**Segmentez les clients d’un site e-commerce**

**Enoncé :**

Vous êtes consultant pour Olist, une entreprise brésilienne qui propose une solution de vente sur les marketplaces en ligne.

Olist souhaite que vous fournissiez à ses équipes d'e-commerce une segmentation des clients qu’elles pourront utiliser au quotidien pour leurs campagnes de communication.

Votre objectif est de comprendre les différents types d’utilisateurs grâce à leur comportement et à leurs données personnelles.

Vous devrez fournir à l’équipe Marketing

> une description actionnable de votre segmentation et de sa logique sous-jacente pour une utilisation optimale,

> ainsi qu’une proposition de contrat de maintenance basée sur une analyse de la stabilité des segments au cours du temps.

Les données :

Pour cette mission, Olist vous fournit 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.

Votre mission :

Votre mission est d’aider les équipes d’Olist à comprendre les différents types d'utilisateurs. Vous utiliserez donc des méthodes non supervisées pour regrouper des clients de profils similaires. Ces catégories pourront être utilisées par l’équipe Marketing pour mieux communiquer.

Vous créez donc un notebook et démarrez votre travail d’analyse exploratoire.

Après quelques premières analyses, vous vous rendez compte qu’Olist ne vous a pas fourni beaucoup de données ; Vous enquêtez donc auprès de l’entreprise pour obtenir quelques informations complémentaires, et vérifier que vous avez bien compris la mission.

Voici sa réponse :

*Bonjour,*

*Pour des raisons de confidentialité, nous ne pouvons pas vous fournir beaucoup de données à ce stade.*

*Ensuite, en raison de ressources limitées, nous avons dû vous fournir l’ensemble des données, alors que seule une partie va vous intéresser. Nos dashboards internes nous indiquent en effet que seuls 3 % des clients du fichier de données partagé avec vous ont réalisé plusieurs commandes.*

*Nous sommes confiants sur le fait que les données à disposition suffiront pour réaliser un premier clustering. Cela a déjà été fait par d’autres prestataires par le passé, avec encore moins de données.*

*La segmentation proposée doit être exploitable et facile d’utilisation par notre équipe Marketing. Elle doit au minimum pouvoir différencier les bons et moins bons clients en termes de commandes et de satisfaction. Nous attendons bien sûr une segmentation sur l’ensemble des clients.*

*Dans un deuxième temps, une fois le modèle de segmentation choisi, nous souhaiterions que vous nous fassiez une recommandation de fréquence à laquelle la segmentation doit être mise à jour pour rester pertinente, afin de pouvoir effectuer un devis de contrat de maintenance.*

*Pour information, le code fourni doit respecter la convention PEP8, pour être utilisable par Olist.*

*À titre d'exemple, je joins à ce courriel la segmentation RFM réalisée par votre prédécesseur il y a quelques années. Mais c’est uniquement à titre d’exemple, j'attends un travail plus approfondi de votre part dans le cadre de ce projet (i.e. RFM, satisfaction et autres possibilités).*

*Bien à vous,*

*Juan de l’équipe Marketing*

Vous réalisez donc trois autres documents afin de préparer une réponse à Olist :

> un notebook avec des essais des différentes approches de modélisation ;

> un notebook de simulation pour déterminer la fréquence nécessaire de mise à jour du modèle de segmentation, afin que celui-ci reste pertinent ;

> une présentation pour un collègue afin d’obtenir ses retours sur votre approche.

**Livrables :**

Un notebook de l'analyse exploratoire (non ‘cleané’, pour comprendre votre démarche).

Un notebook (ou code commenté au choix) d’essais des différentes approches de modélisation (non ‘cleané’, pour comprendre votre démarche).

Un notebook de simulation pour déterminer la fréquence nécessaire de mise à jour du modèle de segmentation.

Un support de présentation pour présenter votre travail à un collègue.

**Référentiel d'évaluation :**

Compétences évaluées :

> Sélectionner et entraîner des modèles d’apprentissage non-supervisé

> Évaluer les performances des modèles d’apprentissage non supervisé

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).

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 sa stratégie d’élaboration d’un modèle pour répondre à un besoin métier.

CE6 Vous avez proposé un nombre de segments et une répartition adaptée 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 hyperparamètres pertinents dans les différents algorithmes.

CE5 Vous avez justifié le choix final d'algorithme et d'hyperparamètres.

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, doc-strings dans les fonctions).