



## Exercice de Data Analyste

### Objectif de l'exercice

L'objectif de l'exercice consiste à évaluer votre capacité à résoudre une problématique de Data Analyse. Différents éléments seront évalués :

- La qualité du code (commentaires, optimisation, propreté, ...)
- La capacité d'analyser les données à disposition (graphes, ...)
- La capacité à implémenter une approche descriptive et prédictive (R, Python, ...)
- Votre capacité de restitution (présentation powerpoint, plotly, beamer ...)

### Quelques détails logistiques

- Les données sont accessibles via le lien suivant : [Used Cars Data - dataset by data-society | data.world](https://data.world/used-cars)
- Vous avez 48h au moment de la réception de l'exercice pour envoyer **uniquement votre code** au recruteur.
- Il est attendu lors de l'entretien **une présentation PPT d'environ 15 minutes** qui synthétise et présente votre travail ainsi que vos résultats. Vous n'avez pas de contrainte temporelle concernant la présentation finale.
- N'hésitez pas à poser vos questions aux recruteurs si vous avez des doutes.

## Enoncé :

Cet exercice est purement fictif.

Votre client est un vendeur de voiture d'occasion qui souhaite se doter d'un outil pour évaluer financièrement le prix d'une voiture par rapport au marché des voitures d'occasion. Il est parvenu à se construire un dataset de référence et souhaite s'en servir pour prédire les prix des nouvelles voitures qui rentrent dans son portefeuille.

Il fait appel à vous afin que vous puissiez l'aider à comprendre les informations présentes dans le dataset et pour construire le moteur de pricing. Après un atelier de brainstorming, vous avez identifié un ensemble de questions à répondre.

- Identifier les champs qui ont des données manquantes. Quelles stratégies préconisez-vous pour imputer les données manquantes.
- Identifier les valeurs extrêmes des variables. Que pouvez-vous faire pour normaliser les valeurs ?
- Pouvez-vous clusteriser les véhicules ? Quel est le profil de chaque cluster ?
- Quelles sont les données corrélées avec la variable « price » ?
- Comment utilisez-vous les variables catégorielles ?
- Est-ce qu'il est possible de réduire le nombre de variable ?
- Quelles sont les variables importantes lors du pricing de la voiture ?
- Est-ce qu'il existe d'autres axes d'analyse à considérer ?

Une fois les questions répondues, il vous sollicite pour une tranche optionnelle visant à implémenter le moteur de pricing. De nature méfiante, il souhaite pouvoir choisir parmi 3 modèles de Machine Learning différents. Il souhaite également dans connaître les variables qui ont impacté le pricing de chaque voiture.

Bon courage dans la réalisation de votre projet