

Projet 5

SEGMENTEZ DES CLIENTS D'UN SITE E-COMMERCE

ETUDIANTE	: BOUCHRA MEKHALDI
MENTOR	: SOUHAIL TOUMDI
EVALUATEUR	: YANN MAILLET
DATE	: 29/06/2022

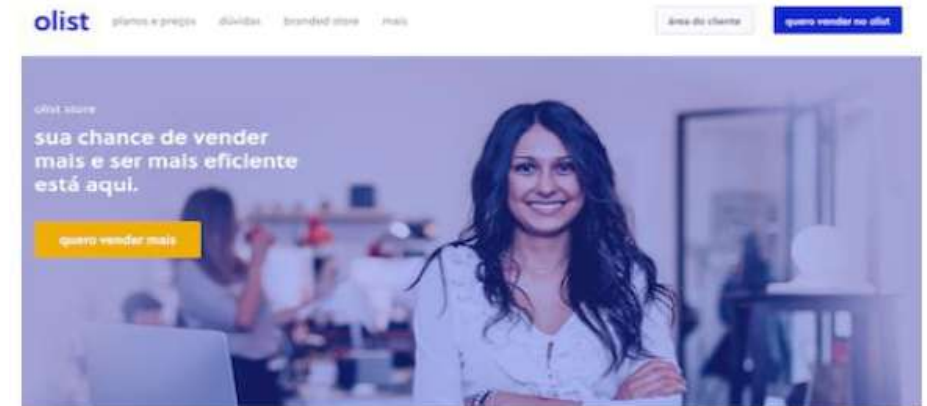
Problématique

Plateforme vitrine de e-commerce au Brésil ,Met en lien acheteurs et vendeurs.

Fournir aux équipes Marketing de l'entreprise une **segmentation des clients** utilisables dans leurs campagnes de communication.

Ce qui est attendu :

- **Comprendre** les différents **types d'utilisateurs**
- **Fournir** une **description actionnable** de la segmentation
- **Proposer** un **contrat de maintenance** basé sur une analyse de la stabilité des segments au cours du temps.



Plan de Présentation



1 _ Préparation du jeu de données

2 _ Piste de modélisation

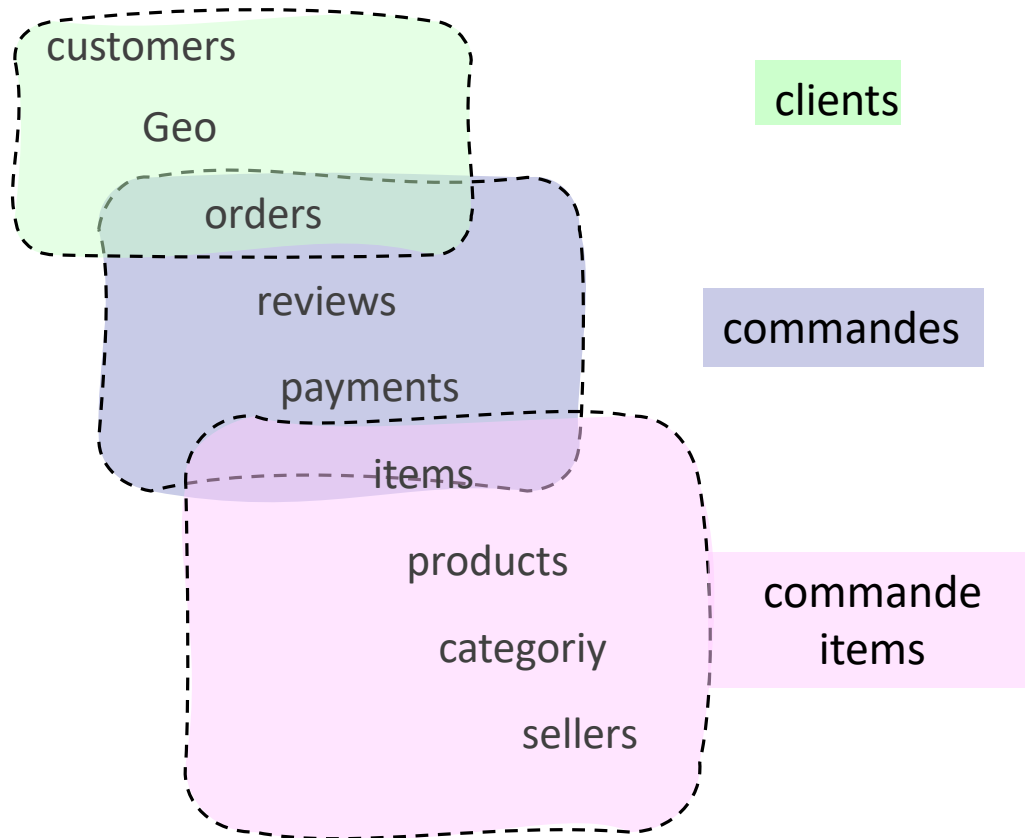
3 _ Contrat de maintenance

4 _ Conclusion

Préparation du jeu de données

La plateforme Olist - tables

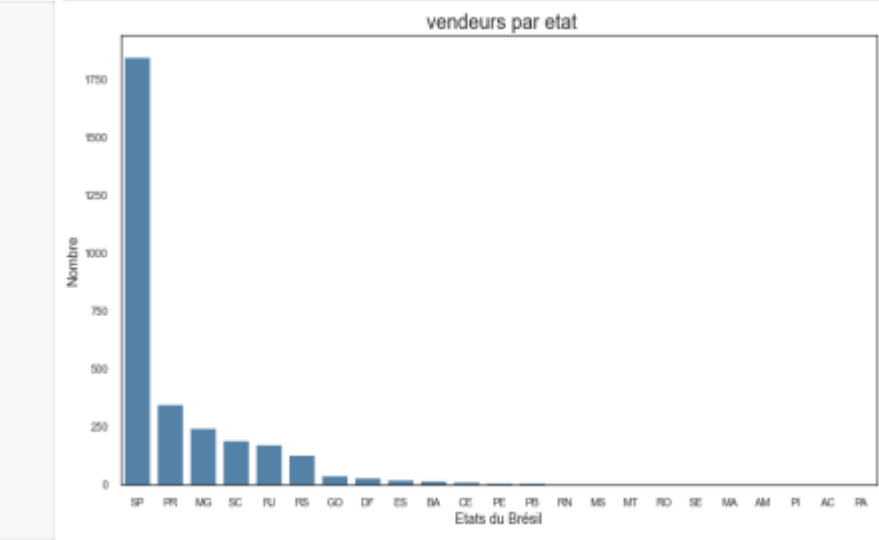
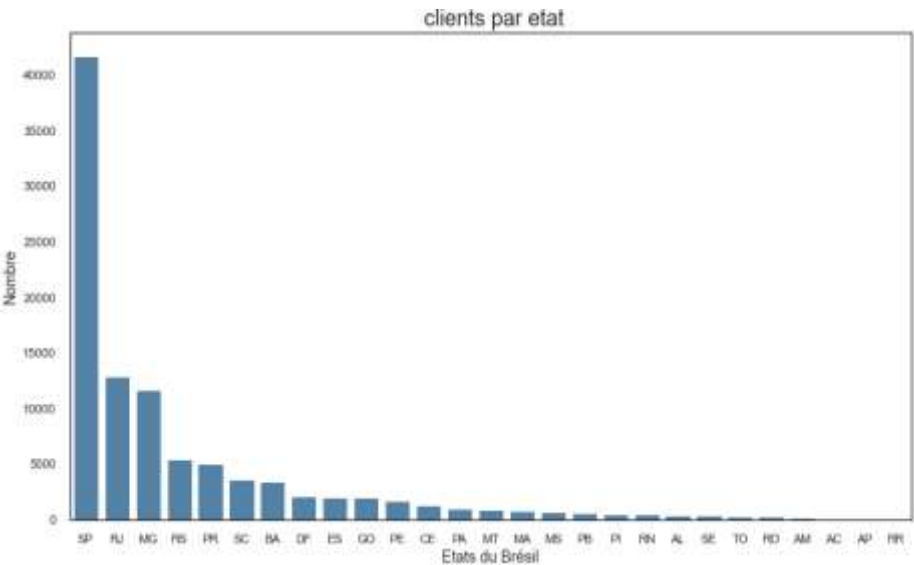
Téléchargeable sur <https://www.kaggle.com/olistbr/brazilian-ecommerce>



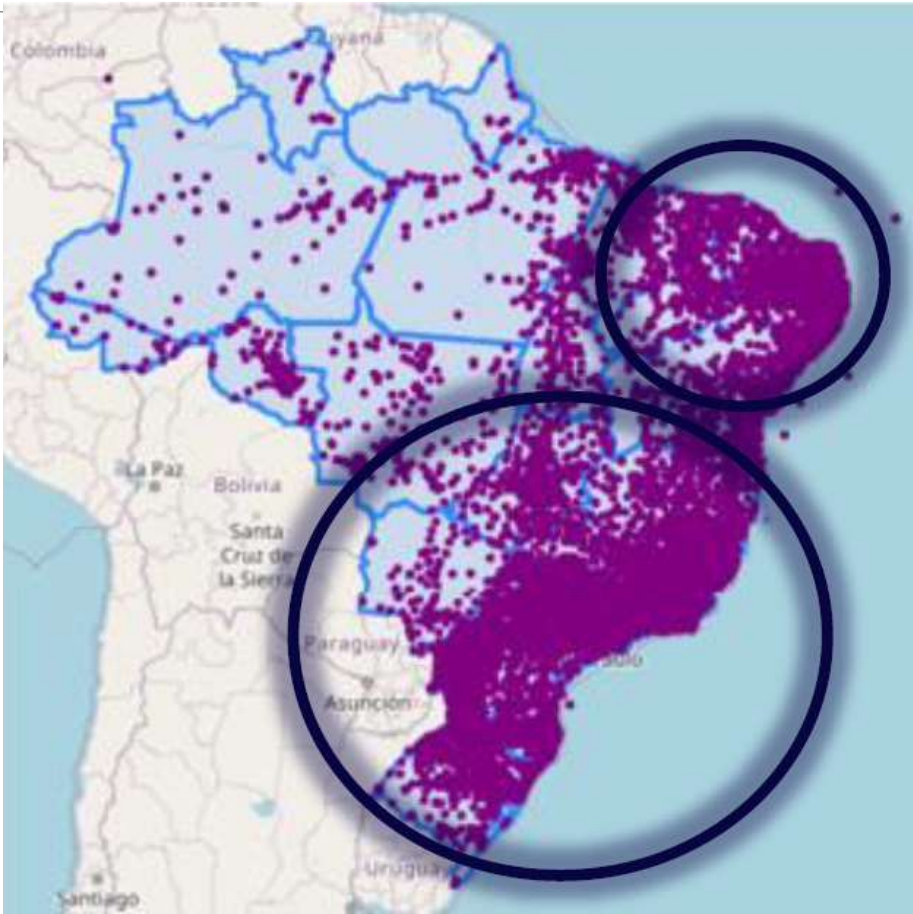
Principales étapes du nettoyage

- Types de données
- Réduction du nombre de catégories de produits (de 72 à 13)
- Assemblage dans une table unique
- Suppression des doublons
- Imputation pour les informations manquantes
- Conservation des commandes 'delivered'
- Suppression des colonnes inutiles
- Création de nouvelles features
- Conservation d'une entrée / client

