



## EXPOSÉ PROJET 5

---

# SEGMENTER DES CLIENTS D'UN SITE E-COMMERCE

---

Le 16 Décembre 2020

Zeineb Guizani

# Plan de la présentation

1. Présentation de l'appel à projet
2. Description et transformation du jeu de données
3. Pistes de modélisation effectuées
4. Modèle final sélectionné
5. Conclusion

# I - Présentation de l'appel à projet

- La solution de vente en ligne « Olist » souhaite segmenter sa base clients selon l'intention d'achat afin de les cibler efficacement.
- Partitionner les données en clusters, de manière non supervisée, grâce aux comportements et aux données personnelles.
- Algorithmes de clustering non supervisés.
  - **fournir à l'équipe marketing une description actionable** de la segmentation.
  - **proposition de contrat de maintenance.**

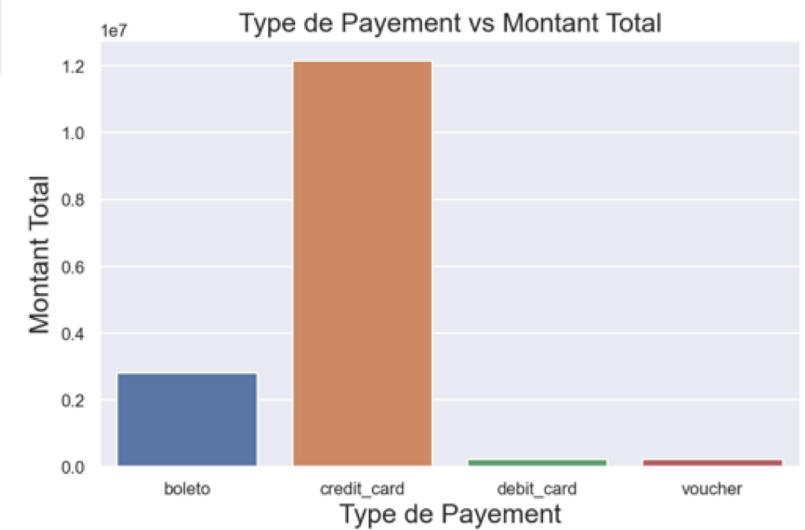
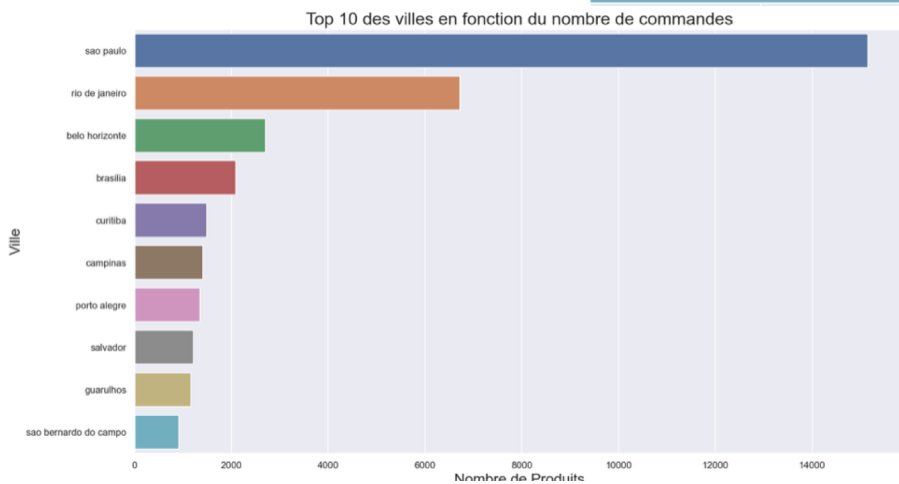
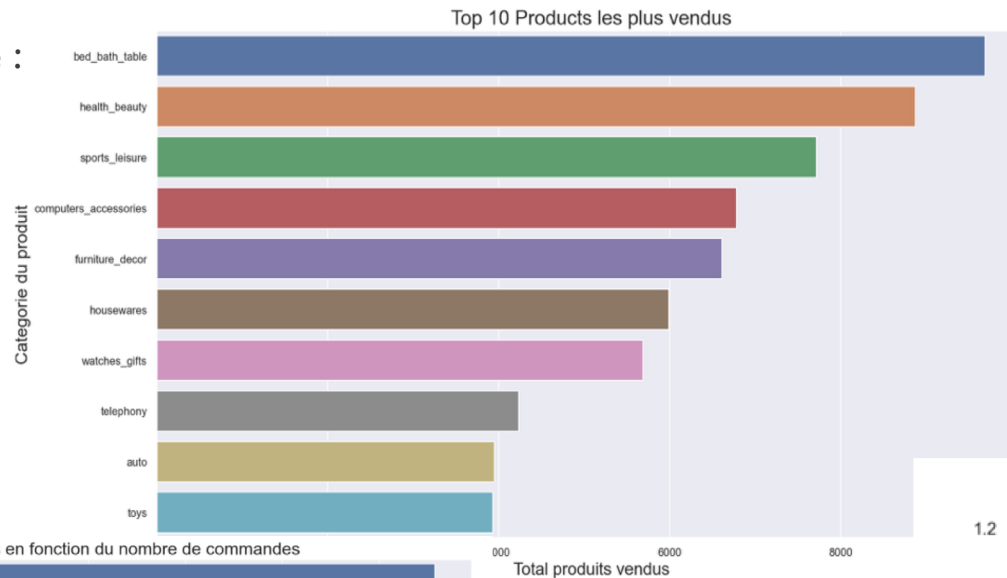
.

