

Scénario

Vous êtes consultant Data Analyst dans une entreprise spécialisée dans la data. Votre entreprise a décroché une prestation en régie au sein de l'Organisation nationale de lutte contre le faux-monnayage (ONCFM).

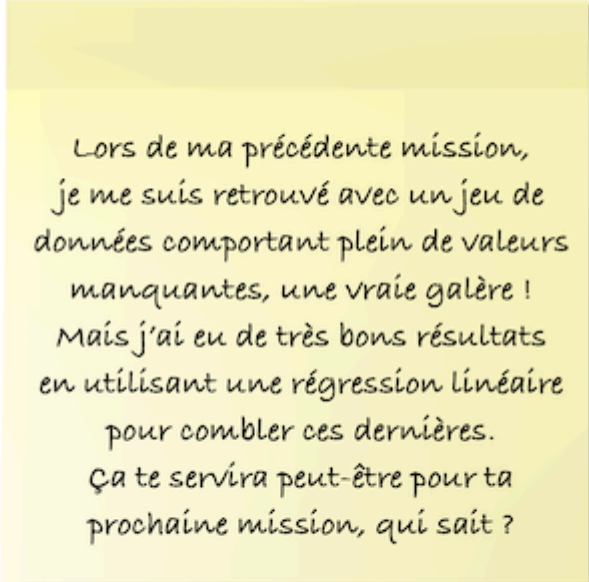


Cette institution a pour objectif de mettre en place des méthodes d'identification des contrefaçons des billets en euros. Ils font donc appel à vous, spécialiste de la data, pour mettre en place une modélisation qui serait capable d'identifier automatiquement les vrais des faux billets. Et ce à partir simplement de certaines dimensions du billet et des éléments qui le composent.

Voici [le cahier des charges de l'ONCFM](#), ainsi que [le jeu de données](#).

Le client souhaite que vous travailliez directement depuis ses locaux sous la responsabilité de Marie, responsable du projet d'analyse de données à l'ONCFM. Elle vous laissera une grande autonomie pendant votre mission, et vous demande simplement que vous lui présentiez vos résultats une fois la mission terminée. Elle souhaite voir quels sont les traitements et analyses que vous avez réalisés en amont, les différentes pistes explorées pour la construction de l'algorithme, ainsi que le modèle final retenu.

Après avoir lu en détail le cahier des charges, vous vous préparez à vous rendre à l'ONCFM pour prendre vos nouvelles fonctions. Vous notez tout de même un post-it qui se trouve sur le coin de votre bureau, laissé par un de vos collègues :



Lors de ma précédente mission,
je me suis retrouvé avec un jeu de
données comportant plein de valeurs
manquantes, une vraie galère !
Mais j'ai eu de très bons résultats
en utilisant une régression linéaire
pour combler ces dernières.
Ça te servira peut-être pour ta
prochaine mission, qui sait ?

Vous embarquez le post-it dans vos affaires et vous dirigez donc à présent vers les bureaux de l'ONCFM pour commencer votre mission.

Livrable

- Votre code en R ou Python contenant :
 - l'ensemble des traitements et des tests effectués ;
 - l'application finale.

Pour faciliter votre passage devant le jury, déposez sur la plateforme, dans un dossier zip nommé "**Titre_du_projet_nom_prénom**", le livrable du projet nommé comme suit : **Nom_Prénom_n° du livrable_nom du livrable_date de démarrage du projet**. Cela donnera :

- *Nom_Prénom_1_code_mmaaaa*

Par exemple, le livrable peut être nommé comme suit :
Dupont_Jean_1_code_012022.

Soutenance

Durant la présentation orale, votre mentor jouera le rôle de Marie, la responsable du projet à l'ONCFM à qui vous présentez vos résultats :

- **Présentation des livrables (20 minutes)**
 - Présentation de l'ensemble de votre cheminement, des traitements et analyses réalisés en amont, des différentes pistes explorées pour la construction de l'algorithme, et du modèle final retenu (15 minutes) ;
 - Test de l'algorithme en direct avec un jeu de données qui vous sera fourni pendant la soutenance, ayant la même forme que le jeu suivant : [FICHIER billets_production.csv](#) (5 minutes).
- **Discussion (5 minutes)**
 - Échange de questions/réponses : l'évaluateur pourra revenir sur certains points pour vous questionner sur vos choix.
 - L'évaluateur vous fera un retour sur votre prestation en soutenance.
- **Debriefing (5 minutes)**
 - À la fin de la soutenance, l'évaluateur arrêtera de jouer le rôle de Marie pour vous permettre de débriefer ensemble.

Votre évaluateur questionnera vos choix, soyez donc prêt à défendre votre travail. À la fin de la session, votre évaluateur reprendra son rôle de mentor pour que vous puissiez débriefer ensemble.

Votre présentation devrait durer 20 minutes (+/- 5 minutes). Puisque le respect des durées des présentations est important en milieu professionnel, les présentations en dessous de 15 minutes ou au-dessus de 25 minutes peuvent être refusées.

Compétences évaluées

- Opérer des classifications automatiques pour partitionner les données
- Réaliser une régression linéaire
- Réaliser une régression logistique
- Réaliser une analyse prédictive