

Segmentation clients

Olist



Sommaire

01 LE CONTEXTE

Problématique et
méthodologie mise en oeuvre

02 LES DONNÉES

Nettoyage, construction de
variables et fait marquants

03 LES PISTES DE MODELISATION

Avantages et inconvénients
des pistes de modélisation

04 MODÈLE FINAL

Segmentation et maintenance

05 CONCLUSIONS





01

LE CONTEXTE

Problématique et actions à mettre en œuvre



Le besoin



SEGMENTER

Des profils types
actionnables



MAINTENIR

Une évaluation de la
fréquence





La méthodologie

NETTOYER

les données

01

02

CONSTRUIRE

des indicateurs

FORMULER

des hypothèses de
modélisation

03

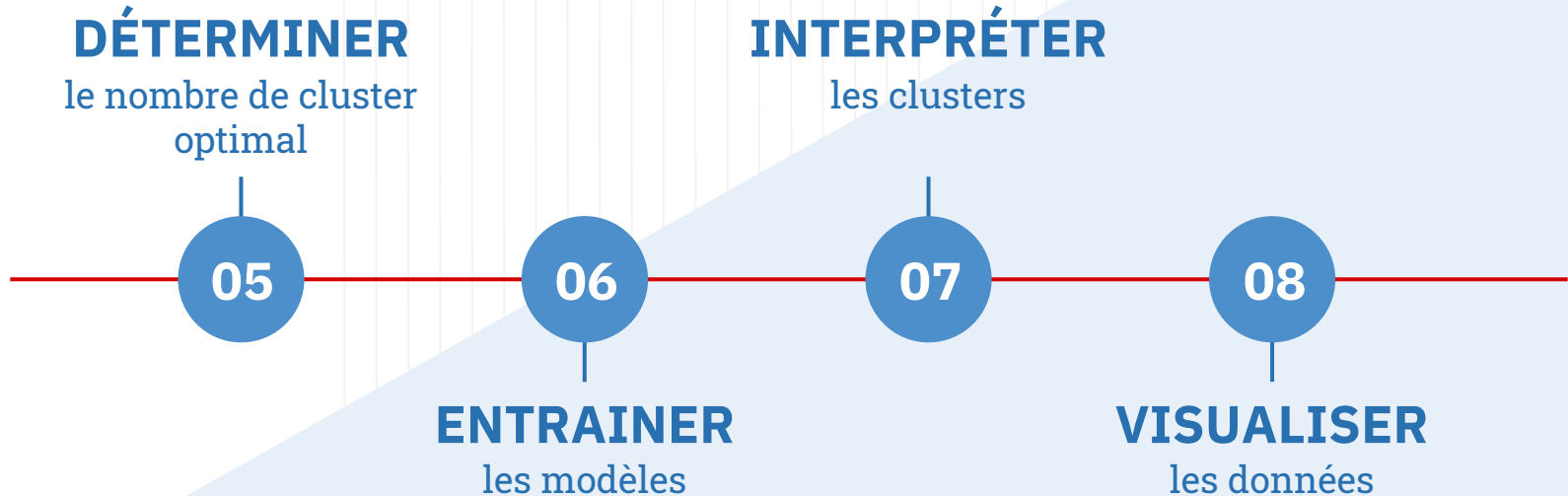
04

RÉDUIRE

les dimensions



La méthodologie





La méthodologie

CHOISIR

le modèle retenu

09

10

ÉVALUER

La fréquence de
maintenance

LES DONNÉES

Nettoyage, construction de variables et fait
marquants

02

Riches mais à reconditionner

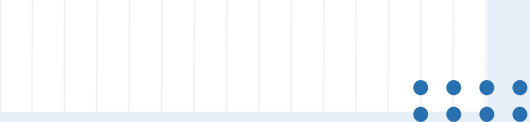
7 extractions



Listing des clients,
localisation, produits
commandés, paiements,
évaluations client,
commandes, vendeurs

1 travail de remise en forme

Disponibles par produits commandés ...
... à ramener à l'échelle du client



Nettoyage

	Données supprimées	Proportion supprimée	Données restantes
Avant nettoyage	0	0%	119 151
Commandes dupliquées	5 726	4,81%	113 425
Statuts différents de “delivered”	3 228	2,84%	110 197
Produits non catégorisés	0 (1 559)	0%	110 197
Commandes sans informations de paiement	3	0,003%	110 194
Commandes avec 0 échéances de paiements	3	0,003%	110 191
Commandes avec 0 échéances de paiements	8	0,007%	110 183