## II- Description et transformation du jeu de données

- Le jeu de données "olist" comporte 9 bases de données :
  - ['customers', 'geolocation', 'items', 'payments', 'reviews', 'orders', 'products', 'sellers', 'category'].
- Nettoyage:
  - suppression des lignes à NAN.
  - Suppression des doublons ['order\_id', 'customer\_unique\_id', 'order\_purchase\_timestamp']
  - Modifier le type de données dans les colonnes de date
- Feature engineering : Rajout de nouvelles variables:
  - review\_score
  - nbr\_produits par clients
  - nb\_cities

## II- Description et transformation du jeu de données

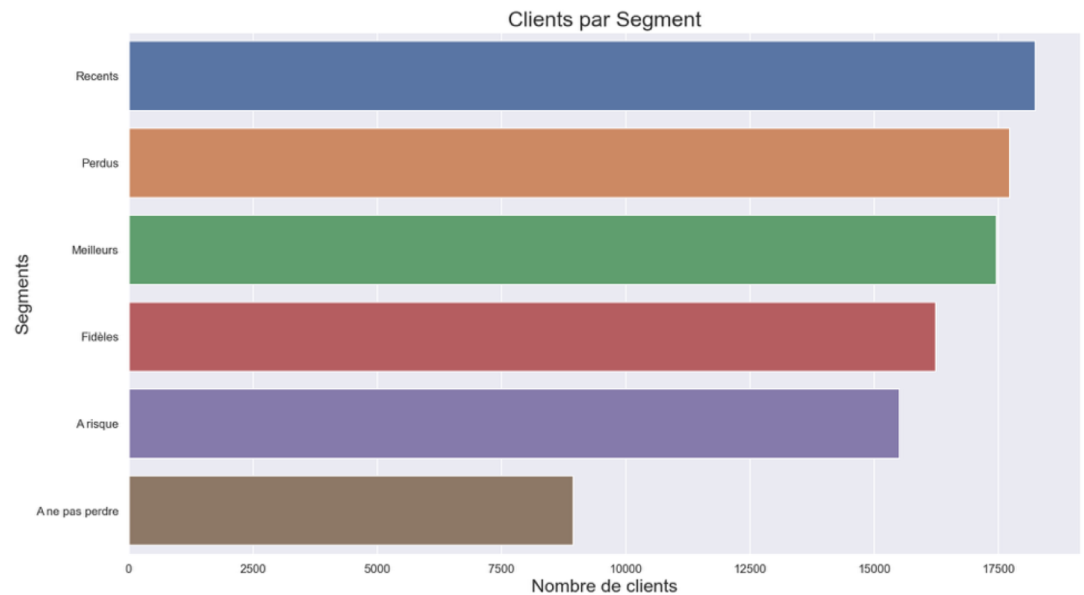
Analyse exploratoire :



