

olist

Cédric - Novembre 2024

# Segmentation des clients



# Mission

Réaliser une segmentation des clients  
pour un usage marketing

Proposer un contrat de maintenance  
basée sur la stabilité des segments

# Les données

Base de données Olist anonymisée :

Historique des commandes

Produits achetés

Commentaires de satisfaction

Localisation des clients

# Mode opératoire


Analyse exploratoire

Approche RFM Métier

Modélisations du RFM

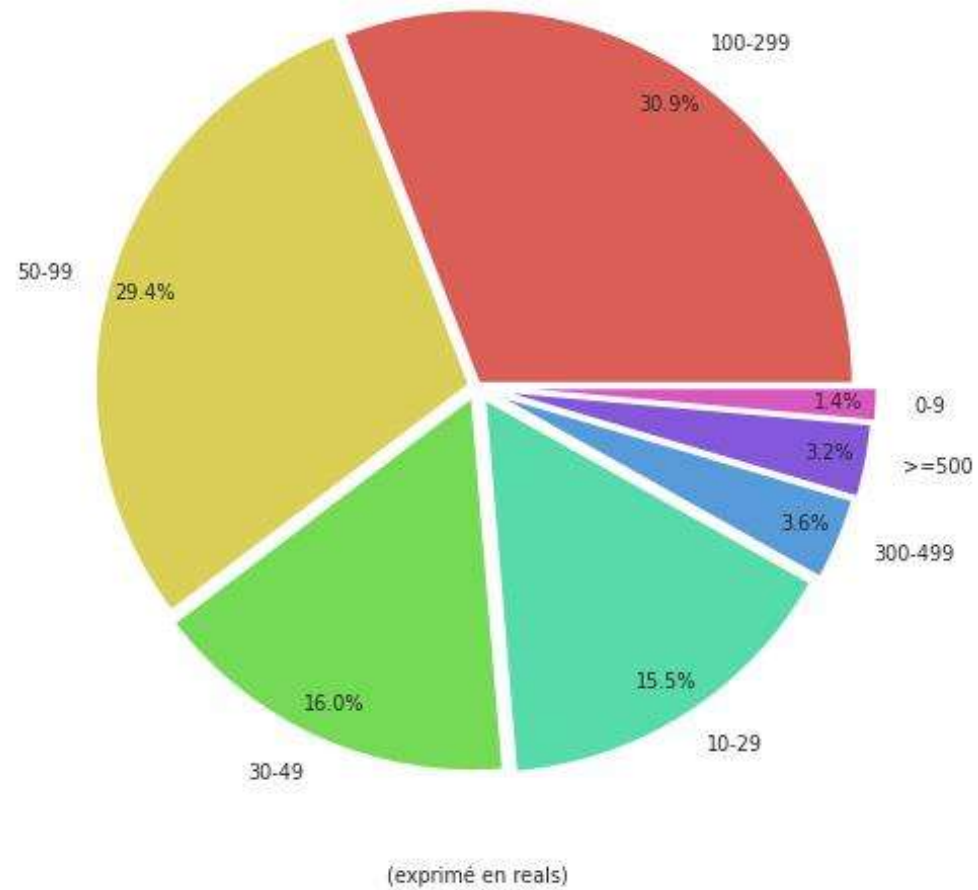
Contrat de maintenance

olist

