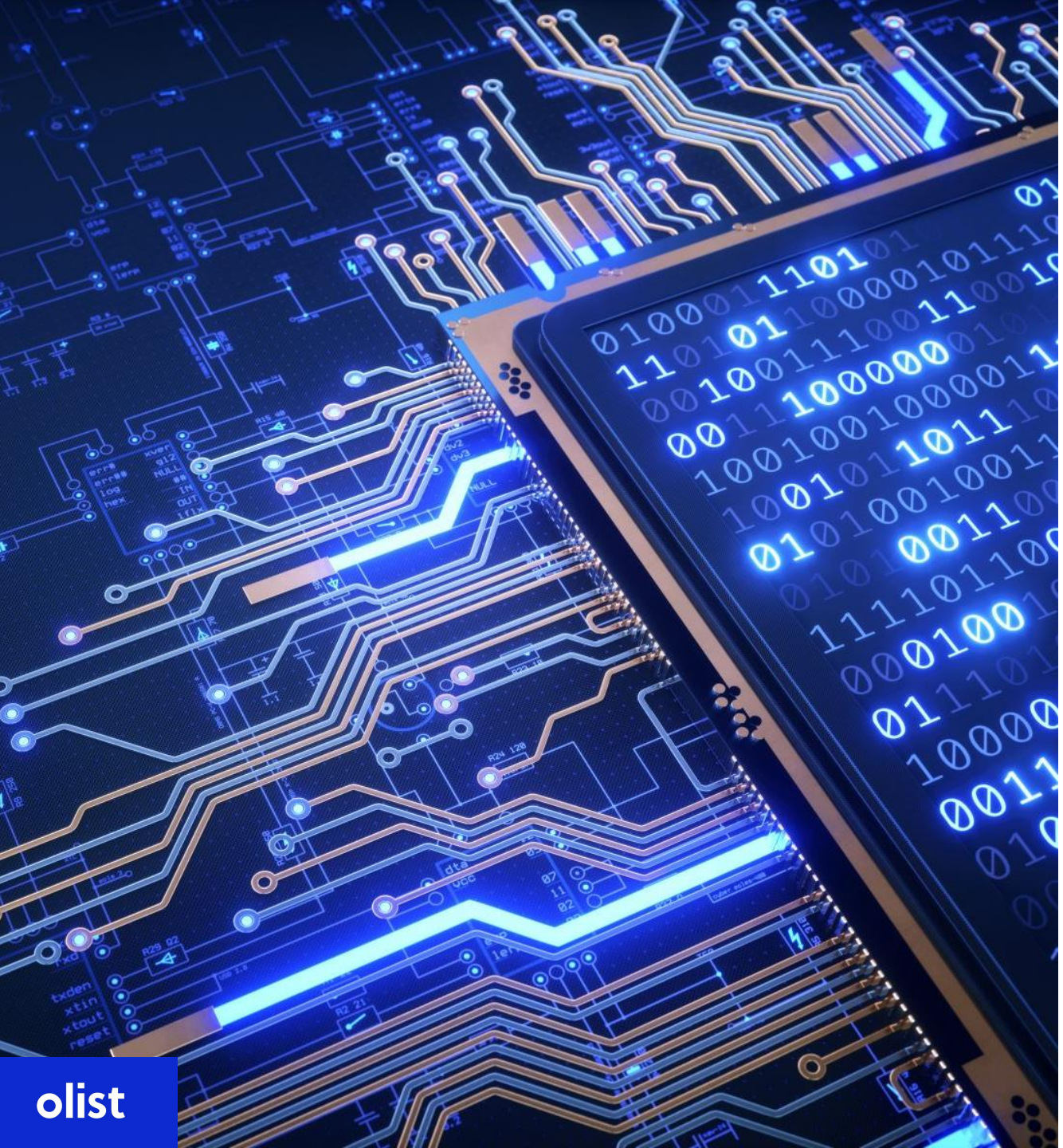
olist

# Segmentation commerciale

SITE DE E-COMMERCE « OLIST »



