
Segmentez des clients d'un site e-commerce

— Anissa TALEB NOV 23 —

Mission : 2 objectifs

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.

recommandation de fréquence à laquelle la segmentation doit être mise à jour pour rester pertinente,

Mission : Segmentation RFM



Récence

La proximité du dernier achat

Ex : durée depuis le dernier achat



Fréquence

Récurrence des achats sur une période

Ex : nombre d'achats sur la dernière année



Montant

Valeur client sur une période

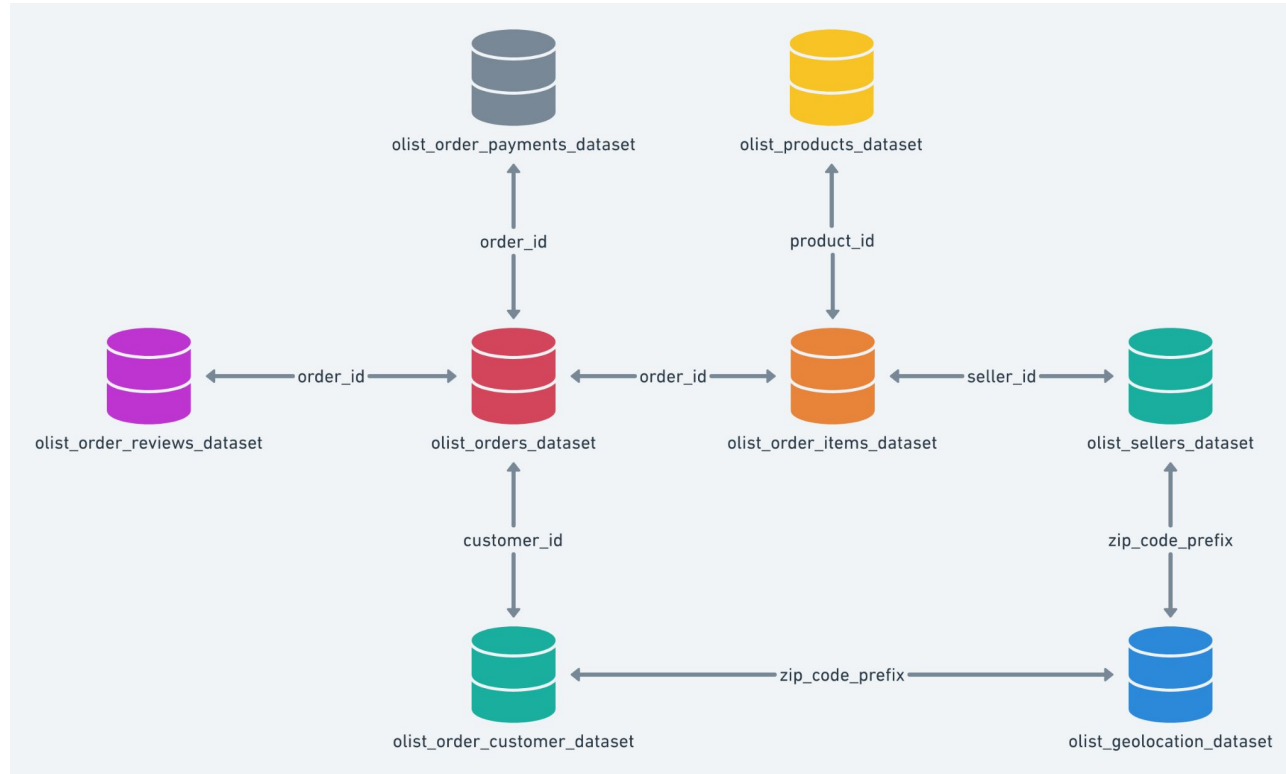
Ex : Somme de tous les montants d'achat sur la dernière année

Démarche :

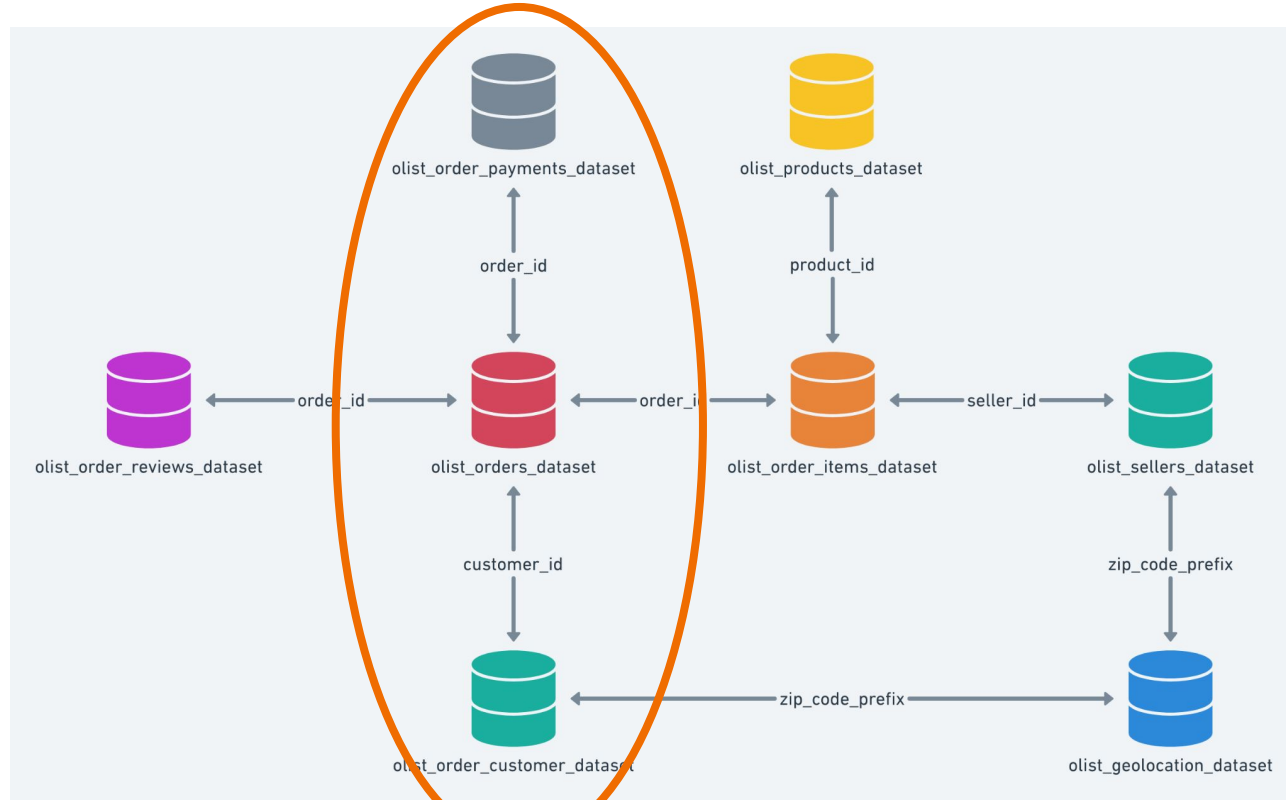
- ⇒ Préparer les données : assemblage des tables
- ⇒ essais des différentes approches de modélisation des méthodes non supervisées.
- ⇒ proposer un contrat de maintenance en basant sur le score ARI

