

[Retour à mon espace](#)

Segmentez des clients d'un site e-commerce

#### CONTENU

- 📄 Mission - Segmente...
- 📄 Livrables et soutena...
- 📊 Évaluation

#### SUPPORTS PÉDAGOGIQUES

- 👤 Companion **Nouveau**
- 📖 Cours
- 📖 Ressources

## Mission - Segmentez des clients d'un site e-commerce



### Comment allez vous procéder ?

Cette mission suit un scénario de projet professionnel.

Vous pouvez suivre les étapes pour vous aider à réaliser vos livrables.

Avant de démarrer, nous vous conseillons de :

- lire toute la mission et ses documents liés ;
- prendre des notes sur ce que vous avez compris ;
- consulter les étapes pour vous guider ;
- préparer une liste de questions pour votre première session de mentorat.

### Prêt à mener la mission ?

28 mai 2024 - le fichier contenant les requêtes SQL est modifié. La question 4 a été changée. Si vous êtes en train de travailler sur le projet, vous pouvez consulter la version précédente des questions sur [ce lien](#). si vous démarrez, suivez directement les consignes ci-dessous.

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



[Retour à mon espace](#)

## Segmentez des clients d'un site e-commerce

### CONTENU

[Mission - Segmente...](#)[Livrables et soutena...](#)[Évaluation](#)

### SUPPORTS PÉDAGOGIQUES

[Companion](#) **Nouveau**[Cours](#)[Ressources](#)

Votre rôle est d'accompagner Olist dans leur projet de monter une équipe Data et leur premier cas d'usage Data Science autour de la segmentation client.

Vous venez de rejoindre Olist dans un contexte de recrutement intensif afin de monter leur équipe Data. En effet, ils font d'abord appel à des consultants comme vous en renfort sur les sujets les plus critiques en attendant d'internaliser les compétences. Avant de réaliser la segmentation client, on vous attribue une première mission pour aider Fernanda, Lead Data Analyst chez Olist.

**De :** Fernanda

**A :** Moi

**Objet :** dashboard Customer Experience

Bonjour,

Bienvenue chez Olist ! Nous sommes ravis d'avoir du soutien en cette période de mise en place de notre écosystème Data.

A ce sujet d'ailleurs, l'un de nos projets Data phares du moment est la construction et la maintenance de notre Dashboard au service des équipes Customer Experience. Nous y exposons les KPIs essentiels pour que les équipes puissent avoir de la visibilité sur les états, les villes, ou les vendeurs qui nécessitent un suivi de près de la part de notre service client.

Nous sommes en train d'alimenter les résultats de ce Dashboard avec des requêtes qui interrogent la base de données SQL à laquelle tu as accès.

C'est sur l'implémentation de certaines requêtes urgentes que tu peux nous donner un coup de main le temps qu'un nouveau Data Analyst rejoigne l'équipe pour prendre le relais.

Je mets en PJ la liste de requêtes SQL que nous avons besoin d'intégrer au Dashboard.

Cordialement,  
Fernanda  
Lead Data Analyst chez Olist

Pieces jointes:

- [Liste de requêtes SQL](#)

Après avoir étudié les éléments partagés par Fernanda vous vous mettez au travail.



[Retour à mon espace](#)

## Segmentez des clients d'un site e-commerce

### CONTENU

[Mission - Segmente...](#)[Livrables et soutena...](#)[Évaluation](#)

### SUPPORTS PÉDAGOGIQUES

[Companion](#) Nouveau[Cours](#)[Ressources](#)

Cette urgence résolue, vous pouvez enfin passer à votre mission principale : 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 actionable** 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 des commandes, les produits achetés, les commentaires de satisfaction, et la localisation des clients depuis janvier 2017.

Vous devez 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.