Présentation des données



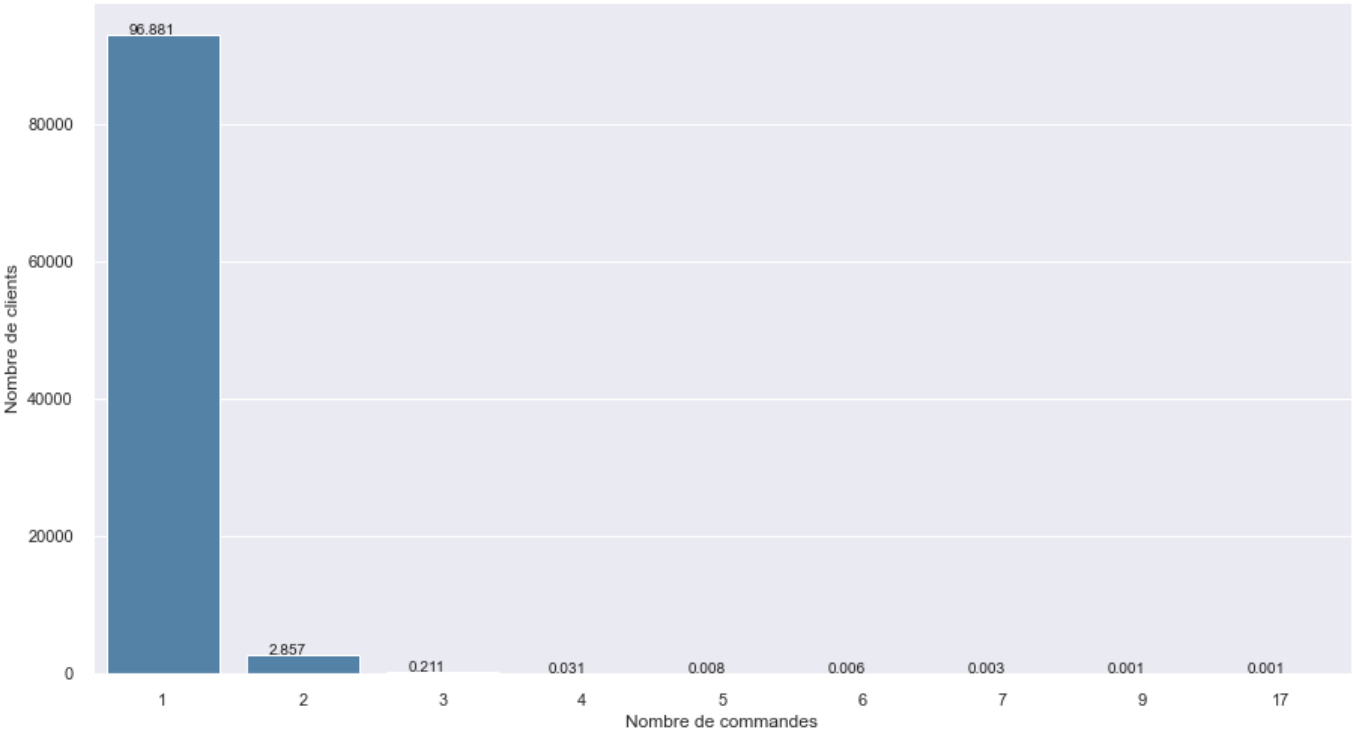
Localisation des clients et vendeurs



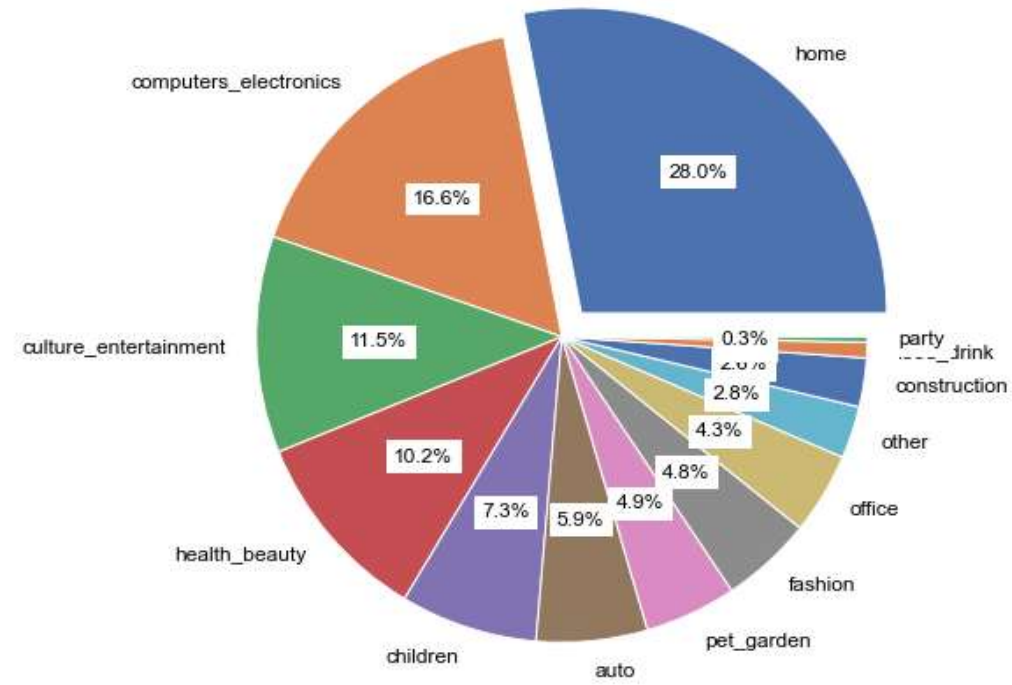
Présentation des données

Produits

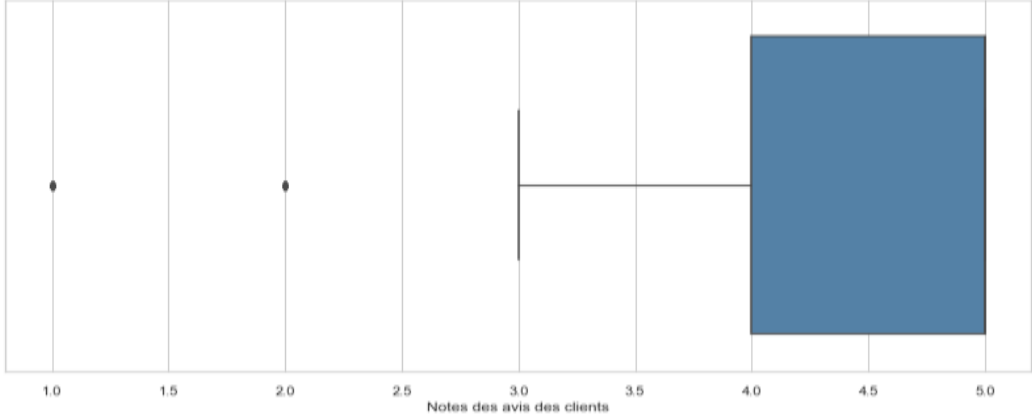
pourcentage de clients par nombre de commandes



