[Fichier](https://drive.google.com/drive/folders/12kj6HfkzOtgfMY1R2vRooKCTBQ_JnJtn?usp=sharing) avec les documents

# Réunion 12/10/2020 : Julie, Gaël, Bénédicte

Jeux de données:

* Adult

Cours [total = 18h]

* 1 Séance Visu (2h) **R**
  + Premiers cours sur des rappels de plot + chargement de données (en R)
  + Coefficient de corrélation veut dire différentes choses
  + Intuition sur les relations conditionnelles
  + Pairplot, jointes entre chaque feature et l’outcome
* 2 Séances réduction de la dimension ACP / ACM / Clustering (anciennement classif) (4h) **R**
  + Visu ACP/ACM

Cours: Julie (première séance - slides type X-HEC qui sont sur la dropbox), TD: Bénédicte (seconde séance - TD décathlon pour ACP, et puis ajouter un clustering)

* 1 Séance fitter et interpréter des modèles linéaires en R (2h) **R**
  + Logistic vs squared loss
  + Interpréter les valeurs de coefficients (attention au scaling des variables
  + Une régression multivarié != plusieurs régressions univariées
* 2 Séances modèles prédictifs (réutiliser du MOOC scikit-learn, 4h) **Python**
  + Overfit + autres
  + Introduction Scikit-learn
* 1 Séance Modèles avancés (2h): **Python**
  + Text
    - Bag of words
    - FastText
  + Image avec deep learning pré entraînés (démystifier le deep, savoir bien l’utiliser)
* 1 Séance Causal (2h): **R**
  + ATE,
  + IPW
  + régression
  + (seulement évoquer le Doublement Robust)
* 2 Séances pratiques: 2x2h (projet de groupe) (langage à choisir par groupe)
  + 3 attaques possible:
    - Modèle descriptif / stats
    - Modèle prédictif de l’outcome
    - Modèle prédictif pour les propensity scores