### 3 jours plus tard

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 :

**De :** João

**A :** Moi

**Objet :** re : segmentation des clients

Bonjour,

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



[Retour à mon espace](#)

## Segmentez des clients d'un site e-commerce

### CONTENU

- Mission - Segmente...
- Livrables et soutena...
- Évaluation

### SUPPORTS PÉDAGOGIQUES

- Companion **Nouveau**
- Cours
- Ressources

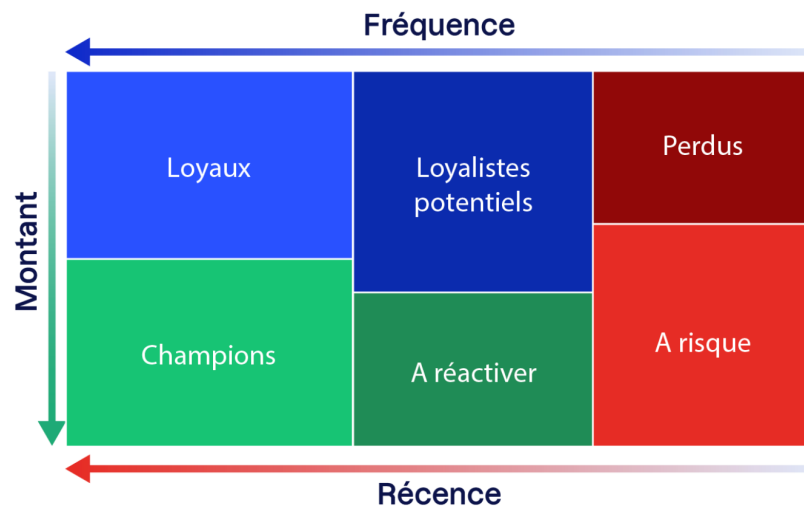
Dans un deuxième temps, une fois le modèle de segmentation choisi, nous souhaiterions que vous nous fassiez une recommandation de fréquence à laquelle la segmentation doit être mise à jour pour rester pertinente, afin de pouvoir effectuer un devis de contrat de maintenance.

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

À titre d'exemple, je joins à cet e-mail 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,  
João de l'équipe Marketing

Pièce-jointe :



Vous réalisez 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

C'est parti pour cette dernière ligne droite !

## Étapes

Étape 1 : Répondez aux requêtes SQL



[Retour à mon espace](#)

## Segmentez des clients d'un site e-commerce

### CONTENU

[Mission - Segmente...](#)[Livrables et soutena...](#)[Évaluation](#)

### SUPPORTS PÉDAGOGIQUES

[Companion](#) **Nouveau**[Cours](#)[Ressources](#)

- approprié les ressources d'apprentissage des bases de SQL ainsi que la présentation du jeu de données sur Kaggle ;
- installé un logiciel de gestion de bases de données (Workbench, DBeaver, extension SQLTools de VSCode etc.)

#### Une fois cette étape terminée, je devrais avoir :

- finalisé un script avec les 4 requêtes SQL demandées par Fernanda.

#### Recommandations :

- Réalisez quelques opérations très basiques pour comprendre le contenu de chaque table.
- Structurez vos requêtes avec la clause WITH AS afin de décomposer les questions en sous-requêtes simples et avoir un script plus lisible.
- Interrogez-vous sur vos choix concernant les types de jointures pour chaque requête (LEFT, INNER, OUTER etc.)

#### Ressources :

- [Cours "Requêtez une base de données avec SQL"](#)
- [Présentation du jeu de données sur Kaggle](#)
- [SQLZoo](#)

### Étape 2 : Transformez des données



#### Avant de démarrer cette étape, je dois avoir :

- suffisamment d'aisance en SQL pour réaliser des agrégations.

#### Une fois cette étape terminée, je devrais avoir :

- achevé un notebook avec la partie feature engineering, incluant la création de features par client.

#### Recommandations :

- Réalisez des requêtes exploratoires pour approfondir votre compréhension du jeu de données.
- Interrogez vous sur la part de features calculées en SQL et la part de features calculée en Python, en fonction du nombre et de la complexité des opérations nécessaires pour calculer la feature.
- Conservez toutes les commandes et focalisez-vous pour commencer sur des features clés comme RFM pour le ciblage marketing.
- Effectuez des transformations de variables (normalizations, encodage des variables catégorielles etc.), en fonction de l'algorithme de clustering que vous souhaitez utiliser et ses conditions mathématiques d'utilisation.

