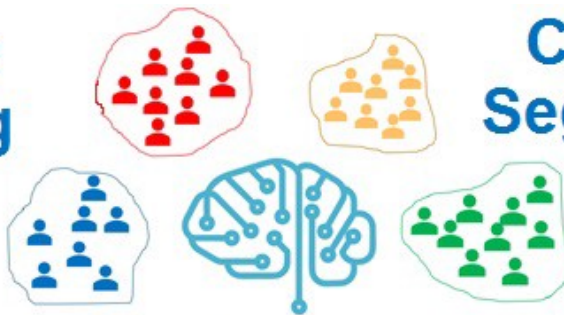


Machine Learning



Customer Segmentation



Le département Marketing et Communication de la division Lubmarine de Total souhaite construire de nouvelles offres adaptées à leurs clients grâce à une analyse fine de ces derniers, via une approche mixte analytique et créative.

Dans ce cadre, votre manager vous a confié une mission d'accompagnement visant à construire une segmentation pertinente basée sur la data et la connaissance terrain. Cette première mission doit aboutir à la définition de personae Clients représentant les segments majeurs de Total Lubmarine. La démarche doit être basée sur du clustering réalisé sur une base consolidée de données clients, ainsi que sur des recommandations d'optimisation des offres de services.

Au début de votre mission, vous n'avez pas accès aux données client. Vous allez utiliser la base de données disponible [ici](#). Il s'agit d'une base de données anonymisée comportant des informations sur l'historique de commandes, les produits achetés, les commentaires de satisfaction, et la localisation des clients depuis janvier 2017.

Enfin, votre client, Total Lubmarine, a spécifié sa demande ainsi :

1. La solution proposée doit être exploitable et facile d'utilisation pour l'équipe marketing.
2. Vous évalueriez la fréquence à laquelle la segmentation doit être mise à jour, afin de pouvoir effectuer un devis de contrat de maintenance.
3. Le code fourni doit être facile à comprendre et utilisable par les data scientists du client

Livrables

1. Un notebook de l'analyse exploratoire
2. Un notebook d'essais des différentes approches de clustering
3. Un support de présentation pour la soutenance.

Compétences	Critères d'évaluation
Sélectionner les variables pertinentes d'un modèle d'apprentissage non supervisé	<ul style="list-style-type: none"> ➤ Les variables pertinentes ont été sélectionnées ➤ Une analyse exploratoire pour mieux comprendre les données
Développer un modèle d'apprentissage non supervisé adapté à la problématique du client	<ul style="list-style-type: none"> ➤ Le nombre de segments est adapté à la problématique métier. ➤ Au moins 3 modèles de clusterings sont étudiés (partie théorique) ➤ Au moins 2 modèles ont été testés ➤ Les hyperparamètres des modèles sont optimisés
Évaluer les performances d'un modèle de clustering	<ul style="list-style-type: none"> ➤ La qualité des clusters est évaluée ➤ La stabilité des clusters est évaluée metrics ➤ La stratégie d'ajout d'un nouveau cluster est expliquée Adding New Observations to Existing Clusters
Qualité des livrables	<ul style="list-style-type: none"> ➤ La convention pep8 est respectée ➤ le notebook est commenté et facile à utiliser ➤ support de présentation clair et destiné à l'équipe marketing

Pour vous entraîner, vous pouvez consulter ce challenge kaggle [Brazilian E-Commerce Public Dataset by Olist](#)

Durée du projet : 5 jours