

# Layüs KICK-OFF Semaine 3

## Piscine DIG14

Par groupe de 5 personnes

### **Durée : 5 jours (toute la semaine)**

- ☞ 4 jours d'exploration, analyses, programmation :
  - 3j explorations/codes/entraînement modèle
  - 1j rédaction (Jupyter Notebook à bien présenter) , feature engineering, interprétation résultats. + Préparation Soutenance
- ☞ 1 jour Soutenance : suite entraînement + présentation projet par groupe

### **THEMATIQUES** : Segmentation Clients - Campagne Marketing

- ☞ Analyse Exploratoire des Données
- ☞ Impact Marketing/Business

### **MISSION :**

Suivi de 2200+ clients

La Direction Marketing d'une multinationale de la grande distribution a entendu vos prouesses de data analyste et vous engage pour tester vos capacités sur l'un de ses magasins.

L'objectif est double :

1. Chiffrer l'efficacité des campagnes marketing
2. Cerner la cible client (Segmenter)

### **Livrables attendus :**

- ☞ **Un Notebook Python (Jupyter Notebook)/R** rassemblant vos analyses exploratoires, vos modèles prédictifs et votre clustering client pour justifier la segmentation choisie.
- ☞ **Vous devez effectuer et justifier votre feature engineering** (signifie expliquer et rationaliser les choix que vous avez faits lors de la manipulation et de la transformation des caractéristiques (ou "features") de votre ensemble de données avant de former un modèle. Le feature engineering est un élément crucial du processus de modélisation des données, car il peut grandement influencer la performance de votre modèle. )  
Eventuellement ajouter des variables pertinentes et interpréter vos résultats en étant conscient de leurs limites éventuelles.
- ☞ **Une soutenance orale 15 min de parole, 10 min d'échanges** avec un jury d'experts pour présenter et défendre vos travaux et conclure sur leur pertinence vis-à-vis des attendus du client.

**Les objectifs sont larges :**

- ☞ Analyse de l'efficacité des campagnes
- ☞ Capacité à prédire la réponse client
- ☞ Possibilité de segmenter les clients en profils clairs pour cibler de manière optimale les campagnes marketing

**COMPLEMENTS :**

- ✓ Pas de technique imposée.
- ✓ En revanche, chaque groupe doit se lancer sur le rendu d'un Notebook Python/R (privilégier Python pour tous).
- ✓ Analyse exploratoire poussée ou des méthodes de classification simples....
- ✓ Selon,....si à l'aise, usage de méthodes type PCA

**Semaine guidée chaque jour :**

**LUNDI :**

- ⇒ Présentation KO
- ⇒ On découvre et on s'initie au mode de fonctionnement de Jupyter Notebook y compris avec les Markdown
- ⇒ En autonomie, on s'initie à Python :
  - Import de librairies
  - Pandas
  - Import d'un csv/excel
  - Affichage du Dataframe des données
  - Savoir comment :
    - Créer de nouvelles colonnes
    - Faire des Opérations simples sur des colonnes
  - Observer s'il y a des données manquantes → S'interroger sur ce que l'on en fait, si suppression, si remplacement et par quoi ?...)
  - Affichage de qqes caractéristiques du dataframe (shape, head, describe, ....)

**Important : on s'aide au sein du groupe, +ChatGPT, + on fait des tests sous Jupyter Notebook.**

**MPJ : Dispo en distanciel de 09h30 à 15h30.**

**MARDI :**

- ⇒ Matin : **Workshop Python** pour check les acquis suggéré le lundi + méthodes Python utiles pour le projet + bonnes pratiques pour nettoyer, encoder, scaler/normliser les données : tous ensemble (participation active)
- ⇒ **Indications (pour avancer sur le projet)** : Suggestions de nouvelles colonnes/variables spécifiques à créer.
- ⇒ Apres : Réalisation de la phase guidée du matin
- ⇒ Affichages de graphes, matrice de corrélation, etc....

**MPJ : Dispo en présentiel de 09h30 à 17h00.**

### MERCREDI :

- ⇒ Matin : Point avancement pour tous les groupes, remise dans la bonne direction si besoin
- ⇒ Matin : **Workshop PCA , KMeans**
- ⇒ Chaque groupe avance
- ⇒ Apres : **WORKSHOP si souhaité sur la méthodologie d'analyse et la technique**
- ⇒ Apres : On tente d'entraîner le modèle via une PCA et un Kmeans....

**MPJ : Dispo en présentiel de 09h30 à 17h30.**

### JEUDI :

- ⇒ Finalisation des analyses
- ⇒ Préparation du Jupyter Notebook (on peaufine, on supprime les affichages inutiles, on organise le document, on numérote les divers paragraphes, on commente les analyses, les résultats.)
- ⇒ Feature Engineering
- ⇒ Préparation Soutenance

**MPJ : Dispo en présentiel de 09h30 à 17h30.**

### L'essentiel est :

- ✓ D'apprendre à étudier les données de manière à choisir les variables pertinentes,
- ✓ De modifier les données de manière à ajouter des variables pertinentes
  - pourquoi garder la date de naissance quand on peut créer une variable d'âge,
  - pourquoi ne pas regrouper les produits,
  - pourquoi ne pas simplifier le niveau d'éducation de manière numérique et catégorielle pour pouvoir créer des corrélations/régressions etc.
  - supprimer les variables d'origine non importantes : deux variables n'ont pas d'explication dans le dictionnaire de variables, c'est normal, à vous de voir quoi en faire et si c'est vraiment nécessaire  
=> indication : en faisant simplement une analyse des quantiles/moyennes sur ces deux variables, même sans en comprendre pleinement la nature, on comprend qu'on peut simplement les exclure de la réflexion sans perte d'information.
- ✓ A vous également d'utiliser les bons sous-ensembles pour lancer les algorithmes prédictifs et savoir les interpréter.

**VENDREDI:** Présentation de vos résultats devant un jury par groupe.

**MPJ : Dispo en présentiel de 09h30 à 17h30.**