## III- Pistes de modélisation effectuées

### ■ I - Création des variables RFM + segmentation manuelle

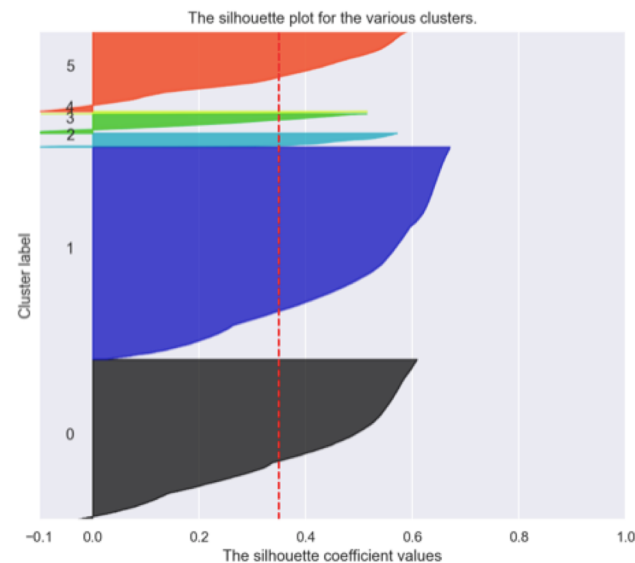
	Segments	Récence	Fréquence + Montant
Meilleurs Clients		3-4	3-4
Clients fidèles		2-4	2-4
Clients récents		3-4	1-2
Clients à risque		1-2	2-3
A ne pas perdre		1	3-4
Clients perdus		1-2	1-2



## III- Pistes de modélisation effectuées

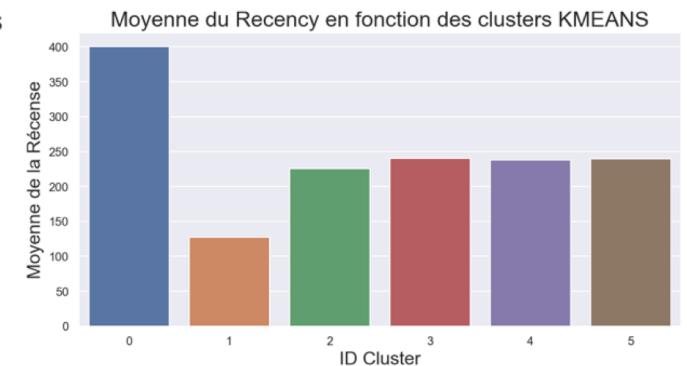
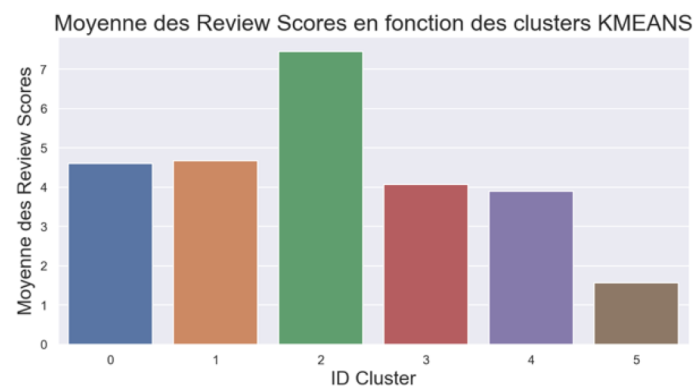
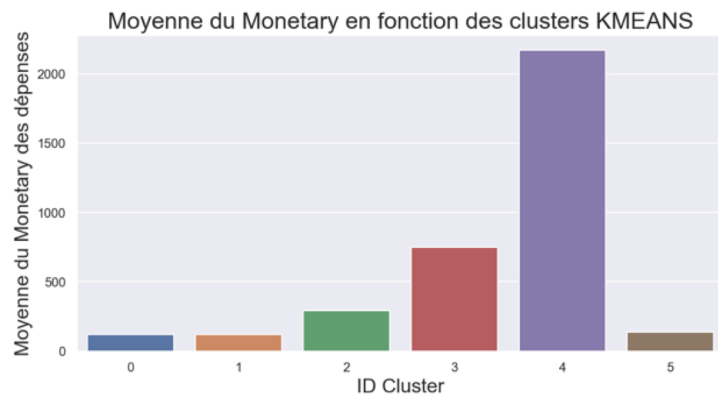
### ■ II - Segmentation avec K-means