EXPLORATION DES DONNÉES

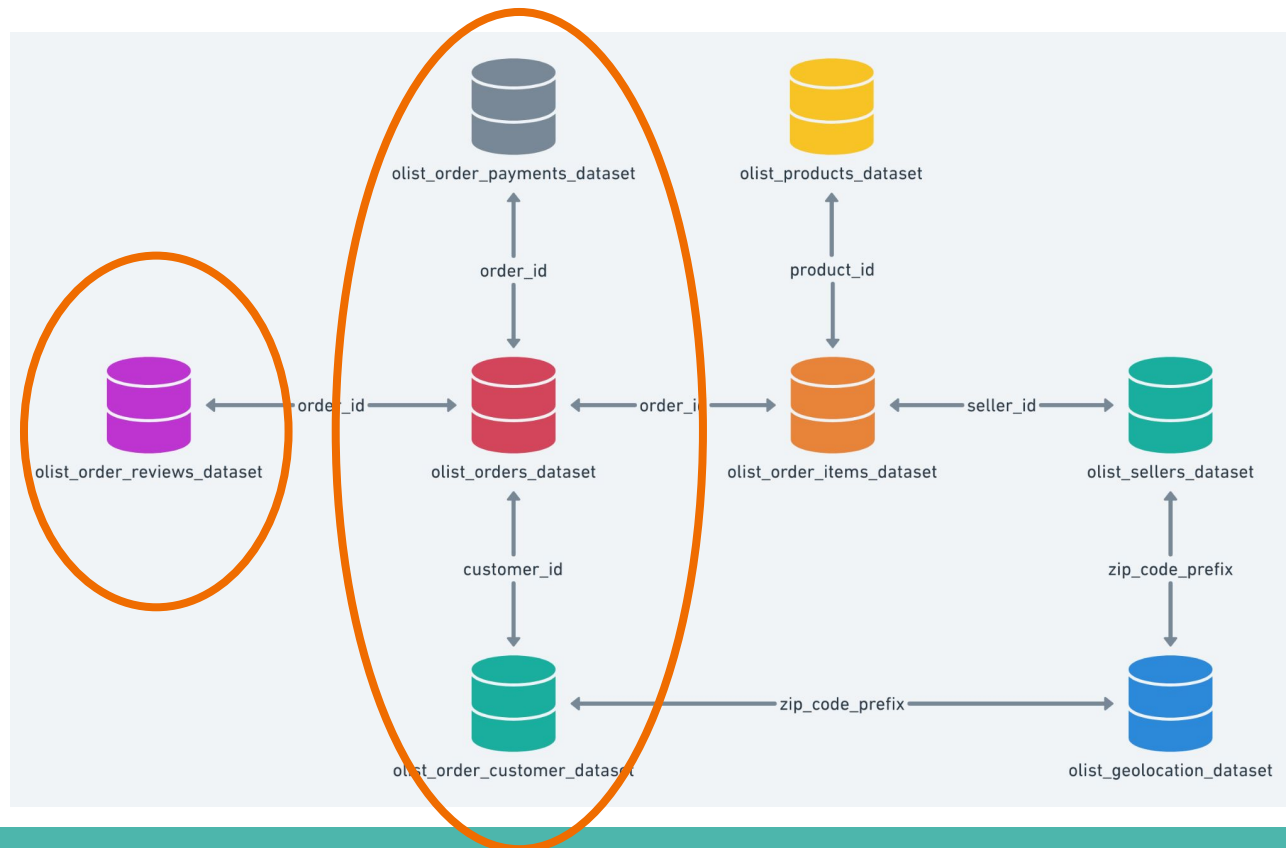
Base de donnée :



Base de donnée : RFM



Base de donnée : RFM + Review score



Techniquement:

Vérifier l'unicité de clé utilisé avant de fusionner :

Fonction Unicity(colonneDataFrame)

Techniquement:

Olist_Orders_dataSet
'order_id', 'customer_id', 'order_status', 'order_purchase_timestamp', 'order_approved_at', ' 'order_delivered_carrier_date', 'order_delivered_customer_date', 'order_estimated_delivery_date'



Orders_dataSet
'order_id', 'customer_id', 'order_purchase_timestamp'

Techniquement:

Olistorder_payments_dataSet

'order_id',
'payment_sequential',
'payment_type',
'payment_installments',
'payment_value'



payments_dataSet

'order_id',
'payment_installments',
'payment_value'

Techniquement:

Olist_customers_dataSet

'customer_id',
'customer_unique_id',
'customer_zip_code_prefix'
'customer_city',
'customer_state'



customers_dataSet

'customer_id',
'customer_unique_id',

Métier:



Récence

La proximité du dernier achat

Ex : durée depuis le dernier achat



Fréquence

Récurrence des achats sur une période

Ex : nombre d'achats sur la dernière année



Montant

Valeur client sur une période

Ex : Somme de tous les montants d'achat sur la dernière année

Métier:



Récence

La proximité du
dernier achat

Ex : durée depuis le
dernier achat

orders_dataSet

```
'order_id',  
'customer_id',  
'order_purchase_timestamp'
```

customers_dataSet

```
'customer_id',  
'customer_unique_id',
```

⇒ Jointure et GroupBy = Colonne durée depuis dernier achat

Métier:



Fréquence

Récurrence des
achats sur une
période

Ex : nombre d'achats
sur la dernière année

orders_dataSet

```
'order_id',  
'customer_id',  
'order_purchase_timestamp'
```

customers_dataSet

```
'customer_id',  
'customer_unique_id',
```

⇒ Jointure et GroupBy = Colonne Nombre de
commande par client

Remarque: seuls 3% des clients ont réalisé plus qu'une commande.

Métier:



Montant

Valeur client sur une période

Ex : Somme de tous les montants d'achat sur la dernière année

orders_dataSet

```
'order_id',  
'customer_id',  
'order_purchase_timestamp'
```

customers_dataSet

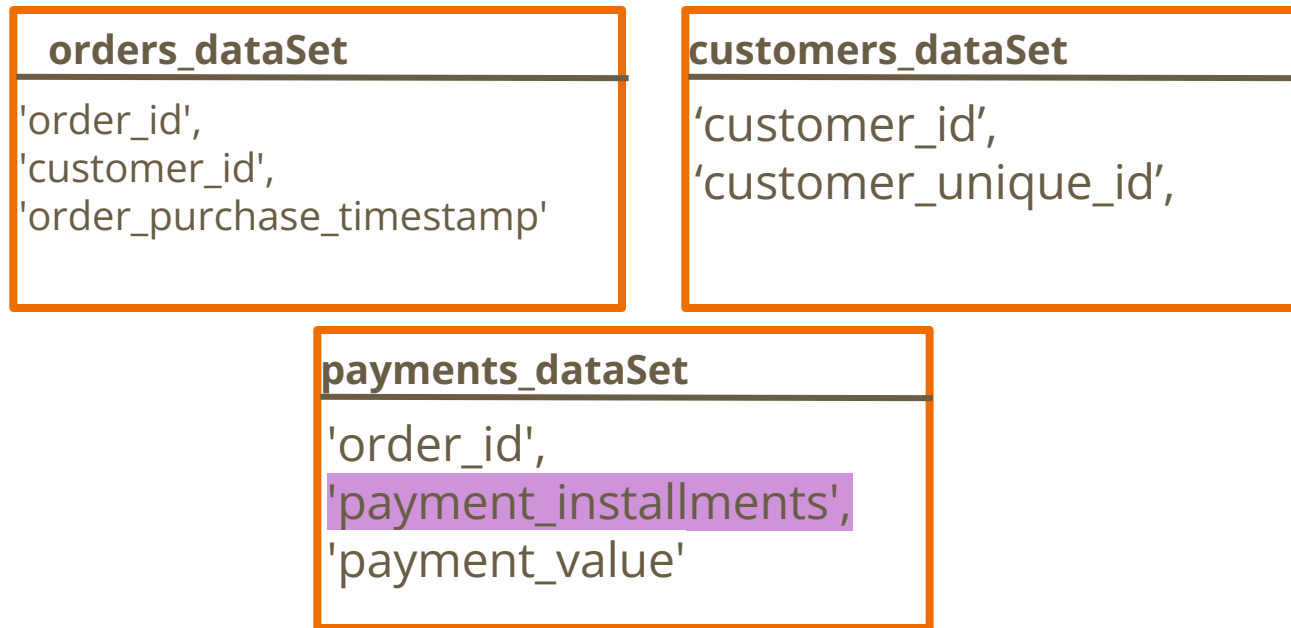
```
'customer_id',  
'customer_unique_id',
```

payments_dataSet

```
'order_id',  
'payment_installments',  
'payment_value'
```

⇒ 2ieme jointure et GroupBy = Colonne Montant

Métier: modification possible - versements échelonnés



⇒ 2ieme jointure et GroupBy = Colonne versements échelonnés

Métier: modification possible - satisfaction clients

Olist_reviews_dataSet
'review_id', 'order_id', 'review_score', 'review_comment_title', 'review_comment_message', 'review_creation_date', 'review_answer_timestamp'

orders_dataSet
'order_id', 'customer_id', 'order_purchase_timestamp'

customers_dataSet
'customer_id', 'customer_unique_id',

payments_dataSet
'order_id', 'payment_installments', 'payment_value'