olist

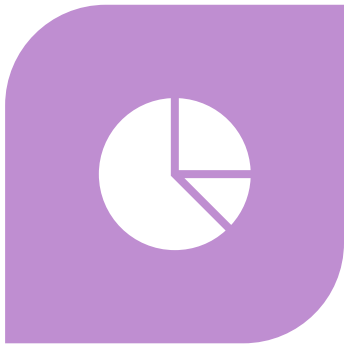
# Sommaire

# Contexte

olist

# Contexte

SITE DE E-COMMERCE

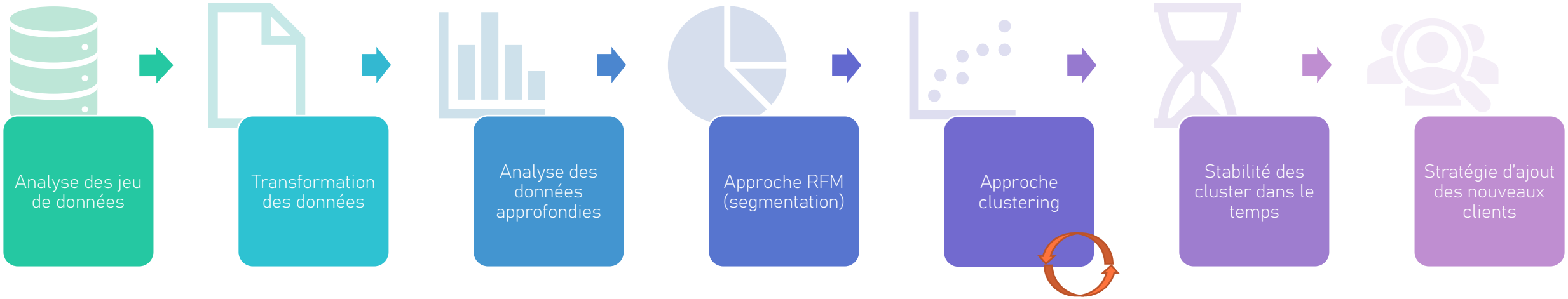


SEGMENTATION COMMERCIALE

CAMPAGNE DE COMMUNICATION

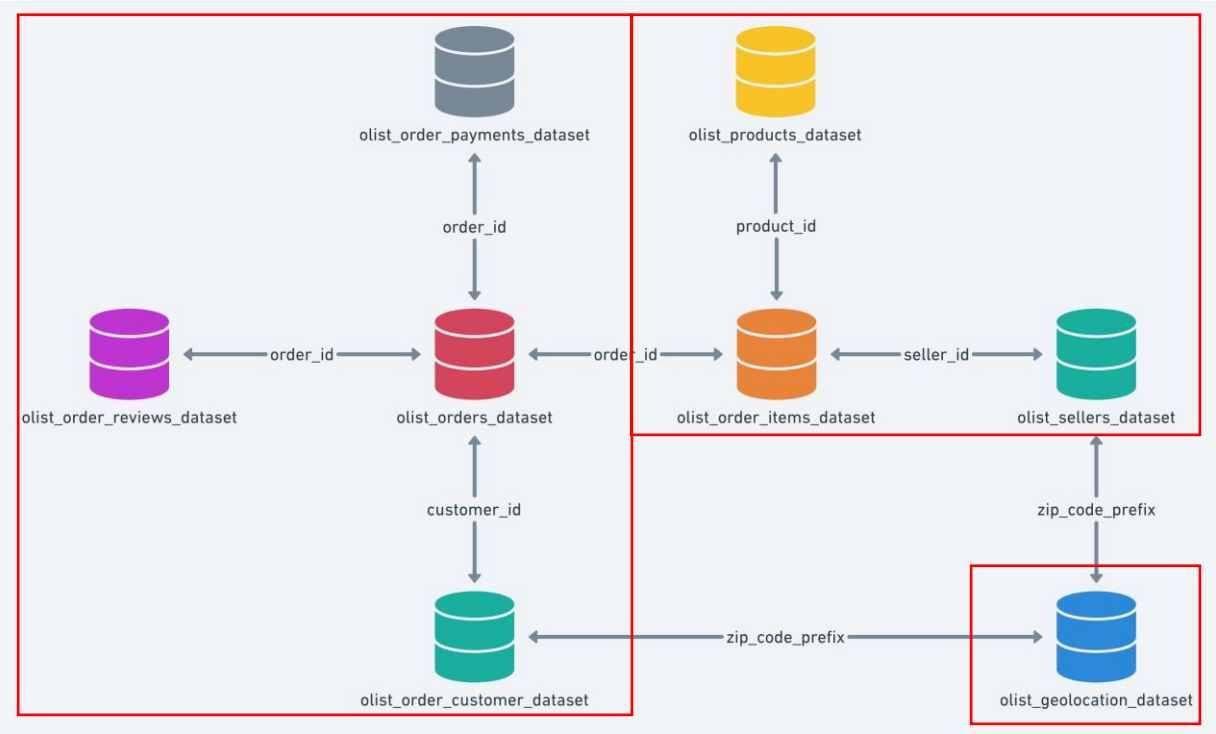
olist

# Approche effectuée



# Présentation / préparation des jeux de données

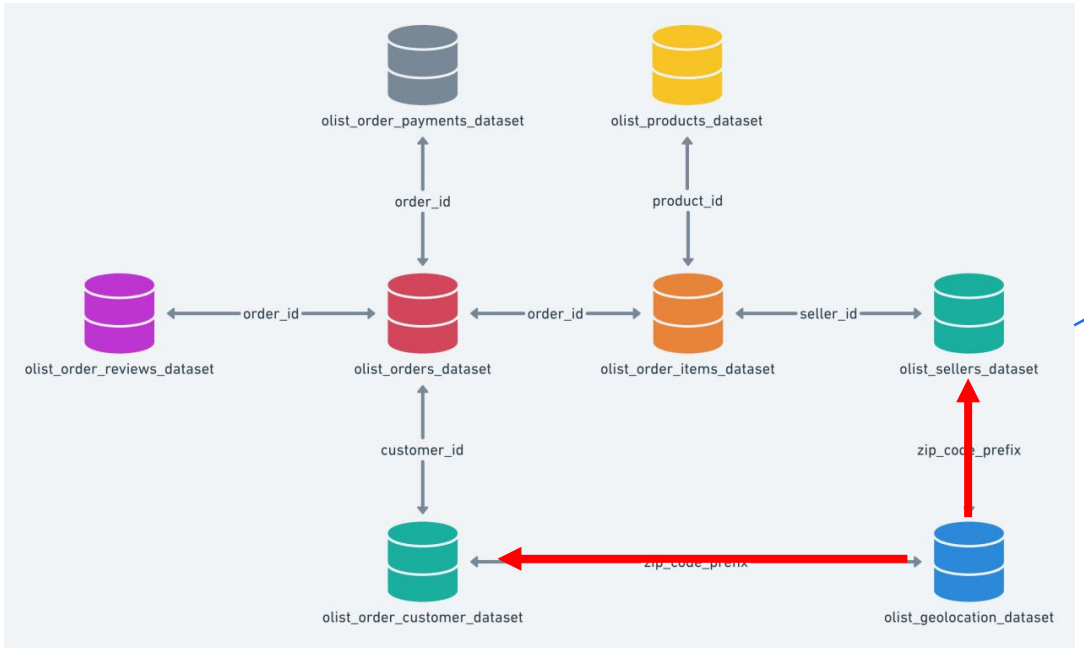
olist

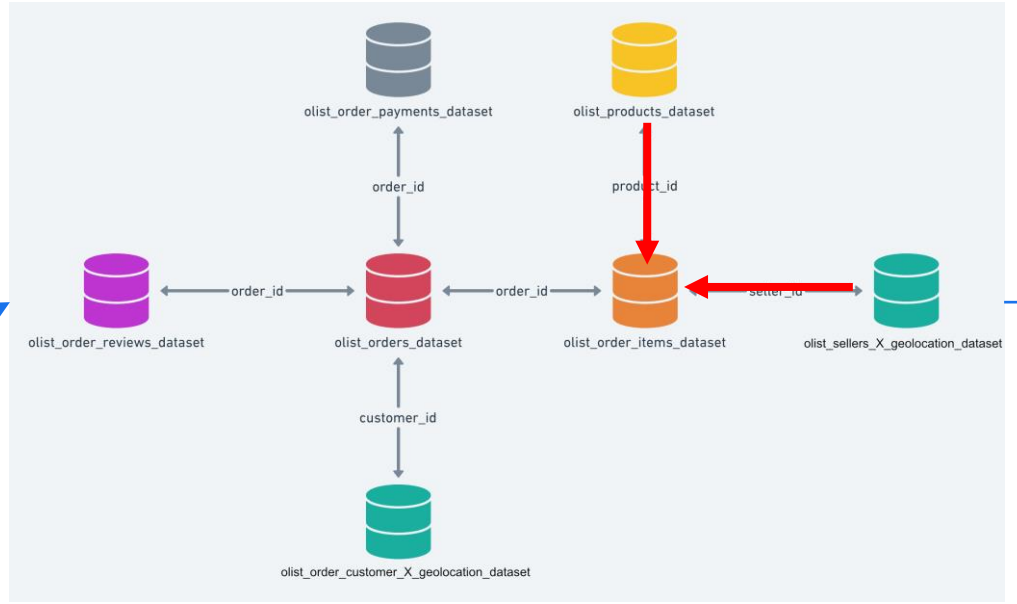


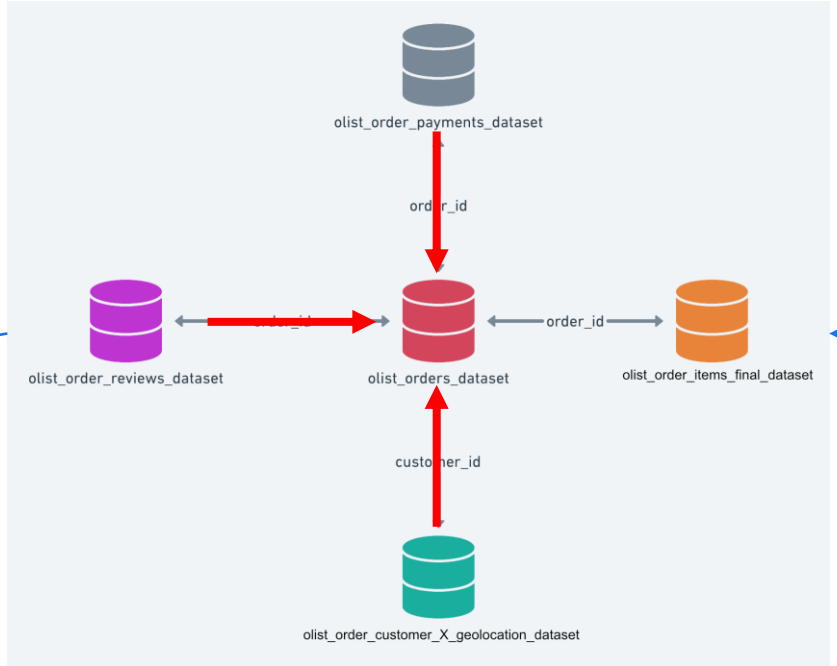

# Présentation des jeux de données

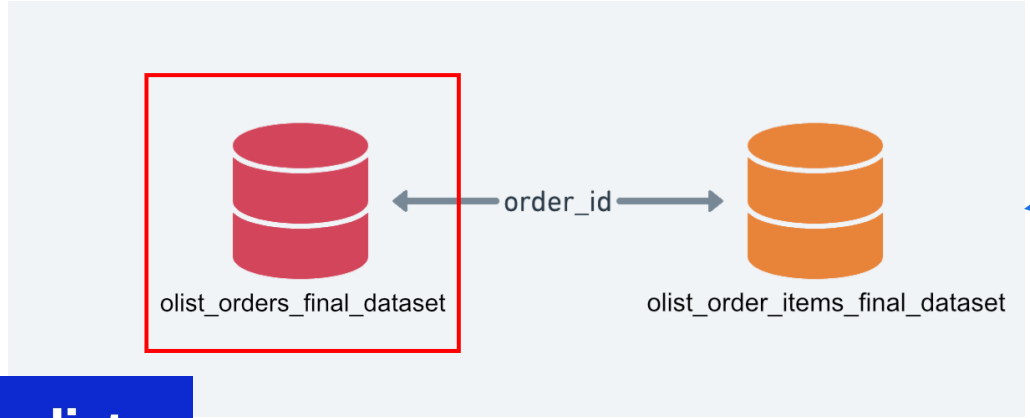
- Produit / Revendeur
- Commande / Client
- Géolocalisation