Distribution des catégories des produits

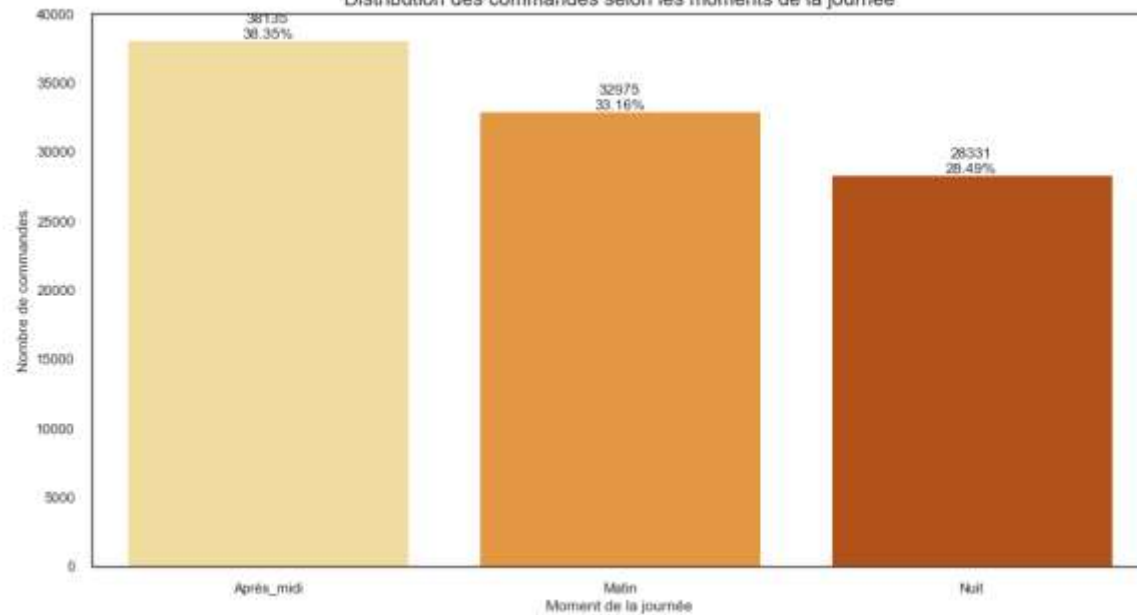


Distribution des notes des avis des client

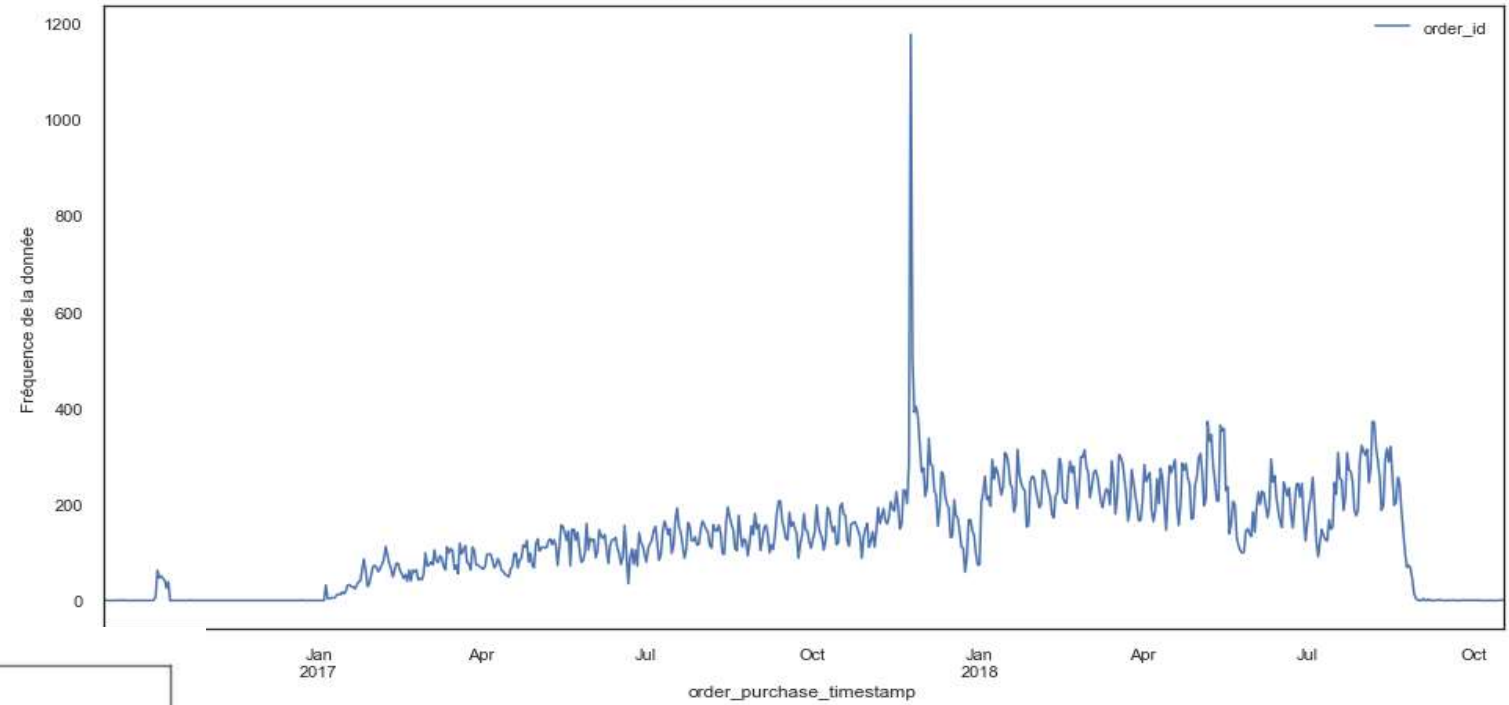


Présentation des données

Distribution des commandes selon les moments de la journée

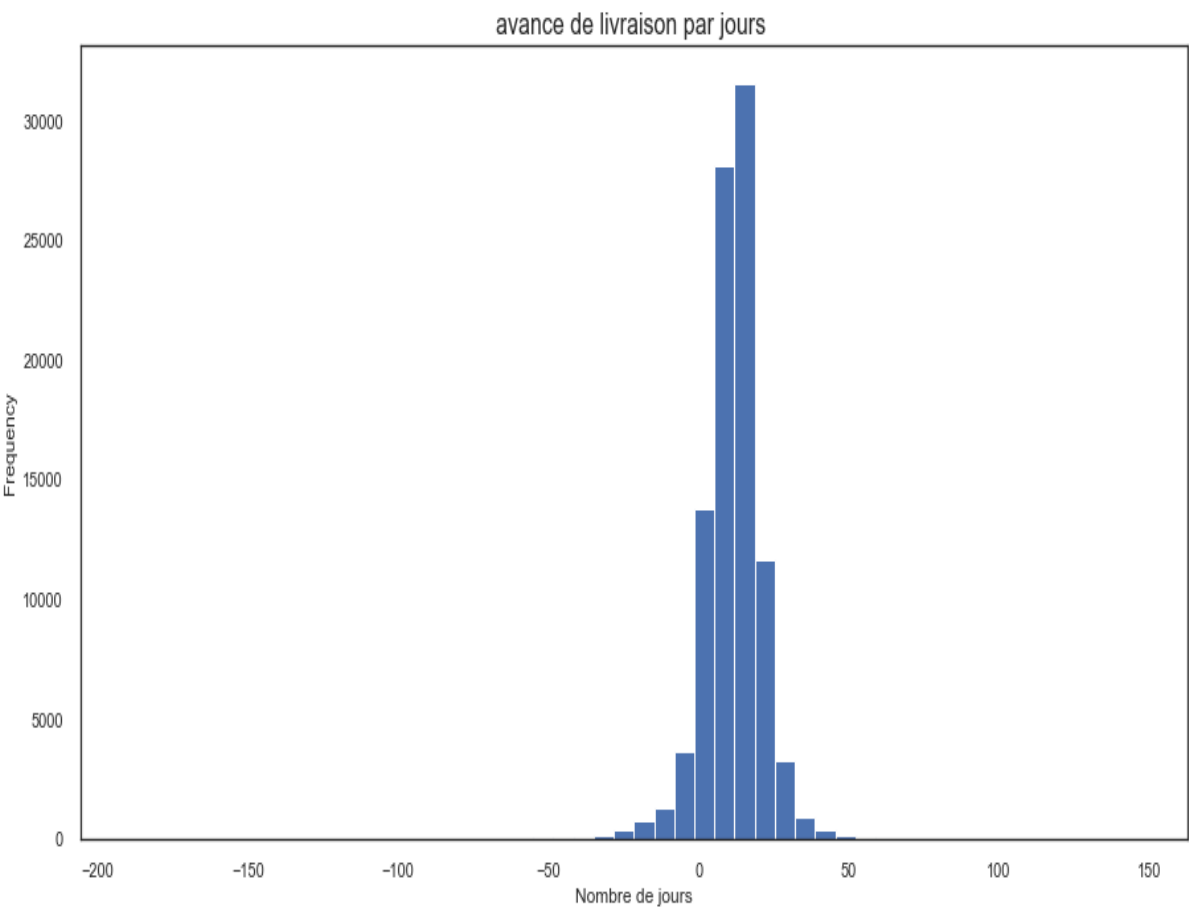
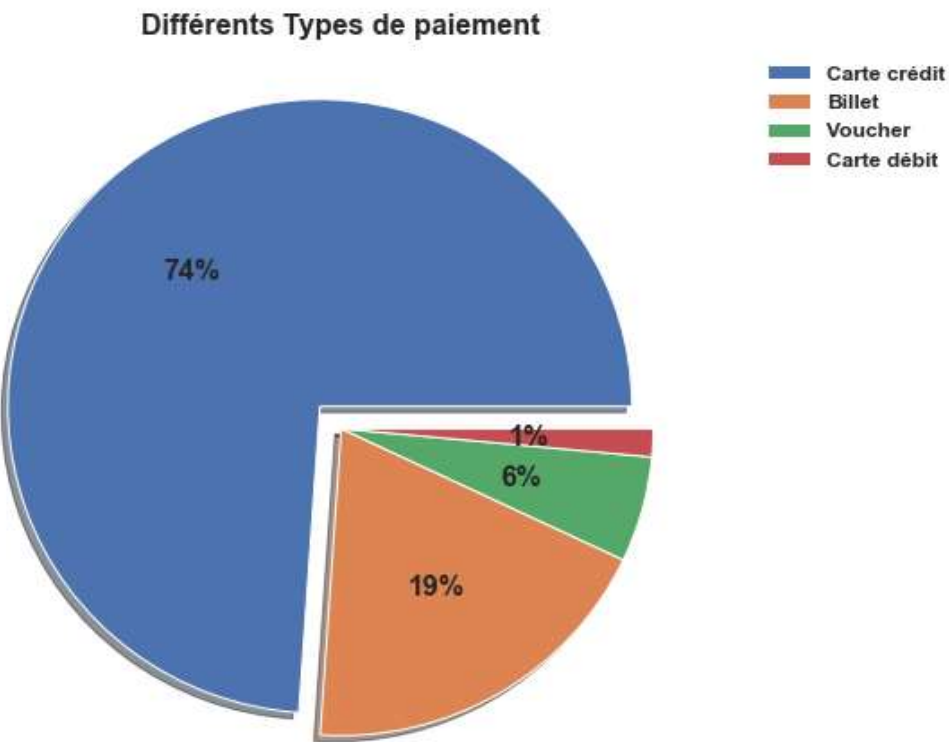