⇒ 3ème jointure et GroupBy = Colonne note clients

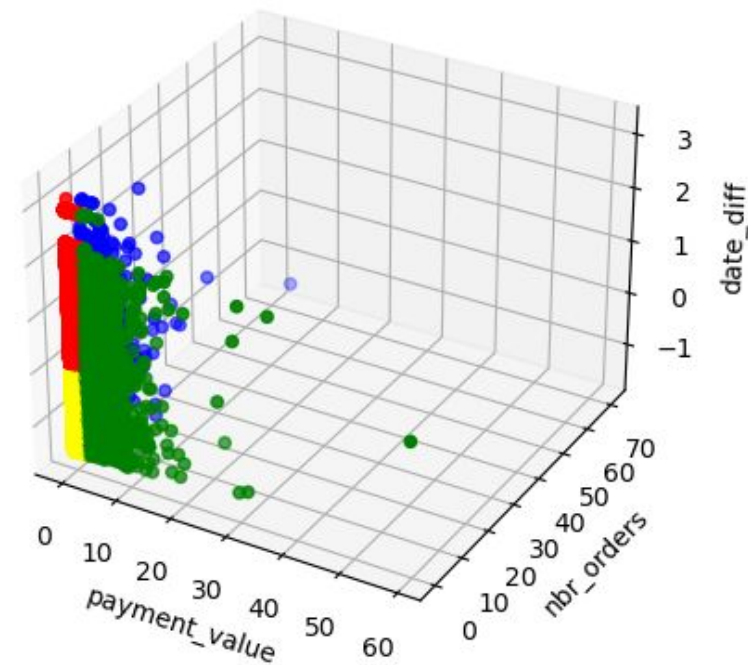
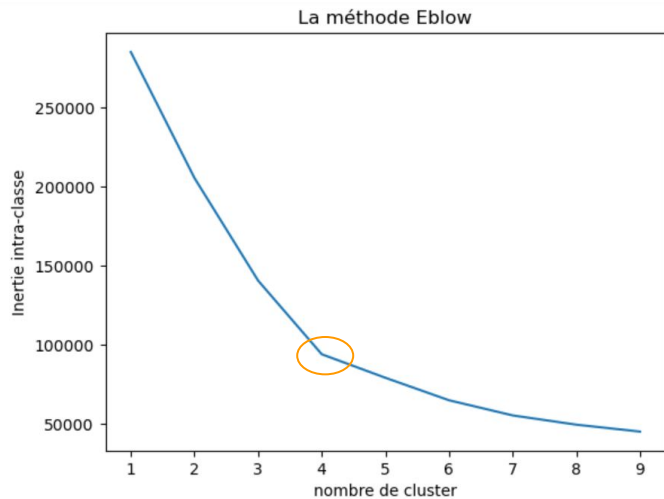
MODÉLISATION: le clustering