# Préparation du fichier final



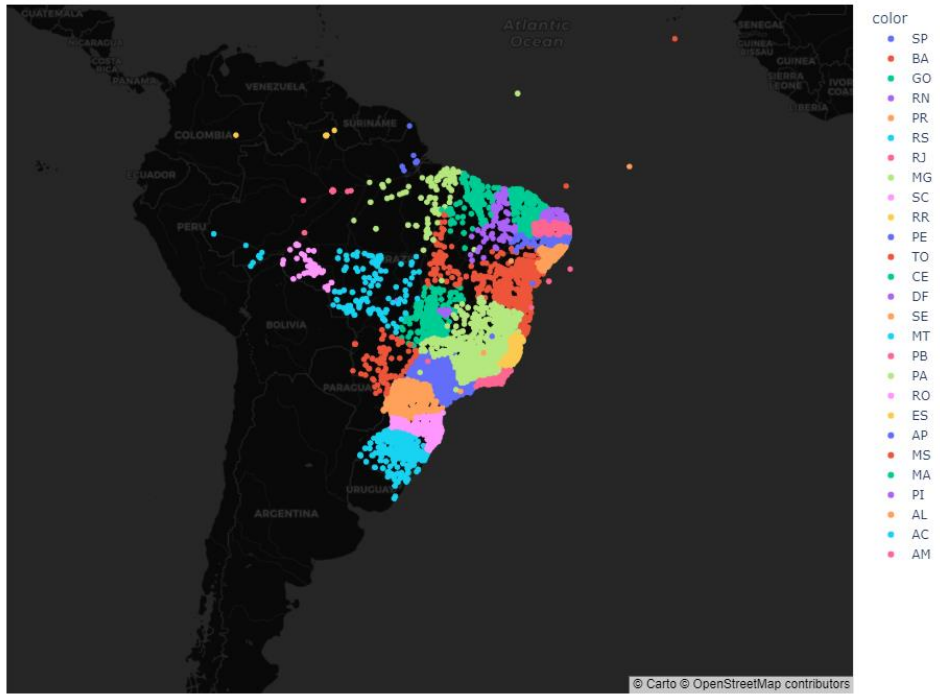



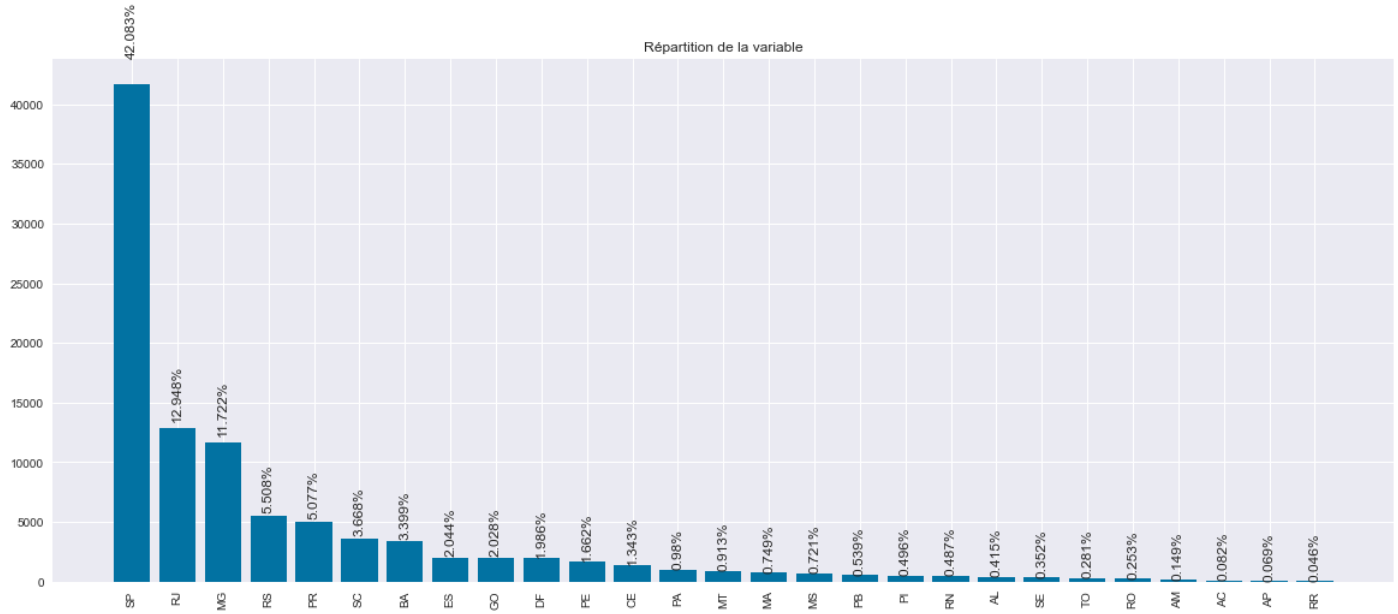



# Analyse du jeu de données

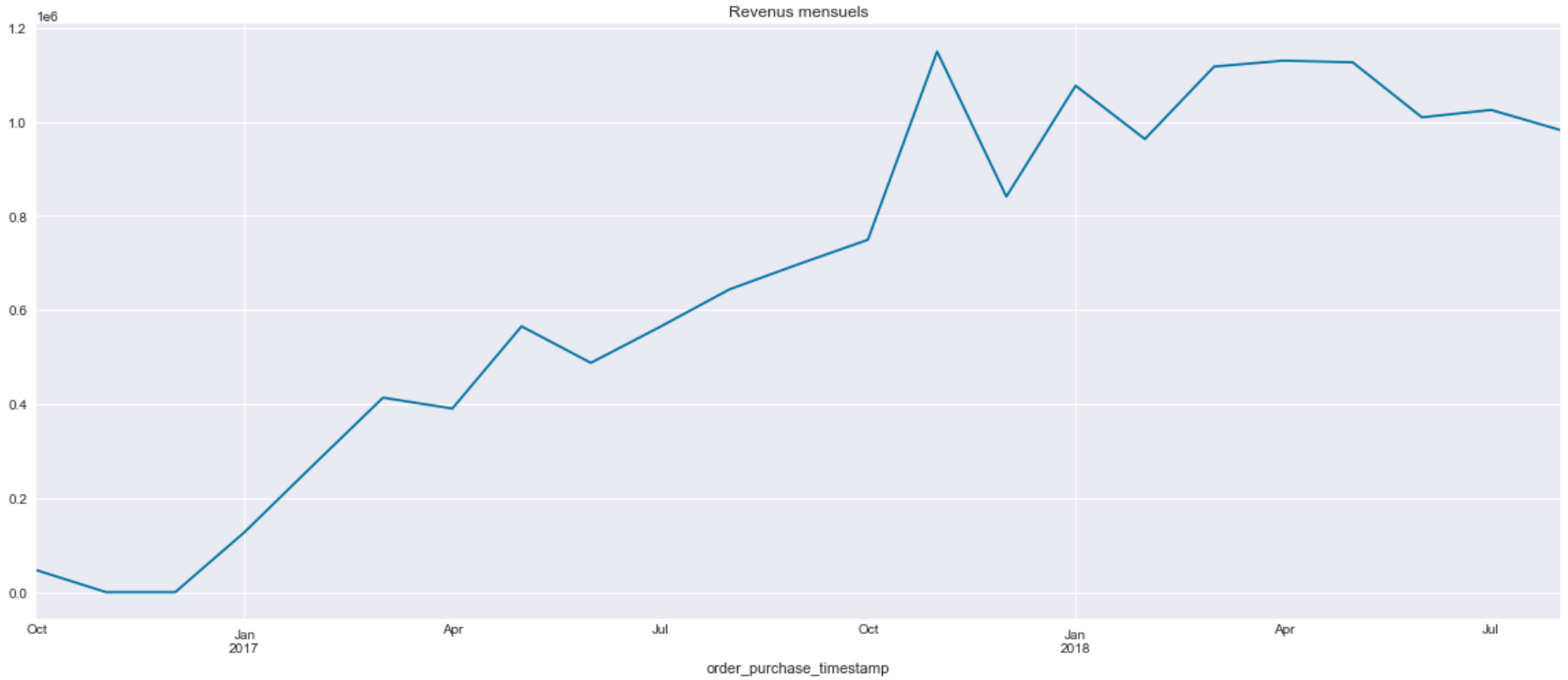
olist

# Analyse du jeu de données (spatial)

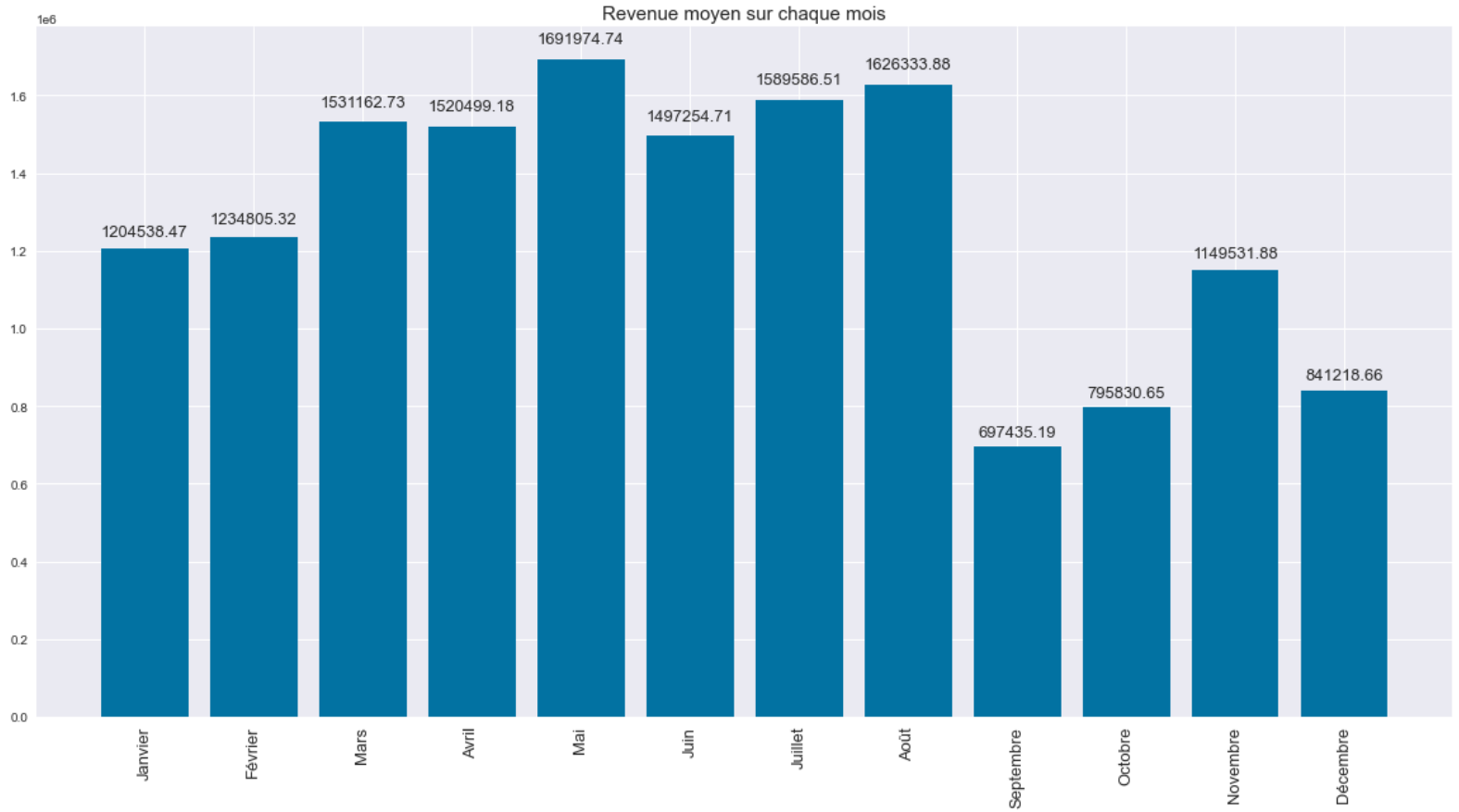


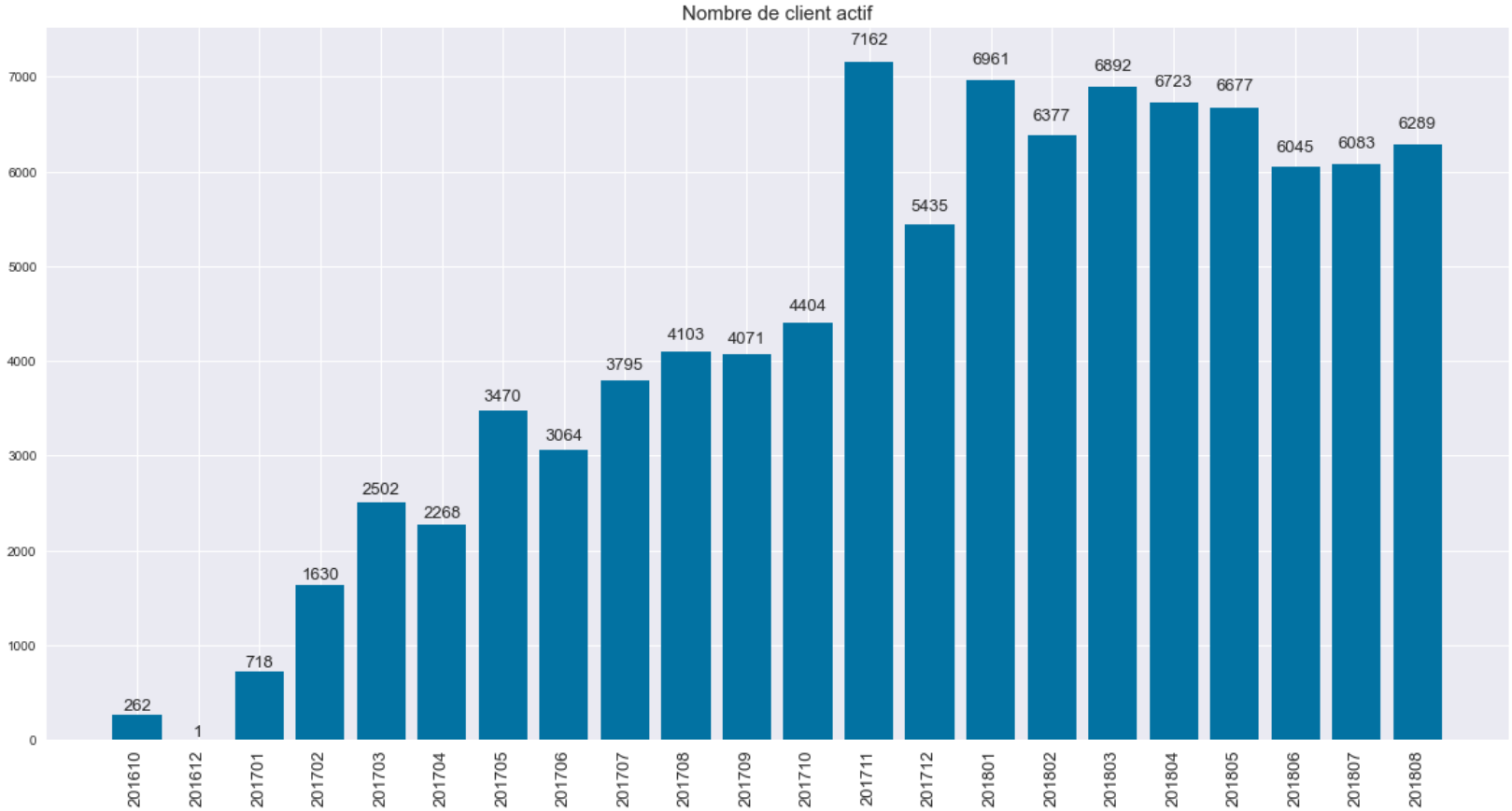
# Données sur les commandes / clients



# Données sur les commandes / clients (bis)

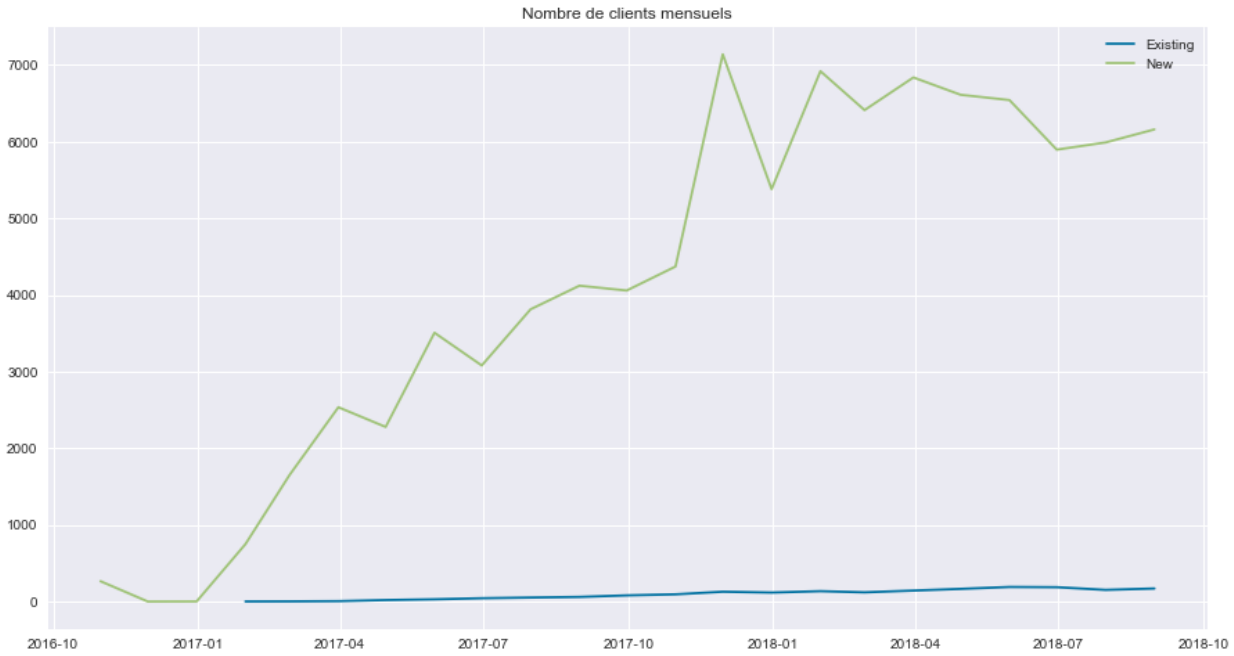


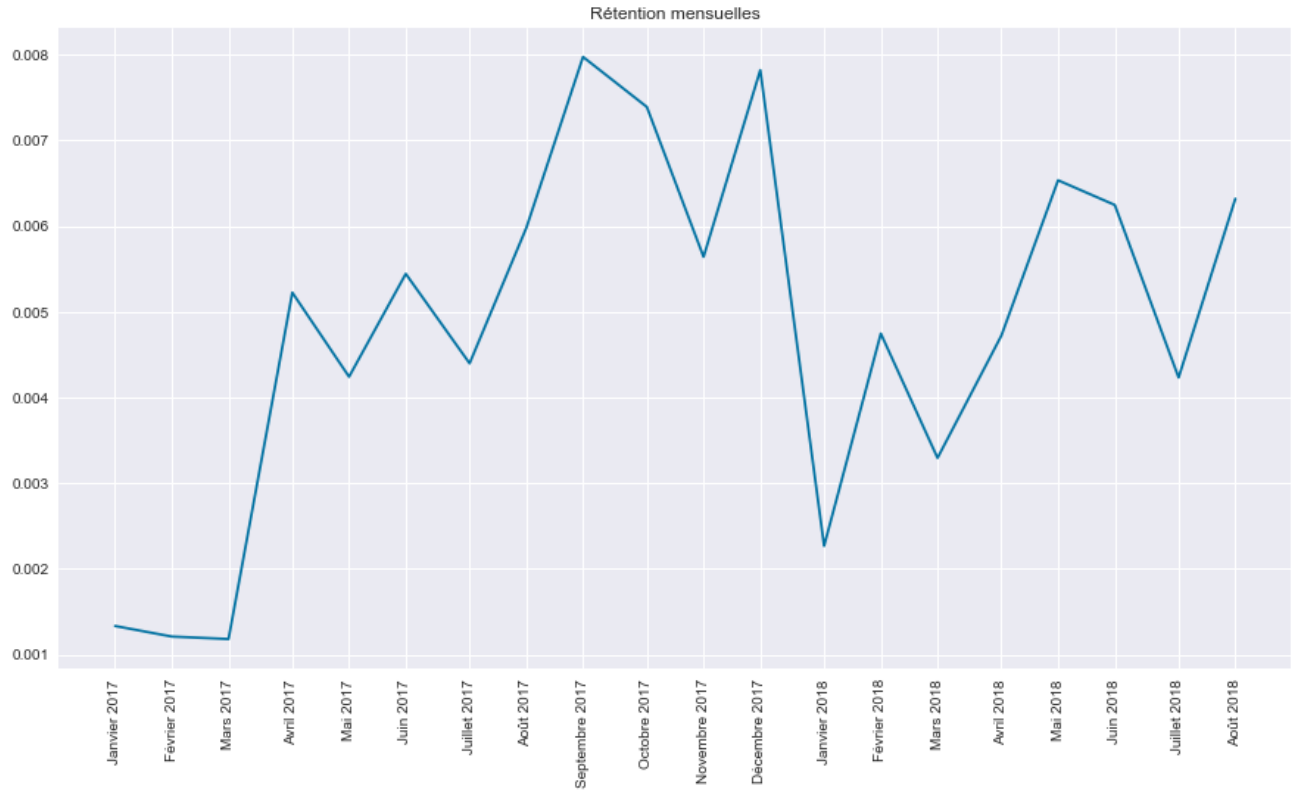