#### Points de vigilance :



[Retour à mon espace](#)

## Segmentez des clients d'un site e-commerce

### CONTENU

[Mission - Segmente...](#)[Livrables et soutena...](#)[Évaluation](#)

### SUPPORTS PÉDAGOGIQUES

[Companion](#) Nouveau[Cours](#)[Ressources](#)

#### Ressources :

- [Documentation sur la méthode RFM](#)

### Étape 3 : Élaborez un modèle de clustering

#### Avant de démarrer cette étape, je dois avoir :

- sélectionné des features pertinentes pour le clustering ;
- acquis les bases théoriques nécessaire pour pouvoir réaliser et interpréter un clustering.

#### Une fois cette étape terminée, je devrais avoir :

- réalisé un notebook avec différentes simulations d'algorithmes de clustering.

#### Recommandations :

- Réalisez d'abord un clustering de type k-means avec un nombre de clusters relativement faible (entre 3 et 6) afin de vous familiariser avec la démarche d'analyse.
- Proposez un nombre optimal de clusters en utilisant une méthode data-driven de votre choix (plusieurs méthodes existent dans la littérature) ainsi qu'en utilisant une méthode plus orientée métier.
- Justifiez clairement et avec des métriques quantitatives la qualité de votre clustering (k-means ou autre).
- Caractérisez vos clusters en établissant le lien entre chaque cluster et vos features en entrée de l'algorithme.
- Il est toujours préférable de légèrement perdre en performance de clustering si cela permet de significativement gagner en tangibilité métier des différents clusters.

#### Points de vigilance :

- Choisissez un nombre de clusters gérable et pertinent du point de vue métier.
- Assurez vous que vos clusters aient un sens métier assez tangible, n'hésitez pas à revoir vos choix de features en fonction

#### Ressources :

- [Réalisez une analyse exploratoire de données](#)
- [Librairie Yellowbrick d'évaluation technique de clusters \(coefficient de silhouette, distorsion\)](#)

### Étape 4 : Simulez un Contrat de Maintenance

#### Avant de démarrer cette étape, je dois avoir :

- sécurisé un clustering final optimisé et interprétable d'un point de vue métier.

#### Une fois cette étape terminée, je devrais avoir :



[Retour à mon espace](#)

## Segmentez des clients d'un site e-commerce

### CONTENU

[Mission - Segmente...](#)[Livrables et soutena...](#)[Évaluation](#)

### SUPPORTS PÉDAGOGIQUES

[Companion](#) Nouveau[Cours](#)[Ressources](#)

#### Recommandations :

- Utilisez l'ARI pour mesurer la divergence des clusters et déterminer quand réentraîner le modèle.
- Il serait intéressant également de regarder l'évolution de la distribution de vos features numériques dans le temps, en utilisant des méthodes graphiques ou un test de Kolmogorov-Smirnov.

#### Points de vigilance :

- Assurez la cohérence entre les données utilisées pour l'entraînement initial et les prédictions ultérieures.

#### Ressources :

- [Documentation sur l'Adjusted Rand Index \(ARI\)](#)

### Étape 5 : Préparez la soutenance

#### Avant de démarrer cette étape, je dois avoir :

- finalisé la partie modélisation dans son entièreté.

#### Une fois cette étape terminée, je devrais avoir :

- la capacité de justifier le choix de mes features, de mon nombre de clusters, de ma démarche d'évaluation de la qualité de mon clustering, du sens métier de chaque cluster et de mon approche pour mesurer la stabilité de l'algorithme dans le temps.

#### Recommandations :

- Préparez une présentation claire de votre modèle de clustering, en mettant l'accent sur les insights métier et la pertinence des segments identifiés.
- Soyez prêt à discuter des choix techniques et méthodologiques effectués au cours du projet.
- Soyez conscient du temps, vous n'allez pas pouvoir présenter toutes les analyses effectuées. Compresses à l'essentiel votre storytelling.

Projet terminé !



[Retour à mon espace](#)

## Segmentez des clients d'un site e-commerce

### CONTENU

[Mission - Segmente...](#)[Livrables et soutena...](#)[Évaluation](#)

### SUPPORTS PÉDAGOGIQUES

[Companion](#) **Nouveau**[Cours](#)[Ressources](#)