- Adapter les hyperparamètres d'un algorithme non supervisé afin de l'améliorer:
- `range_n_clusters = [3, 4, 5, 6, 7, 8]`



# III- Pistes de modélisation effectuées

## ■ II - Segmentation avec K-means



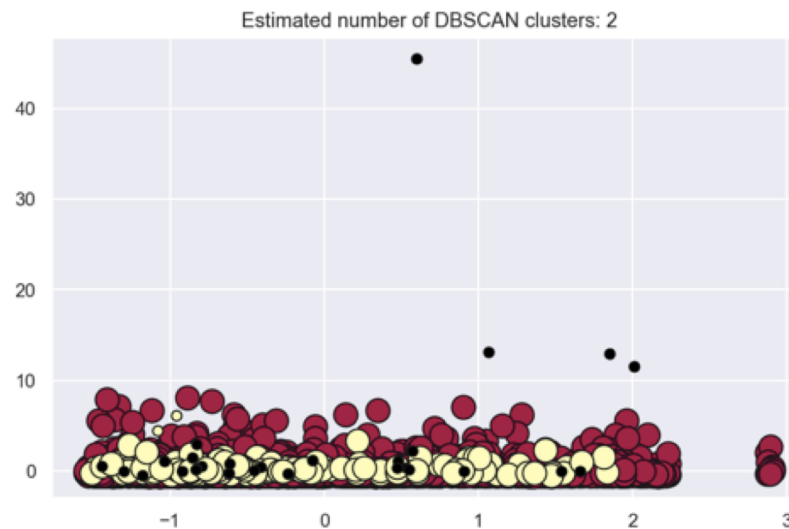
- Cluster 5 : Consommateurs à risque (non satisfaits et n'achètent pas souvent)
- Cluster 4 : Meilleurs consommateurs, il achètent souvent et ils sont actifs
- Cluster 3 : Consommateurs réguliers
- Cluster 2 : Les consommateurs à développer
- Cluster 1 : Consommateurs perdus (la majorité de la base: n'achètent plus et dépensent le minimum)
- Cluster 0 : Nouveaux consommateurs (Ils ont commencé à acheter)



## III- Pistes de modélisation effectuées

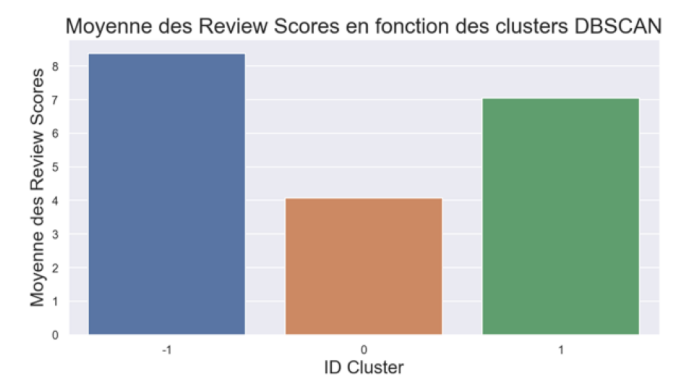
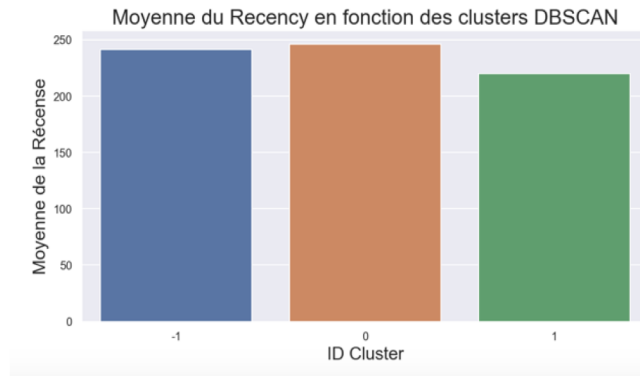
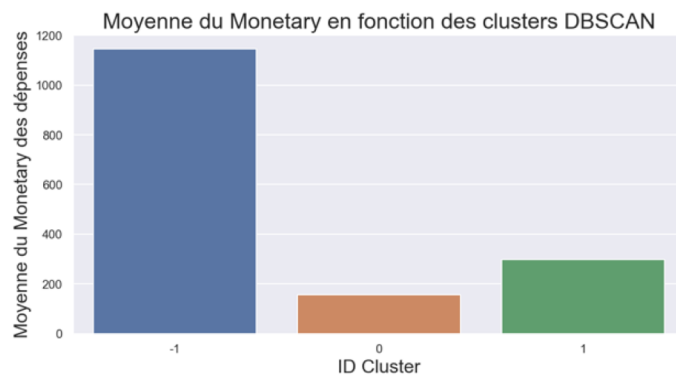
### ■ III - Segmentation avec DBSCAN