# Données sur les commandes / clients (bis)




olist

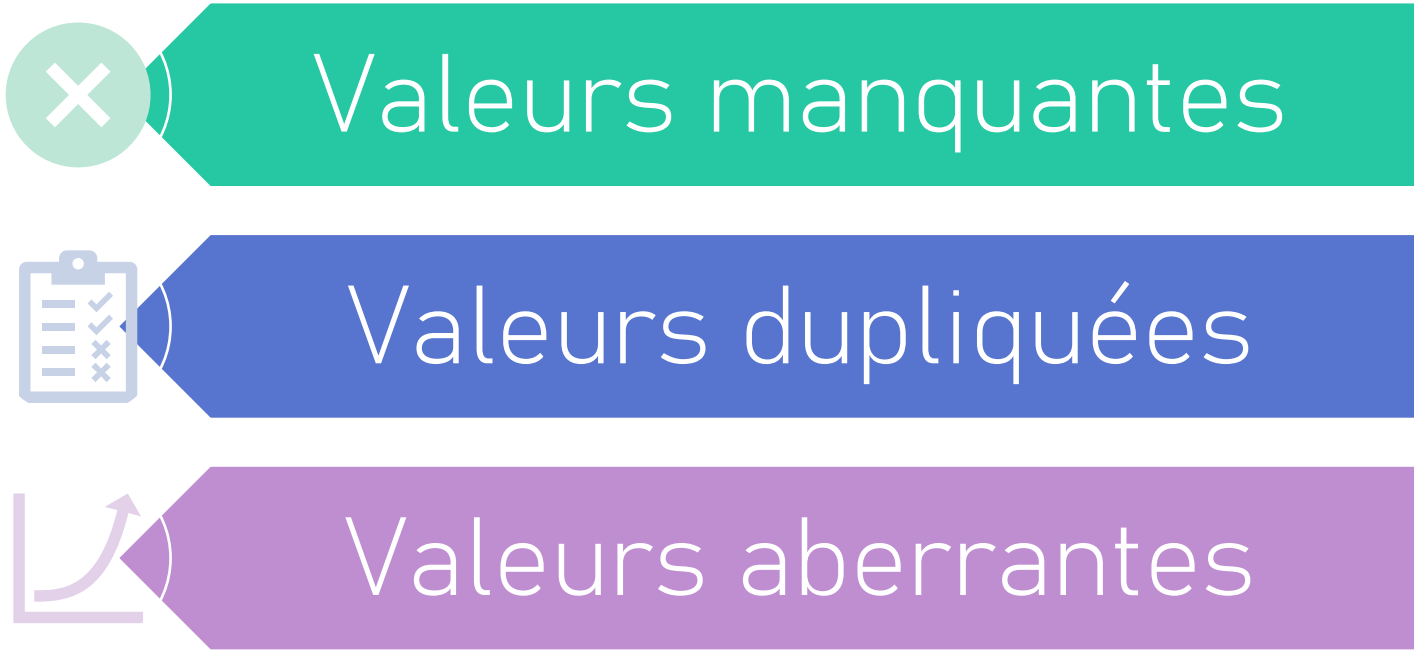
# Nombre de nouveaux clients et rétention




olist

# Nettoyage du jeu de données

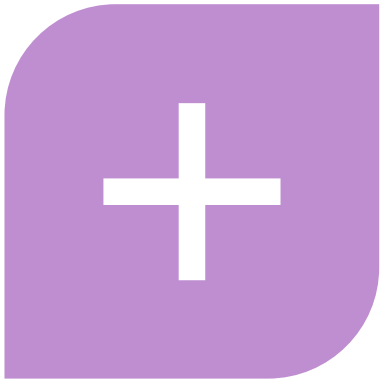


# Modélisation / Segmentation

olist

# Segmentation RFM

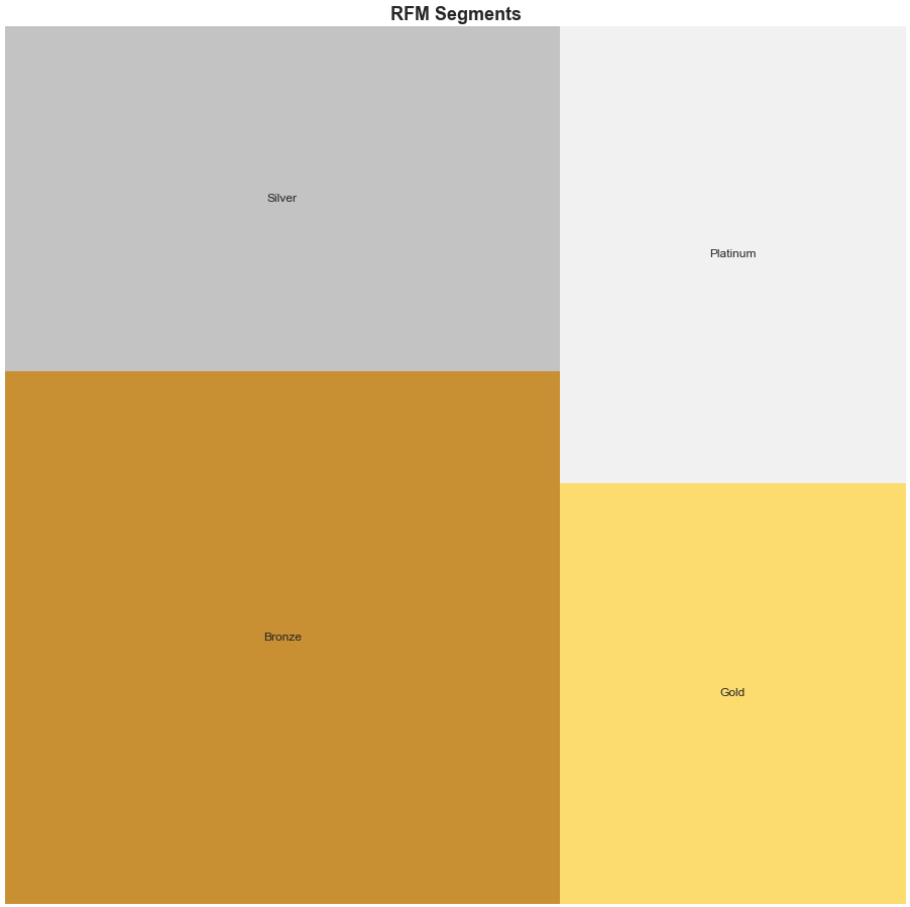
R = RÉCENCE



F = FRÉQUENCE

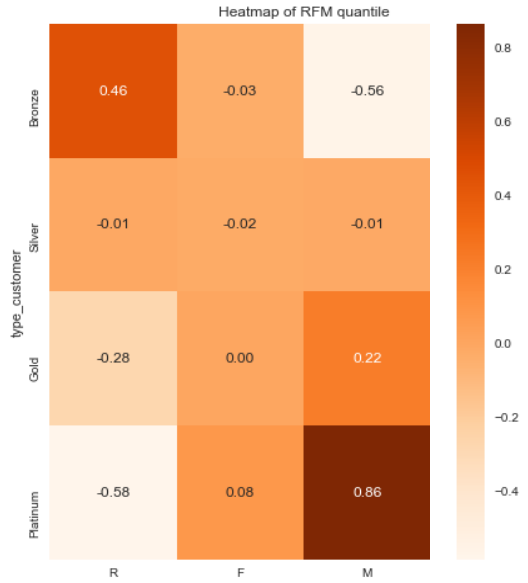
M = MONÉTAIRE
(MONTANT TOTAL)

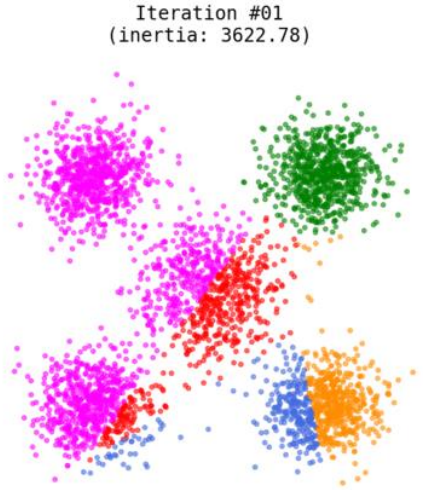
# Segmentation RFM (résultat)




MOYENNE :

