

Instructions

Félicitations! Nous sommes très 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. Nous t'avons transmis un jeu de données provenant de la plateforme Kickstarter (voir fichier « *ks_dataset.csv* »). 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 - password: moovai

Mise en situation

Certains promoteurs de projet tentent de comprendre comment augmenter le taux de réussite de leurs futures campagnes. Ils ont à leur disposition des données historiques de campagnes annoncées sur la plateforme Kickstarter.

Objectif

Développe en **Python** une approche ML (supervisée et/ou non supervisée) pour aider les promoteurs de projet à lancer des campagnes **à fort potentiel de réussite**.

Question #1 : 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 Kickstarter, comment les as-tu résolus?

Question #2 : Identifie des « insights » qui, selon toi, peuvent contribuer à comprendre le succès ou non des campagnes.

- Limite-toi aux trois observations les plus pertinentes selon-toi (appuies ces observations avec un visuel).
- Basé sur les insights mentionnés plus haut, y a-t-il un risque que des « confounding variables » (i.e. facteur de confusion) viennent affecter l'interprétation de tes observations ?
- Est-ce que les « insights » trouvés peuvent être transformés en « features » qui faciliteront l'apprentissage du modèle ML ?

Question#3 : Au niveau de la solution ML:

- 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 lancer des campagnes à haut taux de succès?

Question#4 : 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 évoluées durant la même période. **Selon toi, quel serait une raison qui explique cette situation et comment la ressouderais-tu ?**

Consignes & considérations additionnelles:

- Le but de cet exercice est d'avoir une conversation sur ta méthodologie et ton processus de pensée. **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.
- Nous t'encourageons à aller valider directement sur le site web de Kickstarter ta compréhension de la signification des variables si besoin.
- 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, gardez à l'esprit:
 - La qualité de ton code et la documentation (i.e. commenter adéquatement)
 - Une programmation fonctionnelle et orientée objet
- Si tu es sélectionné pour poursuivre le processus, nous t'inviterons à venir dans les bureaux de Moov afin de présenter ton travail.
- Dans un délai de cinq jours à partir d'aujourd'hui, tu as autant de temps que tu le souhaites pour compléter ce test technique. Prends soin de calculer de façon approximative le temps que tu as passé à le résoudre. Nous recommandons de ne pas passer plus de 6 heures sur ce test.