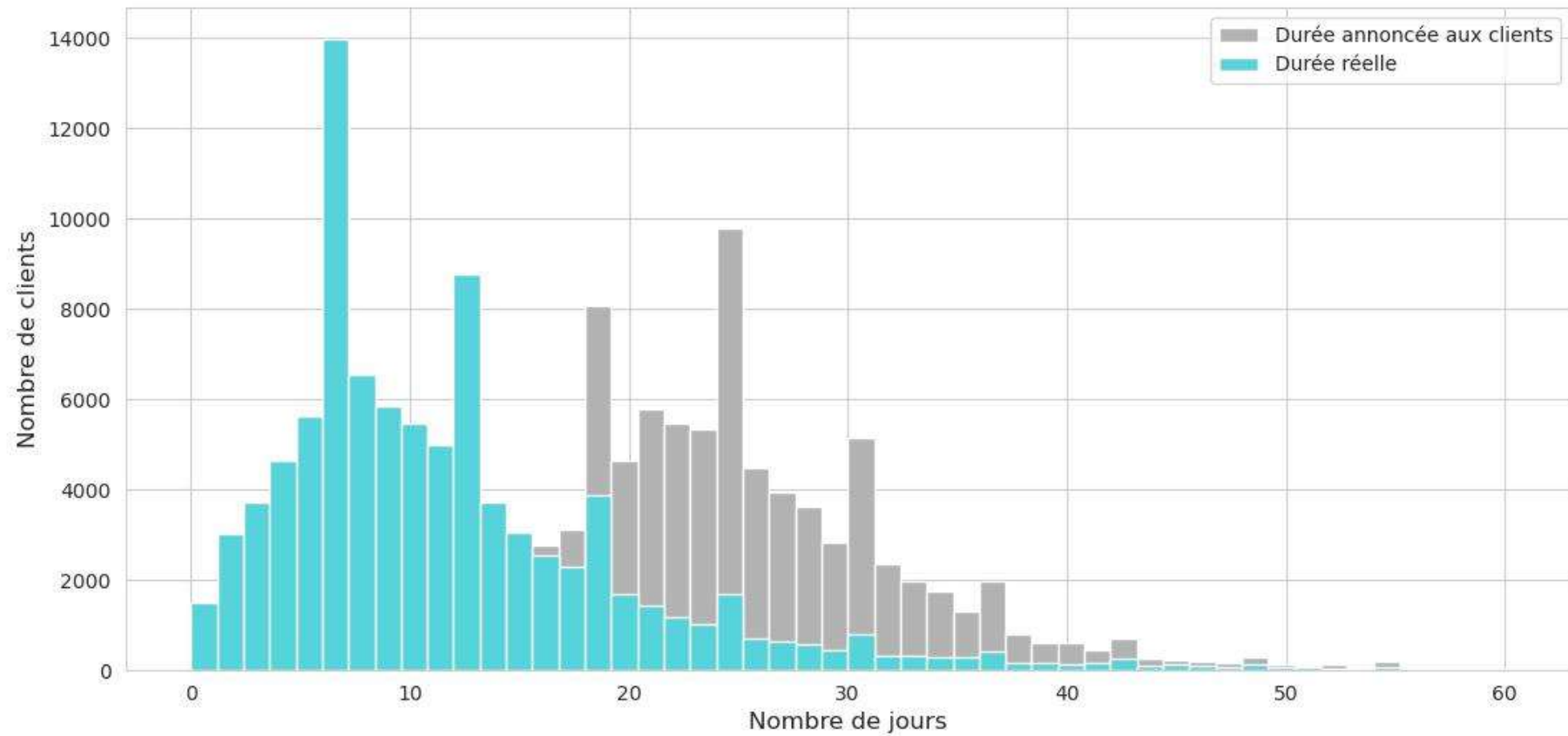


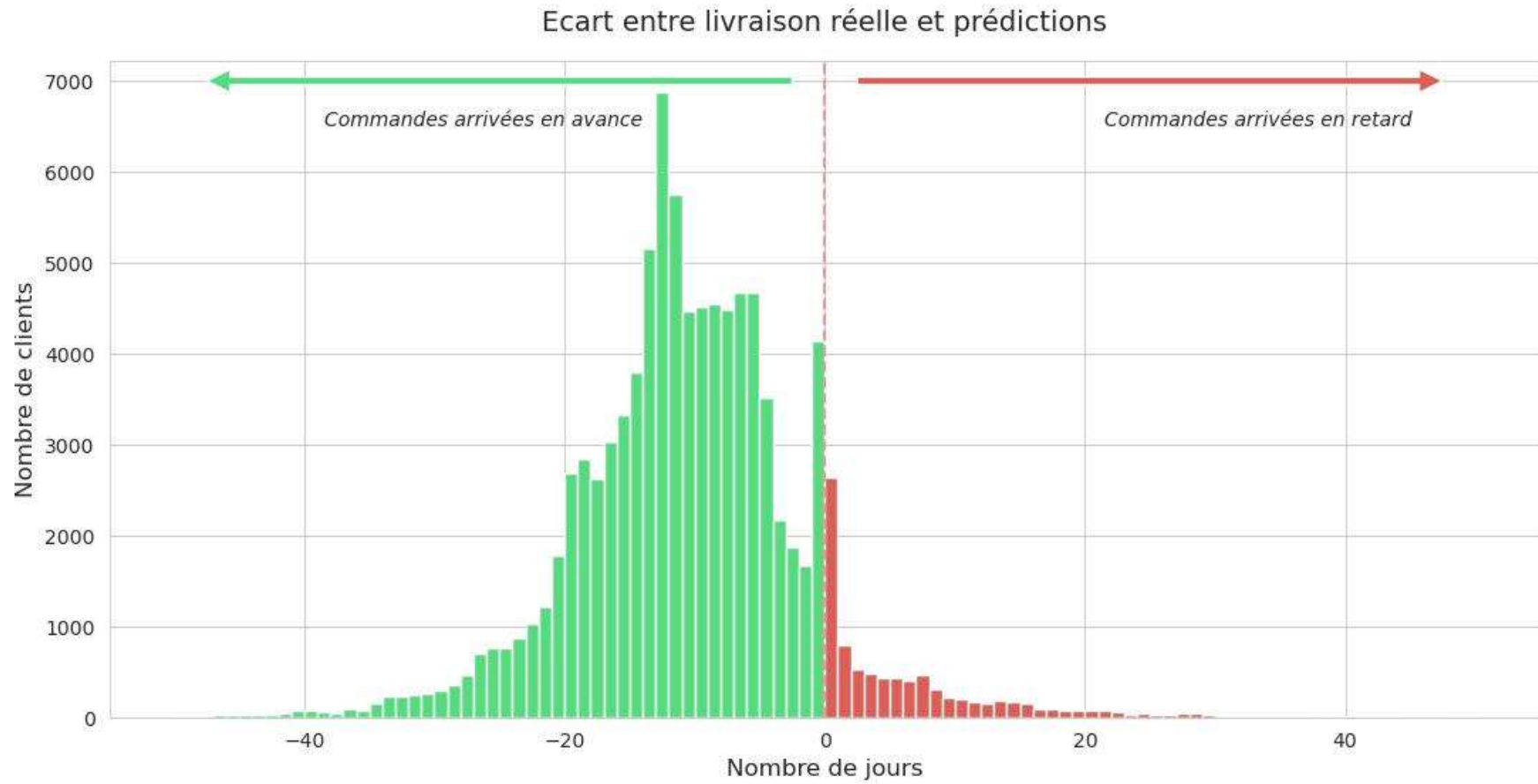
# Analyse Exploratoire

## Montants moyens des commandes par clients



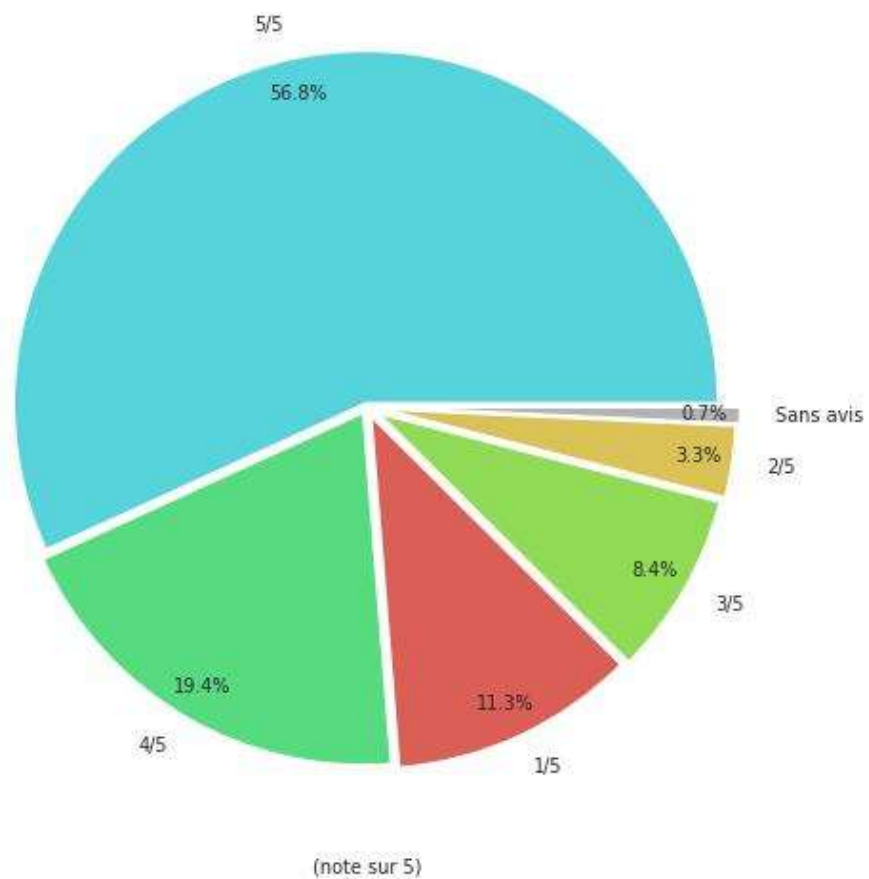
Durée de livraison moyenne des clients

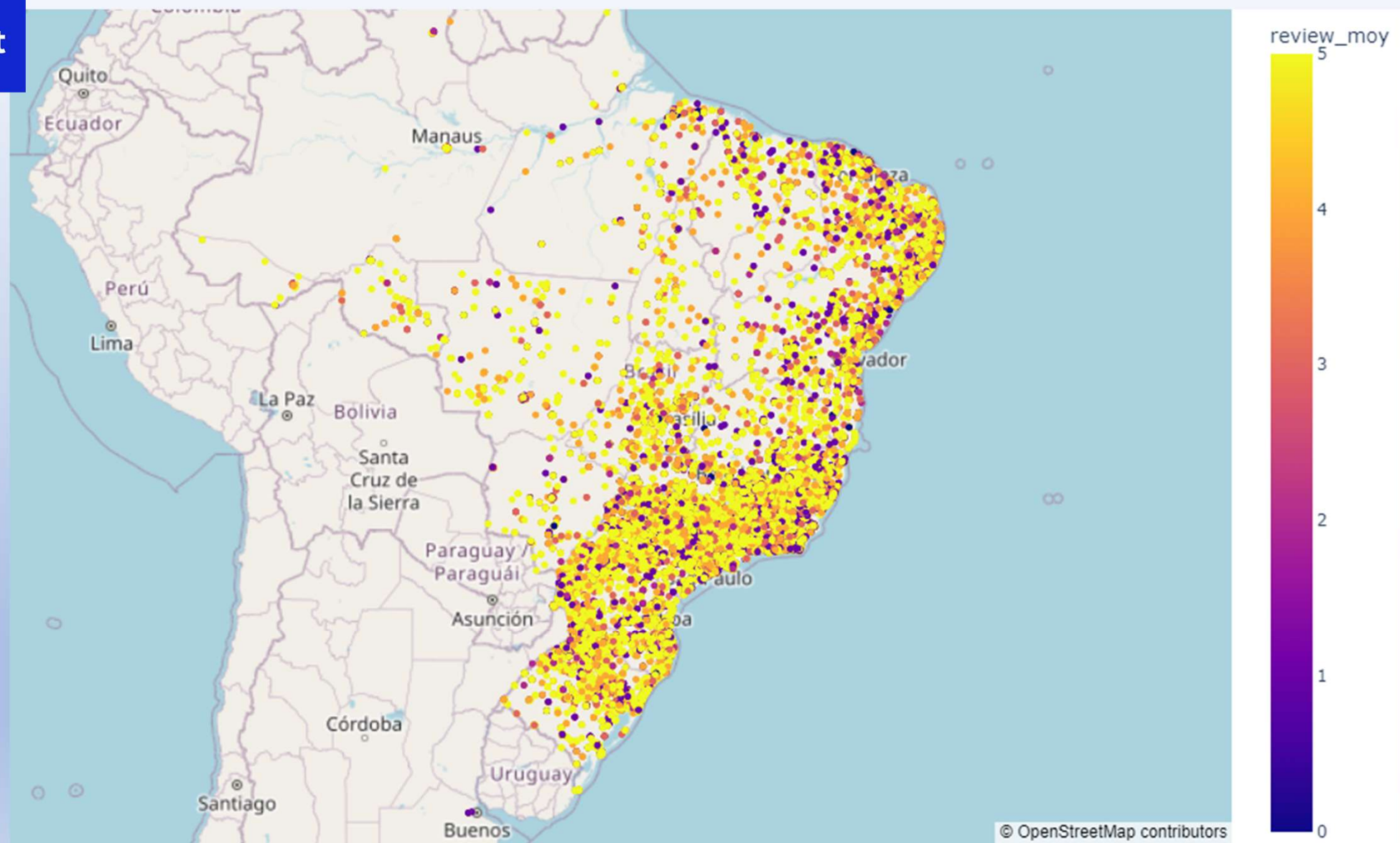




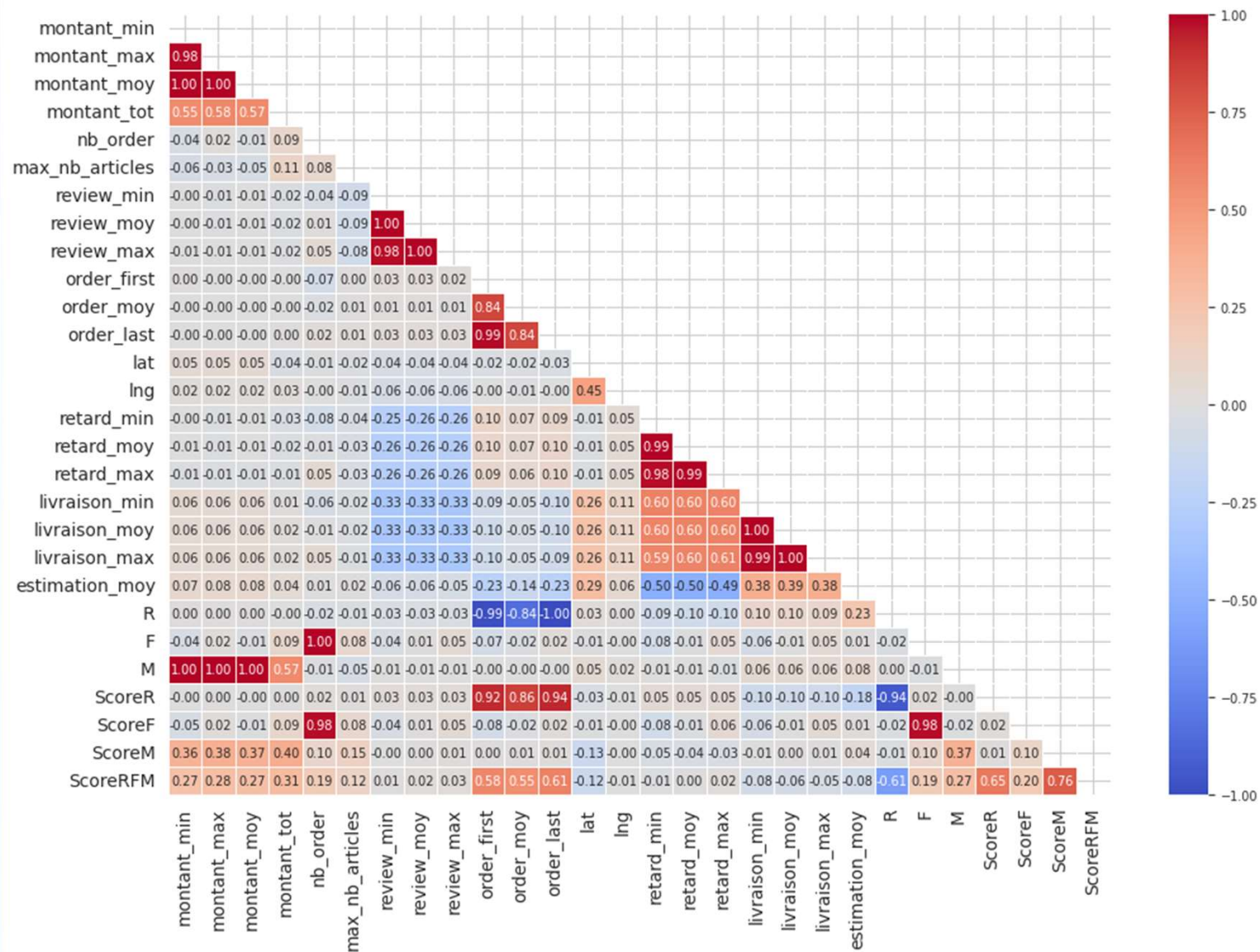


## Répartition des avis clients





Corrélations de pearson entre les features d'origine



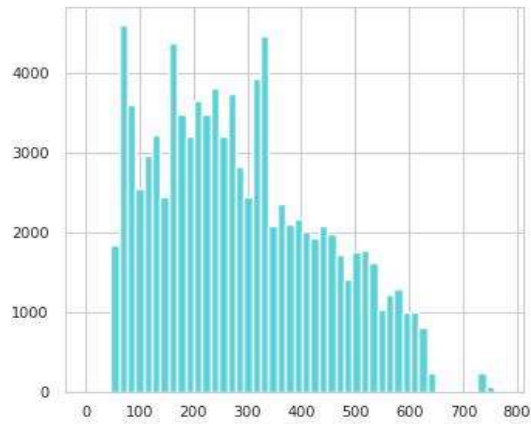
olist

# Approche Métier RFM

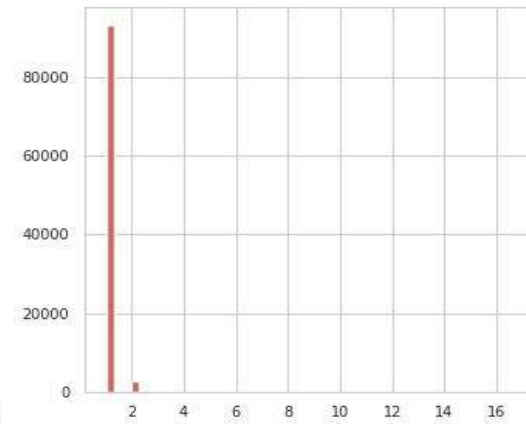