K-Means

Agglomerative Hierarchical Clustering

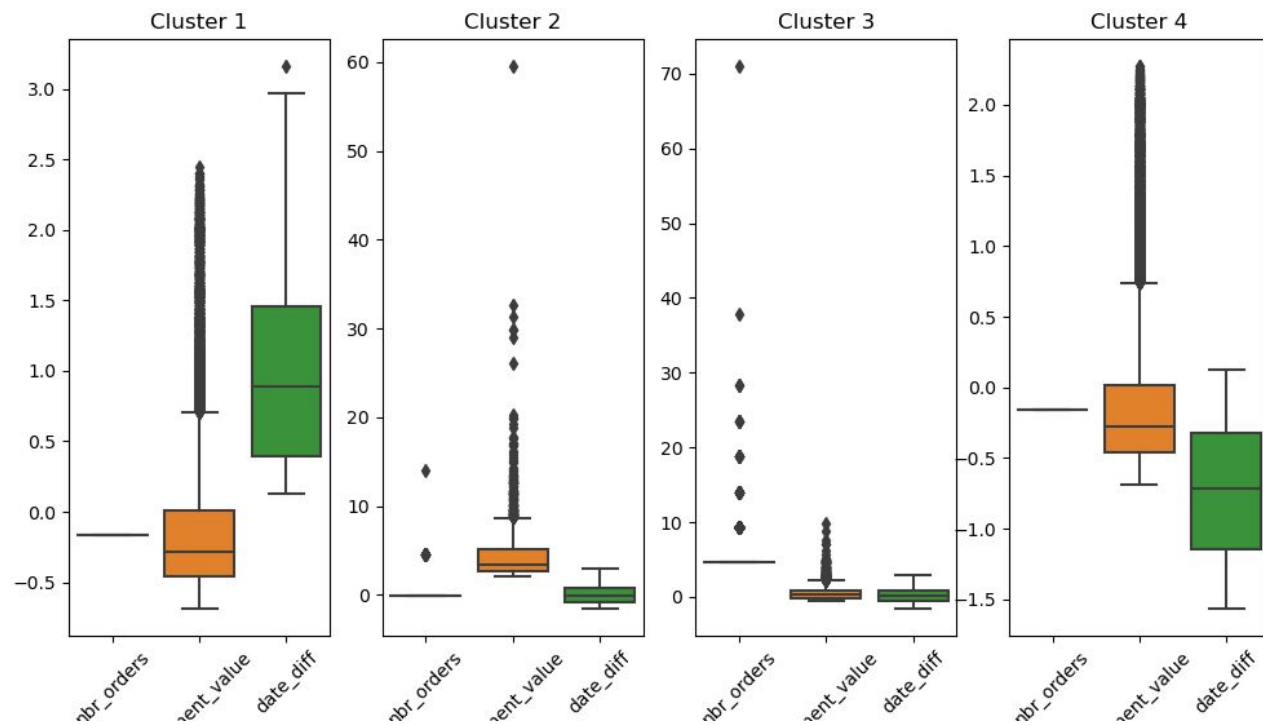
DBSCAN

1 K-means : RFM basique



⇒ Nombre de clusters : 4

1 K-means : RFM basique



1 K-means : RFM basique

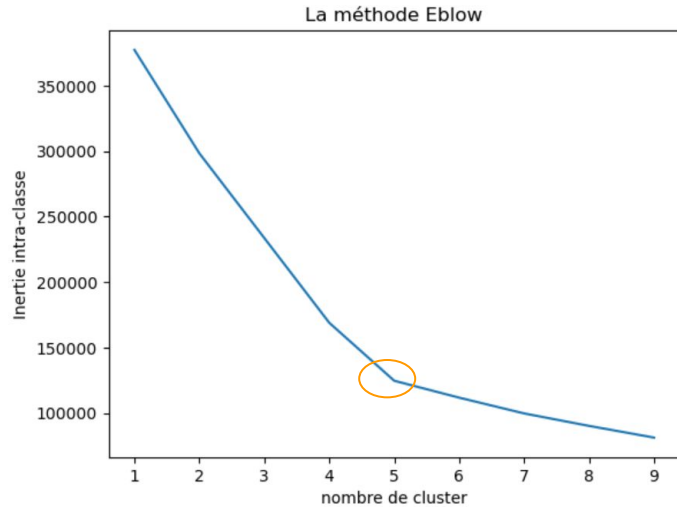
Cluster 1 : Clients qui font des achats en petits montants mais régulièrement

Cluster 2 : Clients qui font des petits achats très rarement (+ 1 an d'écart)

Cluster 3 : Clients qui font des achats en gros montants mais très rarement (+ 1 an d'écart)

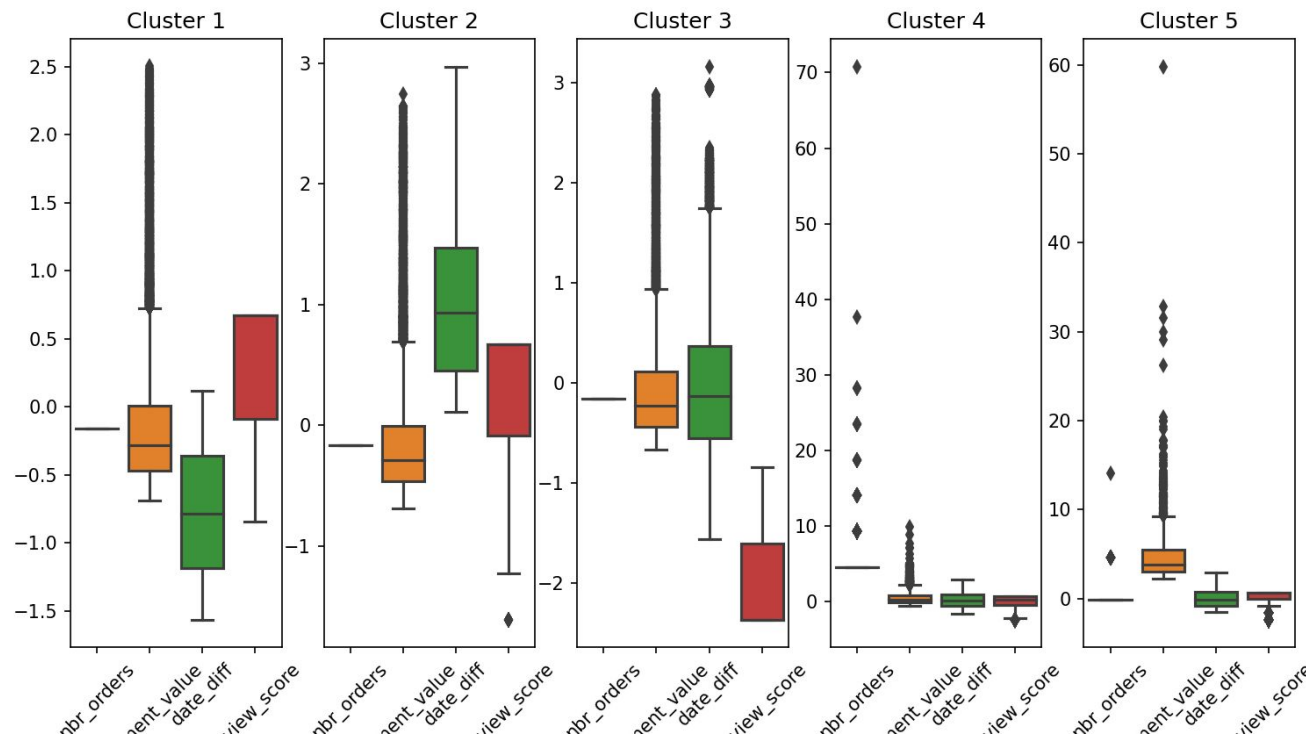
Cluster 4 : Clients qui font des gros achats régulièrement = **VIP**

1 K-means : RFM + Review score



⇒ Nombre de clusters : 5

1 K-means : RFM + Review score



1 K-means : RFM + Review score

cluster 1 : Clients avec des montants d'achats faibles mais très régulièrement et avec une de Review score moyenne.