Distribution de la date d'achat



Présentation des données

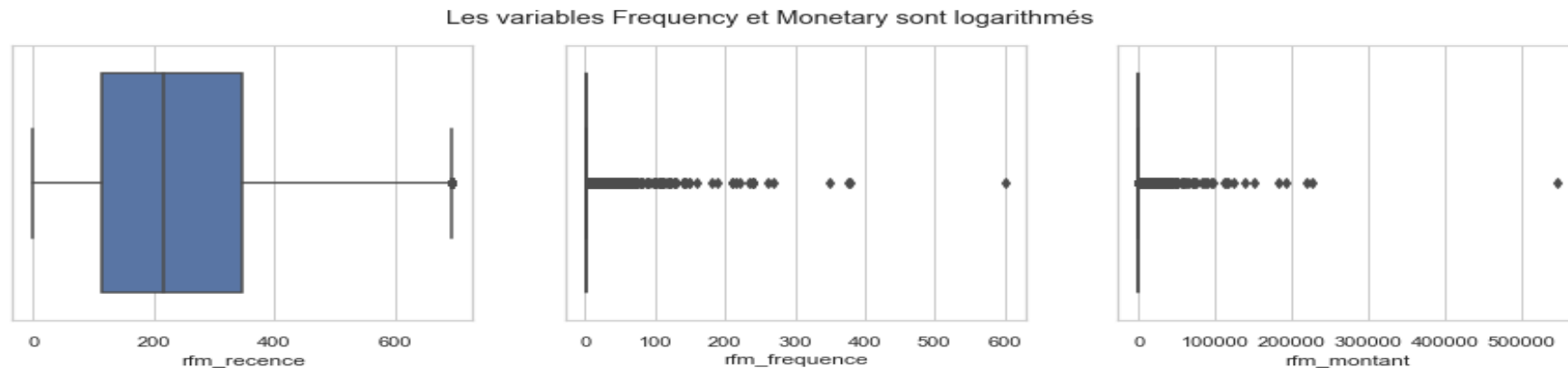
Moyens de paiement



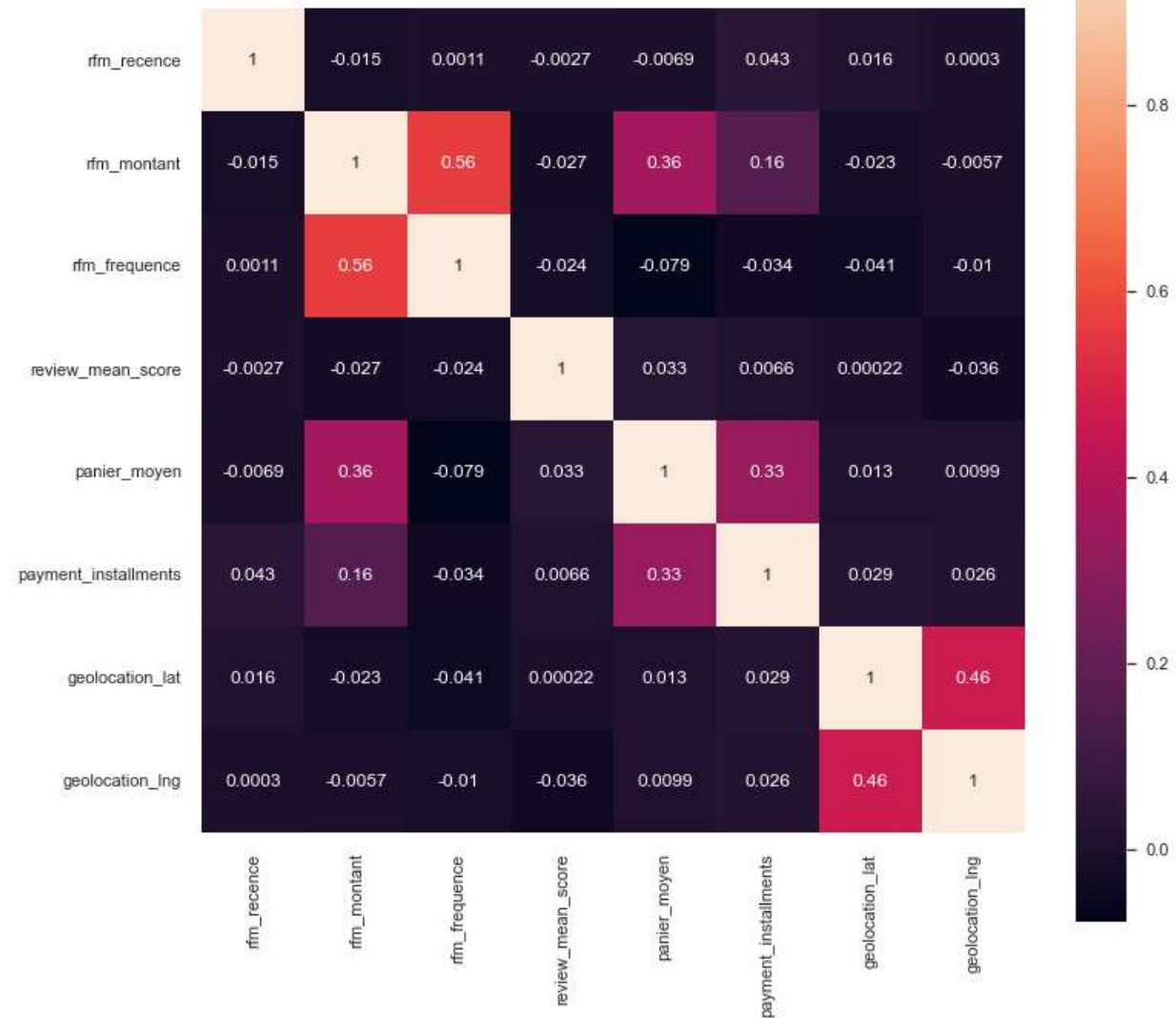
Feature engineering : création de nouvelles variables

- Panier moyen par client (Panier moyen)
- Catégorie la plus achetée (best_category)
- Moyen de paiement préféré (moyen_paiement_preferre)
- Note moyenne des commentaires (review_mean_score)
- Région

Les 3 indicateurs RFM



Matrice de corrélation



Piste de modélisation

Segmentation RFM

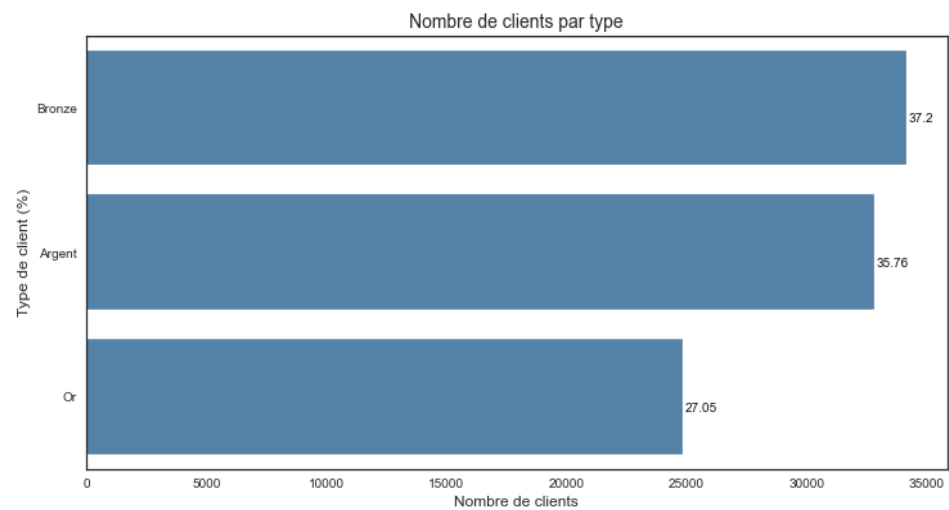
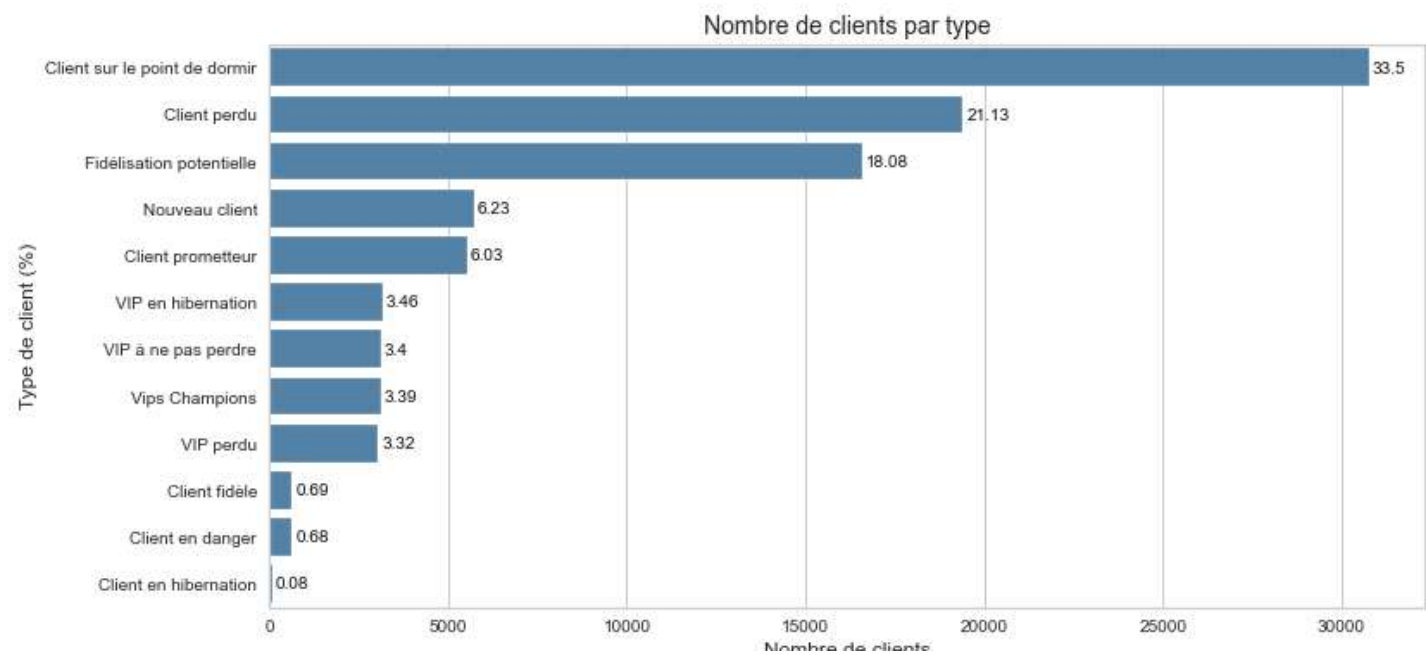
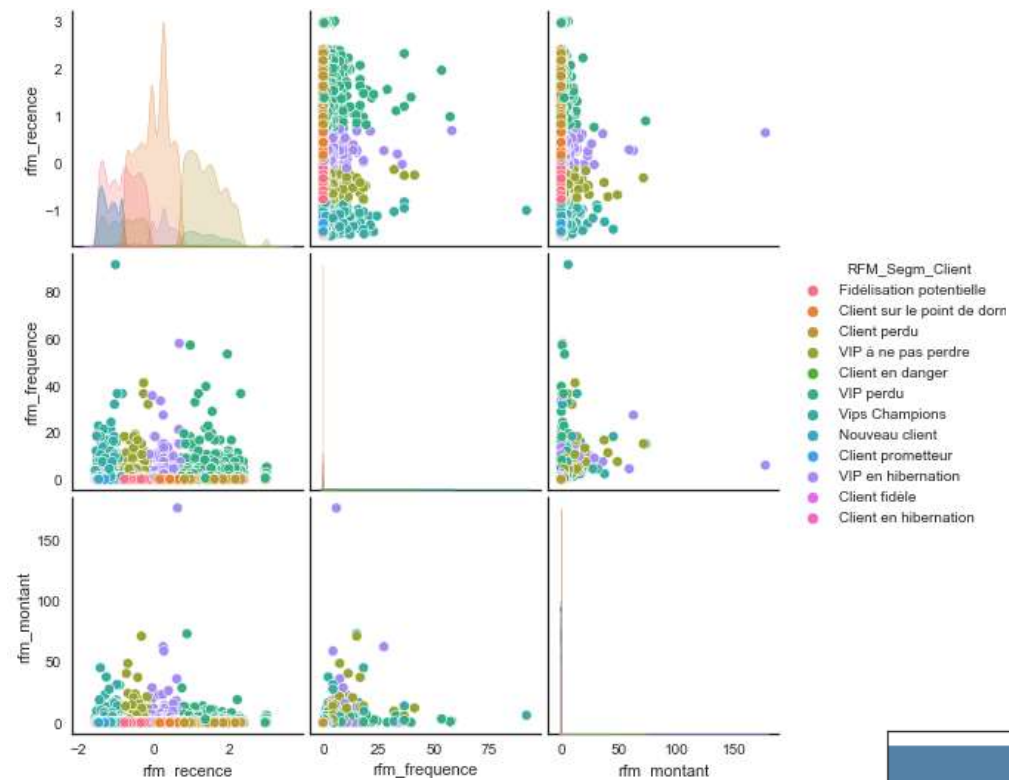
customer_unique_id	rfm_recence	rfm_frequence	rfm_montant	R	F	M	RFM_Segment	RFM_Score	RFM_Segm_Client	RFM_Niveau
0000366f3b9a7992bf8c76cdf3221e2	112	1	141.90	4	1	3	413	8	Fidélisation potentielle	Or
0000b849f77a49e4a4ce2b2a4ca5be3f	115	1	27.19	3	1	1	311	5	Client sur le point de dormir	Bronze
0000f46a3911fa3c0805444483337064	537	1	86.22	1	1	2	112	4	Client perdu	Bronze
0000f6ccb0745a6a4b88665a16c9f078	321	1	43.62	2	1	1	211	4	Client sur le point de dormir	Bronze
0004aac84e0df4da2b147fca70cf8255	288	1	196.89	2	1	3	213	6	Client sur le point de dormir	Argent