R = 236,97 F = 1,03 M = 165,1

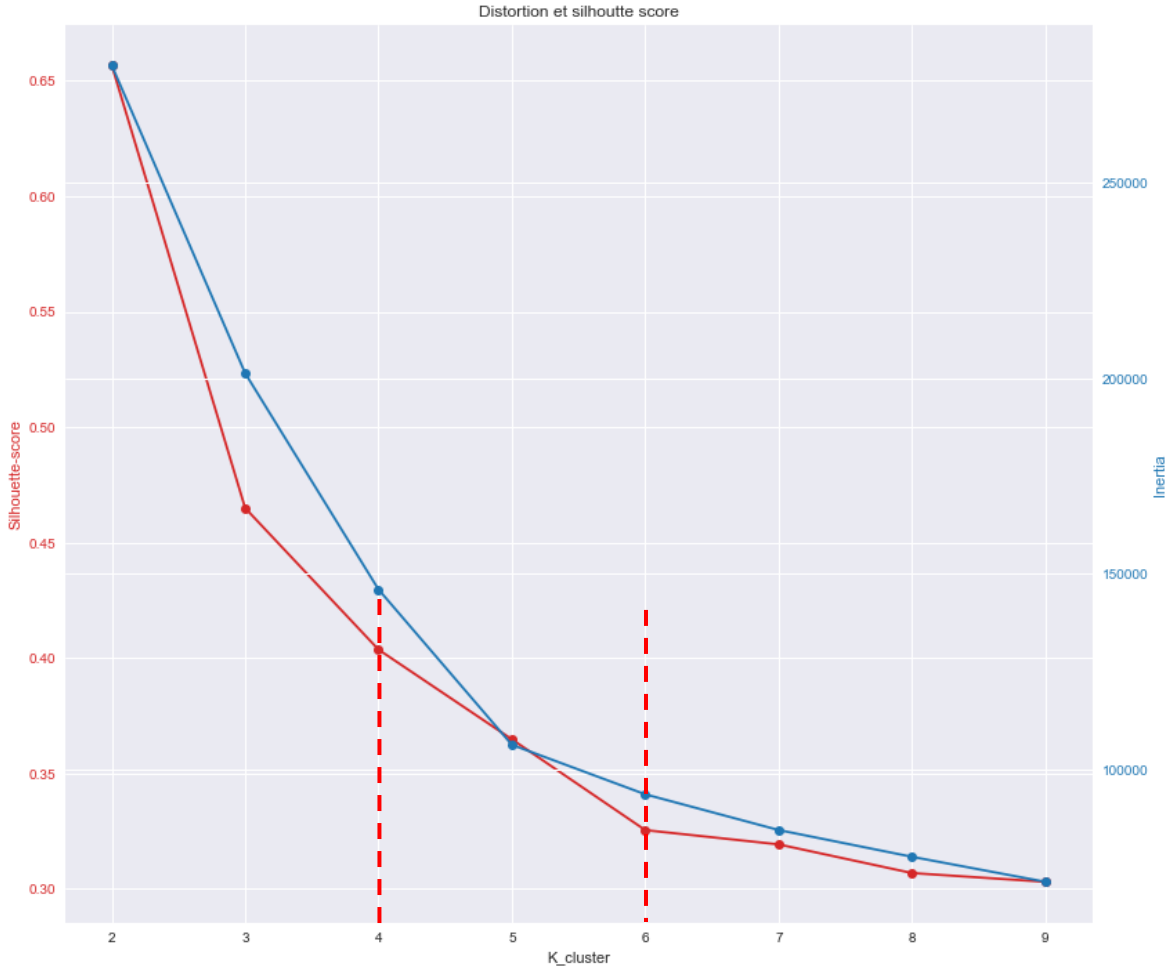




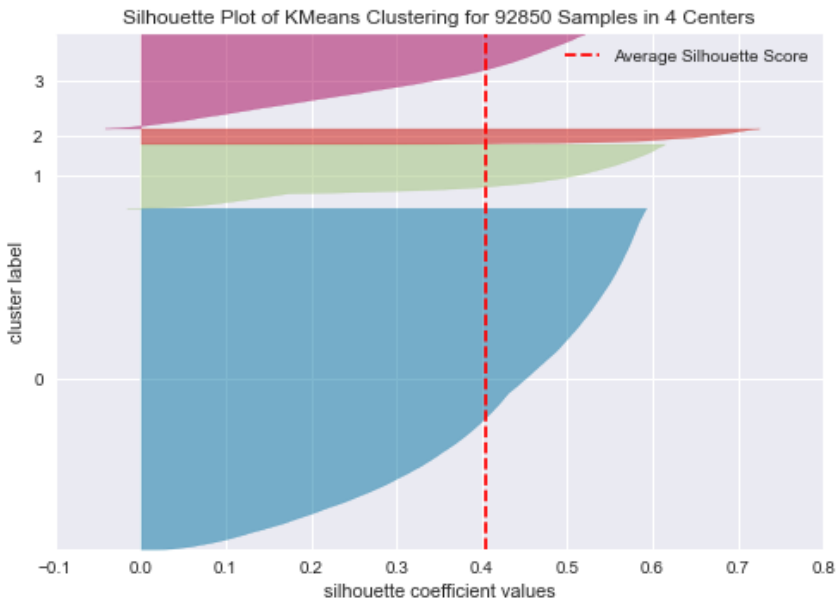

# Kmeans clustering (choix variables)

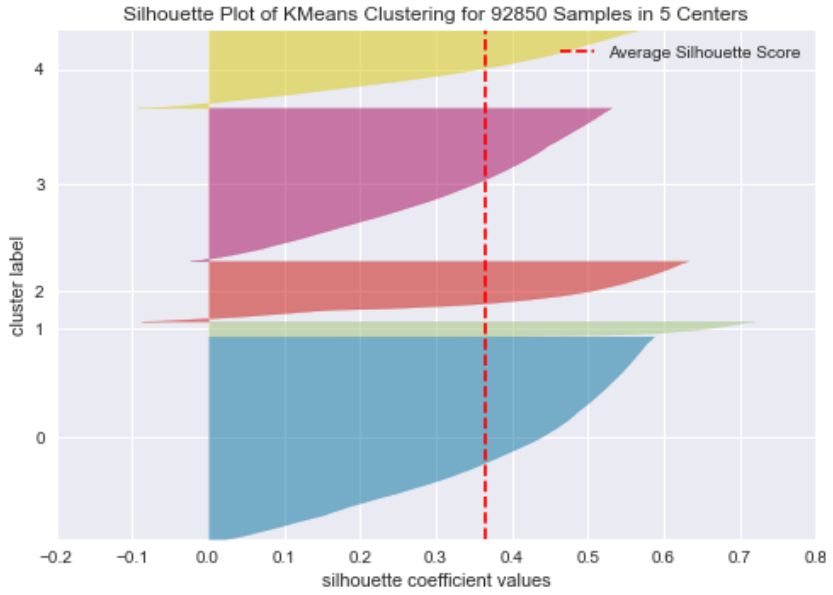
- Récence
- Fréquence
- Montant
- Review_score
- ~~Geolocation_state (one_hot_encoder)~~
- ~~Payment_value_xxxx (credit_card,voucher...)~~

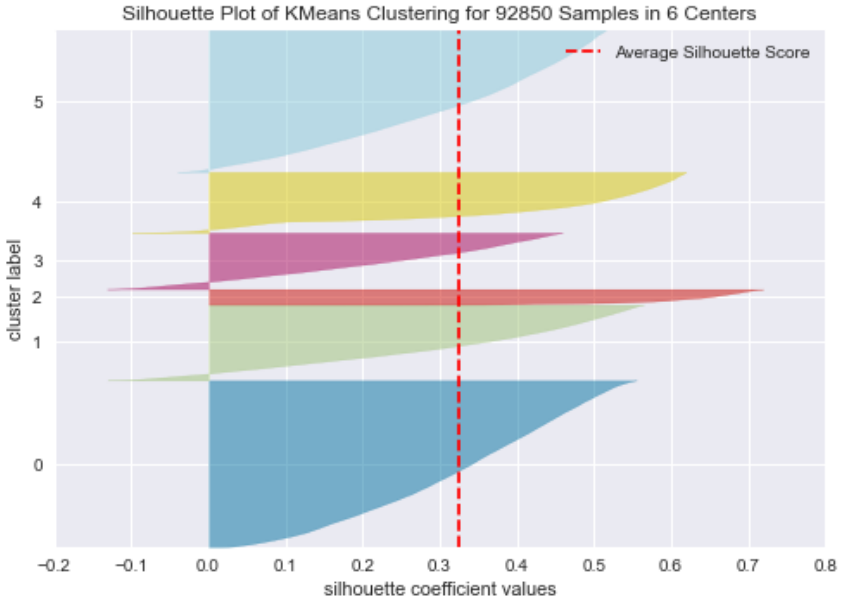
# Kmeans clustering (choix cluster)



