

Effectuer une segmentation client

Contexte

Vous êtes Data scientist dans l'agence DataFlow à Miami depuis quelques mois. Votre responsable vous confie une mission pour un client important, [Olist](#), un site de e-commerce brésilien.



Olist aimerait améliorer l'impact de ses campagnes marketing grâce à une segmentation client. En effet, adapter sa communication en fonction du client permet d'améliorer les performances de ses campagnes.

Voici le mail de votre responsable :

Bonjour Roxane,

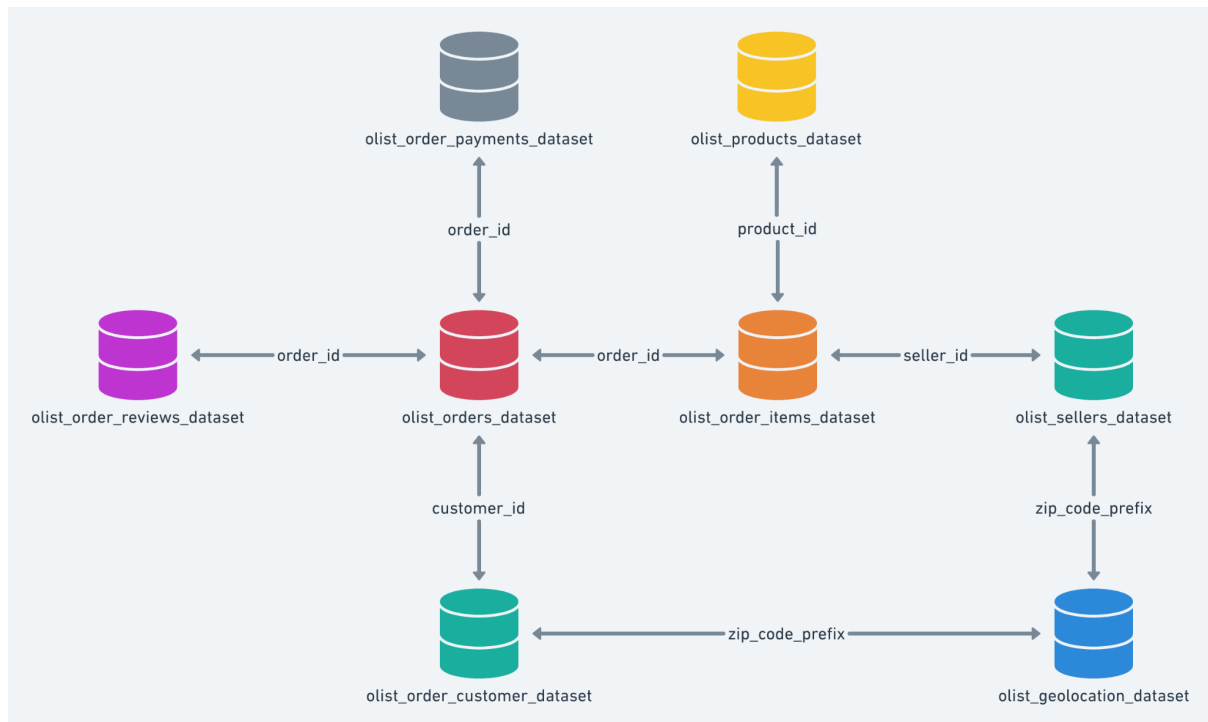
Je te confie le dossier Olist, l'objectif est d' aider leurs équipes marketing à améliorer le taux de retour de leur campagne de communication. En utilisant des méthodes de clustering comme le k-means il faudrait proposer une segmentation des clients basée sur leur comportement d'achat (les données du dataset) mais également à partir d'informations déduites comme les données RFM. Avant de te lancer dans la modélisation, fais également une analyse exploratoire pour analyser les performances de l'entreprise et les tendances globales.

Afin de pouvoir visualiser ces différents clusters, il faudra utiliser des techniques de réduction de dimension, comme l' ACP ou le T-SNE.

Il faudra ensuite utiliser les caractéristiques spécifiques de chaque cluster afin d'émettre des recommandations actionnables par l'équipe marketing. Suite à ton analyse il faudra présenter tes resultats grace a des slides aux équipes d'OLIST.

Il faudra également proposer un contrat de maintenance basé sur l' analyse de la stabilité des segments au cours du temps. En effet, tu devras évaluer la fréquence à laquelle cette segmentation doit être mise à jour (tu peux te baser sur le score ARI).

Tu as à ta disposition une base de données relationnelle anonymisée de 2016 à 2018 dont voici le schéma :



Tu peux récupérer les jeux de données à cette [adresse](#).

Cette analyse doit être prête pour vendredi, ce qui te laisse 5 jours pour réaliser cette analyse. N'oublie pas que tu vas t'adresser à des équipes Marketing et des dirigeants d'Olist, il faudra utiliser le jargon du machine learning avec parcimonie.

Modalités pédagogiques

- Projet individuel.
- Temps pour réaliser le projet : 5 jours.

Critères de performance

- Utiliser des méthodes non supervisées afin de répondre à une problématique métier.
- Méthode supervisée ; feature engineering gestion des hyperparamètres, évaluation des performances, et évaluation des clusters et de la stabilité de ces derniers.
- Adapter son discours en fonction des interlocuteurs.
- Émettre des recommandations marketing pertinentes.

Modalités d'évaluation

- Soutenance orale de vingt minutes
- Tirage au sort

Livrables

- Notebook de nettoyage
- Notebook de modélisation;
- Support de présentation.

Ressources

[Explorez vos données avec des algorithmes non supervisés](#)

[Segmentation RFM](#)

[Indice de Rand](#)