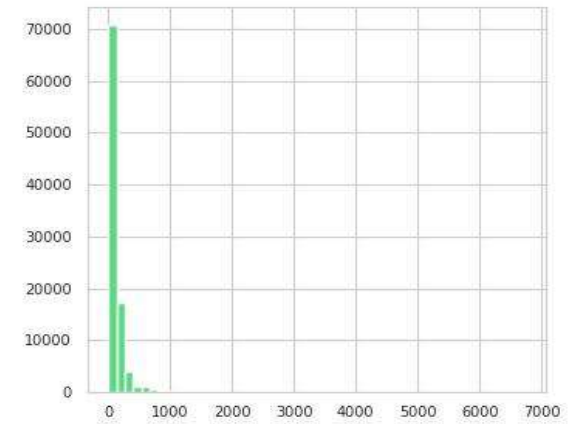
Distribution R



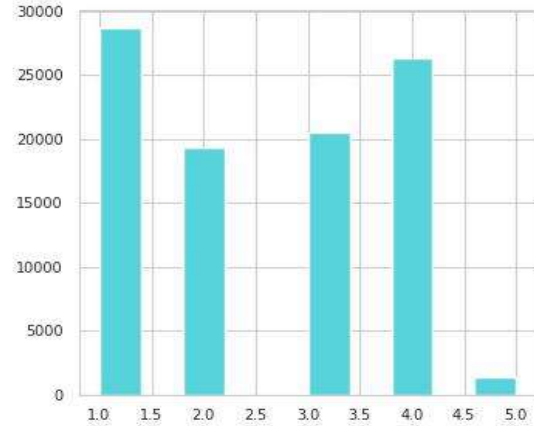
Distribution F



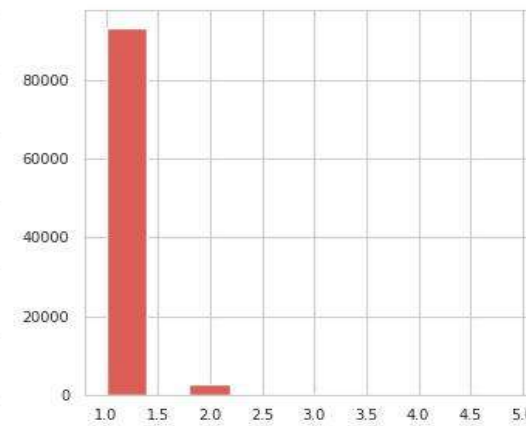
Distribution M



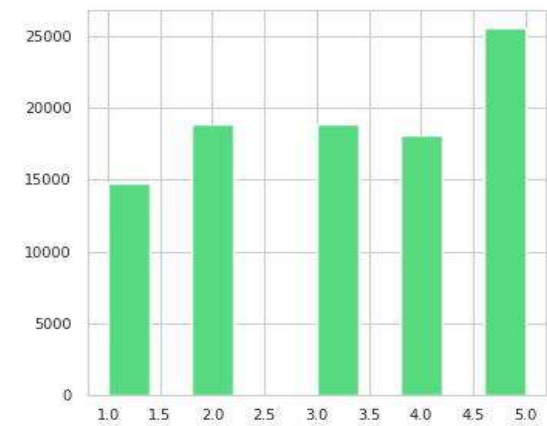
Distribution score R



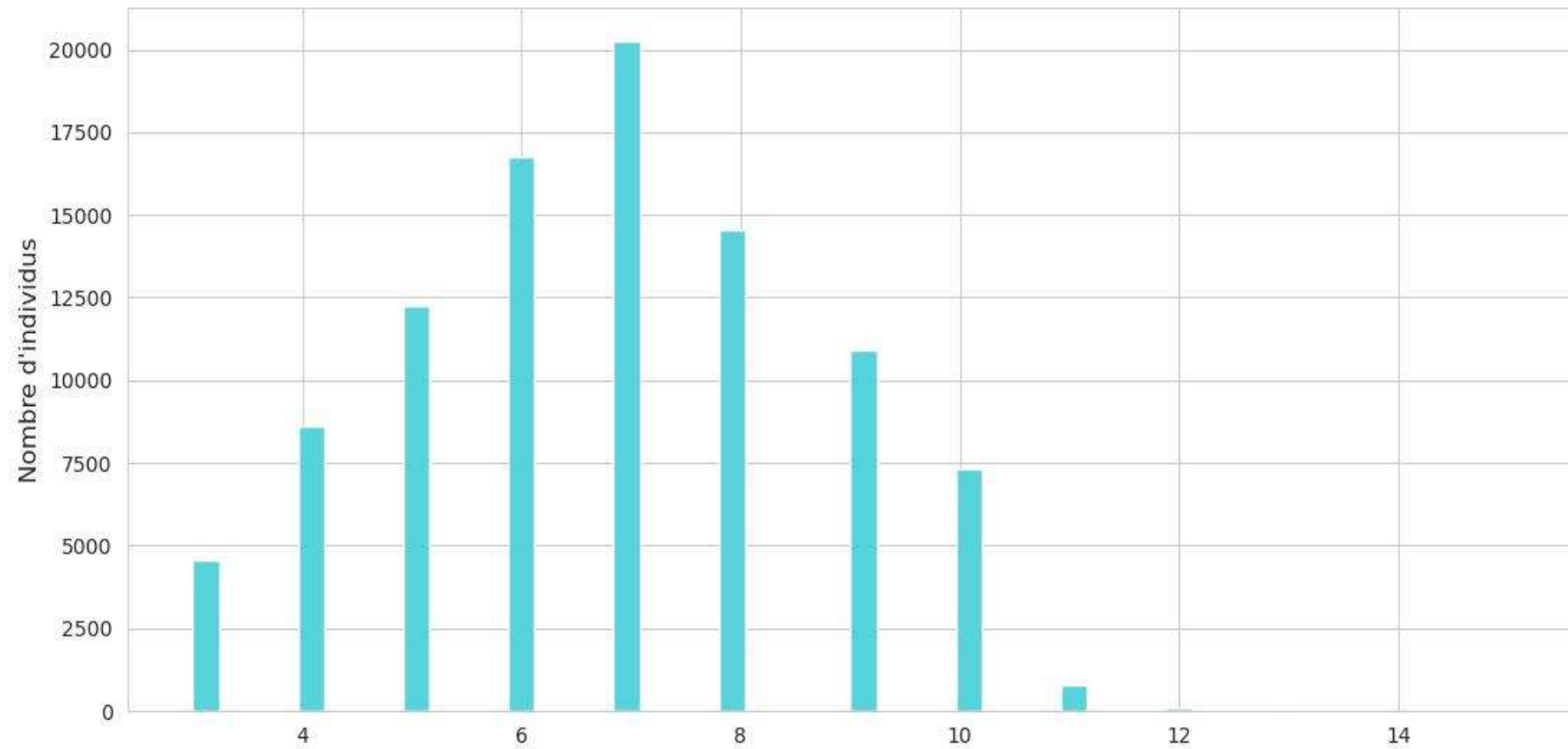
Distribution score F



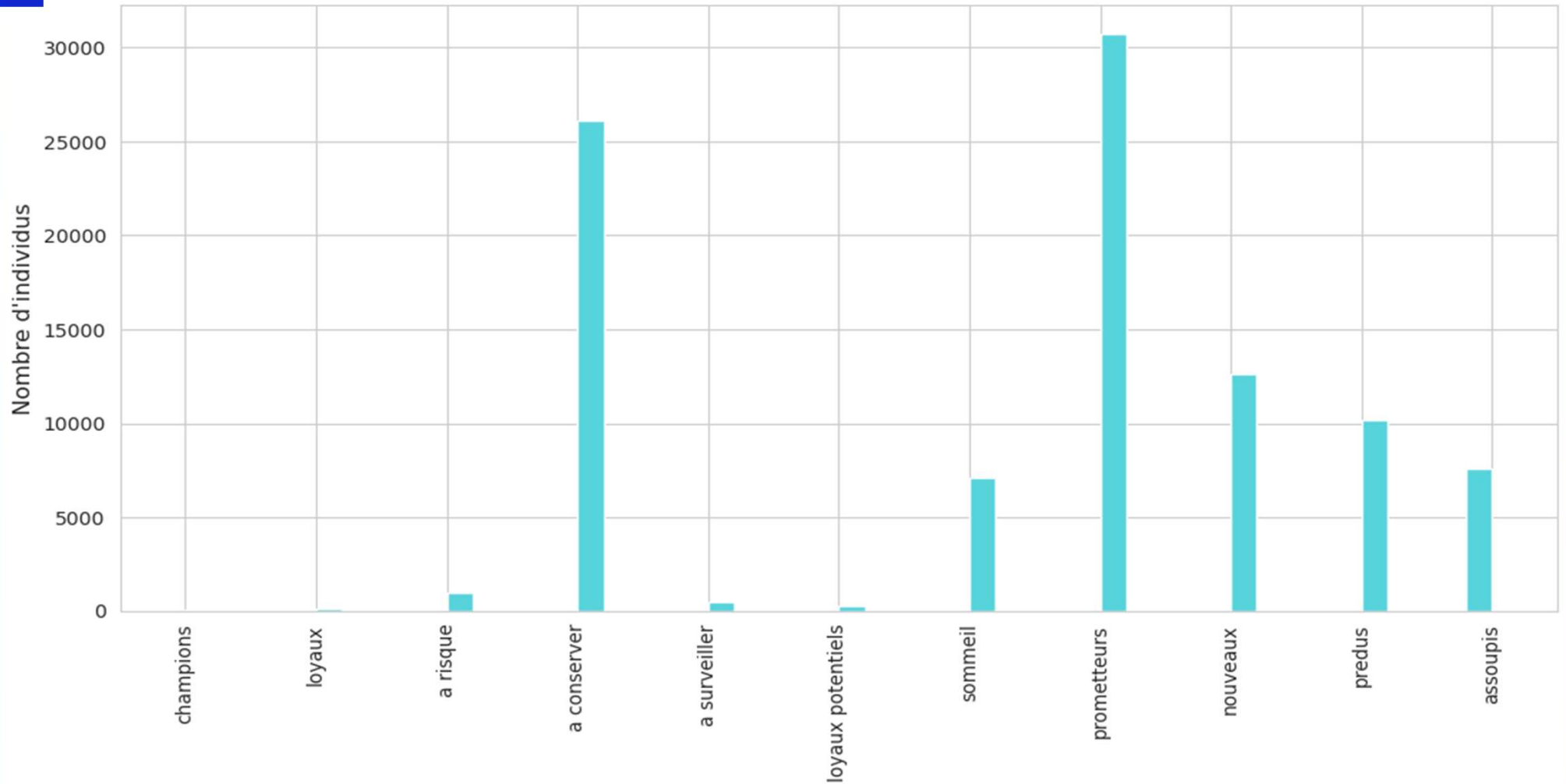
Distribution score M



Distribution score RFM

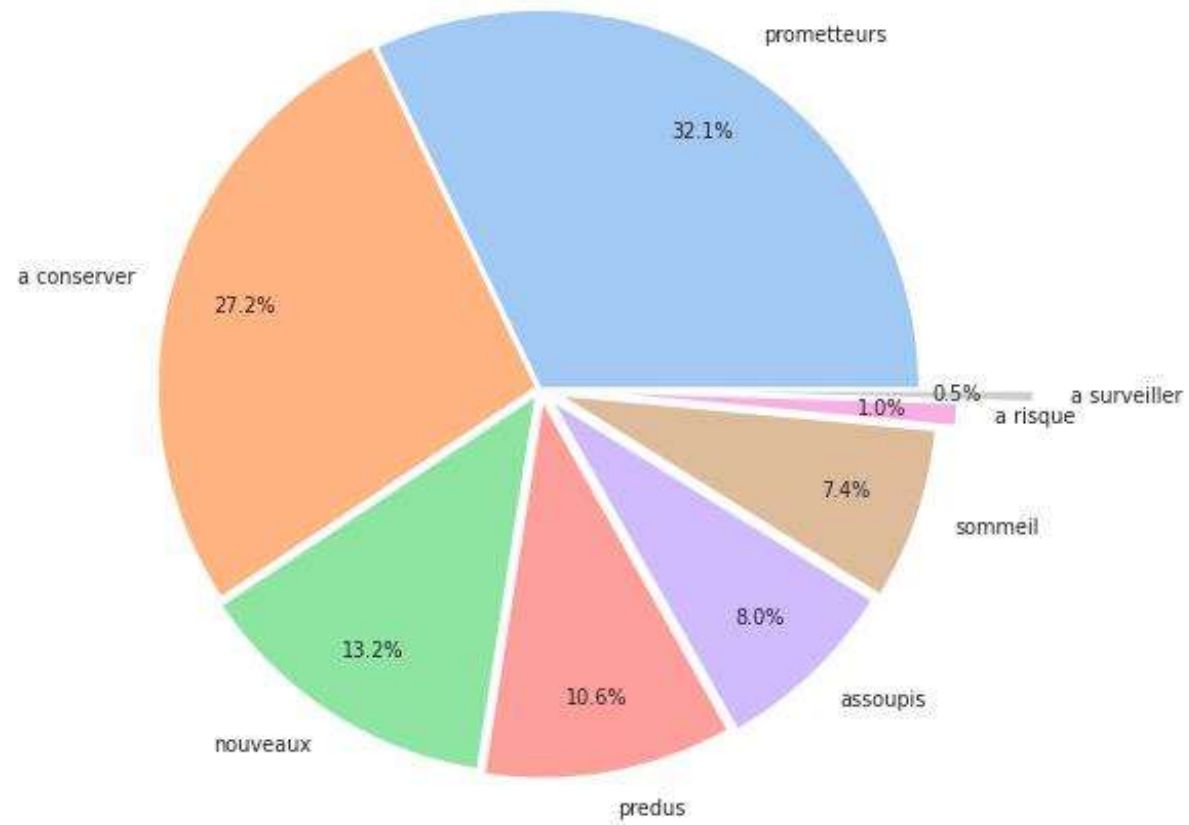


Distribution état RFM



\* Détail des seuils de classification en annexes