Indicateurs généraux



FRÉQUENCE

Dernier achat



VOLUME

Nombre de commandes

Nombre de produits



CA

Panier moyen

Prix moyen



SATISFACTION

Note moyenne

Proportion de
commandes avec reviews



PAIEMENT

Proportion de moyens de
paiement

Nombre d'échéances



LOGISTIQUE

Délais de livraison



LOCALISATION

Densité de clients

Proportion d'achats
locaux



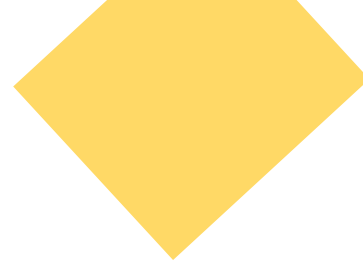
Indicateurs produits spécifiques



**TOP 10 FAMILLES
PRODUITS**



**NOMBRE DE
FAMILLE PRODUIT**



97 %

Commande unique

88 %

Produit unique

98 %

Une seule famille

27 %

Sur les 12 premiers mois



03

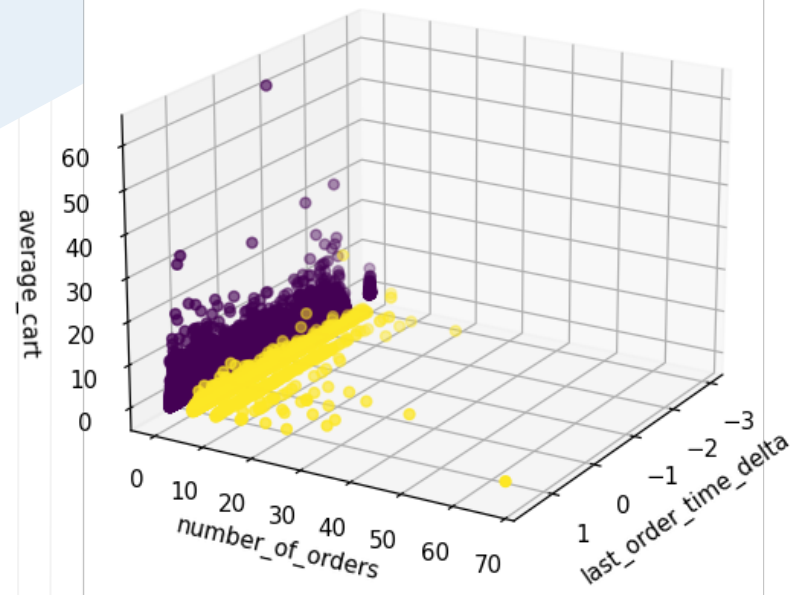
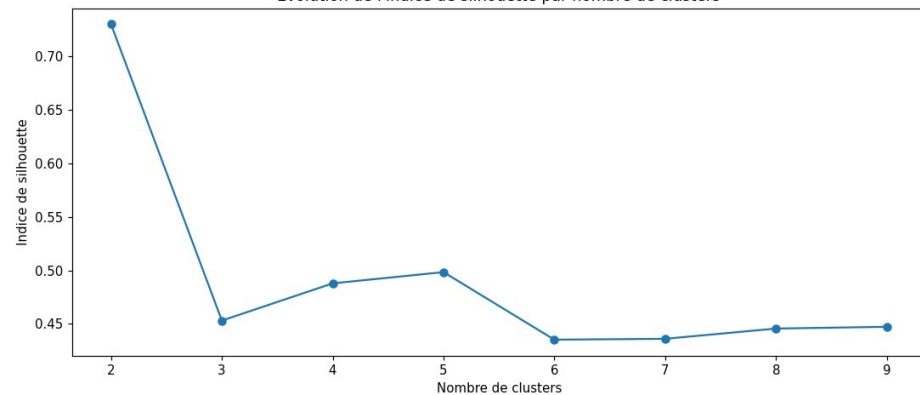
LES PISTES DE MODELISATION

