

## 00\_GetData.ipynb

September 7, 2021

```
[1]: # On importe la bibliothèque os, pour télécharger et décompresser les données
# et pour créer les répertoires de données
import os
```

```
[2]: # On définit ici les chemins relatifs des dossiers de données
# Ainsi que l'URL de téléchargement
source_path = '01_source'
data_path = '02_data'
source_link = 'https://s3-eu-west-1.amazonaws.com/static.oc-static.com/'\
+ 'prod/courses/files/Parcours_data_scientist/'\
+ 'Projet+-+Impl%C3%A9menter+un+mod%C3%A8le+de+scoring/'\
+ 'Projet+Mise+en+prod+-+home-credit-default-risk.zip'
```

```
[3]: # On crée un dossier "données sources", où l'on gardera les données brutes,
# câd non modifiées après leur téléchargement depuis leur source distante
os.mkdir(source_path)
# On télécharge les données depuis le lien distant en se plaçant dans "source"
os.system('cd ' + source_path + ' && wget ' + source_link)
```

[3]: 0

```
[4]: # On crée le dossier "données" qui contiendra une copie des données brutes
# que l'on utilisera pour visualiser, explorer et modifier les données à traiter
os.mkdir(data_path)
os.system('cd ' + source_path
+ ' && unzip -d ../' + data_path
+ ' ' + '`ls | grep .zip`')
```

[4]: 0