- Adapter les hyperparamètres d'un algorithme non supervisé afin de l'améliorer:
- $\epsilon$  = [0.2, 0.25, 0.3, 2, 4, 5]



## III- Pistes de modélisation effectuées

### ■ III - Segmentation avec DBSCAN

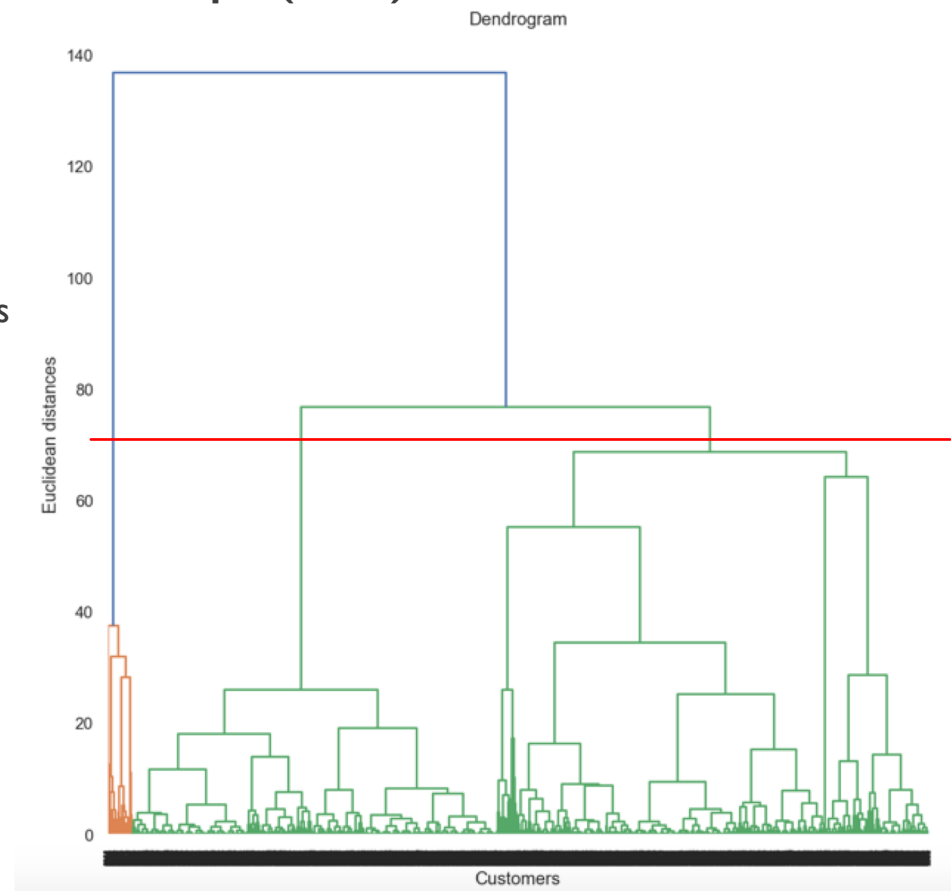


- - Cluster -1: bruit
- - Cluster 0 : nouveaux consommateurs.
- - Cluster 1 : consommateurs standards