Avantages et inconvénients

RFM

Clustering RFM

Evolution de l'indice de silhouette par nombre de clusters



Top 10 familles de produits

Evolution de l'indice de silhouette par nombre de clusters

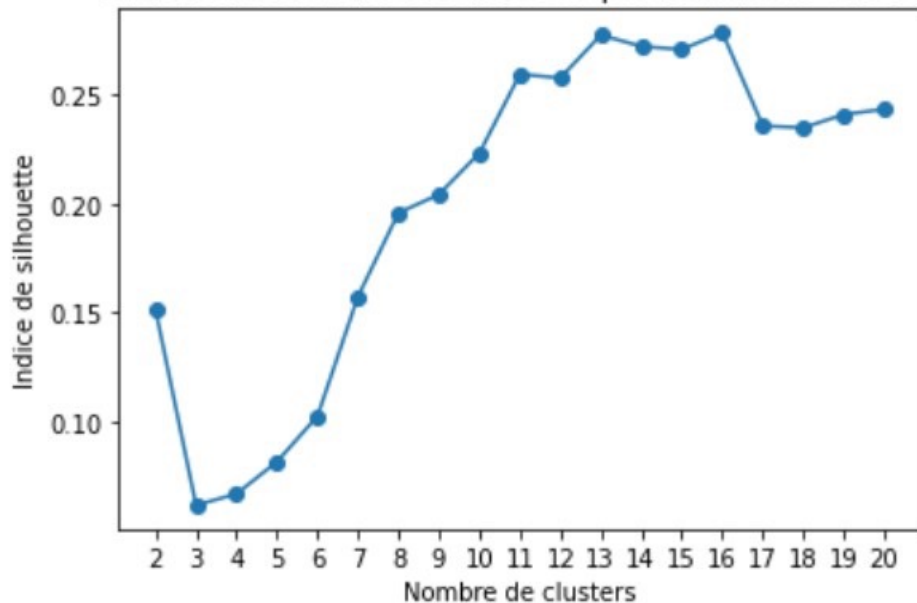
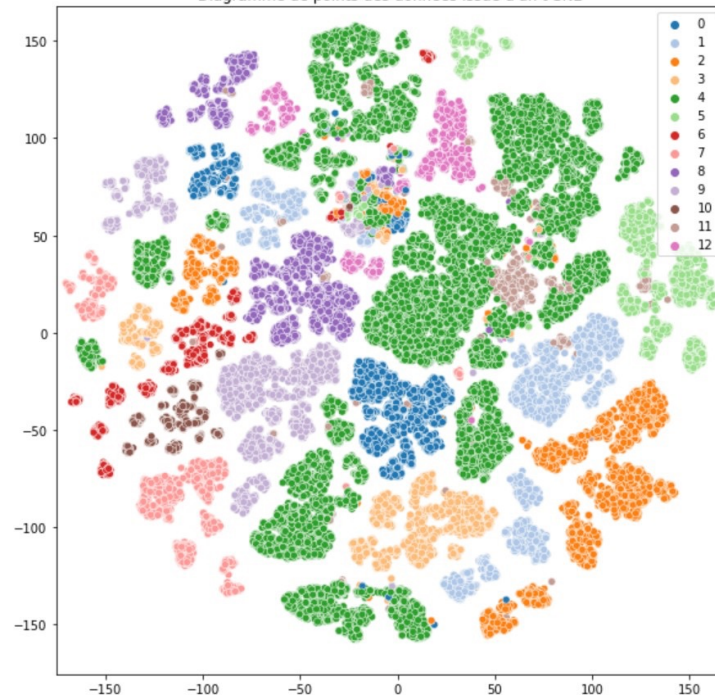