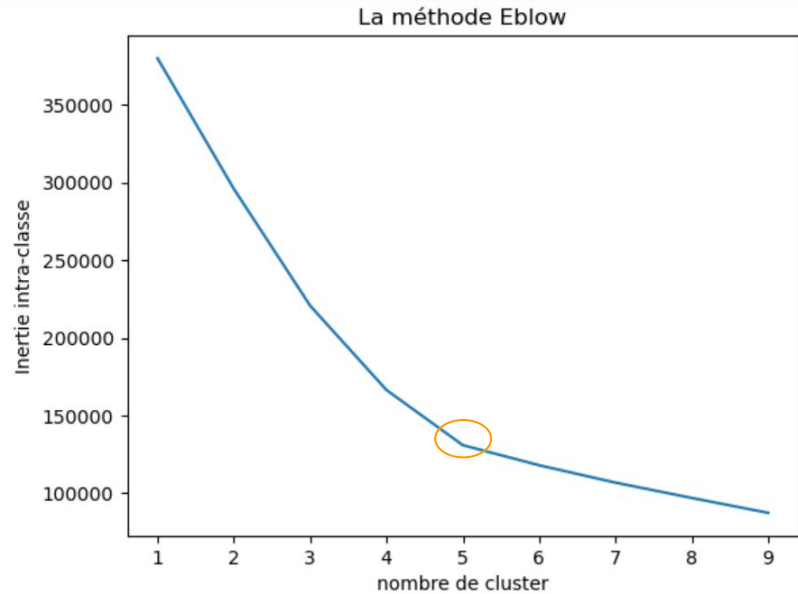
cluster 2 : Clients avec des montants d'achats très faibles et très rarement et avec une de Review score extrêmement bas.

cluster 3 : Clients avec des montants d'achats importants et régulièrement et avec une de Review score élevée. **=VIP**

cluster 4 : Clients avec des valeurs de paiement très élevées et très régulièrement et avec une de Review score moyenne.

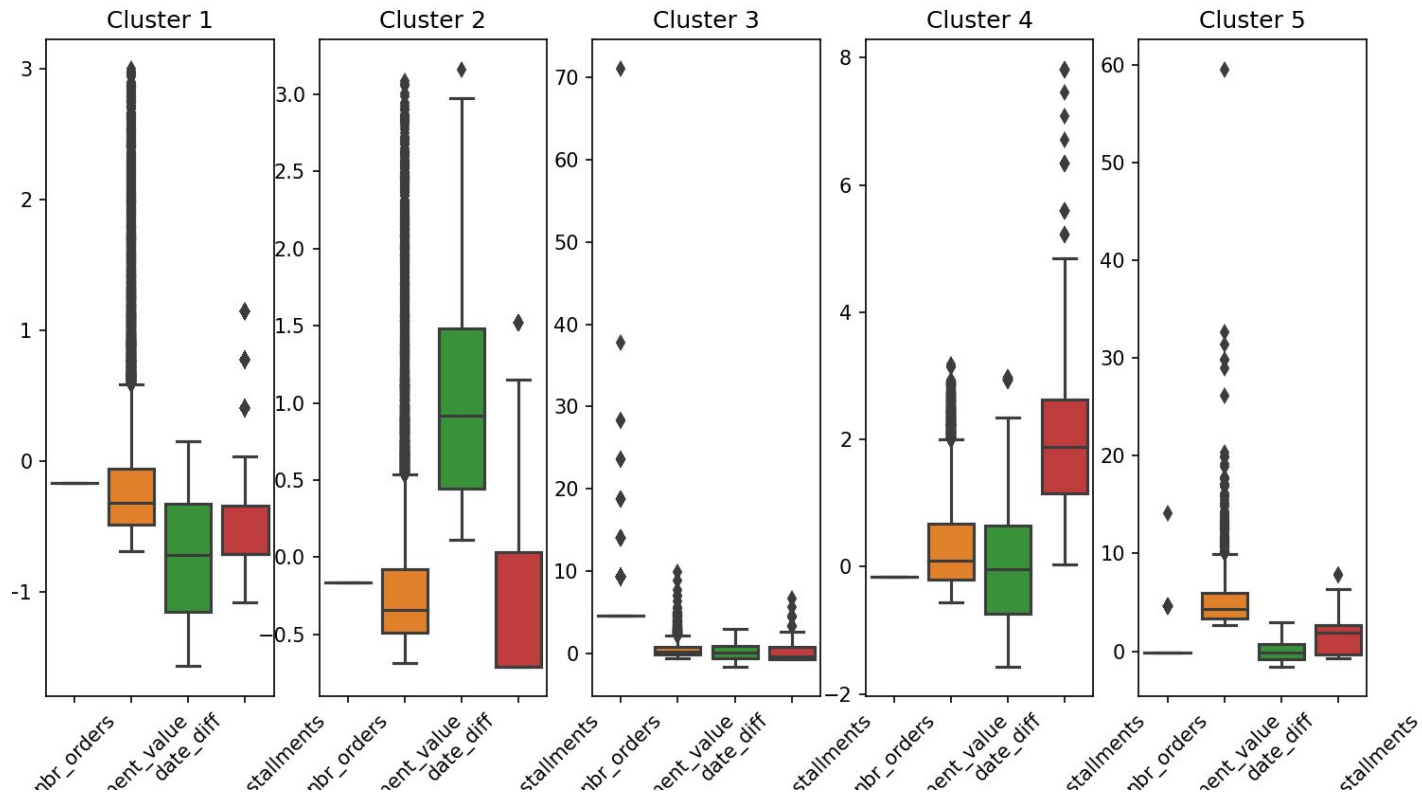
cluster 5 : Clients avec des montants d'achats moyennes mais régulièrement et avec une de Review score moyenne.

1 K-means : RFM + Paiement échelonné



⇒ Nombre de clusters : 5

1 K-means : RFM + Paiement échelonné



1 K-means : RFM + Paiement échelonné

cluster 1 : Clients avec des montants d'achats faibles mais très régulièrement et avec un recours aux paiements échelonnés faible.

