marché est relativement simple et le but de l'exercice est essentiellement d'évaluer la démarche. Il servira ainsi de support pour un éventuel entretien afin de discuter des choix effectués et des éventuelles pistes d'amélioration que vous envisageriez.

**NB2 :** Il n'y a pas de contrainte de langage de programmation pour effectuer cet exercice. Il sera cependant nécessaire de savoir justifier le choix effectué.