## Répartition des classes de clients

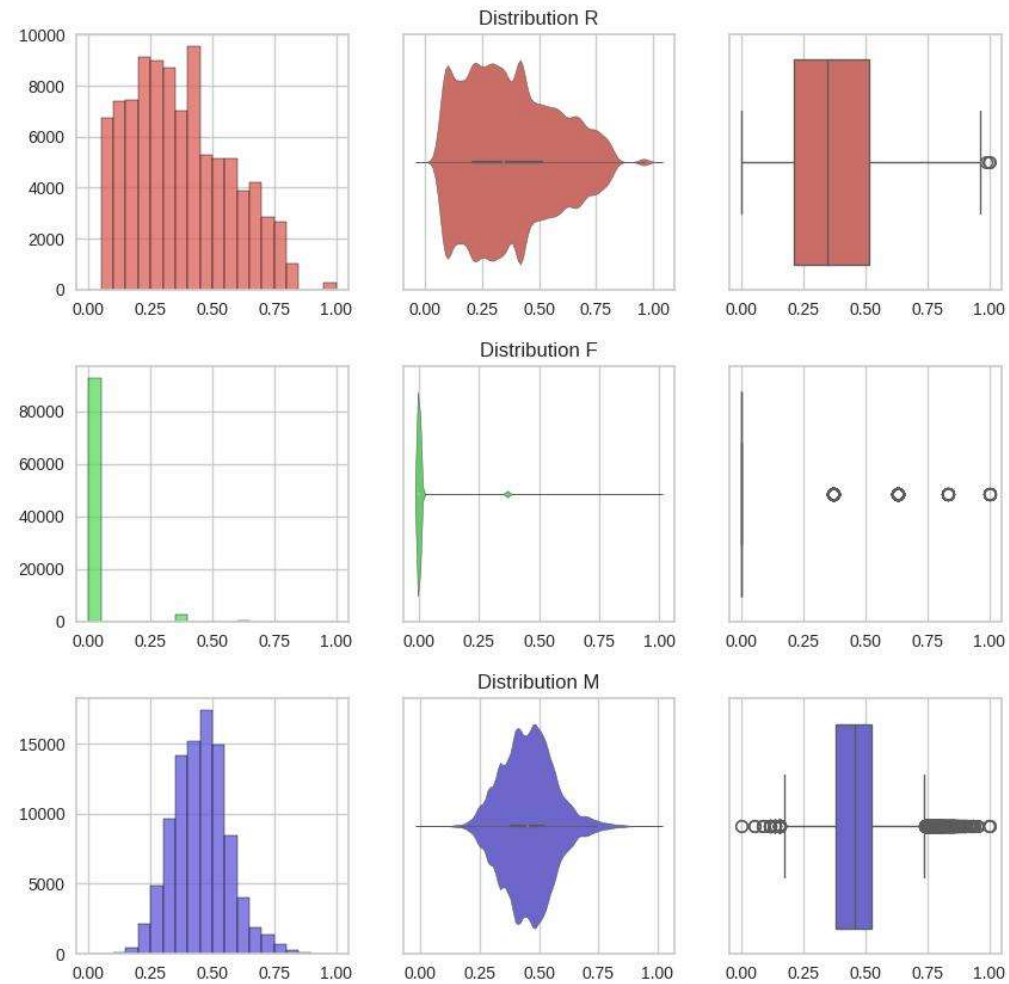


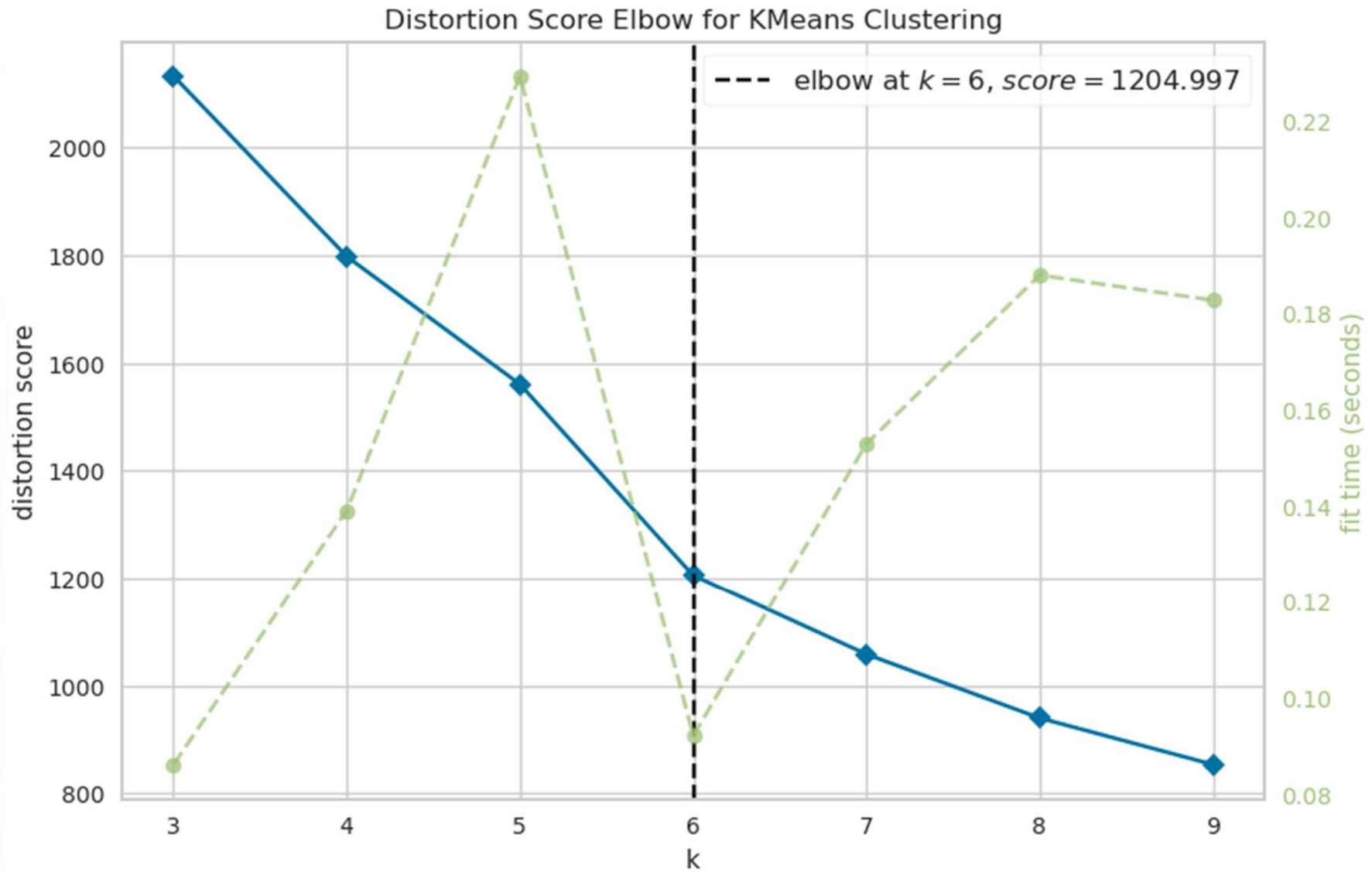


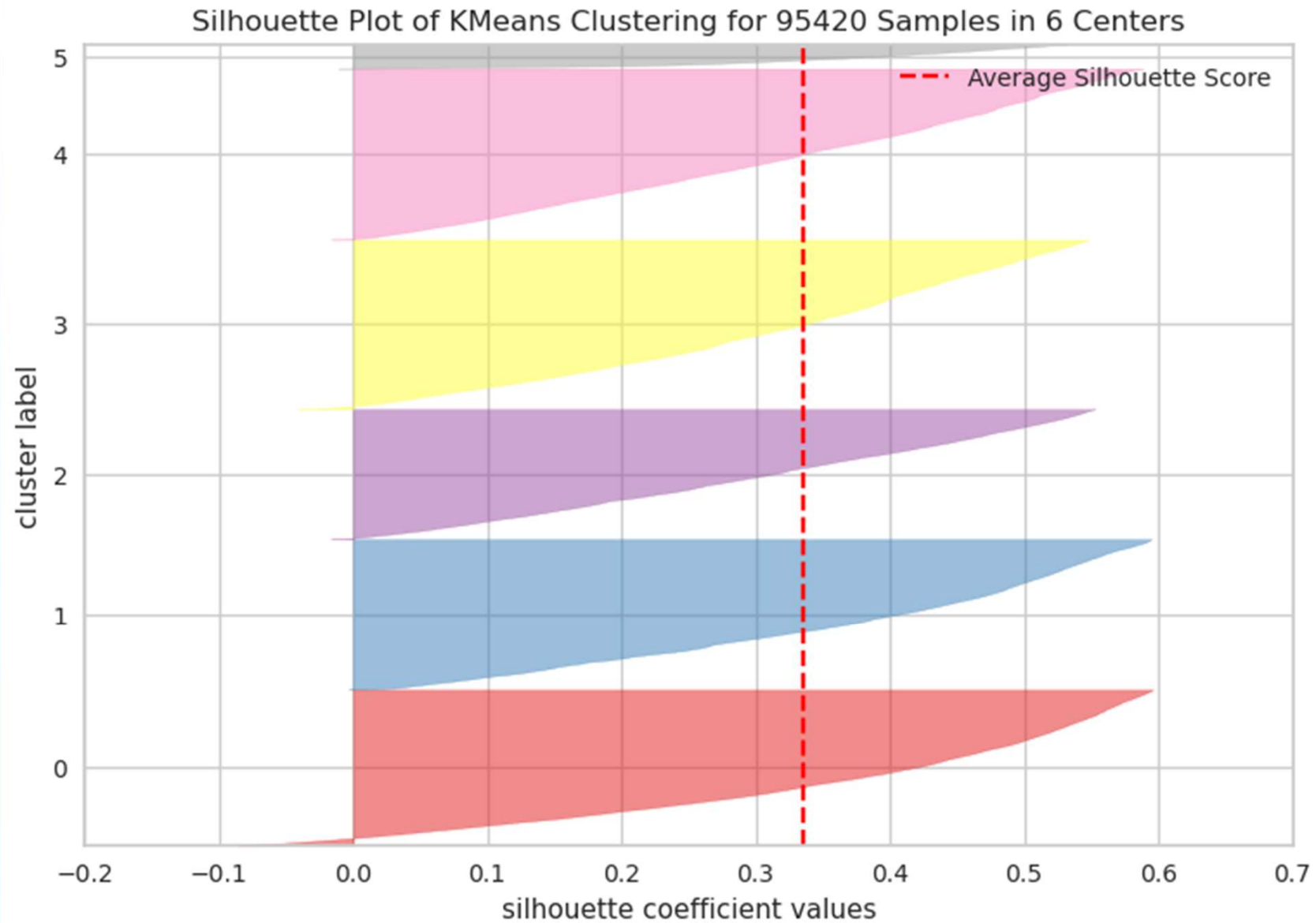
olist

Modélisation RFM Kmeans

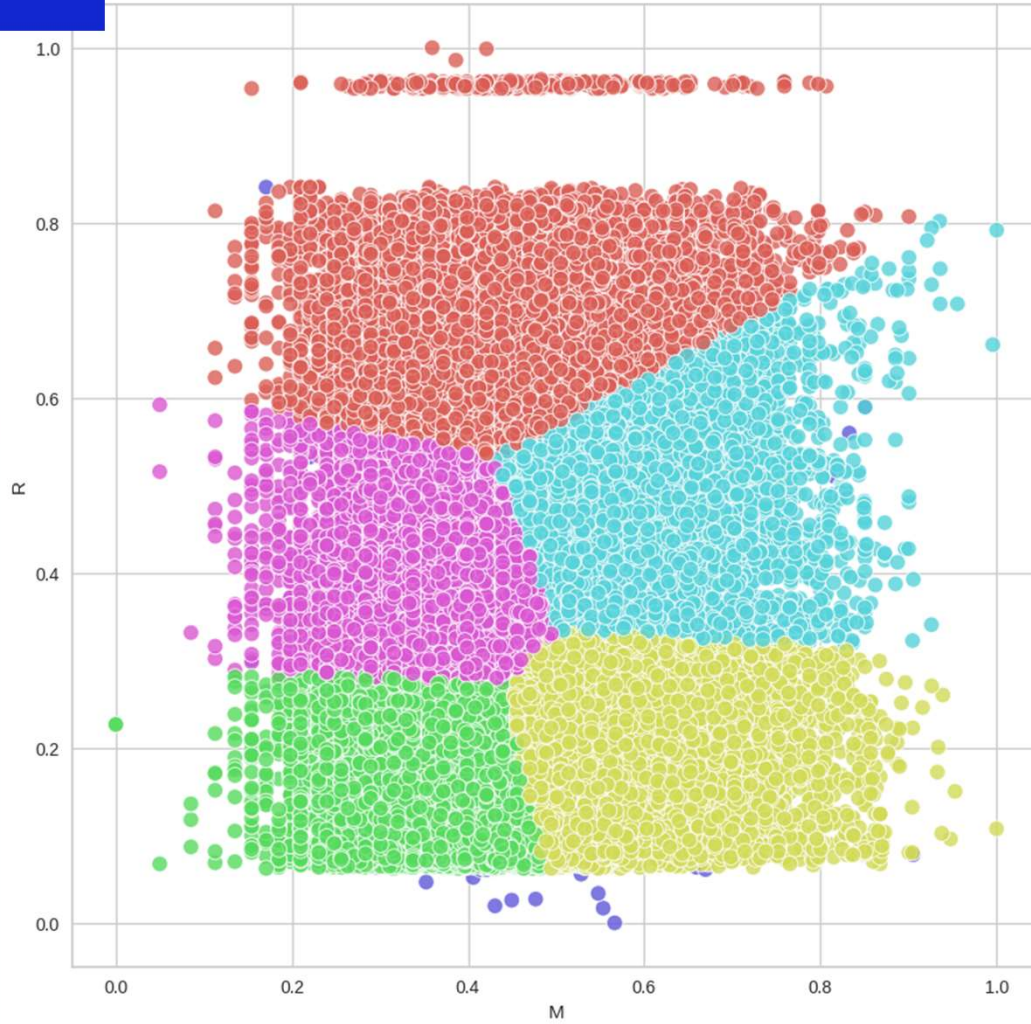
## Distributions des variable avec Log et MinMax



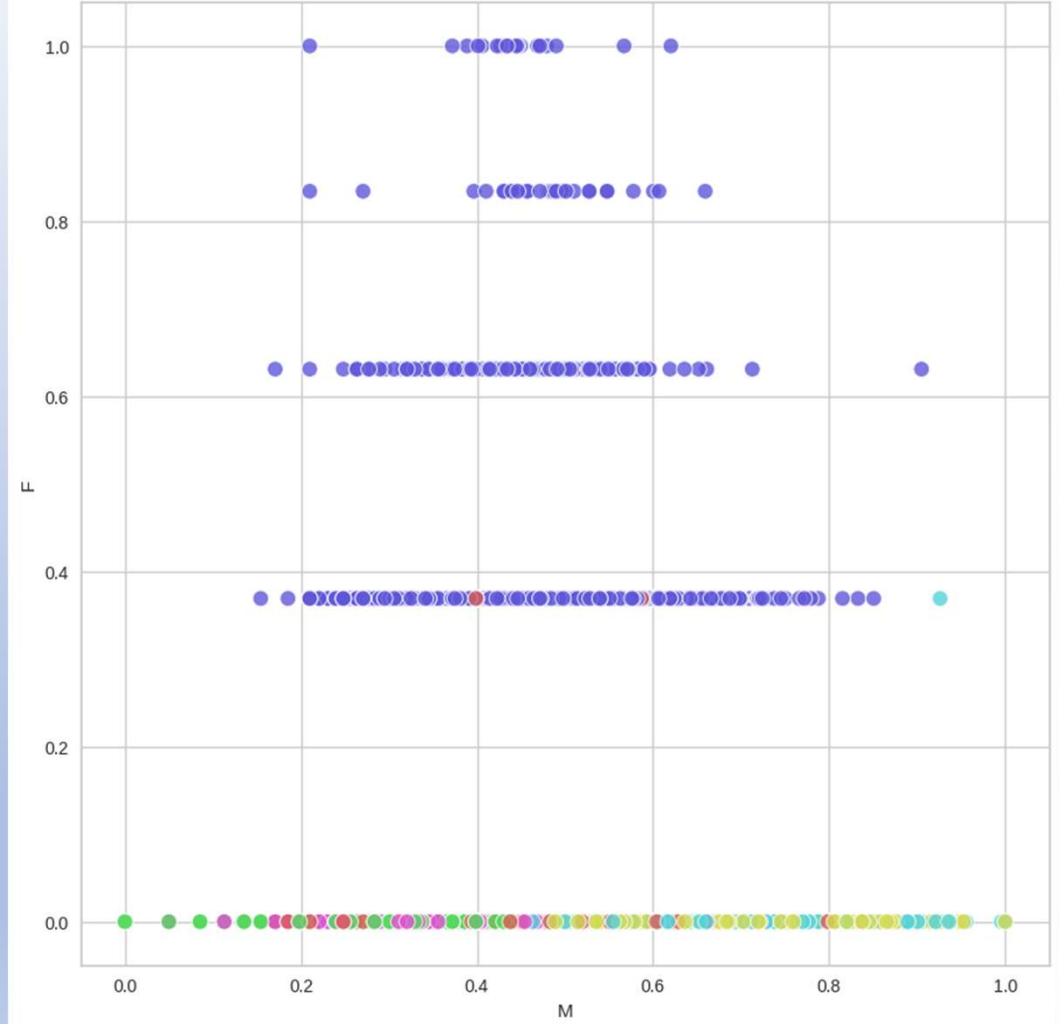




Kmeans - Clusters en fonction de la récence et du montant



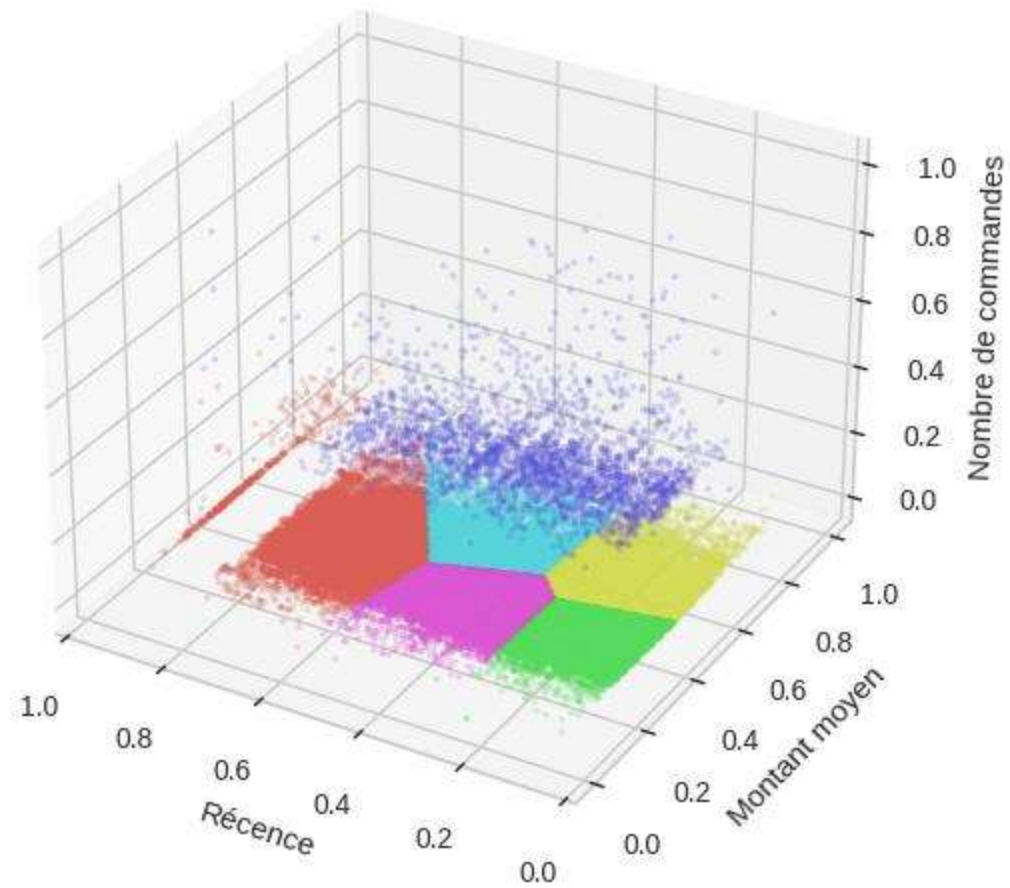
Kmeans - Clusters en fonction du nombre de commandes et du montant