Récence, Fréquence et Montant
découpées en 1, 2, 3 ou 4
par quartiles

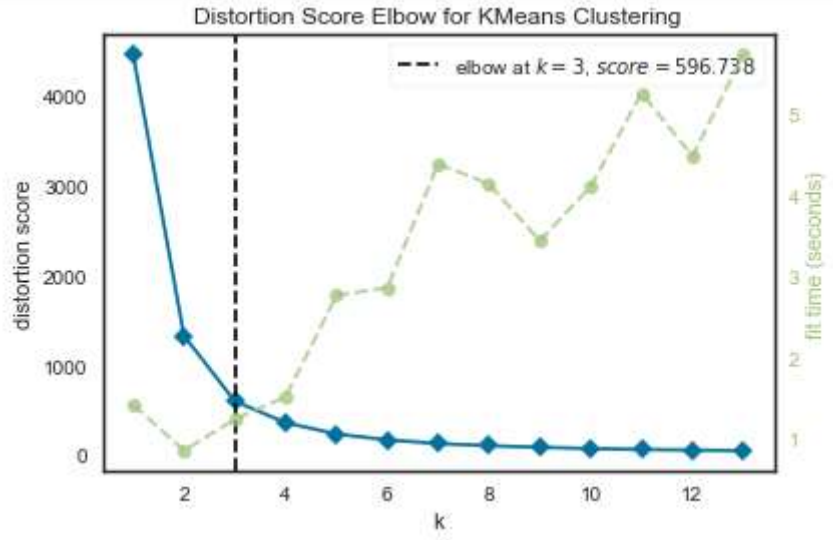
Concaténation

Scoring

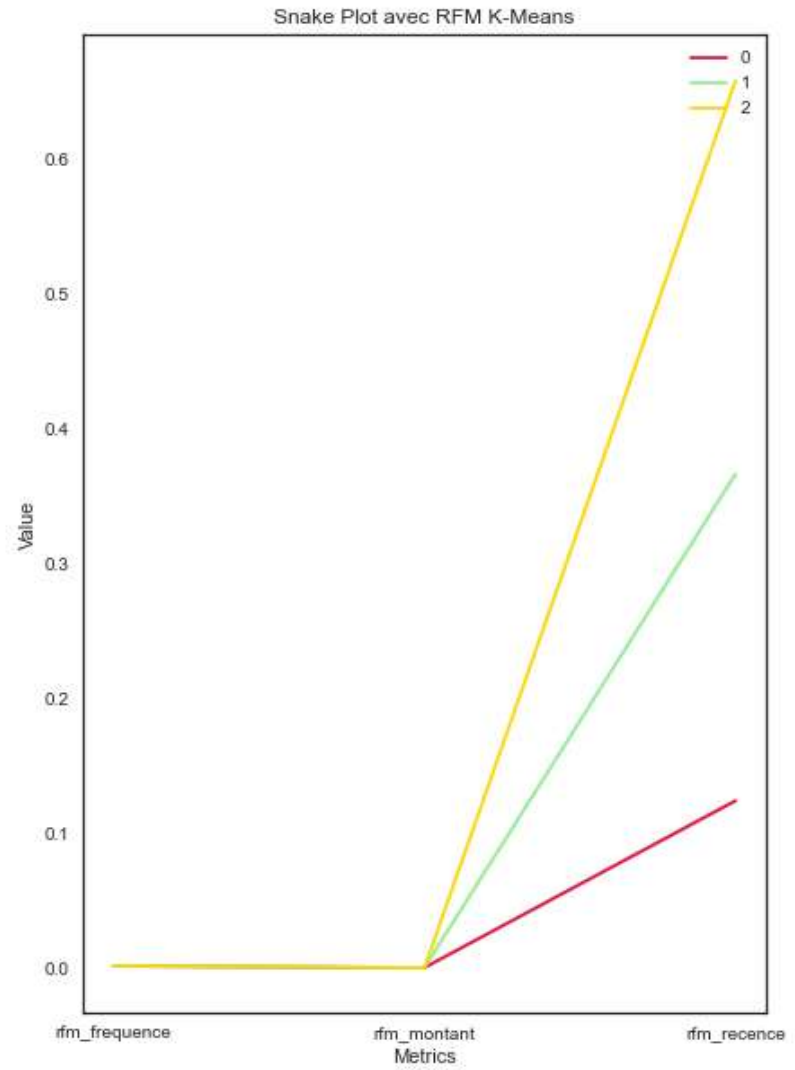
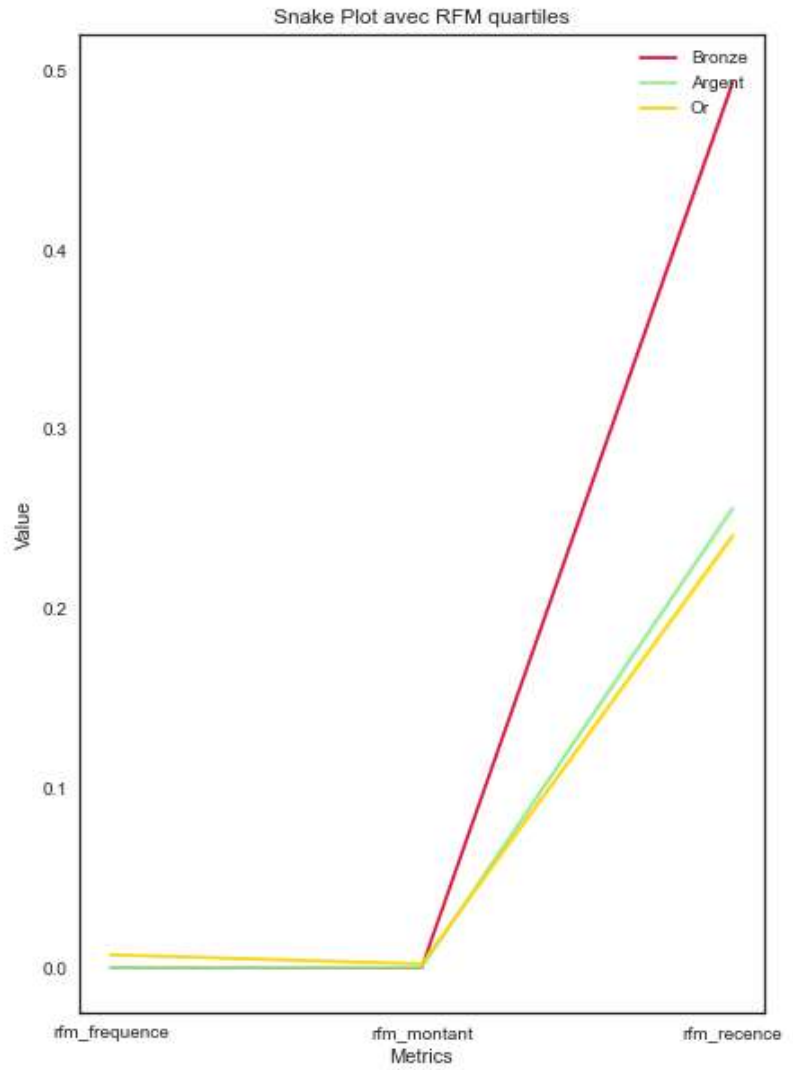
Segmentation de clientèle



Segmentation RFM - K-Means



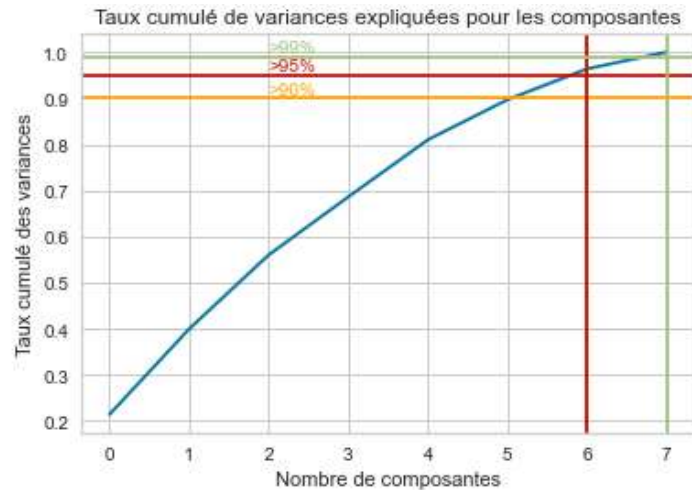
0= Or
1= Argent
2= Bronze



Segmentation RFM+ autres variables

```
['rfm_recence', 'rfm_montant', 'rfm_frequence', 'review_mean_score', 'panier_moyen', 'payment_installments',  
'geolocation_lat', 'geolocation_lng']
```

réduction de nombre de dimensions (ACP)