Diagramme de points des données issue d'un t-SNE



Nombre de familles de produits

Evolution de l'indice de silhouette par nombre de clusters

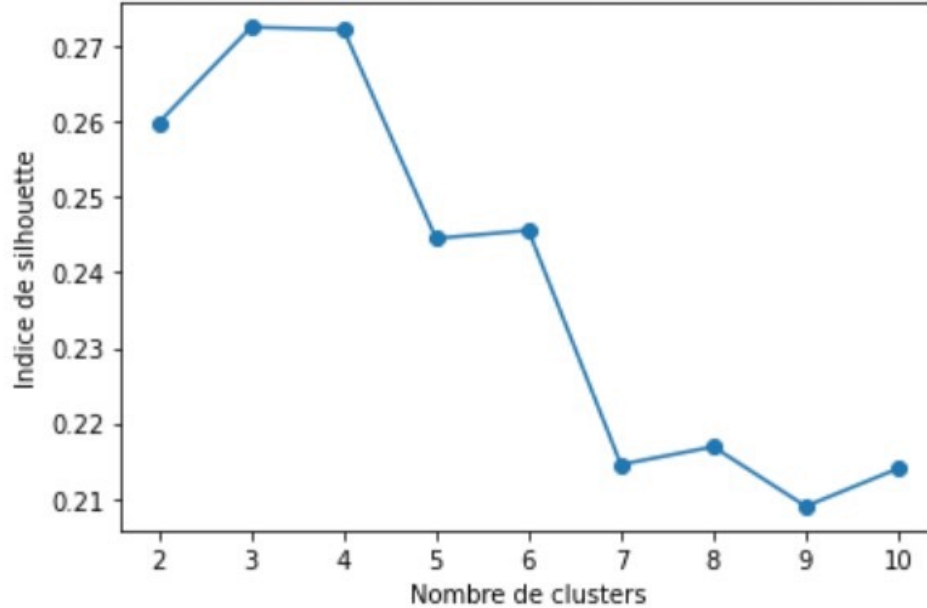
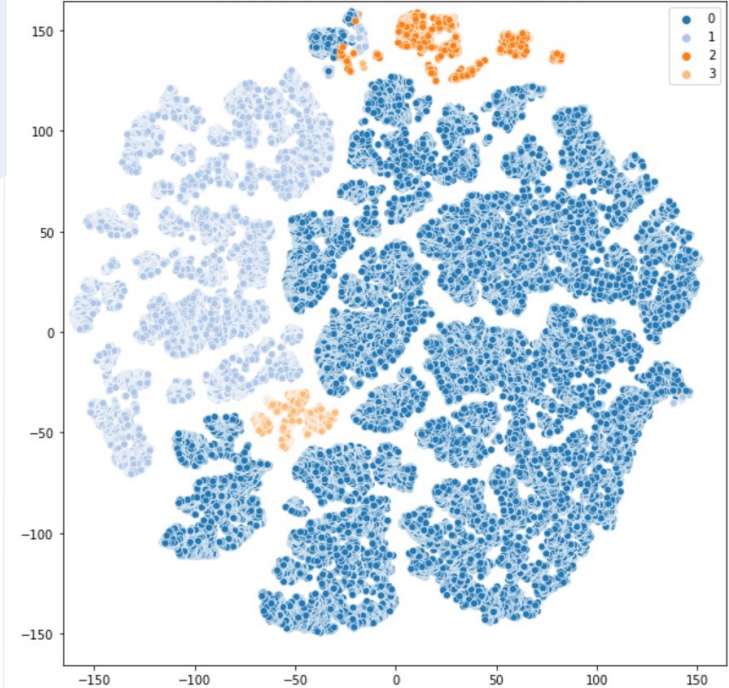




Diagramme de points des données issue d'un t-SNE



Avantages / Inconvénients

		
RFM	<ul style="list-style-type: none">• Très simple à déployer• Très simple à interpréter• Segmente de manière nette	<ul style="list-style-type: none">• Une segmentation trop pauvre pour être actionnable
Top 10 familles produit	<ul style="list-style-type: none">• Segmente fortement sur les familles de produits	<ul style="list-style-type: none">• Réduction de dimensions peu efficace• Perte des indicateurs hors produit• Nombre trop élevé de clusters• Des zones floues• Même résultat possible sans machine learning
Nombre de familles produit	<ul style="list-style-type: none">• Une segmentation efficace sur plusieurs type d'indicateurs• Nombre de clusters satisfaisant• Une segmentation relativement nette	<ul style="list-style-type: none">• Dimension familles de produits absente• Deux macros clusters représentent la majorité des clients

MODÈLE FINAL

Segmentation et maintenance

04

Segments



Éléphants

- Peu de commandes
- Peu de produits
- Produits chers
- Panier moyen élevé
- Carte de crédit
- Nombre élevé de mensualités
- Peu de commentaires



Tortues

- Peu de commandes
- Panier moyen peu élevé
- Produits peu chers
- Boletto ou voucher
- Peu de mensualités
- Zones peu denses en clients
- Délais de livraison longs
- Peu de commentaires
- Notes élevées