## III- Pistes de modélisation effectuées

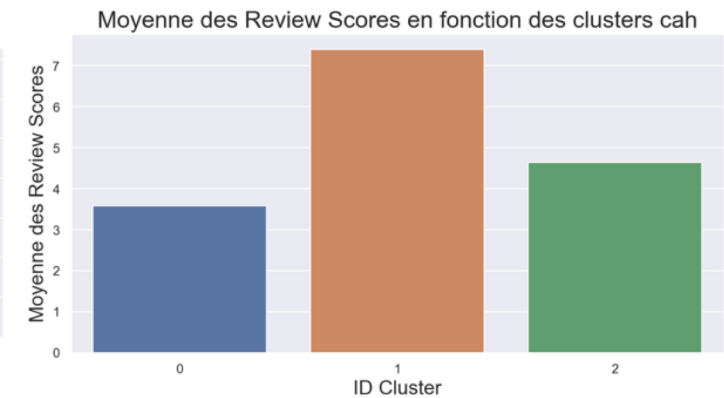
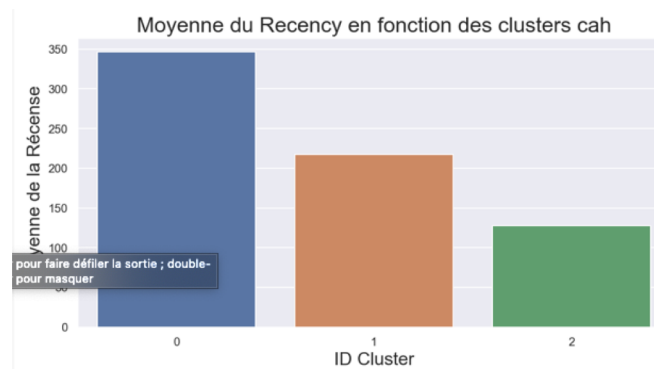
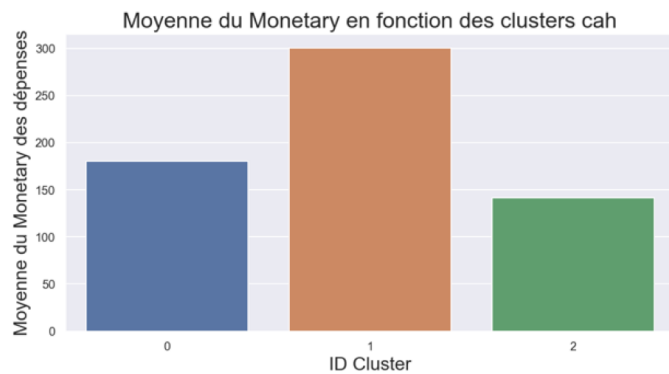
### ■ IV - Segmentation avec Classification ascendante hiérarchique (CAH)

- Le choix du seuil à 70 (presque  $140/2$ ).
- Le dendrogramme « suggère » un découpage en 3 groupes



## III- Pistes de modélisation effectuées

### ■ IV - Segmentation avec Classification ascendante hiérarchique (CAH)



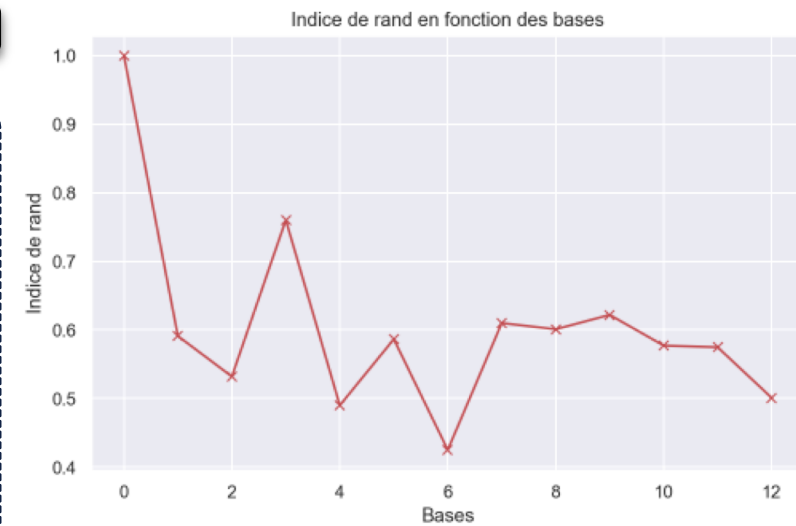
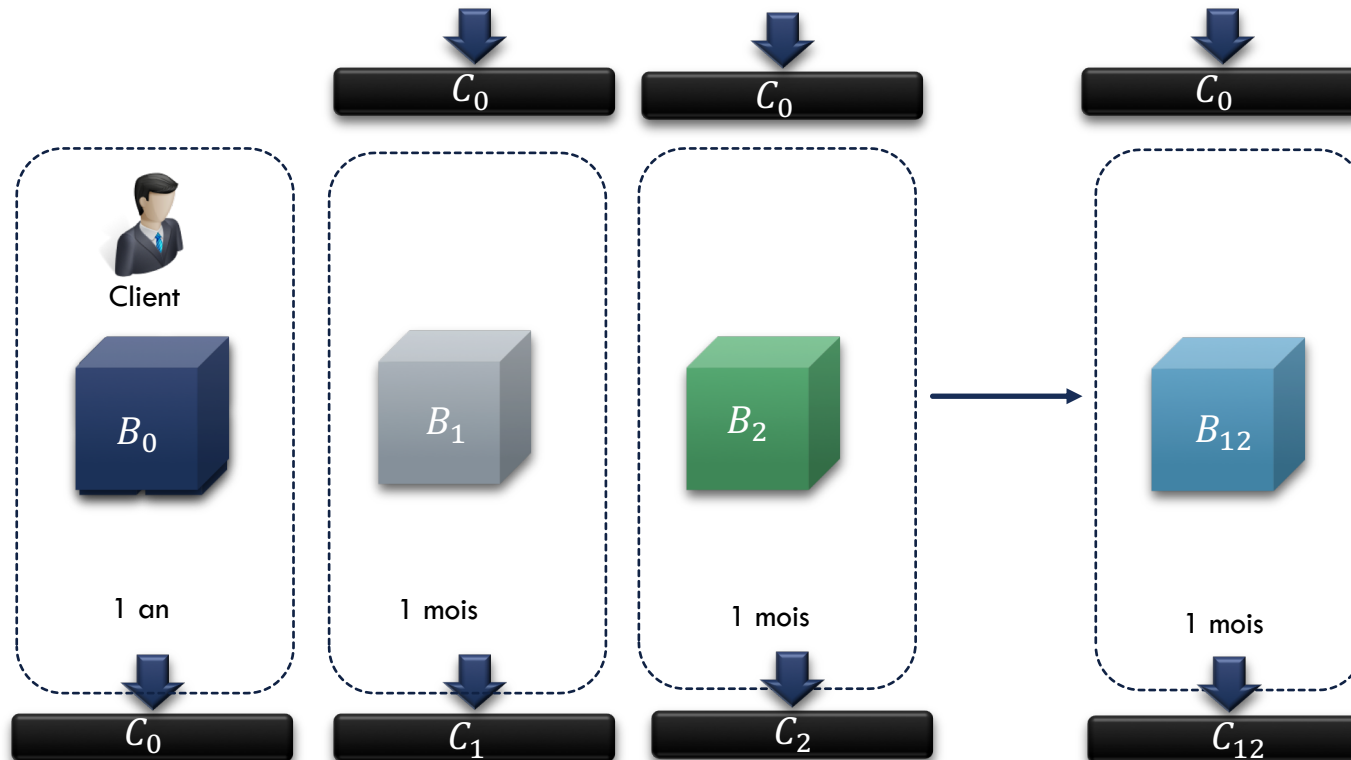
- Cluster 0: Consommateurs réguliers
- Cluster 1: Meilleurs consommateurs
- Cluster 2: Clients à risque (presque perdus)

# V- Modèle final sélectionné

## ❖ Modèle sélectionné:

- Segmentation avec K-means

- Nclusters= 6



# Conclusion

## Conclusion :

- Relever les défis de la segmentation non supervisée des clients à travers :
  - La transformation des variables pertinentes d'un modèle d'apprentissage non supervisé
  - L'adaptation des hyperparamètres d'un algorithme non supervisé afin de l'améliorer
  - L'évaluation des performances d'un modèle d'apprentissage non supervisé
  - L'étude de la stabilité



MERCI DE VOTRE ATTENTION