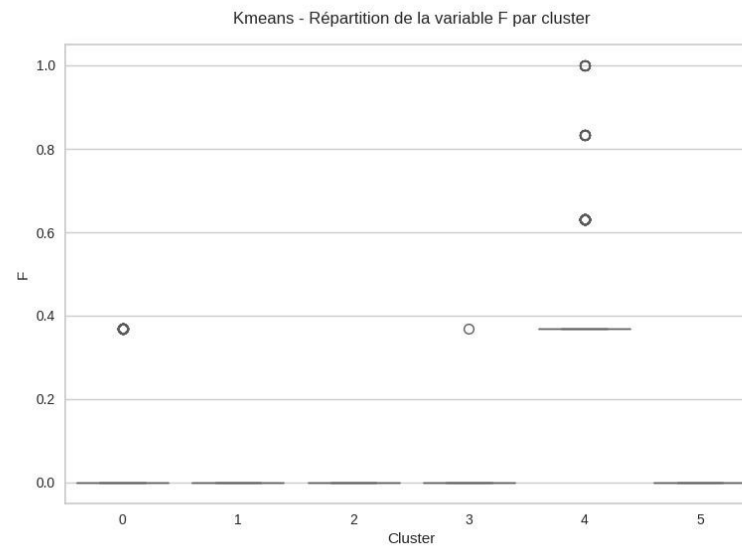
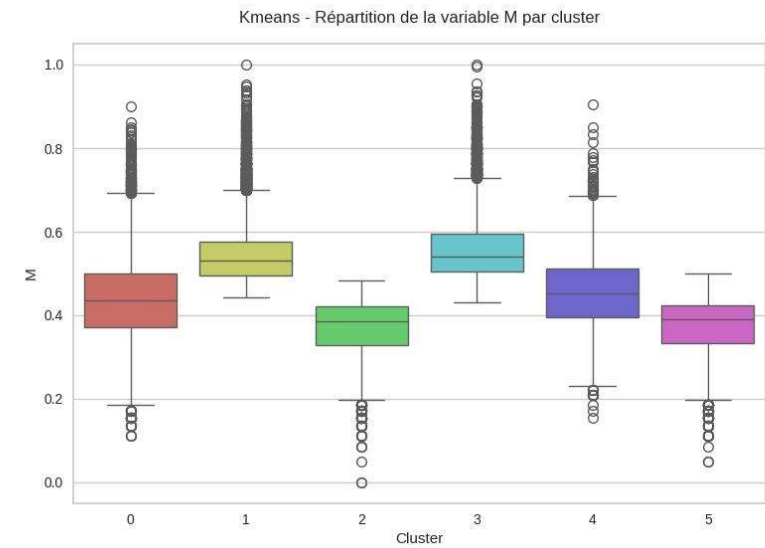
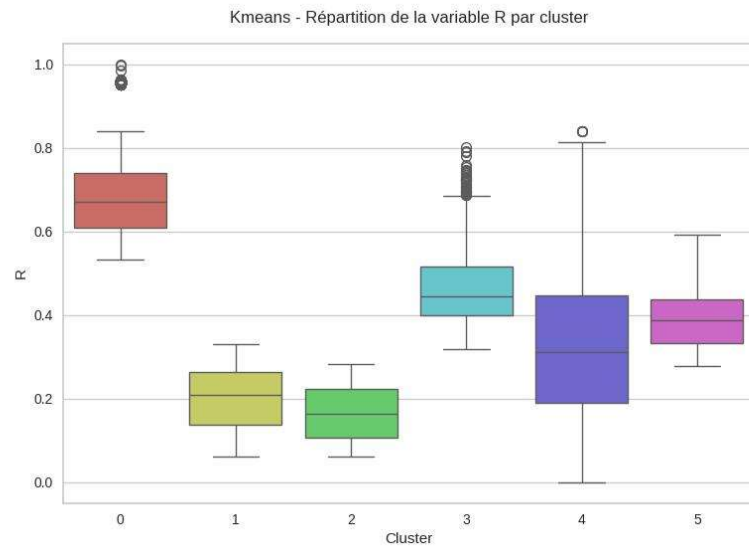


\* Détail des clusters en annexes<sup>21</sup>




## Kmeans - Visualisation RFM 3d





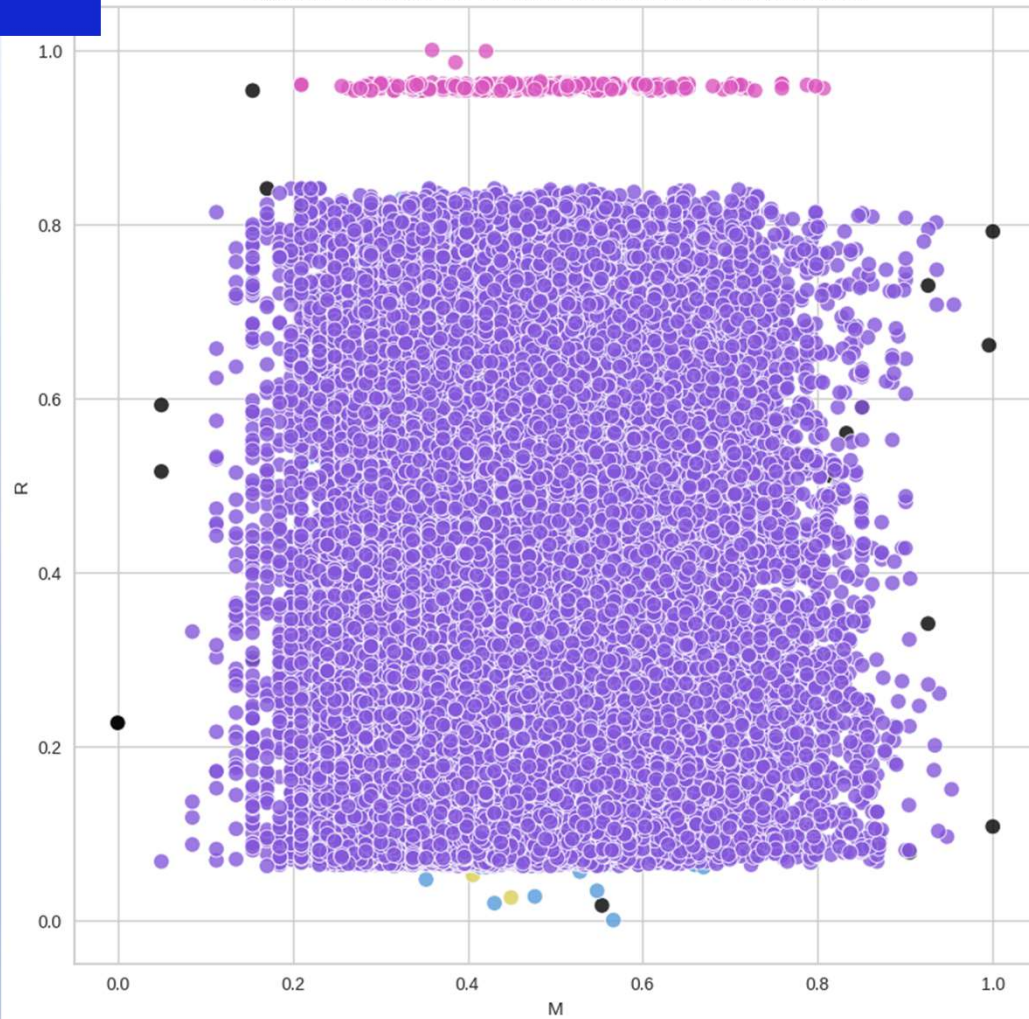
olist



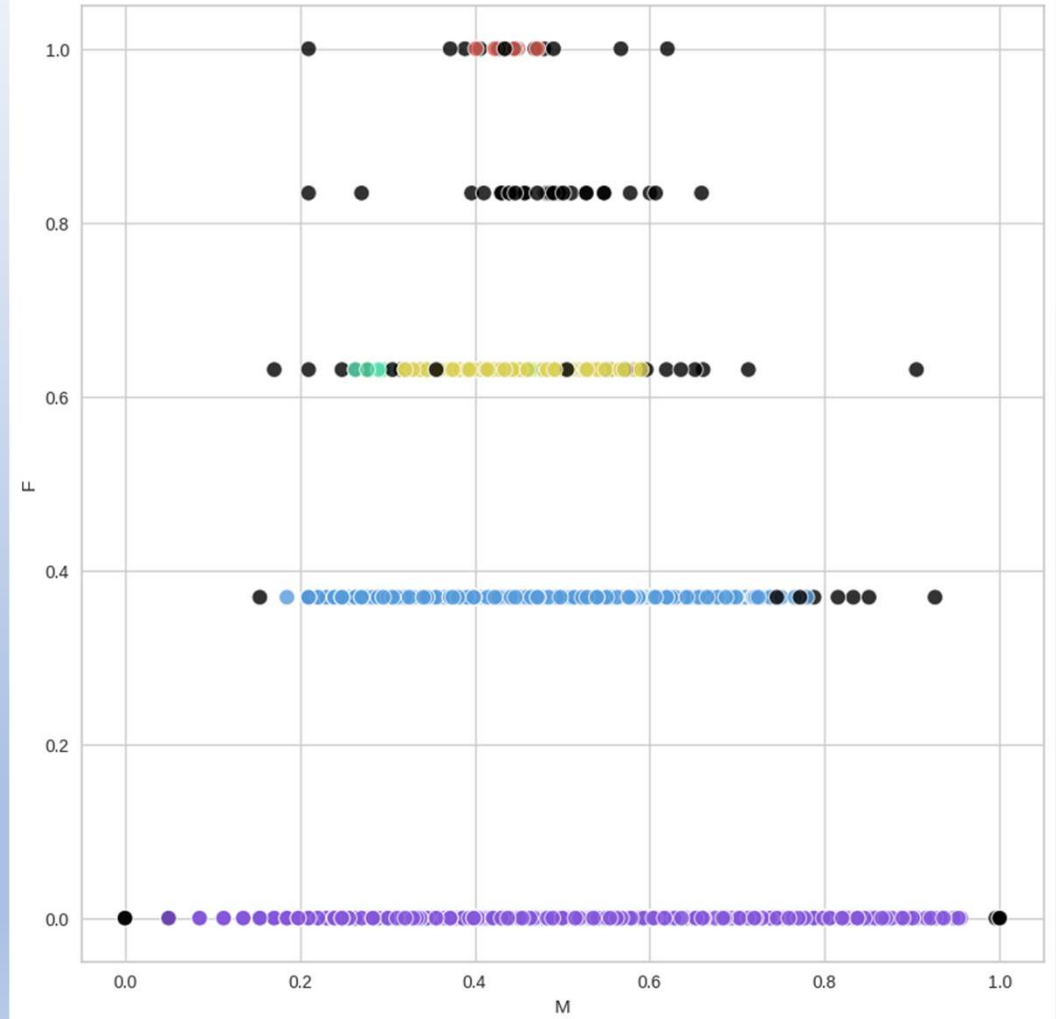
# Modélisation RFM Dbscan



DbScan - Clusters en fonction de la récence et du montant

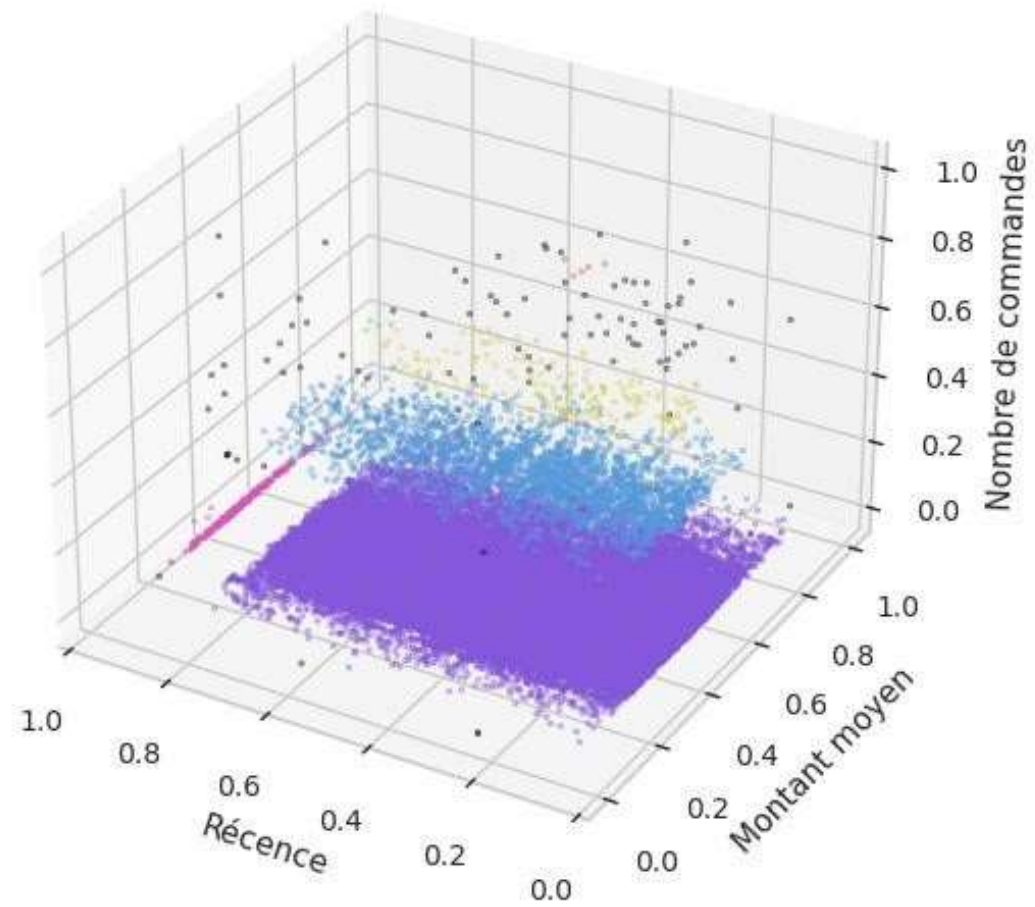


DbScan - Clusters en fonction du nombre de commandes et du montant



\* Détail des clusters en annexes 25

## DbScan - Visualisation RFM 3d



olist

Choix de la modélisation

K-means meilleure option

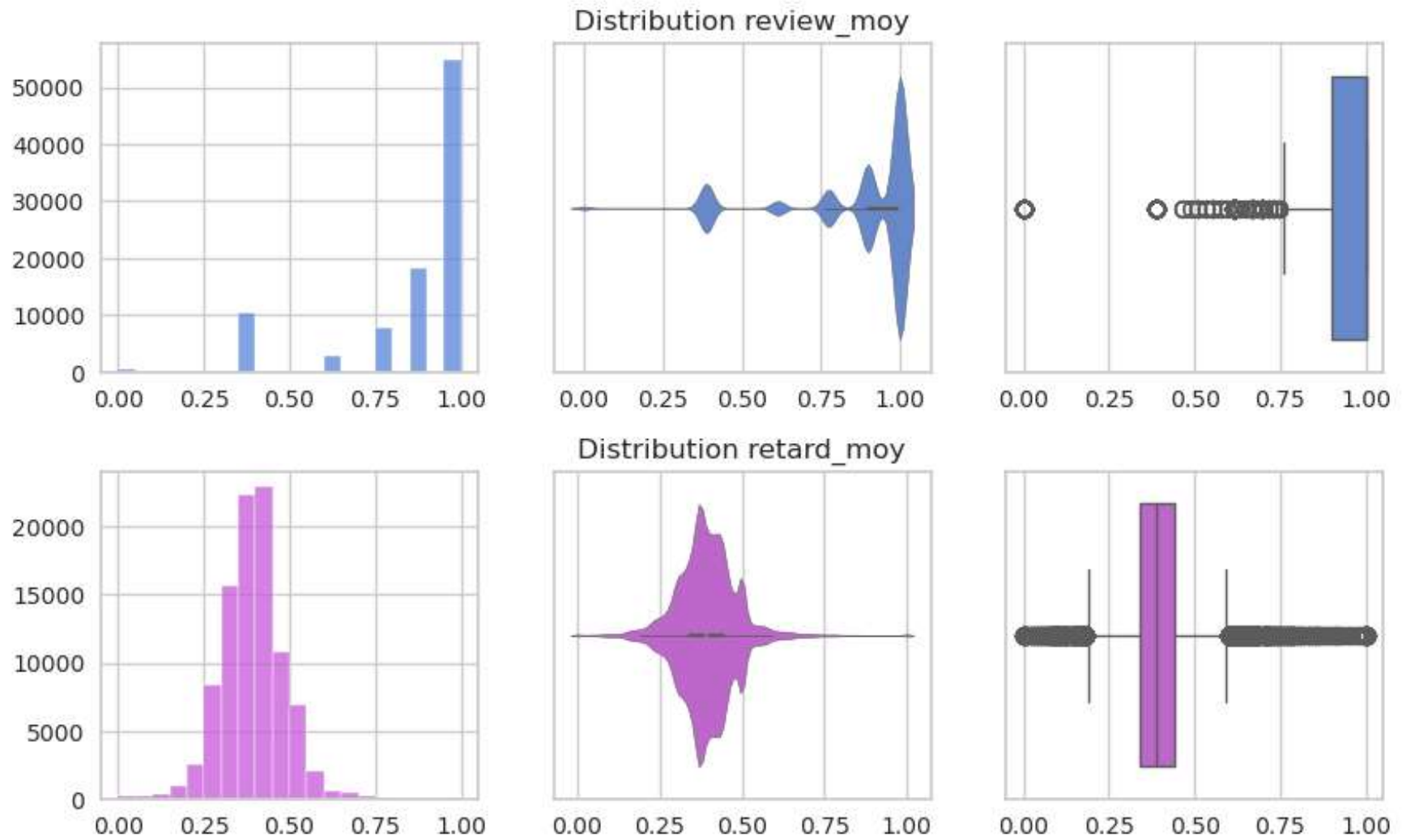
segmentation des clients  
avec une seule commande

