

# Instructions

Félicitations! Nous sommes intéressés par ton profil et nous avons hâte de poursuivre le processus d'embauche avec toi. Ton prochain défi sera de nous démontrer certaines de tes compétences au moyen d'un court cas en science des données. Tu pourras trouver un lien vers un jeu de données plus bas. En utilisant ces données, nous t'invitons à résoudre la problématique ci-dessous. Une fois résolue, renvoie-nous tes réponses aux questions ainsi que le code que tu as produit. Bonne chance!

[Lien vers les données](#)

## Mise en situation

Certains gestionnaires de magasins tentent de comprendre comment augmenter les ventes de leurs magasins. Ils ont à leur disposition des données historiques de ventes de différents magasins.

## Objectif

Développe en Python une approche ML (supervisée et/ou non supervisée) pour aider les gestionnaires de magasins à prévoir les ventes futures.

## Question #1 : Préparation des données

Comme c'est souvent le cas dans les projets, le jeu de données peut nécessiter quelques manipulations pour être utilisable par une approche ML.

- Si tu rencontres des problèmes de qualité des données durant ta manipulation des données de ventes, comment les as-tu résolus?
- Limite-toi aux trois enjeux les plus pertinents selon toi (appuie-toi avec un visuel).
- Est-ce que les insights trouvés peuvent être transformés en features qui faciliteront l'apprentissage du modèle ML?

## Question #2 : Insights et interprétation

Identifie des insights qui, selon toi, peuvent contribuer à comprendre les variations des ventes.

- En tenant compte des parties prenantes visées par ta solution, comment interprètes-tu les résultats produits par ta solution ML? Comment cette solution ajoute-t-elle de la valeur pour ces parties prenantes?
- Selon toi, comment envisage-tu que les parties prenantes vont utiliser ta solution pour tenter de comprendre comment augmenter les ventes?

### Question #3 : Solution ML

Décris l'approche ML que tu as utilisé pour prévoir les ventes. Pourquoi as-tu choisi cette approche? Quels sont les avantages et les inconvénients de cette approche?

### Question #4 : Dégradation de la performance

Imaginons que ta solution est déployée et roule maintenant en production. Tu remarques que la performance de ton modèle se dégrade progressivement depuis les derniers mois. De plus, tu identifies également certaines variables dont les valeurs semblent avoir évolué durant la même période. Selon toi, quelle serait une raison qui explique cette situation et comment la résoudrais-tu?

### Question #5 : Intégration de l'IA générative

Comment l'IA générative pourrait-elle être intégrée dans un système de prévision des ventes pour soutenir les utilisateurs finaux de ce modèle?

- Donne un exemple d'architecture où l'IA générative pourrait être utilisée.
- Propose un exemple de prompt qui pourrait être utilisé pour interagir avec l'IA générative.

### Consignes & considérations additionnelles:

- Le but de cet exercice est d'avoir une conversation sur ta méthodologie et ton processus de réflexion. Il n'est pas attendu de toi de résoudre complètement le problème.
- Maintiens une structure claire et concise qui explique bien ton plan d'action.
- Ne complique pas trop tes solutions, ni le traitement des données. Cherche plutôt des solutions simples et laisse des commentaires sur ce que tu aurais aimé faire.
- Renvoie-nous tes réponses aux questions ainsi que le code ou « repository » que tu as développé. Assure-toi que nous soyons en mesure de reproduire tes résultats.
- Tu es libre d'utiliser le support de ton choix pour présenter ta solution (Notebook, Scripts, PDF, PowerPoint, Github, etc.).
- Dans la mesure du possible, garde à l'esprit:
  - La qualité de ton code et la documentation (i.e. commenter adéquatement)
  - Une programmation fonctionnelle et orientée objet
- Nous recommandons de ne pas passer plus de 6 heures sur ce test.