# Kmeans clustering (choix cluster) (bis)

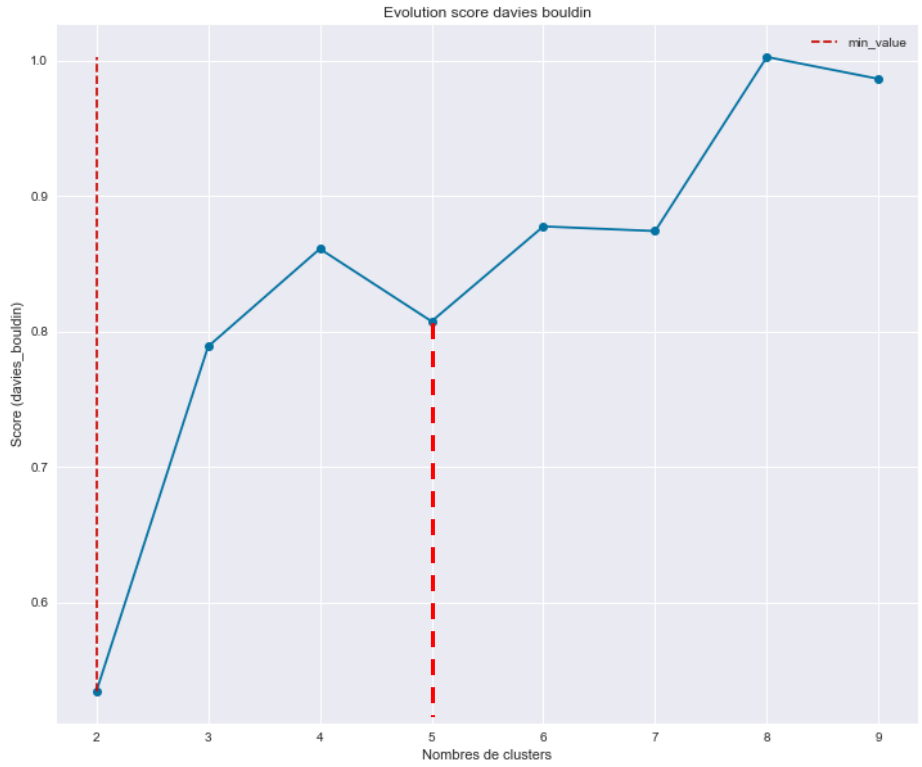


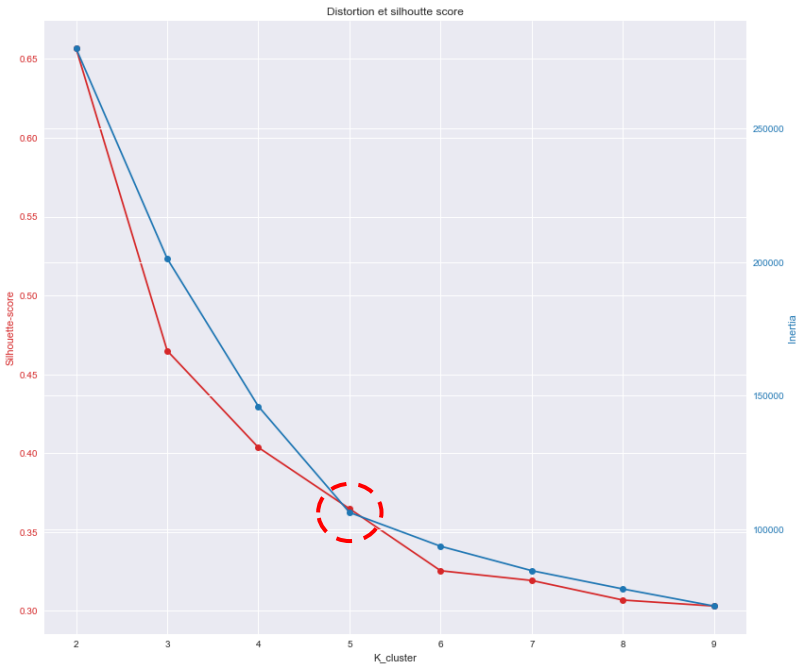



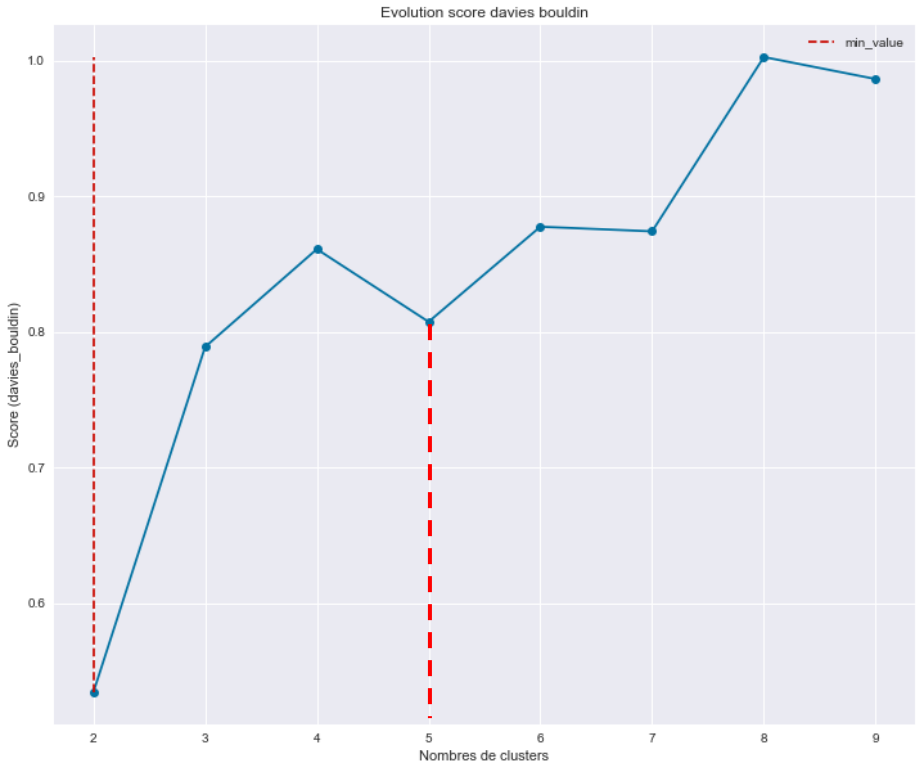

olist

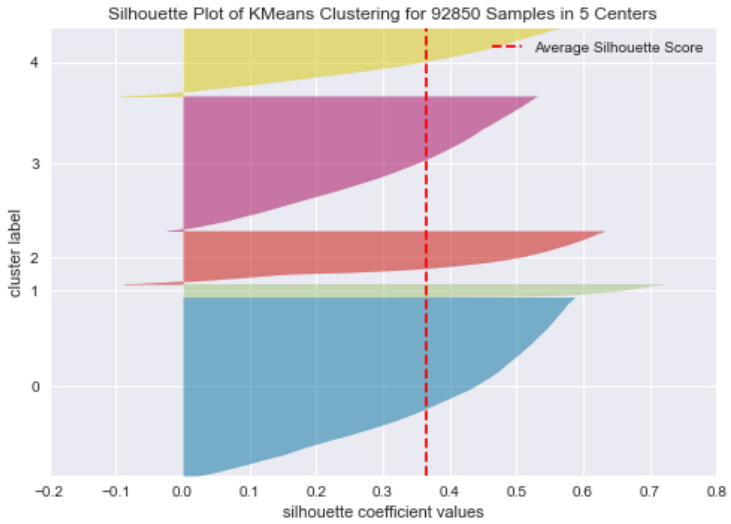
# Kmeans clustering (choix cluster) (bis)



# Kmeans clustering (choix cluster) (bis)




olist

# Kmeans clustering (résultat)

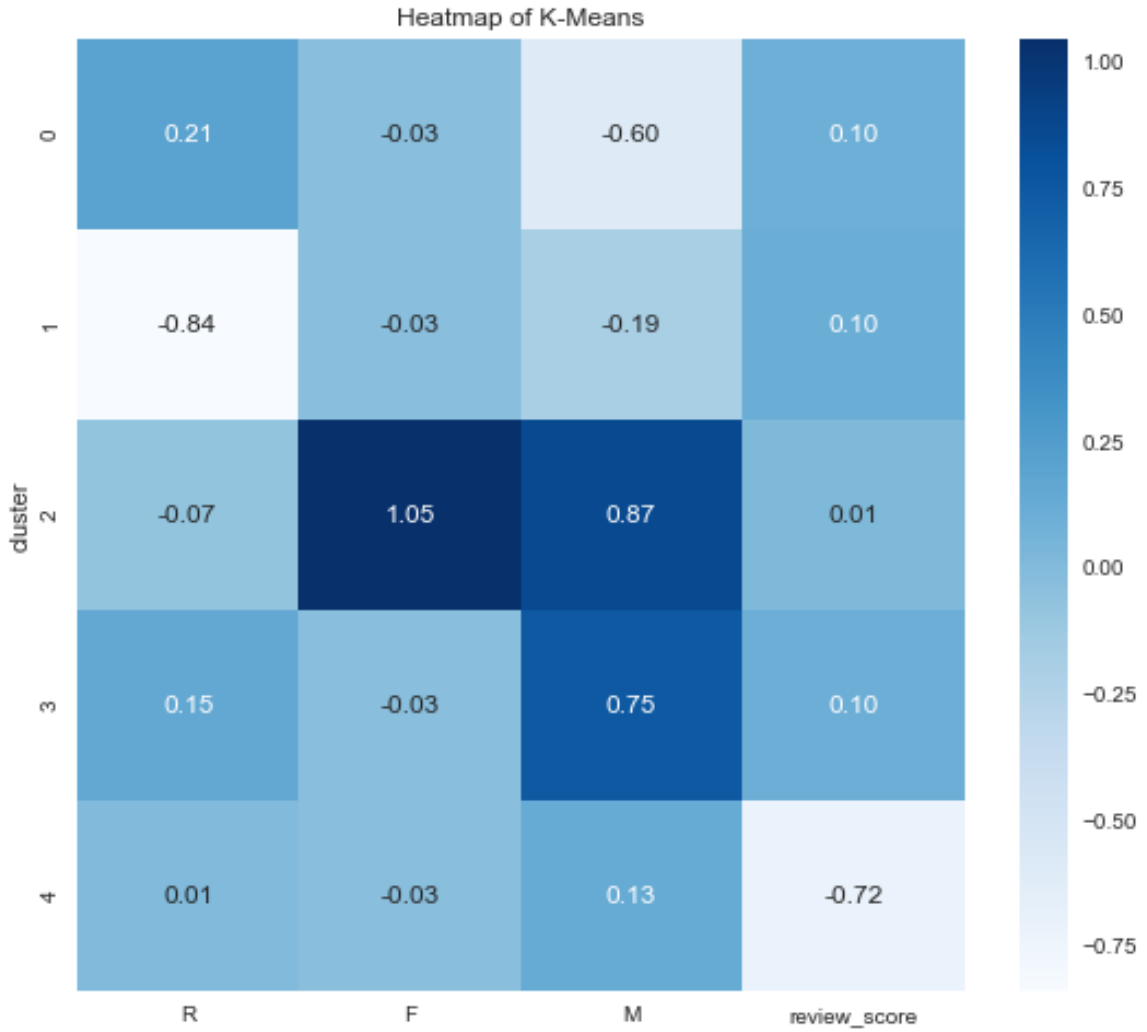
MOYENNE :

