

Collecte de fonds

- Devoir en équipe, de deux à quatre personnes par équipe.
- Les données sont disponibles sur ZoneCours.
- Vous devez faire vos analyses avec R, et fournir votre code.
- Le devoir compte deux parties:
 - Rapport expliquant votre analyse, date de remise le 16 octobre 2024, sur ZoneCours, valeur 10 %.
 - Compétition de prévisions, avant 11:55PM le 4 novembre 2024, valeur 5 %.

Contexte

Vous aidez une fondation caritative dans sa campagne de financement 2024. La fondation compte des membres sur lesquelles elle a de l'information à laquelle vous aurez accès pour votre mandat. La fondation a l'opportunité d'envoyer une trousse de remerciement à ses membres, tout en les incitant à donner pour la campagne. Une entreprise locale vous offre de préparer des trousseaux à rabais. Normalement, la préparation et l'expédition d'une trousse coûte 25 \$ par personne, mais grâce à la collaboration de cette entreprises, la fondation peut en envoyer jusqu'à 60 000 au modique coût de 5 % chacune. Au-delà de 60 000, la fondation devra payer le coût total, soit 25 \$ l'unité.

Votre rôle consiste à déterminer quels individus devraient recevoir la trousse. Vous devez donc choisir combien de trousseaux, et pour qui. Votre objectif d'affaire est de maximiser les dons reçus pour la campagne. Les membres sont tous en droit de donner à la campagne. Ils peuvent donner ou non à la campagne, qu'ils reçoivent une trousse ou non.

Données

Un fichier zip contenant des données est disponible sur ZoneCours. Les données proviennent de différentes sources, et se trouvent donc dans plusieurs fichiers:

- *MembersList.csv*: Contient les informations de base au sujet des membres de la fondation à jour à la fin de 2023. Chaque membre possède un identifiant unique.
- *DonationHistory.csv*: Une liste de tous les dons reçus jusqu'en 2023. Chaque personne ne peut faire qu'un seul don par an.
- *NewsLetterRead.csv*: La fondation envoie une infolettre aux membres, puis trace qui a consulté la lettre ou non. Les données de 2023 vous sont fournies. Tous les membres sont inscrits à l'infolettre, mais ils ne la lisent pas tous.
- *SocialNetworkUsage.csv*: La fondation qui est active sur les réseaux sociaux a colligé des informations sur ses activités en ligne en 2023. Elle a réussi à identifier que certains usagers des médias sociaux étaient aussi membre, mais n'a pas cette information pour tous.

Étape 1 : Prévisions et rapport

Le problème pour ce devoir d'apparente plus à un vrai problème en entreprise qu'à un exercice académique. Ni la variables cible, ni les prédictors ne sont identifiées pour vous. Vous devez les définir vous-même. Plusieurs stratégies sont acceptables. Vous n'êtes pas tenus d'utiliser toutes les sources de données. Nous vous demandons toutefois de :

- Utiliser l'apprentissage supervisé (faire des prévisions),
- Décrire les prédictors et la variables cible que vous avez définis et expliquer pourquoi ce sont de bons choix,
- Faire une analyse exploratoire des variables créées,
- Évaluer la performance de votre modèle avec une méthodologie appropriée, et décrire votre approche,
- Dire combien de personnes devraient recevoir la trousse. Leur identité sera utile pour la partie 2.

Votre rapport doit être produit avec R Markdown ou Quarto. Avant la date limite, remettez sur ZoneCours (une copie par équipe) :1) Votre rapport en pdf, et 2) votre code R qui a servi à créer le rapport.

Étape 2 : Compétition

Le devoir est basé sur un problème d'affaire, et votre succès sera également évalué sur un résultat d'affaires. Préparez un fichier CSV avec une liste des IDs des membres qui recevront la trousse. Téléchargez votre décision sur le site <https://dsqgame.hec.ca/play> (les instructions de connexion paraîtront sur ZoneCours bientôt). Vous devez créer un profil par équipe. Lorsque vous téléchargez votre décision, vous verrez automatiquement combien d'argent vous avez levé (dons totaux moins le coût des troussees).

La plateforme permet de nombreux téléchargements (maximum 99 par équipe), et le concours restera ouvert après la remise du rapport. Vous avez le droit de faire évoluer votre modèle et d'essayer d'autres options. Vous n'êtes pas obligé de vous limiter au modèle présenté dans votre rapport écrit. À la fin du devoir, nous discuterons en classe des différentes stratégies que vous aurez adoptées.

Sur la plate-forme, vous verrez vos résultats, mais également ceux de toutes les équipes des deux groupes. Pendant que la partie est en cours, les résultats sont ceux du « interim leaderboard. » Vous pouvez faire de nombreux téléchargement, mais vous devrez choisir votre décision finale avant la fin. À la fin de la partie, le « real-life leaderboard » sera révélé. Afin d'établir vos résultats finaux. Le « real-life leaderboard » joue le rôle d'échantillon test. C'est une copie des données gardée secrète jusqu'à la fin du processus qui servira à mesurer la vraie performance de toutes les équipes.

Évaluation

Le rapport de la partie 1 est évaluée sur la qualité et l'exactitude de la méthodologie adoptée.

La note de la partie 2 est entièrement basée sur la performance. Le montant de base « baseline » correspond à l'argent récolté si aucune trousse n'est envoyée. Il suffit de battre le « baseline » pour obtenir la note de passage.

Chaque équipe aura une note de :

- 0% - si vous êtes sous le « baseline » du « interim leaderboard », mais au moins 50% si votre profit est supérieur.
- 100% - si vous êtes dans le top 10 % des équipes du « real-life leaderboard. »
- Les autres notes sont calculées par interpolation selon l'équation suivante selon le « real-life leaderboard » :
$$50\{1 + (x - B)/(T - B)\}$$
où x est votre profit, B le « baseline », et T est le profit de la dernière équipe parmi le top 10% qui a obtenu 100% selon le critère ci-dessus.

Dictionnaire de données

Les noms des variables dans les différentes tables sont évocateurs. Si vous avez des questions, demandez à votre enseignant.