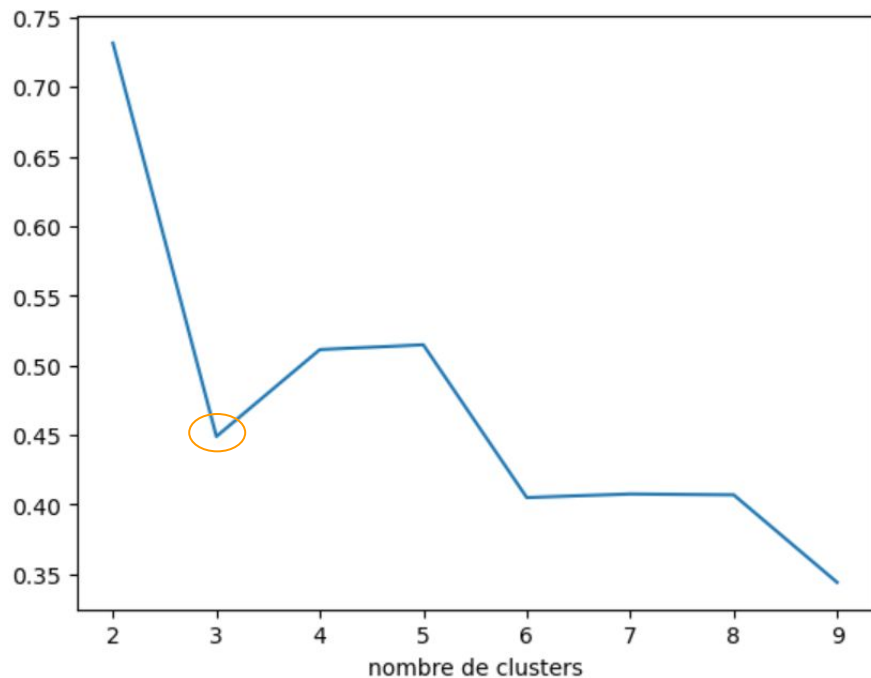
cluster 2 : Clients avec des valeurs de paiement très élevées et très régulièrement et un recours aux paiements échelonnés moyen = **VIP**

cluster 3 : Clients avec des montants d'achats moyens mais régulièrement et avec un recours aux paiements échelonnés moyen.

cluster 4 : Clients avec des montants d'achats très faibles et très rarement et avec un recours aux paiements échelonnés faible.

cluster 5 : Clients avec des montants d'achats importants et régulièrement et paiements échelonnés très réguliers = **VIP**

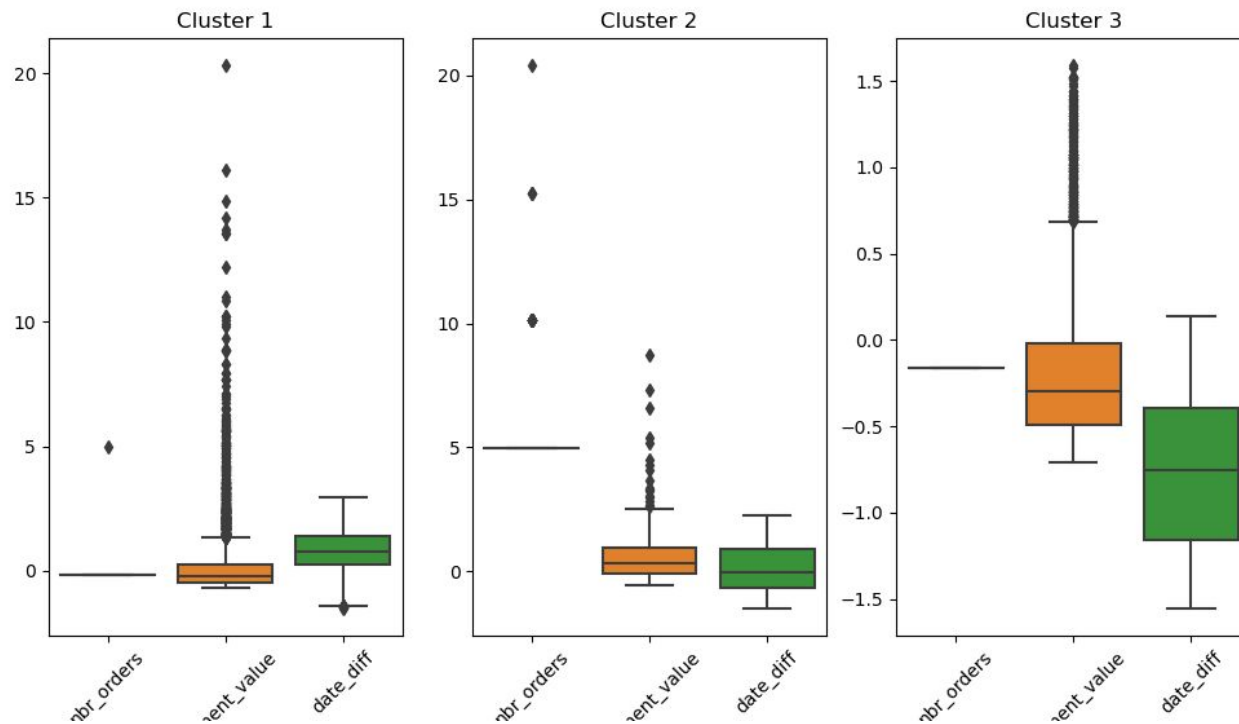
2 Agglomerative Hierarchical Clustering : RFM basique



⇒ analyse sur 10000 individus !