R = 236,97

F = 1,03

M = 165,19

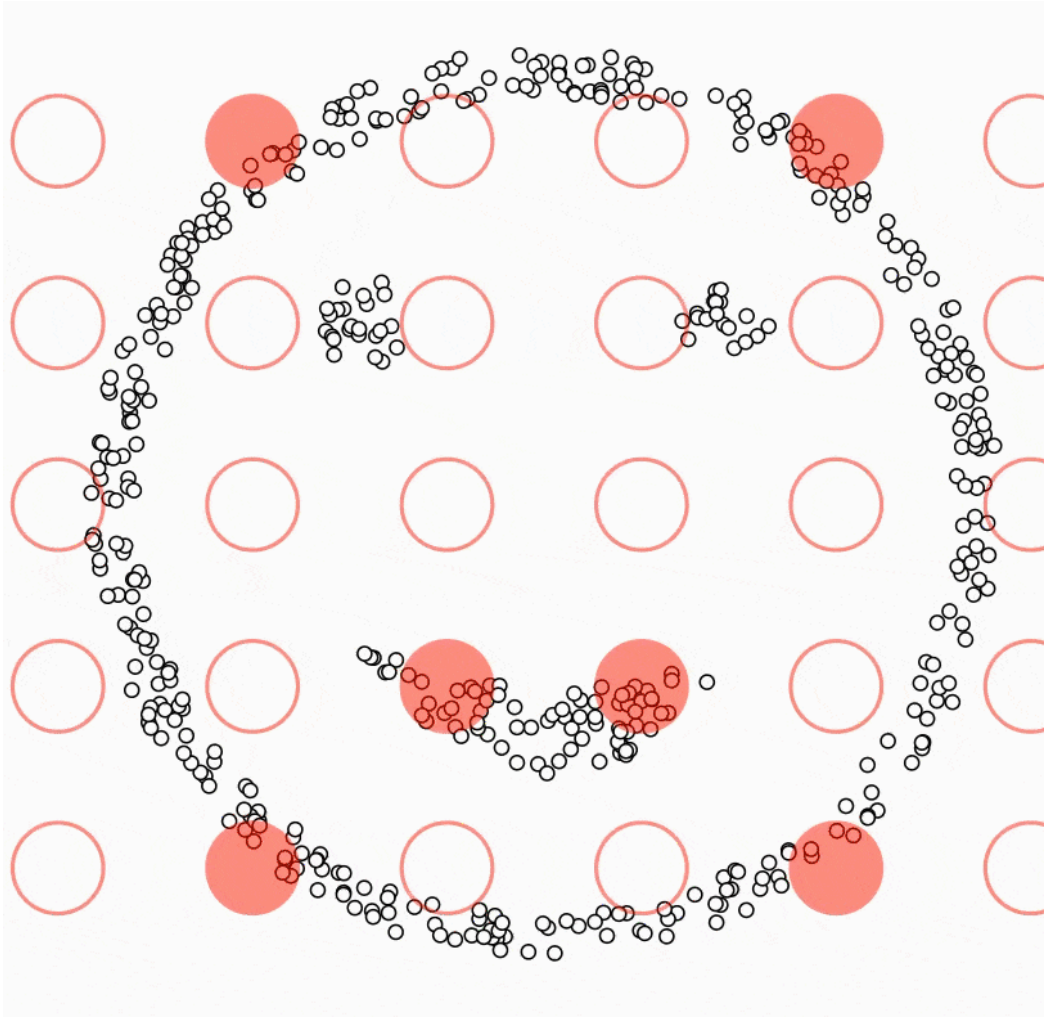
REVIEW_SCORE = 4,14



# DBScan

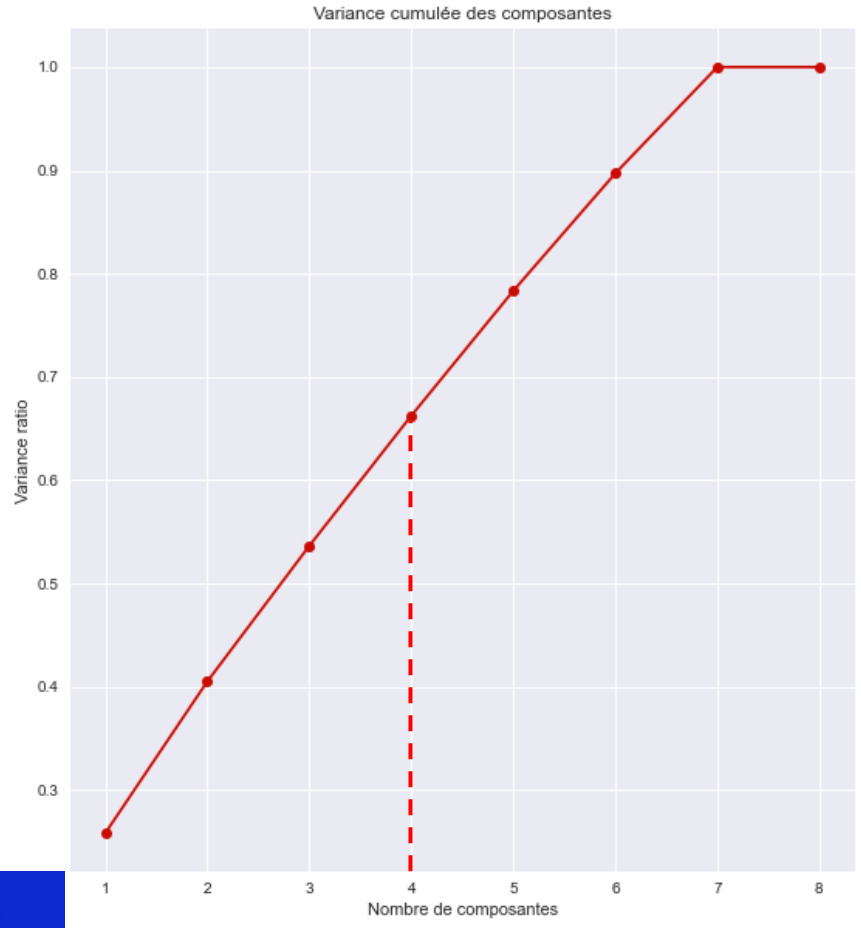
Silhouette-score : 0,48

| cluster | R_mean | F_mean | M_mean | Review_score | GroupSize |
|---|---|---|---|---|---|
| -1 | 132.212121 | 4.363636 | 1288.427273 | 3.686195 | 33 |
| 0 | 236.697519 | 1.000000 | 156.719776 | 4.491350 | 80987 |
| 1 | 232.420595 | 1.000000 | 194.095607 | 1.000000 | 9080 |
| 2 | 220.628874 | 2.000000 | 291.281295 | 4.177128 | 2549 |
| 3 | 213.994220 | 3.000000 | 435.929769 | 4.435453 | 173 |
| 4 | 175.142857 | 4.000000 | 678.645238 | 4.535714 | 21 |
| 5 | 150.714286 | 5.000000 | 610.208571 | 4.228571 | 7 |



olist

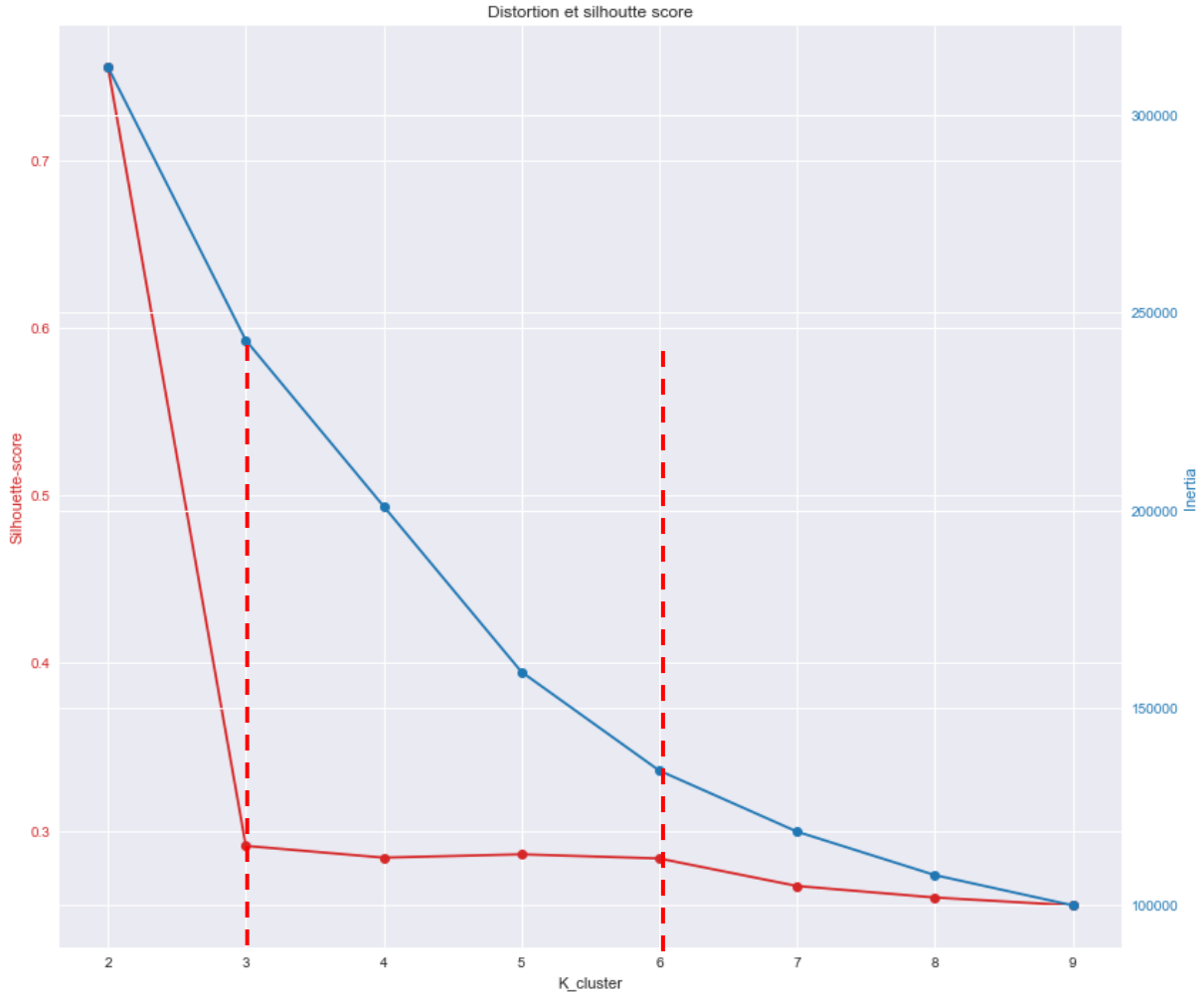
# Kmeans ACP



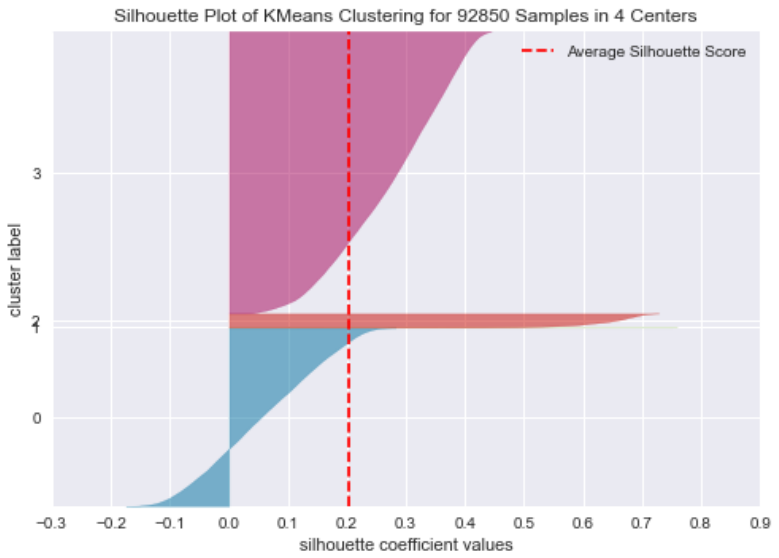

- Récence
- Fréquence
- Montant
- Review_score
- Geolocation_state (population)
- Time_to_answer (days)
- Score_evolution
- Nb_review

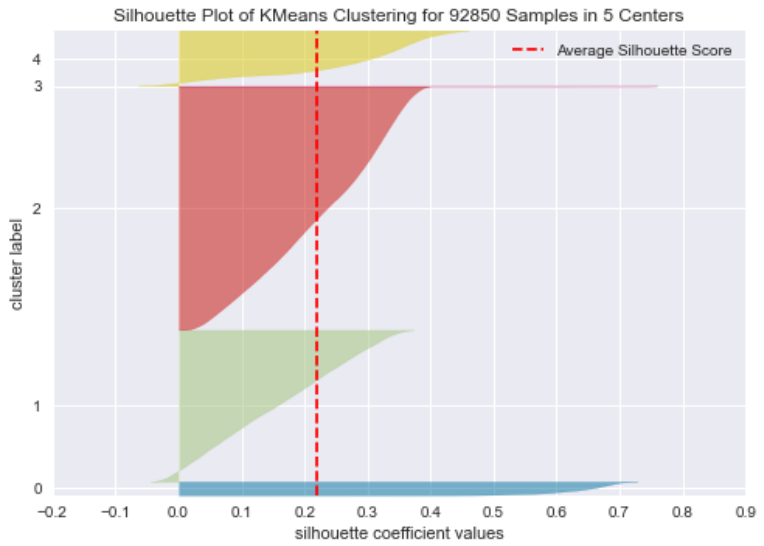
olist

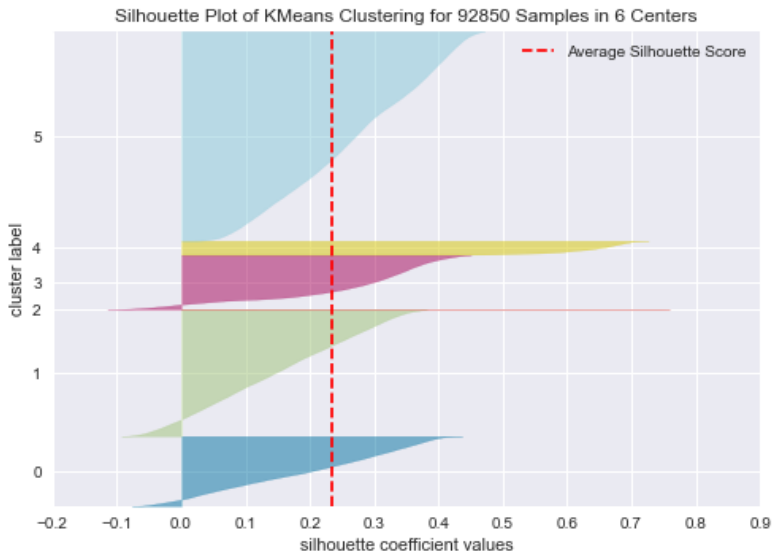
# Kmeans ACP (bis)



