PROJECT 4

Questions :

* Min samples for optics and dbscan 100 ? Nickel, rien d-autre a faire
* Vous avez explicité la stratégie d'ajout de nouveaux clients. ???
  + Un nouveau client n’a pas encore acheté de produits et j’ai des données de comportement d’achat uniquement ici, donc pour l’instant il ne va pas entrer dans un cluster. Je peux l’ajouter dans un cluster à part. En vrai tu as d’autres métriques genre utm\_campaign, navigation sur notre site, os du client (Apple clients va mettre plus d’agent)… Mais là pas. Quand achat then tu peux predict son cluster id.

Le dendograme a 5 couleurs parce qu’il te propose 5 comme bon nombre de clusters, la règle est online (barre verticale la plus haute?).

# Current method: en dessous de 0,7 il faut faire de la maintenance donc recreer un nouveau modele.  
# Vu mes resultats ca veut dire ou je fais cette maintenance vers 9  
# semaines ou vers 7-8 mois.  
  
# Current methode est meilleure quand la saisonabilite est tres forte genre tu as des soldes en automne,  
# sinon la nouvelle est meilleure. Cest pour ca que j'ai 2 sunken  
# courbes, genre 2 saisons speciales.

Mes methodes sont liees a la distribution des achats donc cest pas super representatif.

##### Définir la stratégie d’élaboration d’un modèle d’apprentissage non supervisé, sélectionner et entraîner des modèles adaptés à une problématique métier afin de réaliser une segmentation ou une réduction de données.

* CE1 Les variables catégorielles identifiées ont été transformées en fonction du besoin (par exemple via OneHotEncoder ou TargetEncoder). → THERE IS NONE IN THIS PROJECT.
* CE2 Vous avez créé de nouvelles variables à partir de variables existantes.
* CE3 Vous avez réalisé des transformations mathématiques lorsque c'est requis pour transformer les distributions de variables.
* CE4 Vous avez normalisé les variables lorsque c'est requis.
* CE5 Vous avez défini la stratégie d’élaboration d’un modèle pour répondre à un besoin métier.
* CE6 Vous avez proposé un nombre de segments et une répartition adaptés à la problématique métier.
* CE7 Vous avez explicité la stratégie d'ajout de nouveaux clients.
* CE8 Vous avez pris en compte la nature des variables d'entrée dans le choix des algorithmes et de la distance.
* CE9 Vous avez testé et comparé plusieurs algorithmes.

##### Évaluer les performances des modèles d’apprentissage non supervisé selon différents critères (scores, temps d'entraînement, etc.) en adaptant les paramètres afin de choisir le modèle le plus performant pour la problématique métier.

* CE1 Vous avez choisi des métriques adaptées pour évaluer les valeurs candidates pour le nombre de segments (selon algorithme, par exemple : coefficient silhouette / elbow pour le KMeans, eps/min samples pour le DBScan).
* CE2 Vous avez évalué la forme des clusters.
* CE3 Vous avez évalué la stabilité des clusters à l’initialisation.
* CE4 Vous avez optimisé les hyper-paramètres pertinents dans les différents algorithmes.
* CE5 Vous avez justifié le choix final d'algorithme et d'hyper-paramètres.
* CE6 Vous avez analysé la stabilité du modèle dans le temps afin d’en assurer sa maintenance
* CE7 La convention PEP8 est respectée. Le code est commenté (commentaires réguliers, docstrings dans les fonctions) .