⇒ Nombre de clusters : 3

2 Agglomerative Hierarchical Clustering : RFM basique



2 Agglomerative Hierarchical Clustering : RFM basique

cluster 1 = clients avec achats de montants moyens mais très régulièrement

cluster 2 = clients ayant fait plusieurs commandes avec achats de montants importants et souvent. = **VIP**

cluster 3 = clients avec peu d'achat et rarement.

⇒ résultats assez semblable à ceux obtenu avec le Kmeans.

3 DBSCAN : RFM basique

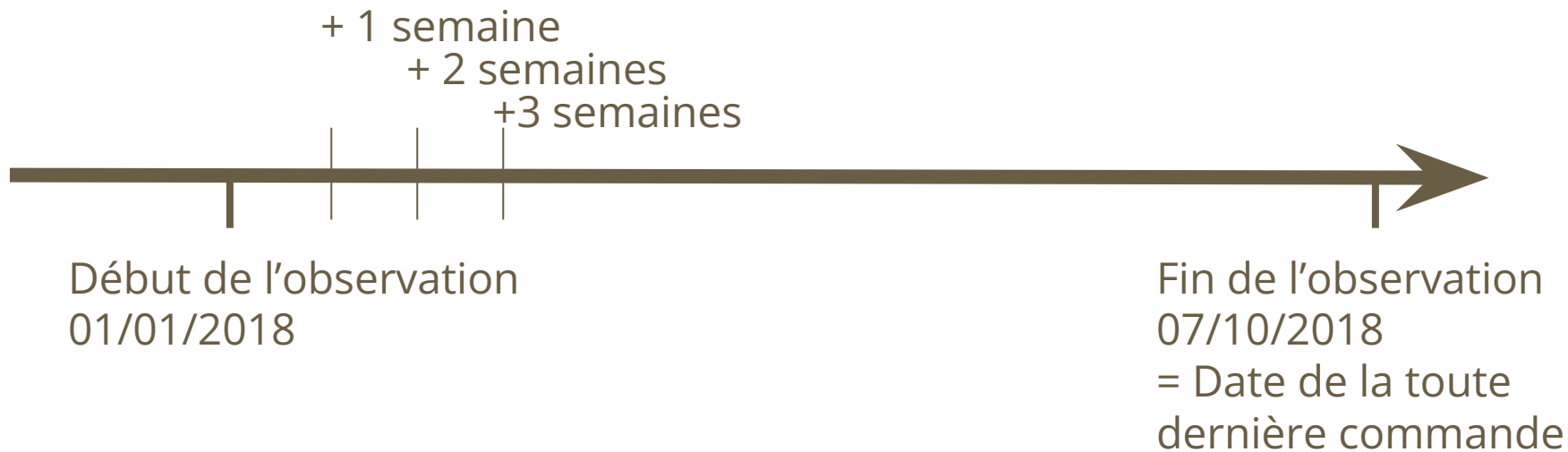
⇒ l'analyse nous donne qu'un seul cluster !

⇒ modèle pas adapté vu que uniquement 3% de données sont utiles pour le groupement

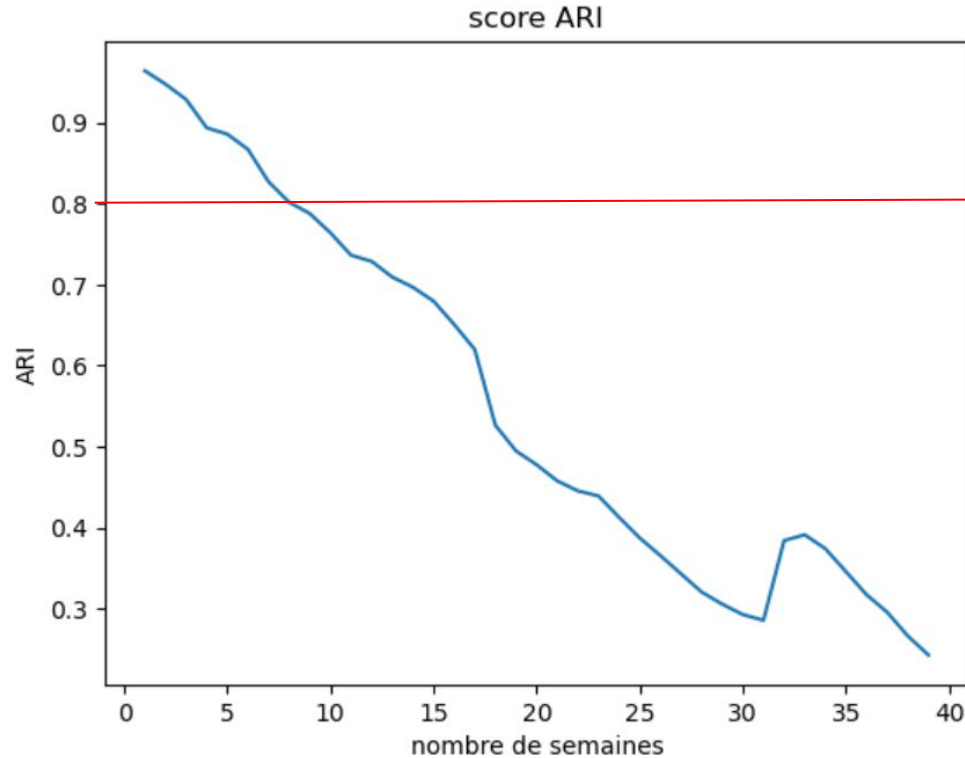
Contrat de maintenance



Contrat de maintenance



Contrat de maintenance



⇒ Il est recommandé de réaliser une nouvelle analyse toutes les 10 semaines environ.

