

**Objectif:** Interface Shiny / Dashboard pour l'exploration de données.

- Le projet doit être réalisé en binôme. Merci de bien préciser les nom et prénom de chaque membre du binôme et de mettre en copie votre binôme lors de l'envoi des scripts et de votre rapport.
  - Les résultats de vos analyses doivent être commentés et reprendre les différents notions vues en cours.
  - Vous devez me faire parvenir votre projet (scripts et rapport séparés) pour le 12/03/2021 à 23h59 ([severine.affeldt@u-paris.fr](mailto:severine.affeldt@u-paris.fr)).
- 

## 1. Données pour ce mini-projet

Pour ce mini-projet, vous analyserez au choix les données suivantes:

- [Credit fraud](#)
- [Bank marketing](#)
- [Employee attrition](#)

## 2. Réalisation attendue

Créer une **shinyapp** permettant l'exploration du jeu de données choisi. Votre analyse devra être mise en ligne *via* votre compte shiny sur **shinyapps.io**. Elle reprendra les différentes explorations vues en cours dans la partie *churn*.

### 1. Onglet de présentation des données

- (a) Prévoyez un onglet de présentation des données.
- (b) Insérer l'analyse univariée des variables (choix dynamique des variables).
- (c) Insérer l'analyse bivariée des variables (choix dynamique des paires de variables).

### 2. Onglet spécifique à l'exploration des variables vis-à-vis du churn

- (a) Affichage (dynamique) de Customer attrition in data, Variables distribution in customer attrition.
- (b) Affichage ciblé pour une variable a priori d'intérêt (comme pour *tenure* dans le cours).

### 3. Entraînement de modèles

- (a) Choisissez deux modèles de classification supervisée (eg. logistic regression, KNN...).
- (b) Entraînez vos modèles sur vos données et faites leur évaluation.
- (c) Affichez les résultats de vos modèles dans l'interface.
- (d) Résumez clairement vos évaluations.

### 4. Afin de réduire le problème du déséquilibre des classes, le tutoriel ci-dessous propose des approches de oversampling, qui consiste à augmenter/diminuer le nombre d'instances de la classe minoritaire/majoritaire. Familiarisez-vous avec ces approches et utilisez-les pour réduire le déséquilibre des classes churn/non churn.

- [Resampling strategies for imbalanced datasets](#)
- (a) Entraînez vos modèles précédent sur votre jeu rééquilibré et faites leur évaluation.
- (b) Affichez les résultats de vos modèles dans l'interface.
- (c) Résumez clairement vos évaluations.