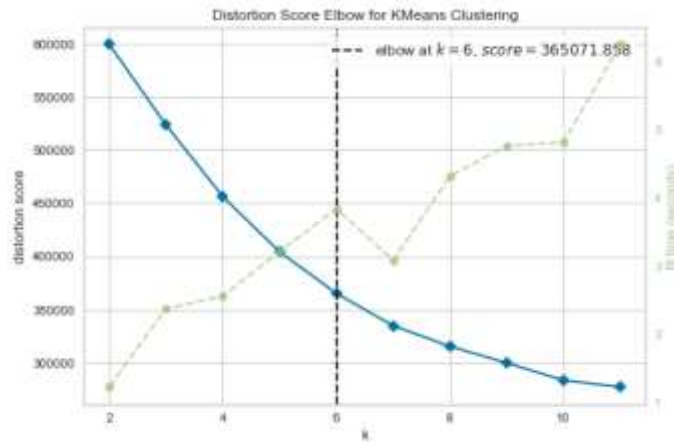
ACP pour visualiser les résultats

F1 représente les clients dépensiers avec des commandes fréquentes

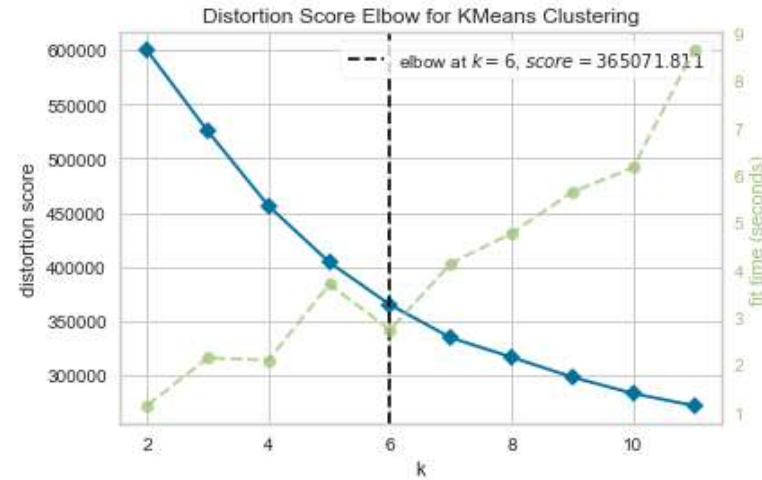
F2 géolocalisation

F3 : nombre de commande et satisfaction

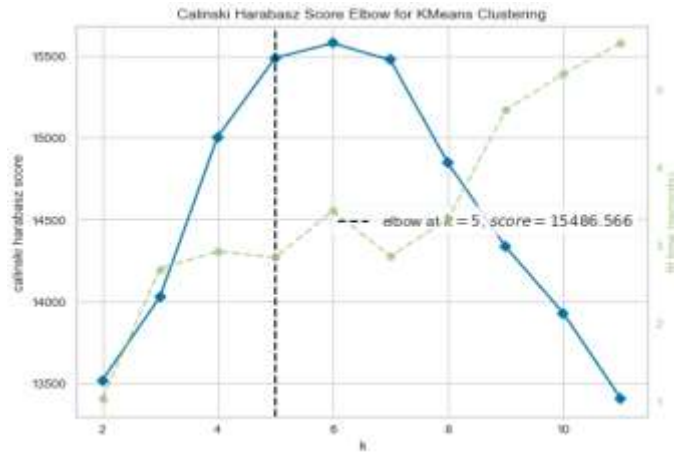
choix du nombre de clusters (k)



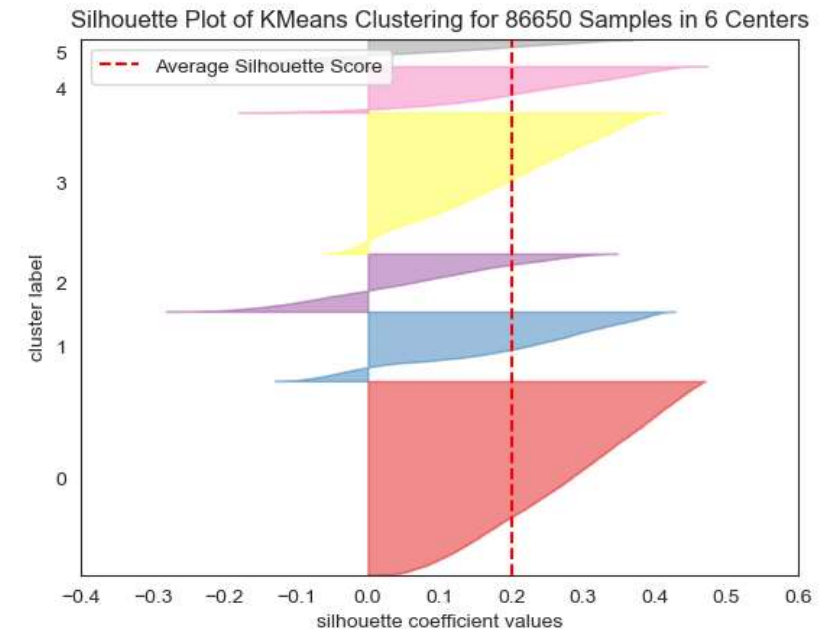
Méthode Elbow



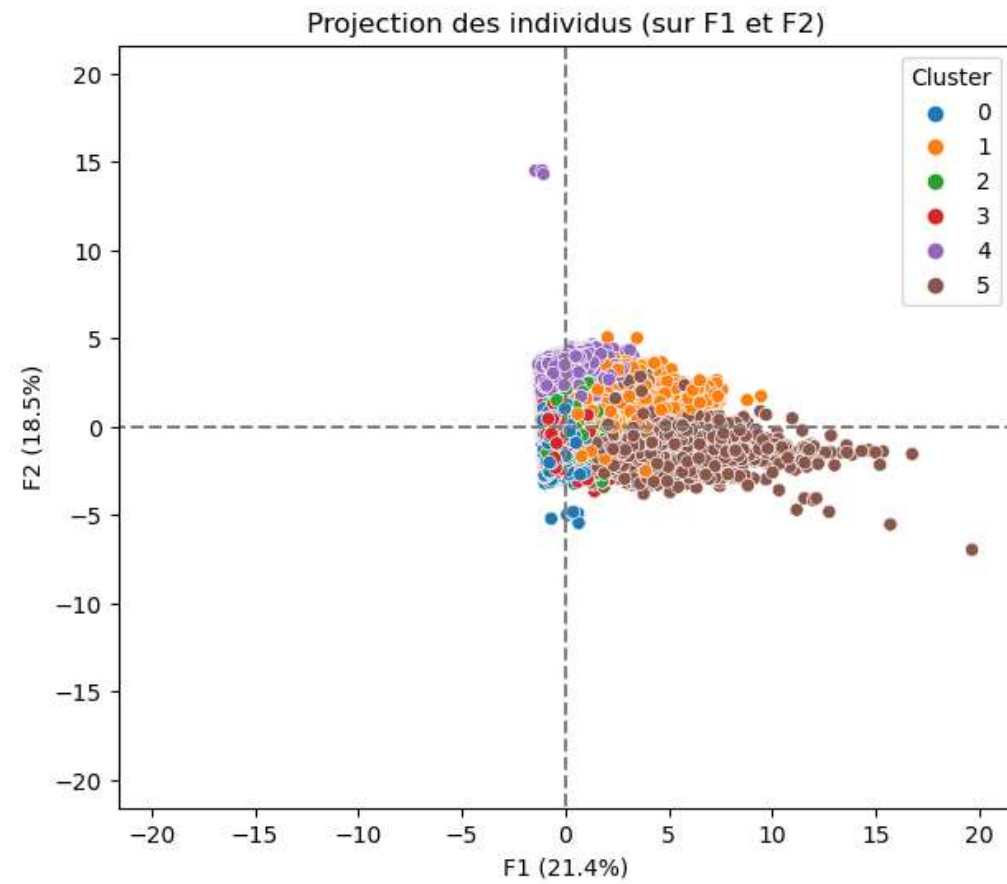
Inertia (distortion)