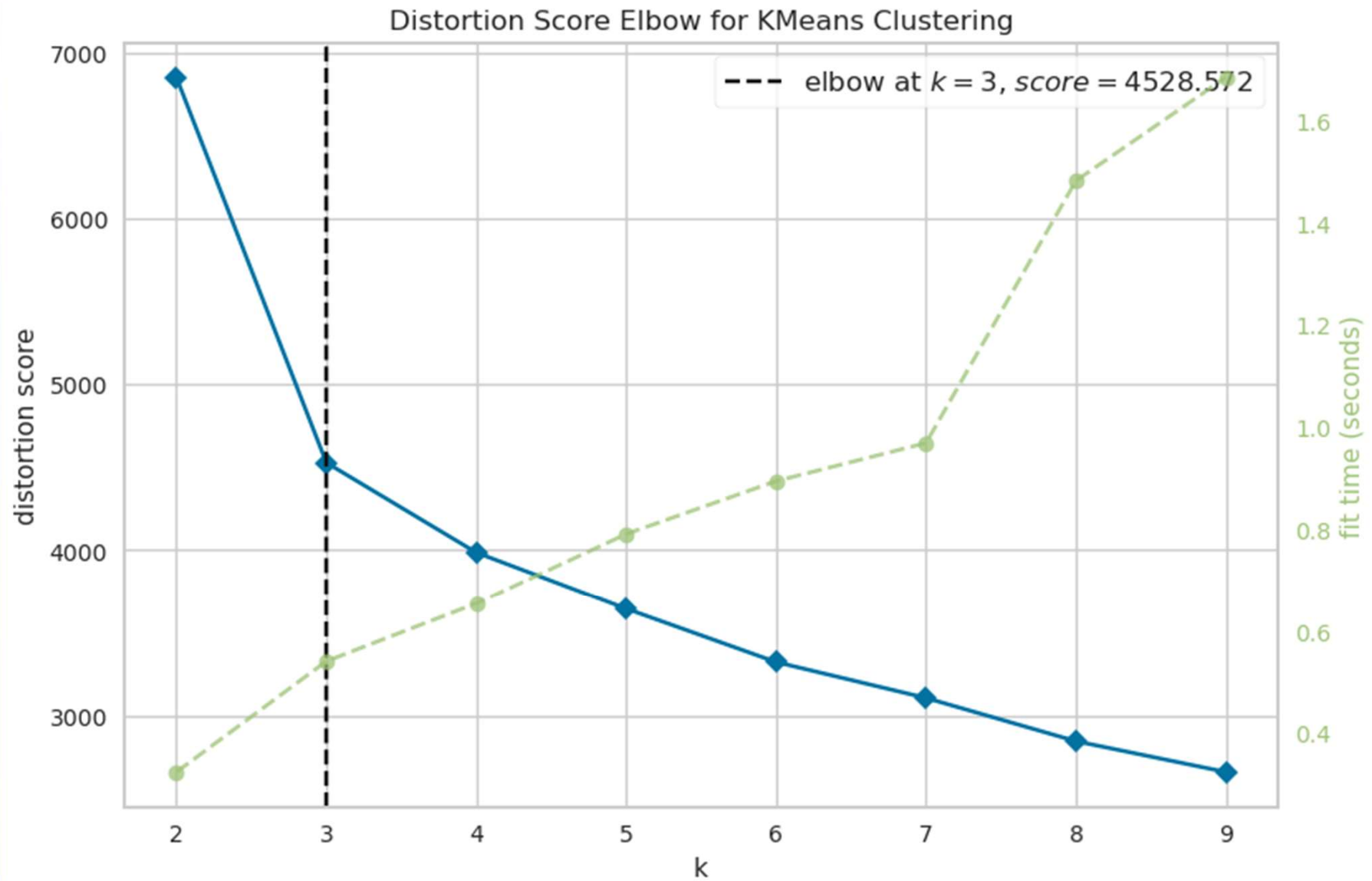


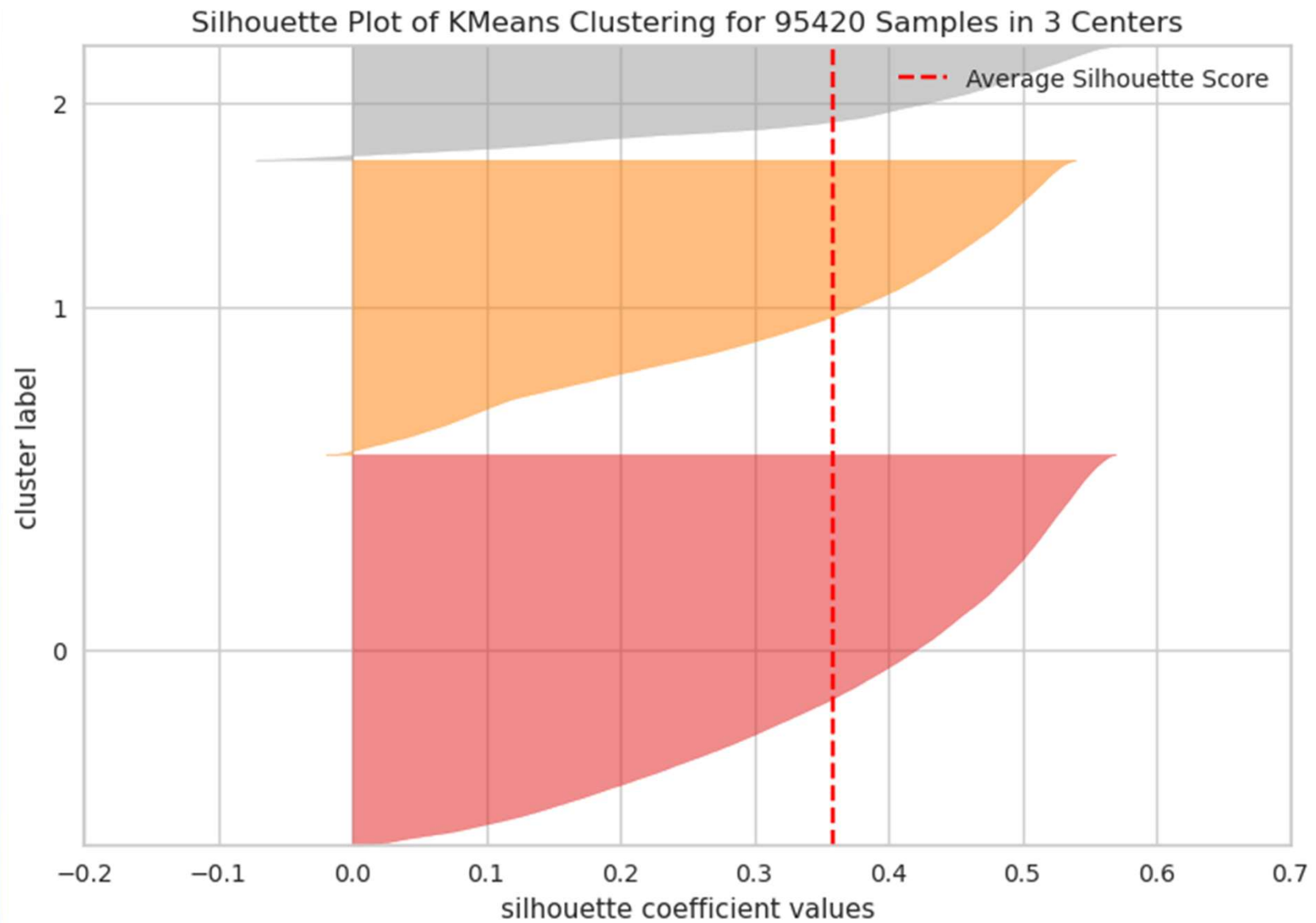
olist

Augmentation des features

## Distributions des variable avec Log et MinMax

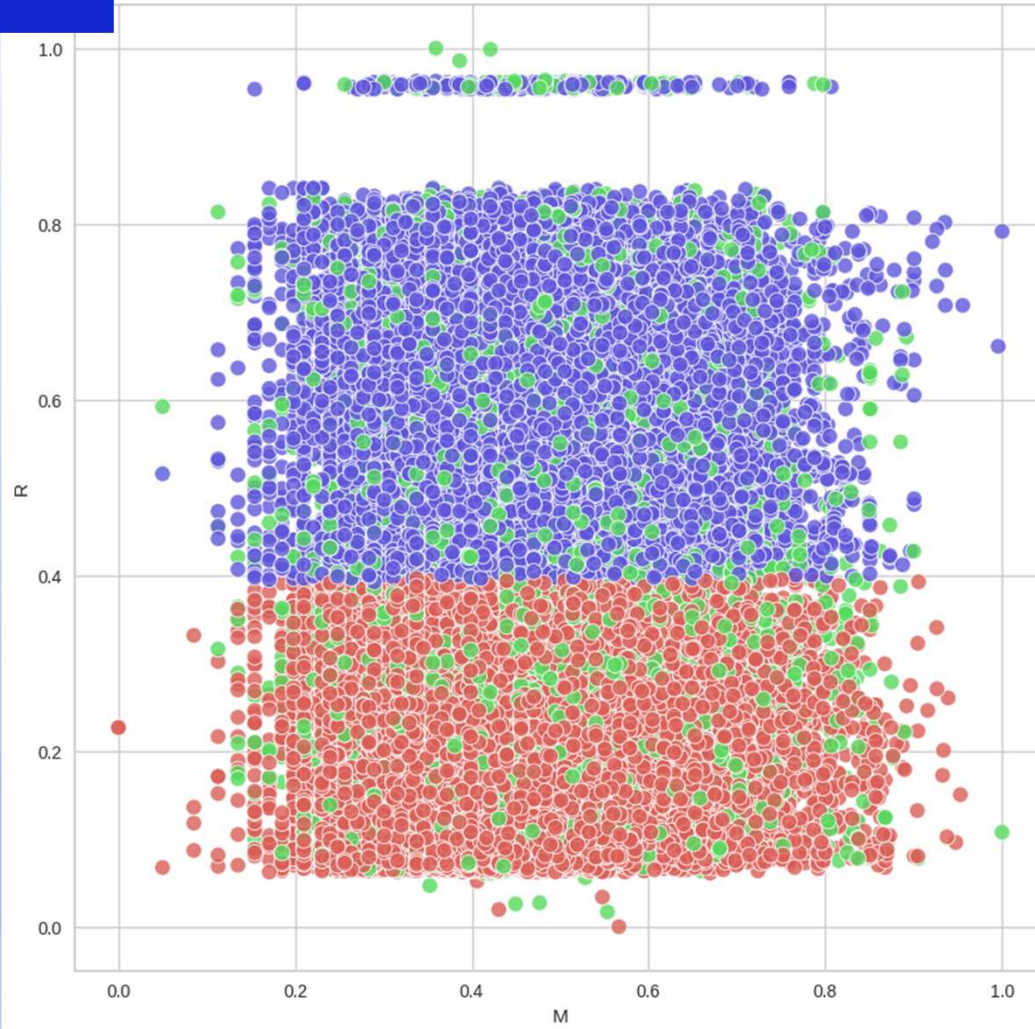




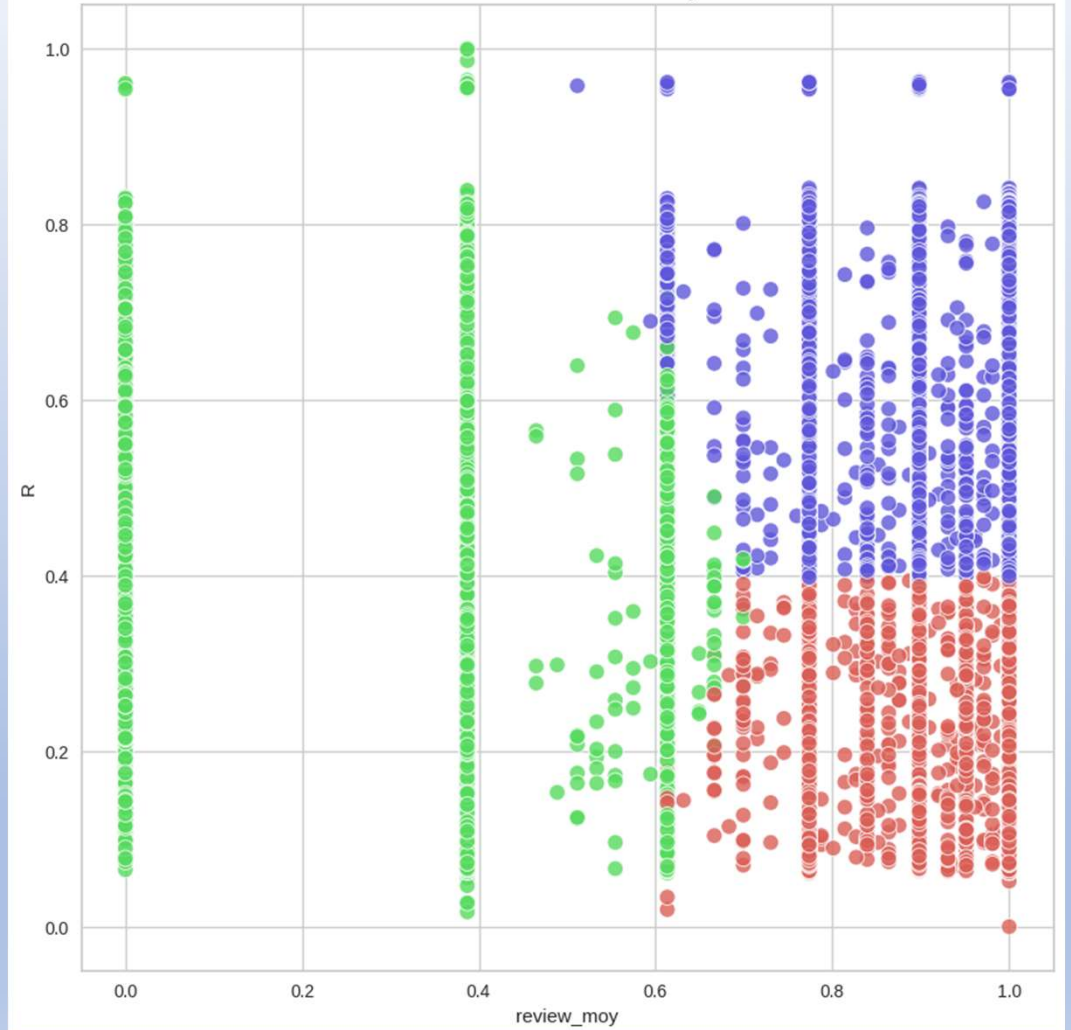




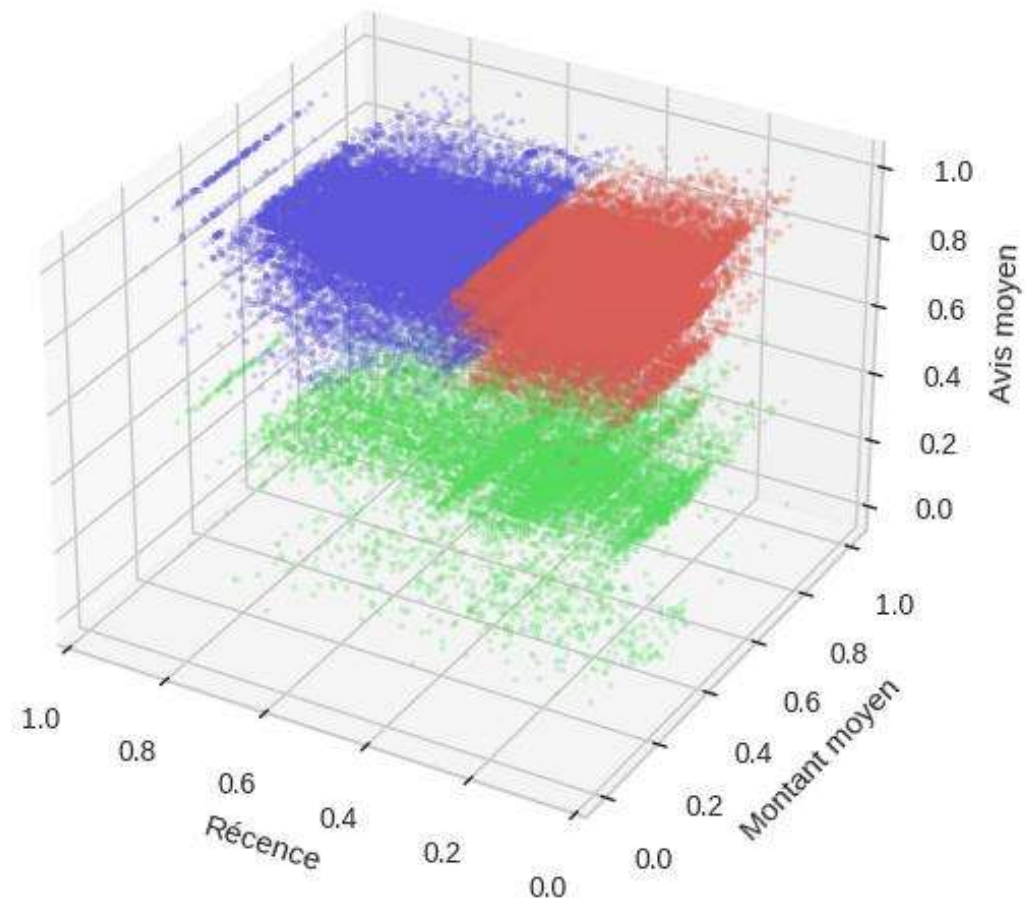
Kmeans - Clusters en fonction de la récence et du montant



Kmeans - Clusters en fonction de l'avis moyen et de la récence



## Kmeans - Visualisation 3d RFM + features



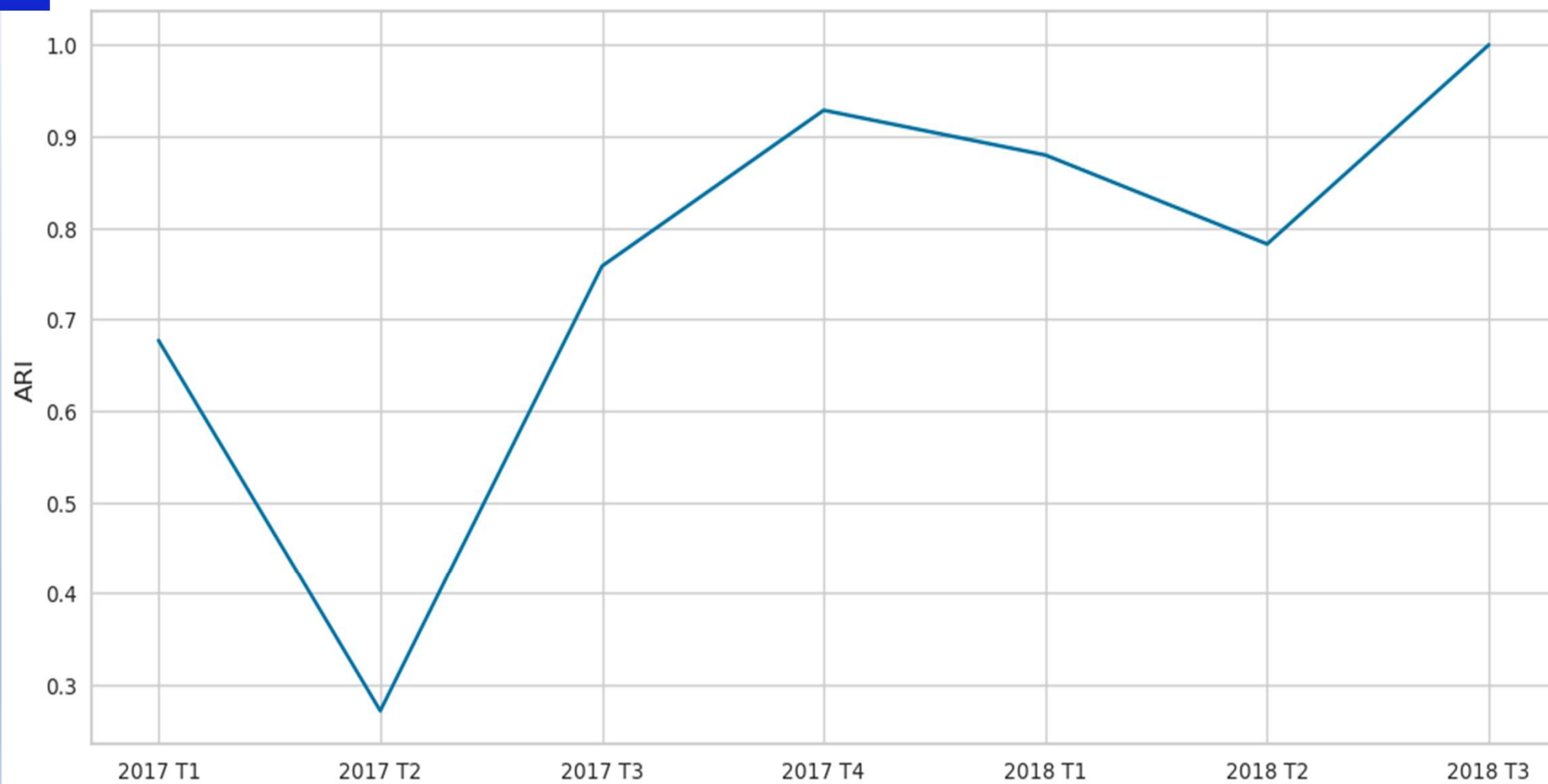
olist

Contrat de maintenance

# Adjusted Rand Index

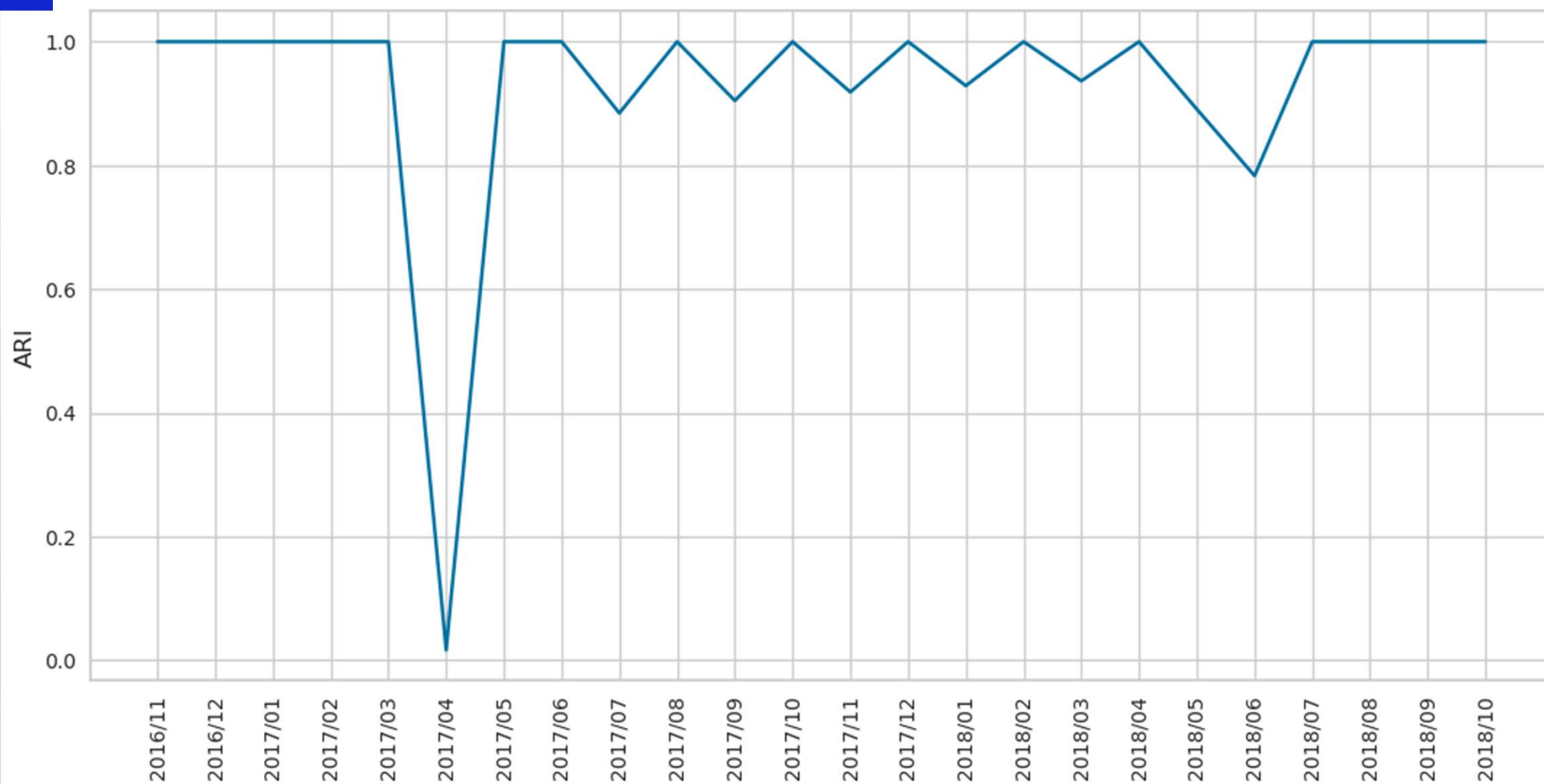
Mesure d'évolution des clusters dans le temps

Evolution du score Adjusted Rand Index

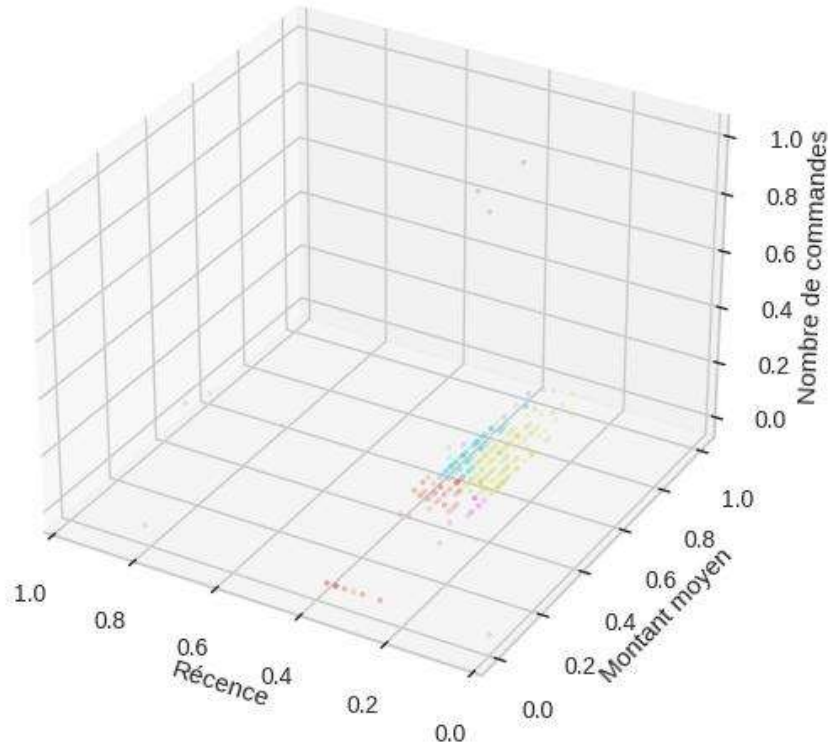




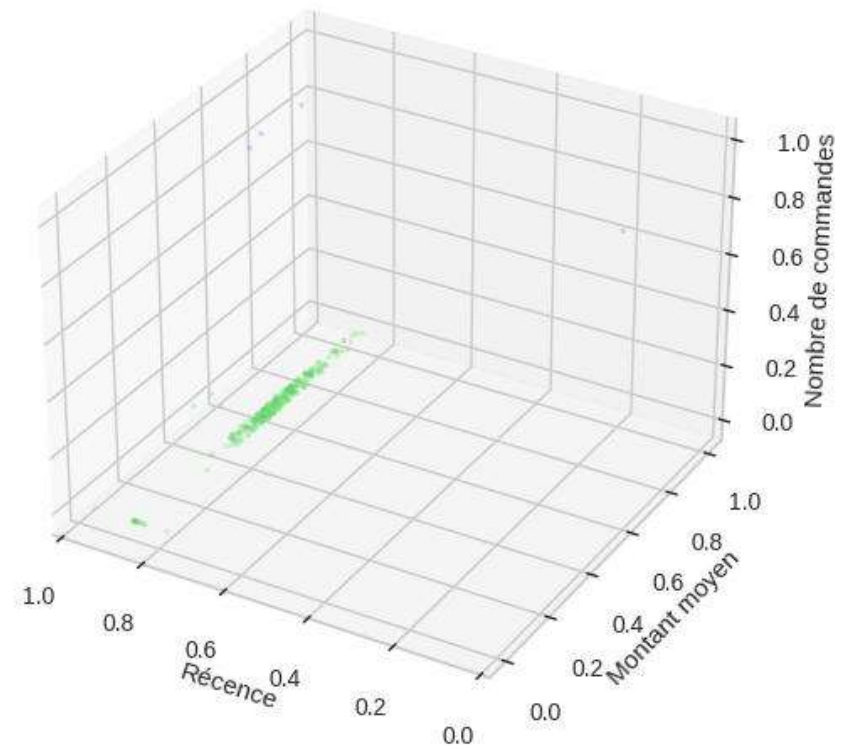
Evolution du score Adjusted Rand Index du modèle



Kmeans - Clasification RFM 3d du 11-2016



Kmeans - Clasification RFM 3d du 04-2017



Chute importante après 6 mois

Recommandation mise à jour mensuelle



olist

Annexes

# Seuils de classification Métier du score RFM

## Score R

60j > 5 points  
180j > 4 points >= 60j  
270j > 3 points >= 180j  
365j > 2 points >= 270j  
1 point >= 365j

## Score F

5 points >= 5 commandes  
4 points = 4 commandes  
3 points = 3 commandes  
2 points = 2 commandes  
1 point = 1 commande

## Score M

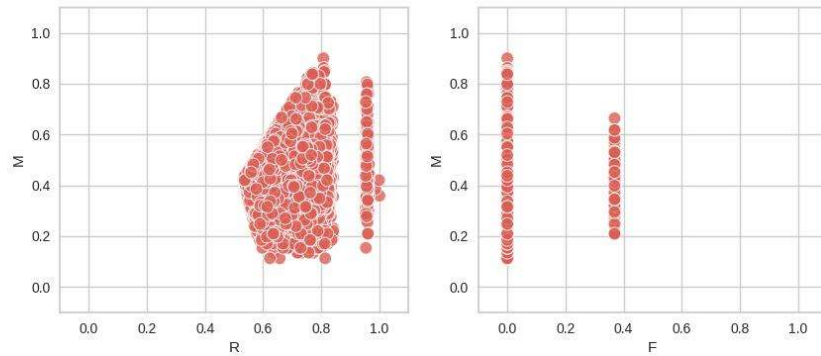
20000 > 5 points  
10000 > 4 points >= 20000  
5000 > 3 points >= 10000  
2000 > 2 points >= 5000  
1 point >= 2000

# Seuils de classification Métier RFM

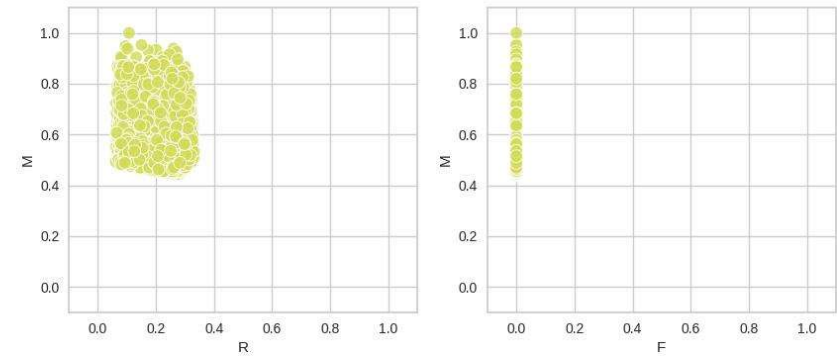
```
classification['champions'] = [555,554,544,545,454,455,445]
classification['loyaux'] = [543,444,435,355,354,345,344,335]
classification['loyaux_potentiels'] = [553,552,551,541,542,533,532,531,453,452,451,442,441,433,432,431,423,353,352,351,342,341,333,323]
classification['nouveaux'] = [512,511,422,421,412,411,311]
classification['prometteurs'] = [525,524,523,522,521,515,514,513,425,424,415,414,413,315,314,313]
classification['a_surveiller'] = [535,534,443,434,343,334,325,324]
classification['a_conserver'] = [155,154,144,214,215,115,114,113]
classification['assoupis'] = [331,321,312,221,213]
classification['a_risque'] = [255,254,245,244,253,252,243,242,235,234,225,224,235,234,225,224,153,152,145,143,142,135,134,133,125,124]
classification['sommeil'] = [332,322,231,241,251,233,232,223,222,132,123,122,212,211]
classification['predus'] = [111,112,121,131,141,151]
```

Les séries de chiffres représentent les scores de RFM de chaque classe

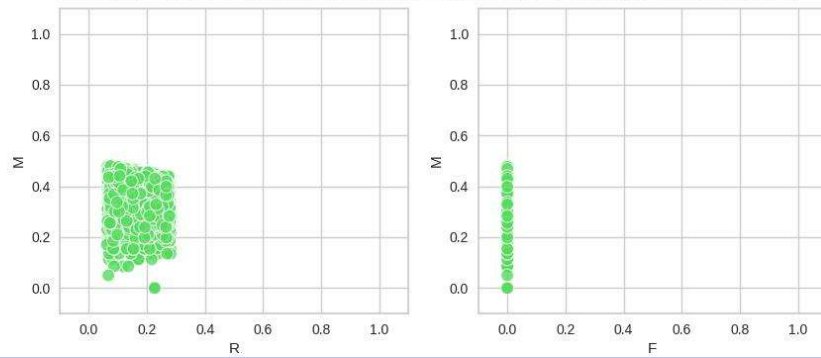
Kmeans - Cluster 0 - montant en fonction de la récence et du nombre de commandes



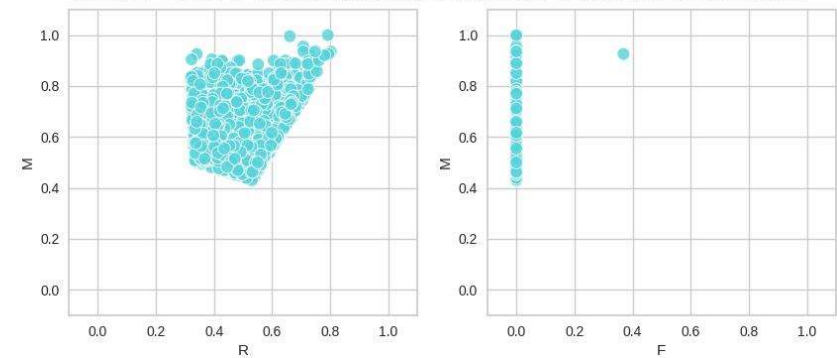
Kmeans - Cluster 1 - montant en fonction de la récence et du nombre de commandes



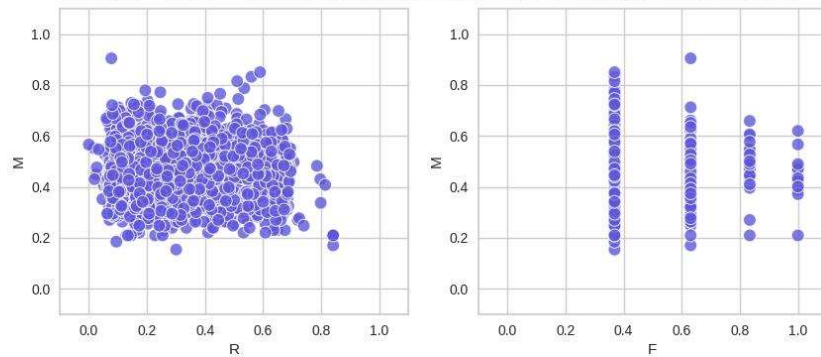
Kmeans - Cluster 2 - montant en fonction de la récence et du nombre de commandes



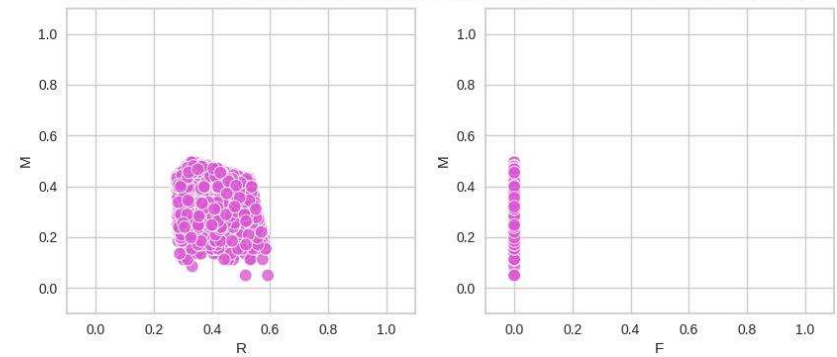
Kmeans - Cluster 3 - montant en fonction de la récence et du nombre de commandes



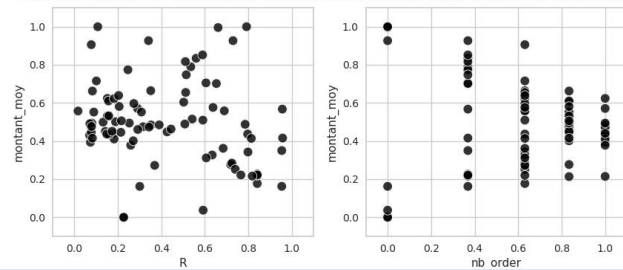
Kmeans - Cluster 4 - montant en fonction de la récence et du nombre de commandes



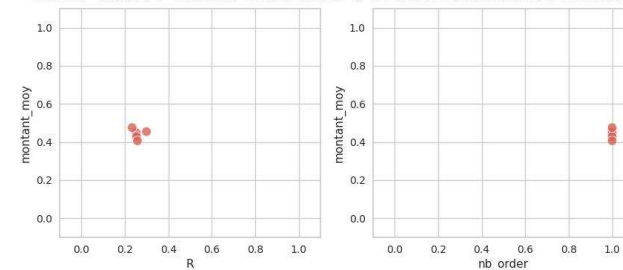
Kmeans - Cluster 5 - montant en fonction de la récence et du nombre de commandes



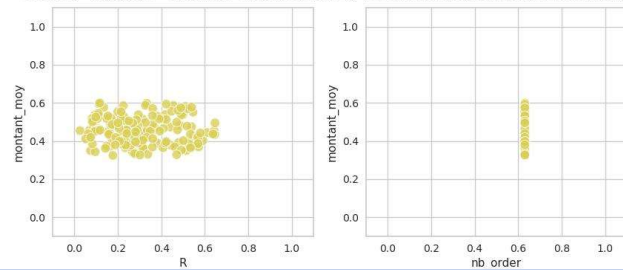
Dbscan - Cluster -1 - montant en fonction de la récence et du nombre de commandes



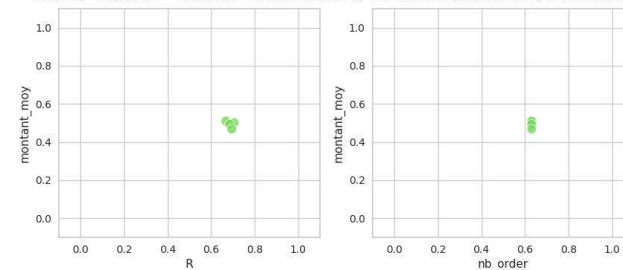
Dbscan - Cluster 0 - montant en fonction de la récence et du nombre de commandes



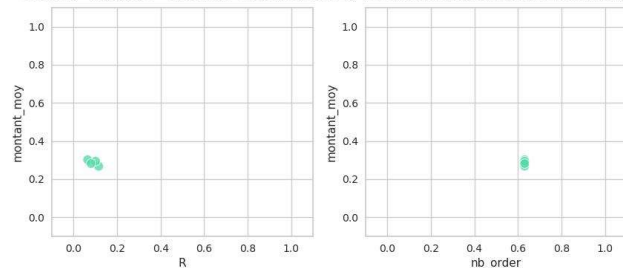
Dbscan - Cluster 1 - montant en fonction de la récence et du nombre de commandes



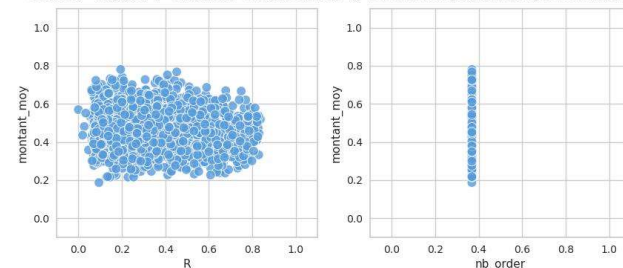
Dbscan - Cluster 2 - montant en fonction de la récence et du nombre de commandes



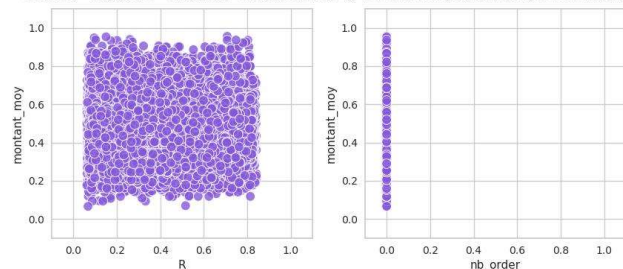
Dbscan - Cluster 3 - montant en fonction de la récence et du nombre de commandes



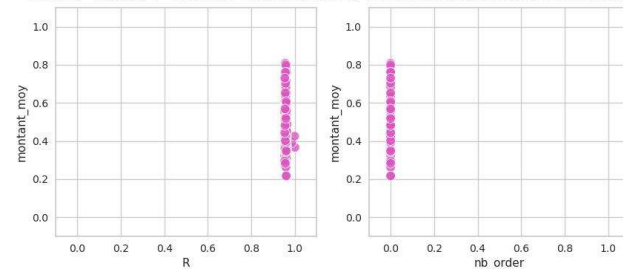
Dbscan - Cluster 4 - montant en fonction de la récence et du nombre de commandes



Dbscan - Cluster 5 - montant en fonction de la récence et du nombre de commandes



Dbscan - Cluster 6 - montant en fonction de la récence et du nombre de commandes



Merci pour votre attention