Ratons laveurs

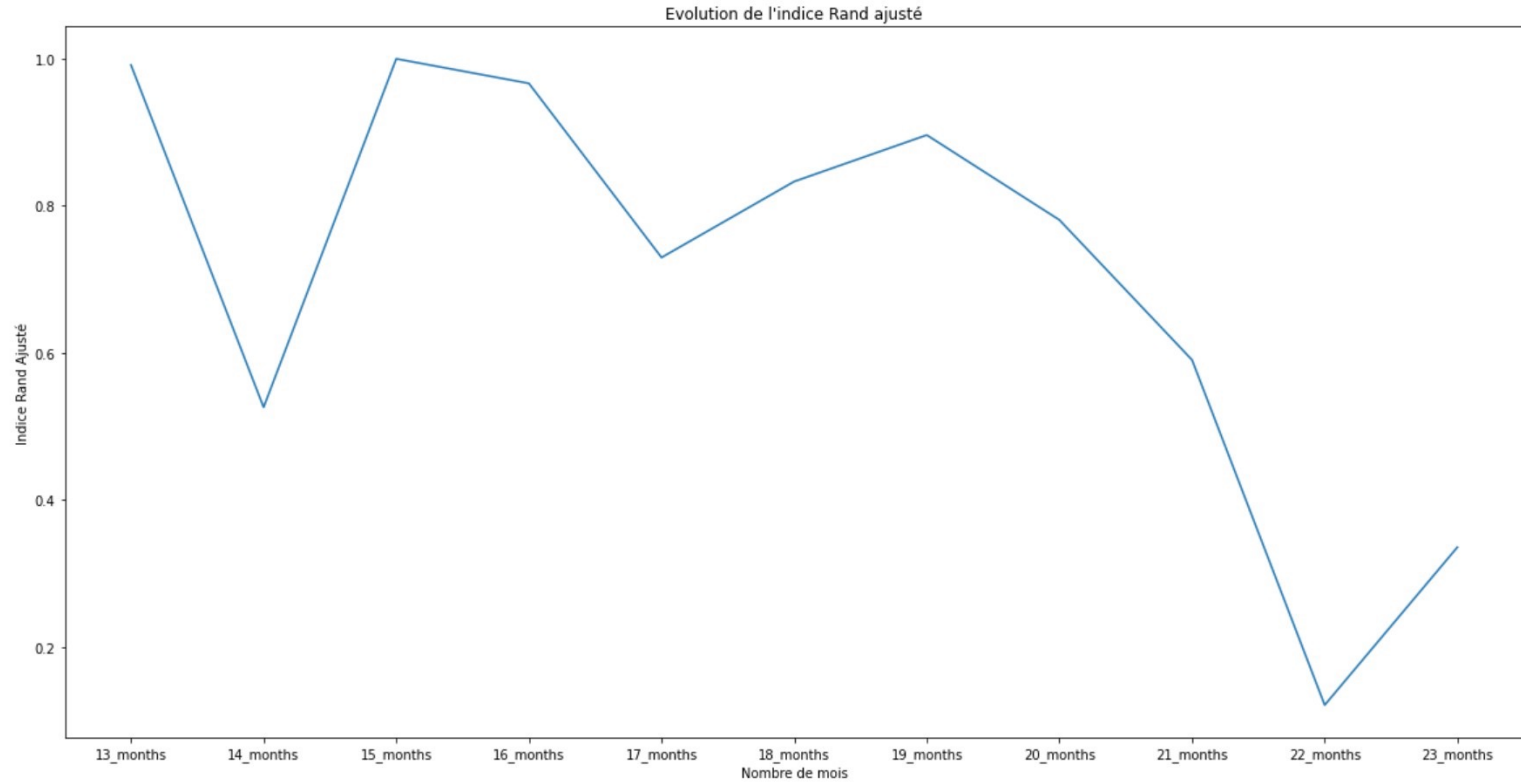
- Beaucoup de commandes
- Nombreuses familles de produits
- Paniers moyens élevés
- Produits peu chers
- Carte de crédit ou voucher
- Nombre élevés de mensualités
- Achat récent
- Zones denses en clients
- Délais de livraison long
- Beaucoup de reviews
- Notes basses



Écureuils

- Peu de produits
- Panier moyen peu élevé
- Carte de débit
- Achat récent
- Longs délais de livraison
- Achats locaux
- Zones denses en clients
- Notes élevées

Maintenance





05

CONCLUSIONS



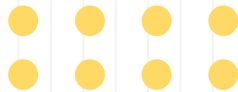
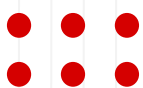
- Comportements utilisateurs
- Pertinence du machine learning



- Quatre cluster identifiés
- Une action possible: récompenser les rats laveurs
- Évolution avec une dimension produit identifiée



- Pas de dimension produit
- Deux très gros segments



Merci!

contact

cedricsoares@me.com

06 09 25 47 45

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution