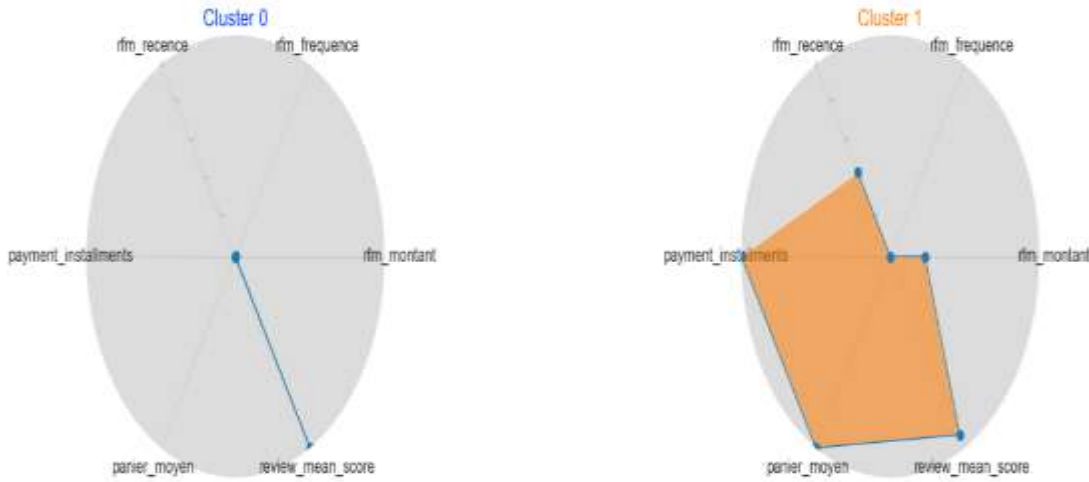
calinski_harabasz



Projection géographique



Interprétation des clusters



Cluster 0

Clients récents

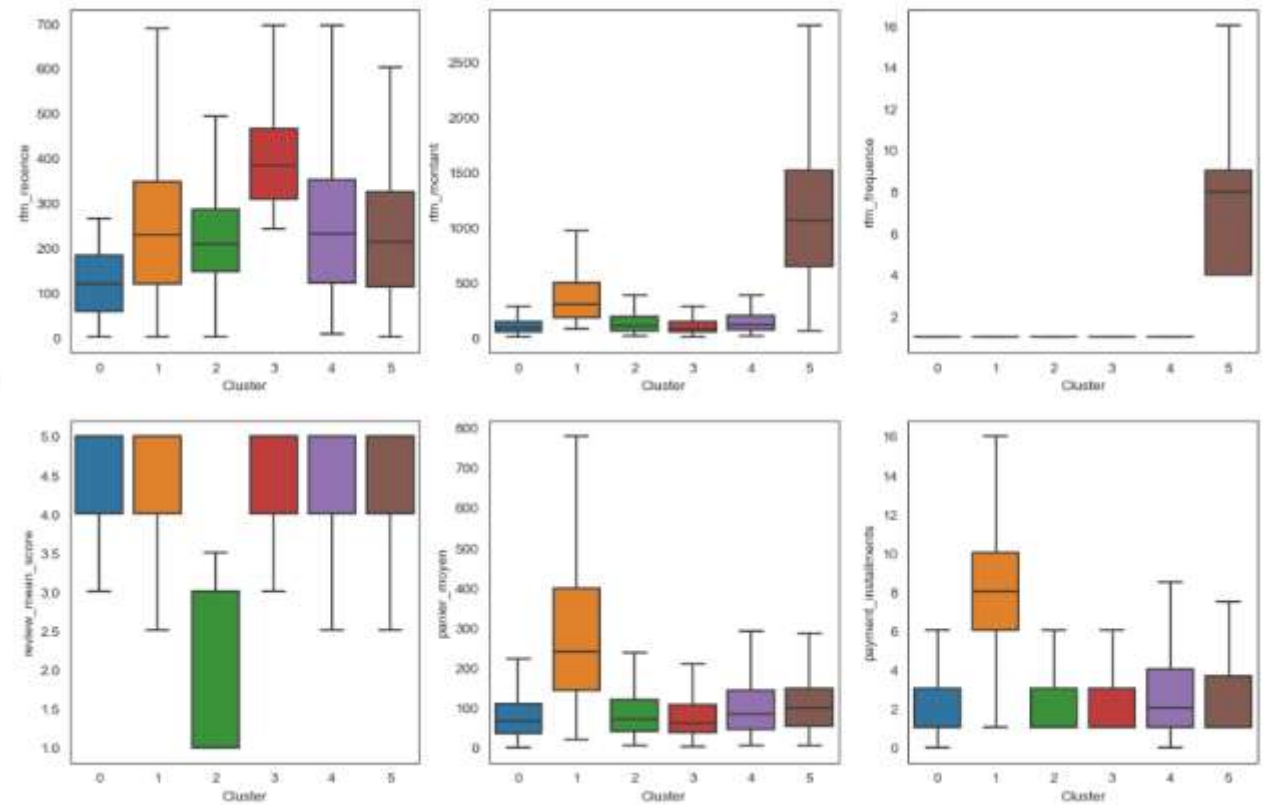
1 seul achat

Montants faibles

satisfaits de leur achat

Régler en 2 ou 3 fois en général

Région = sud et sud – est



Cluster 1

Clients un peu plus récents

1 seul achat

Montants faibles

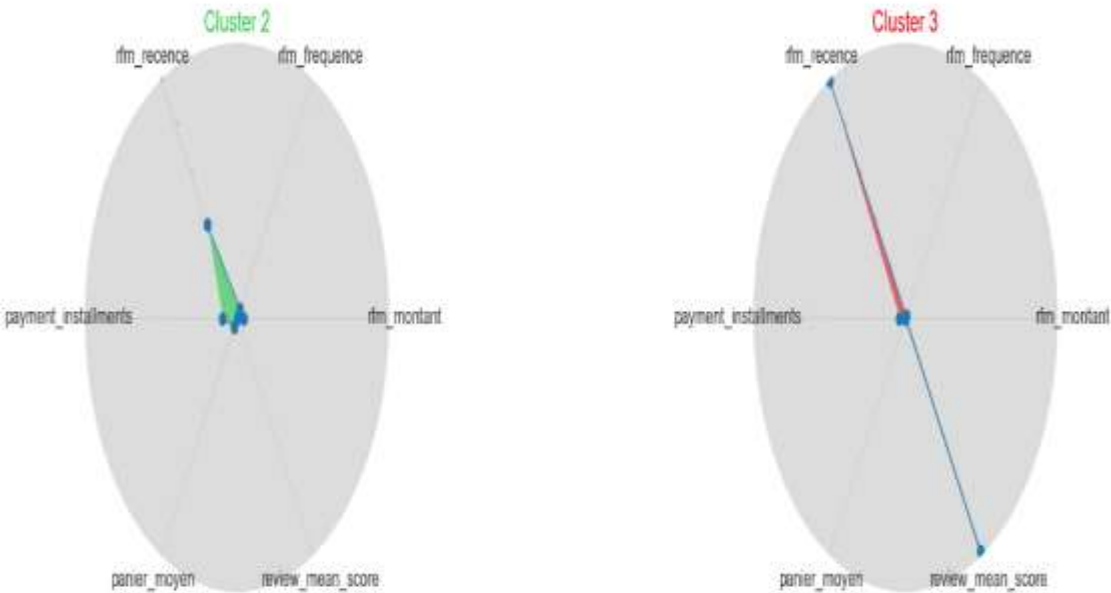
satisfaits de leur achat

Régler en plusieurs fois une fois

Un panier moyen relativement élevé

Région = sud

Interprétation des clusters



Cluster 2

Clients peu récents

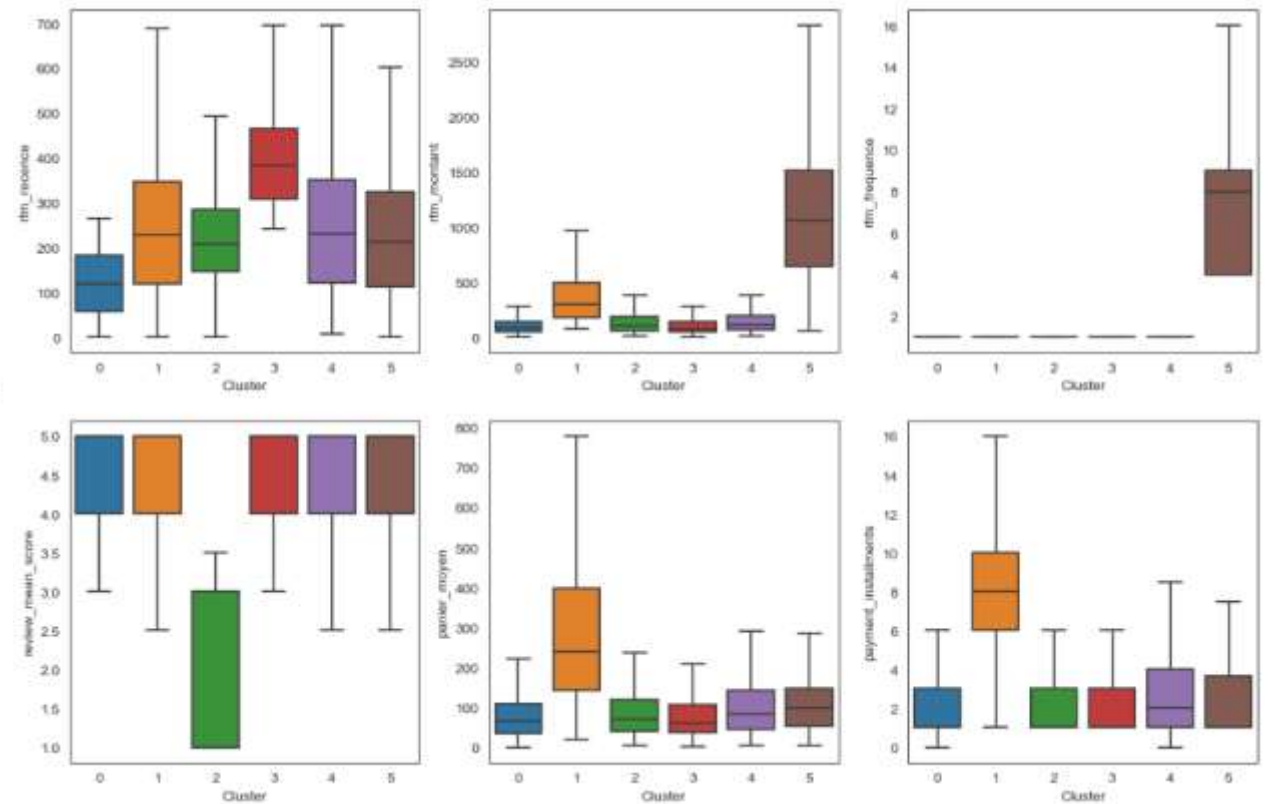
1 seul achat

Montants faibles

pas satisfaits de leur achat

Régler en 2 ou 3 fois en général

Région = centre ouest



Cluster 3

Clients anciens

1 seul achat

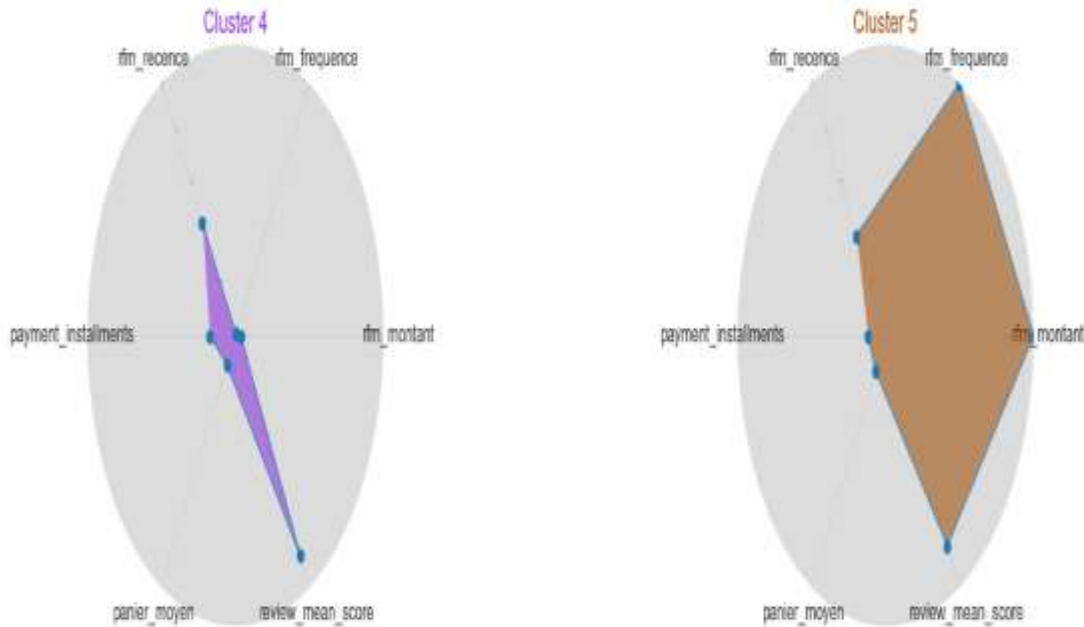
Montants faibles

satisfaits de leur achat

Régler en plusieurs fois une fois

Région = sud

Interprétation des clusters



Cluster 4

Clients peu récents

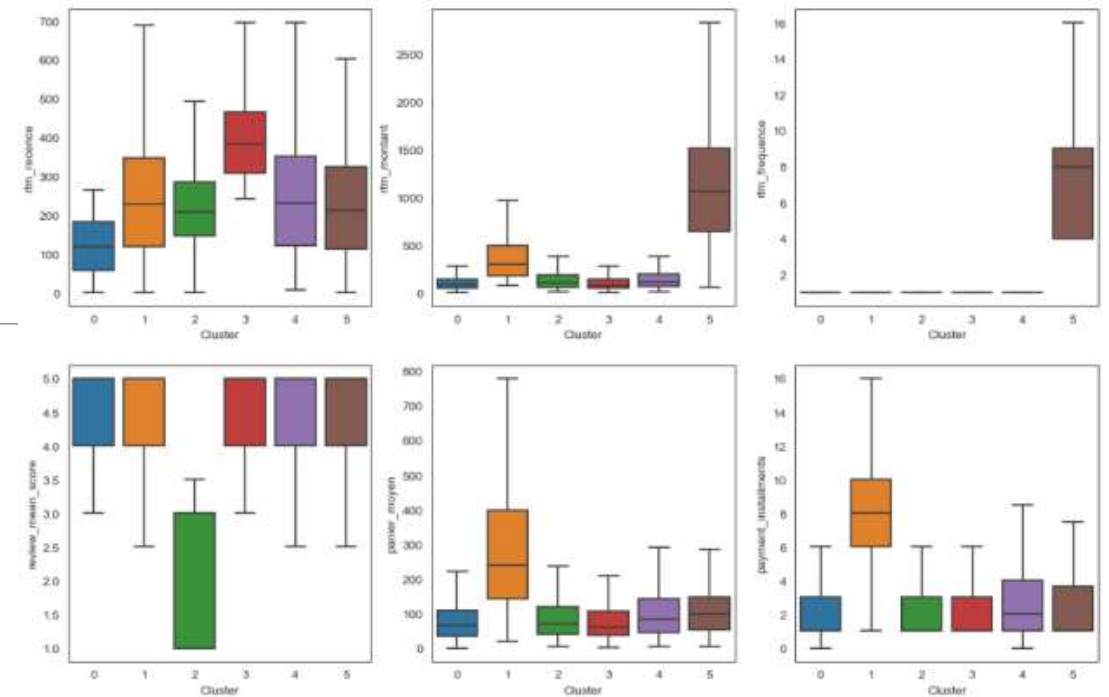
1 seul achat

Montants faibles

satisfaits de leur achat

Régler en 2 ou 3 fois en général

Région = nord est



Cluster 5

Clients peu récents

plusieurs achat

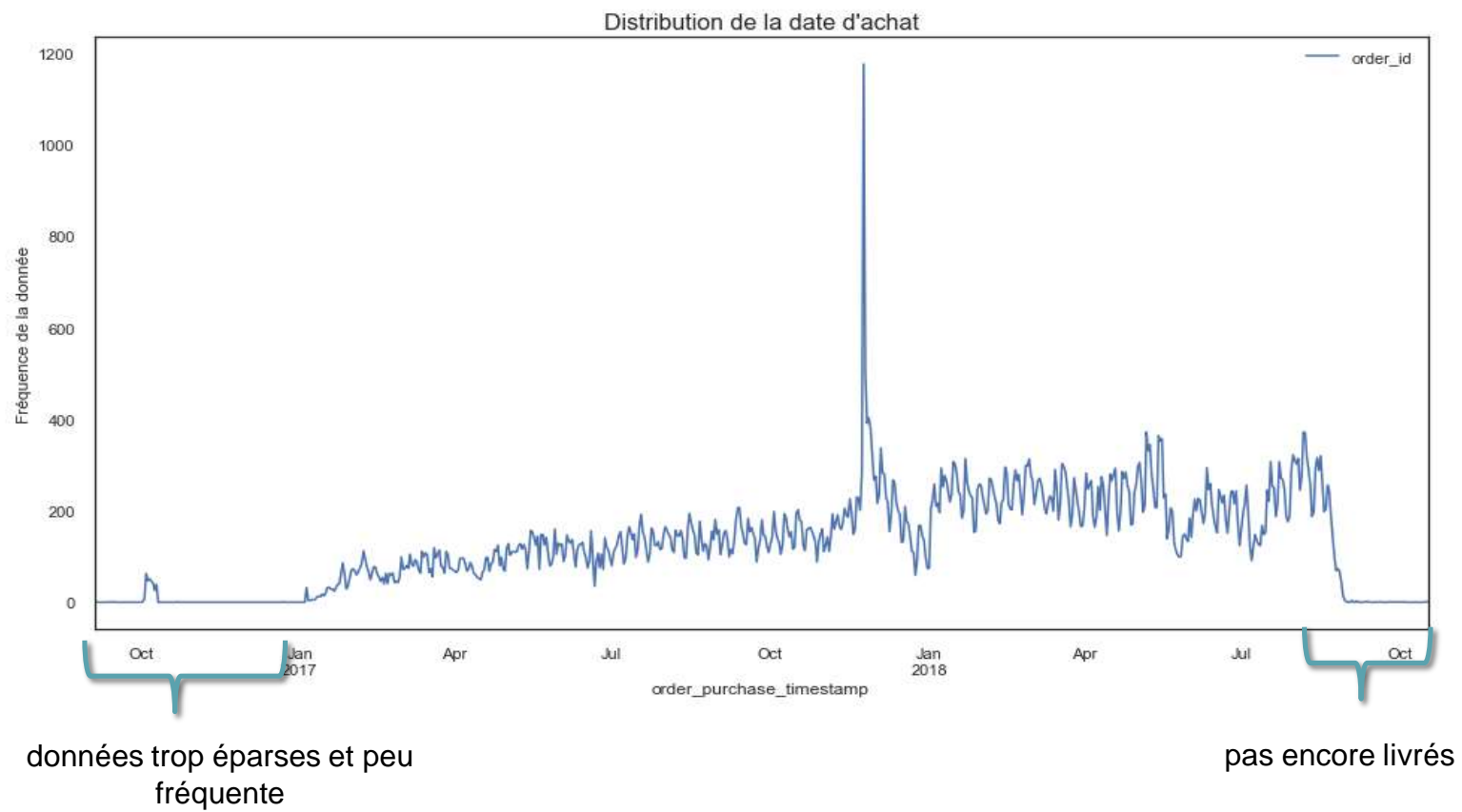
Montants élever

satisfaits de leur achat

Régler en 2 à 4 fois une fois

Région = sud est

Contrat de maintenance

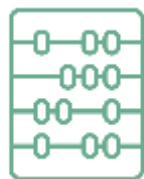


Période d'étude de 01/01/2017 à 01/08/2018



Analyse de la stabilité des clusters au cours du temps

Point de départ:
Base de données avec
12 mois d'existence



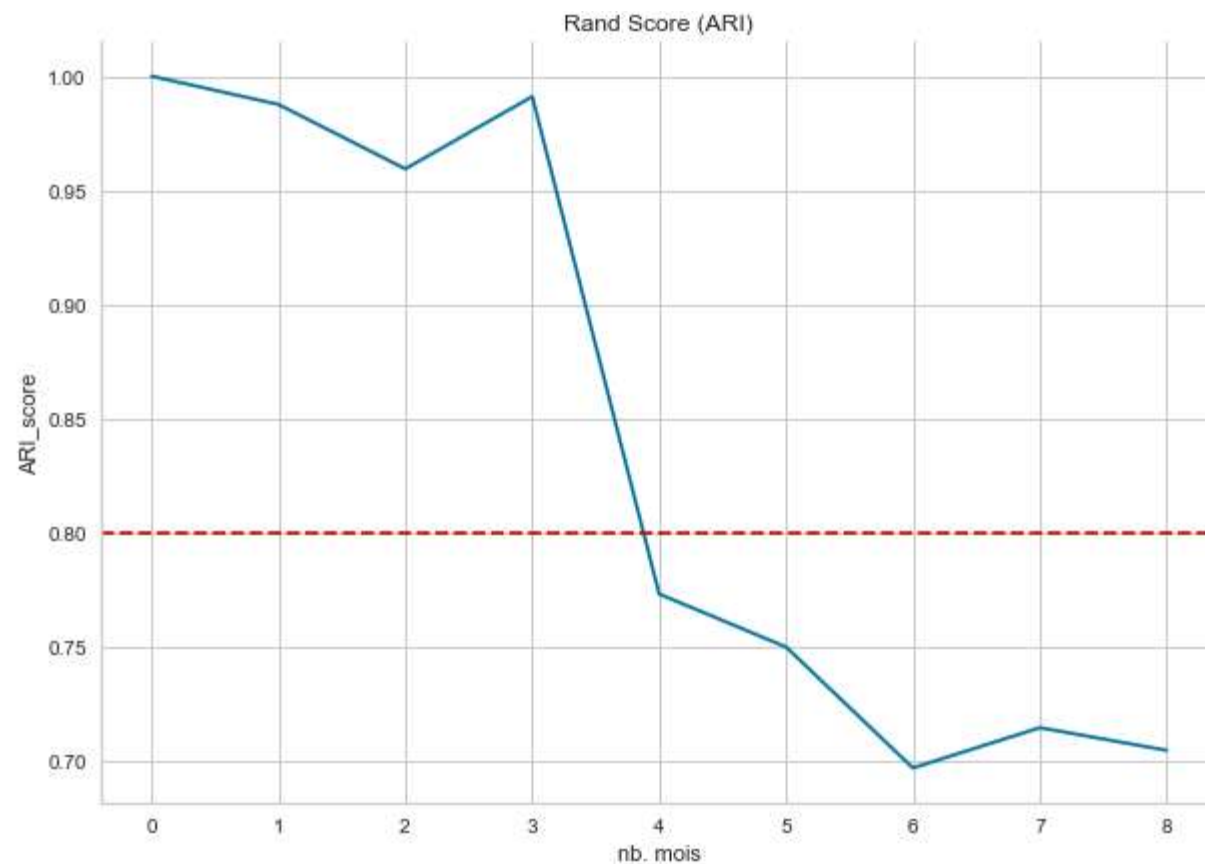
Recalcul des Clusters
tous les mois

Calcul de l'ARScore entre

- FIT d'un nouveau clustering à date
- Un PREDICT de la période d'origine



Répétition de l'opération
7 fois en décalant
de 1 mois à chaque
itération



Conclusion

- La segmentation RFM permet d'obtenir une segmentation rapide et facilement interprétable
- La segmentation KMEANS permet de créer 6 groupes de clients interprétables
- Une mise à jour de la base de données est nécessaire tous les 3 mois
- Il manque des informations dans le jeu de données (sexe, âge, profil socio-professionnel, ...).
- Possibilités d'amélioration Changer le nombre de clusters en accord avec le service Marketing

MERCI