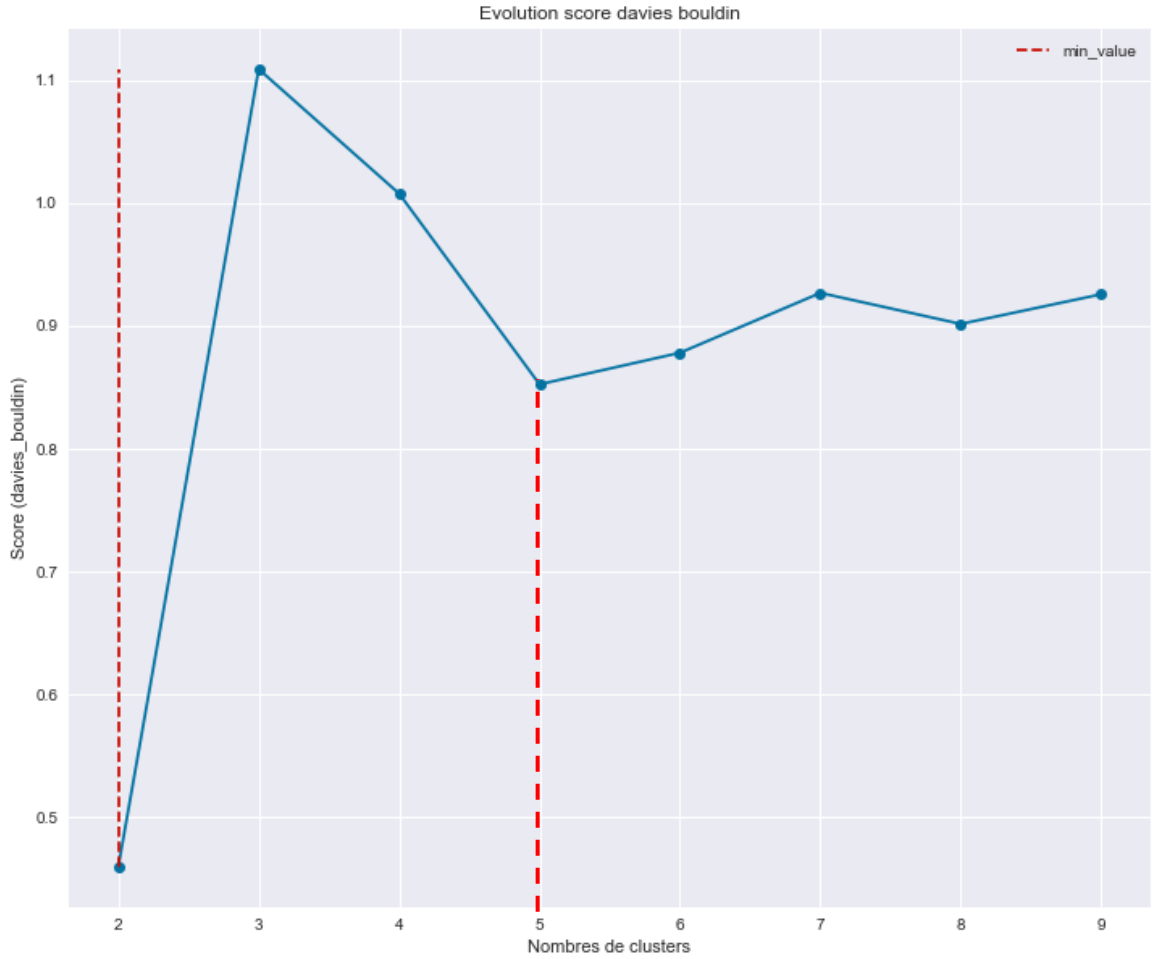
# Kmeans ACP (bis)



# Kmeans ACP (bis)




olist

# Kmeans ACP (résultat)

| cluster | monetary | review_score | review_score_evolution | nb_review | frequency | recency | time_to_answer | state_population | GroupSize |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 97.585912 | 4.567887 | -0.000563 | 1.000000 | 1.000000 | 195.341817 | 45.738369 | 30548.754208 | 42593 |
| **1** | 253.305598 | 2.315349 | -0.000061 | 1.000000 | 1.000000 | 233.455080 | 36.931801 | 13310.517398 | 16496 |
| **2** | 308.264214 | 4.210412 | -0.022251 | 2.113744 | 2.113744 | 217.760481 | 69.921144 | 21747.152752 | 2743 |
| **3** | 198.215649 | 4.521922 | -0.000646 | 1.000000 | 1.000000 | 294.006974 | 136.873190 | 11540.474947 | 30974 |
| **4** | 306.350682 | 2.662879 | 1.367424 | 1.954545 | 1.954545 | 298.704545 | 77.034880 | 16698.818182 | 44 |

# Modélisation finale

olist

# Modélisation finale

Kmeans

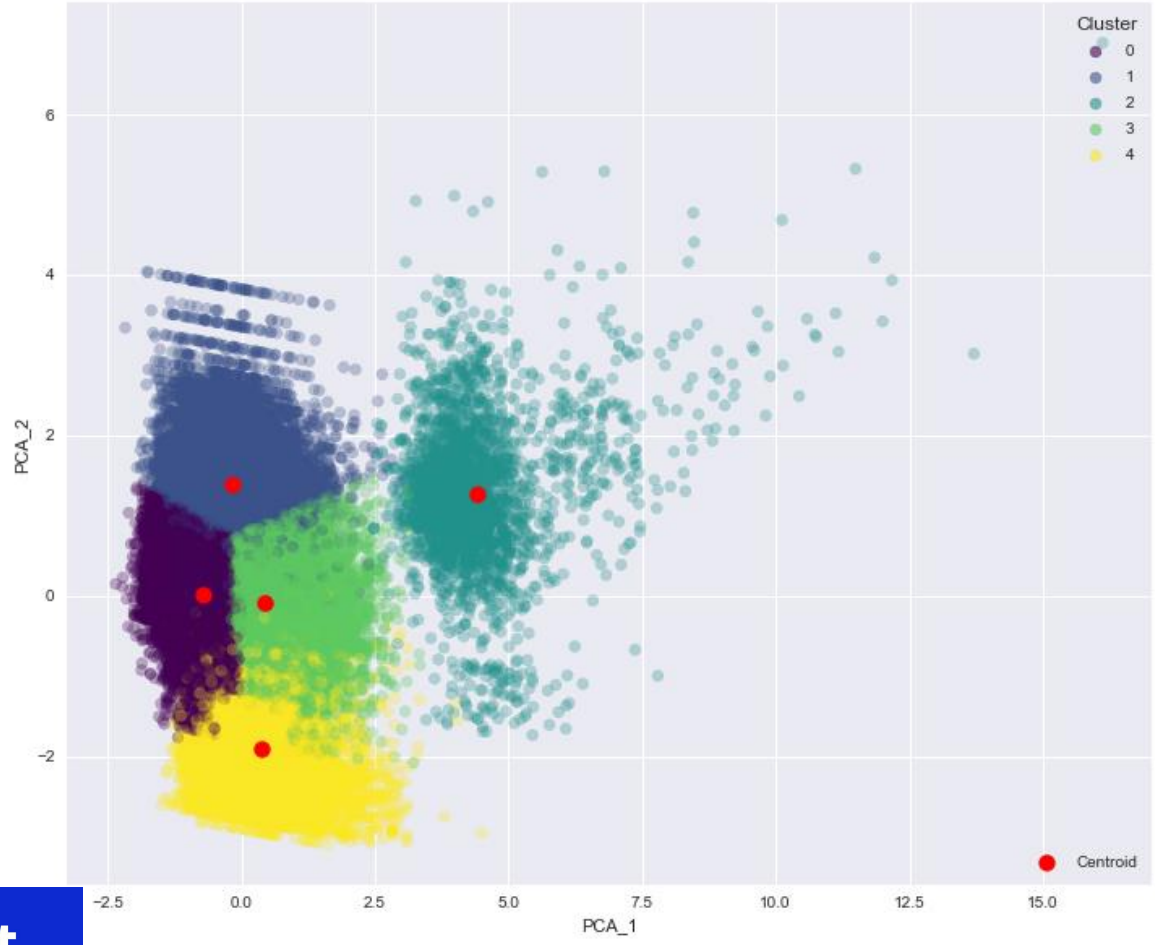
- Récence
- Fréquence
- Montant
- Review_score

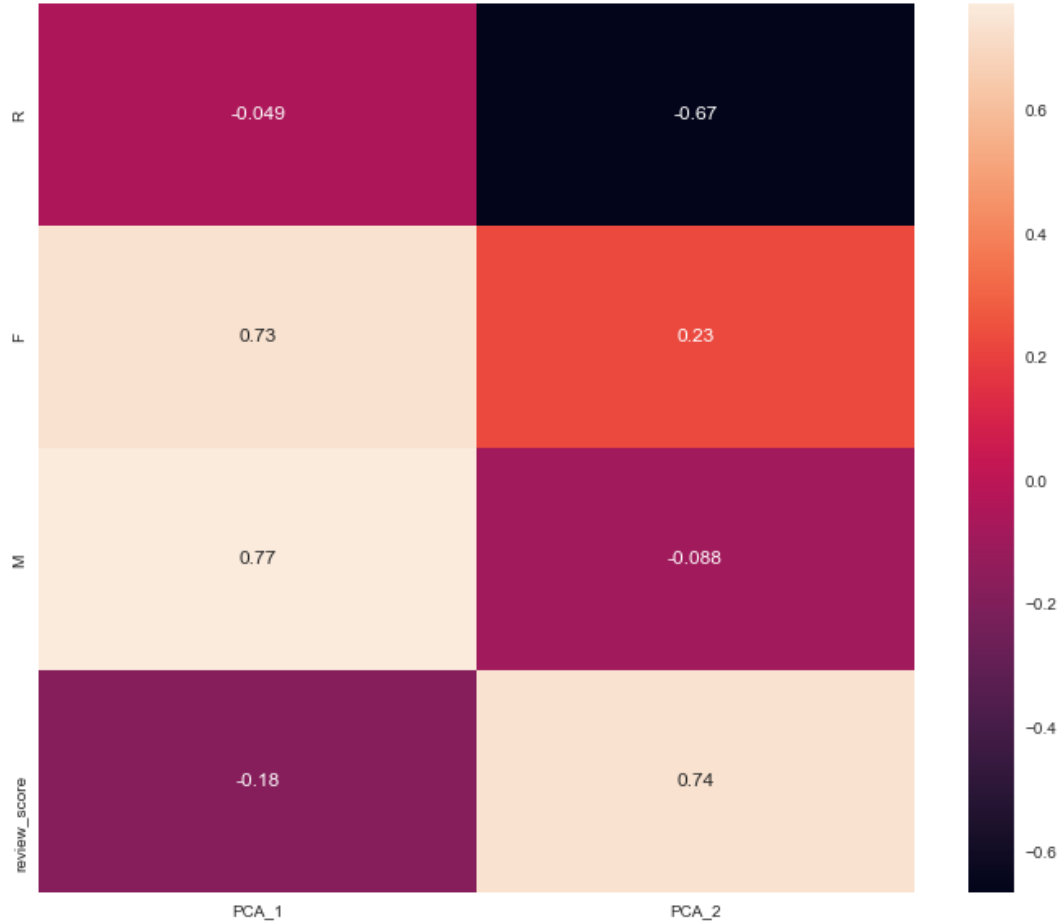
5 cluster

Intérprétable

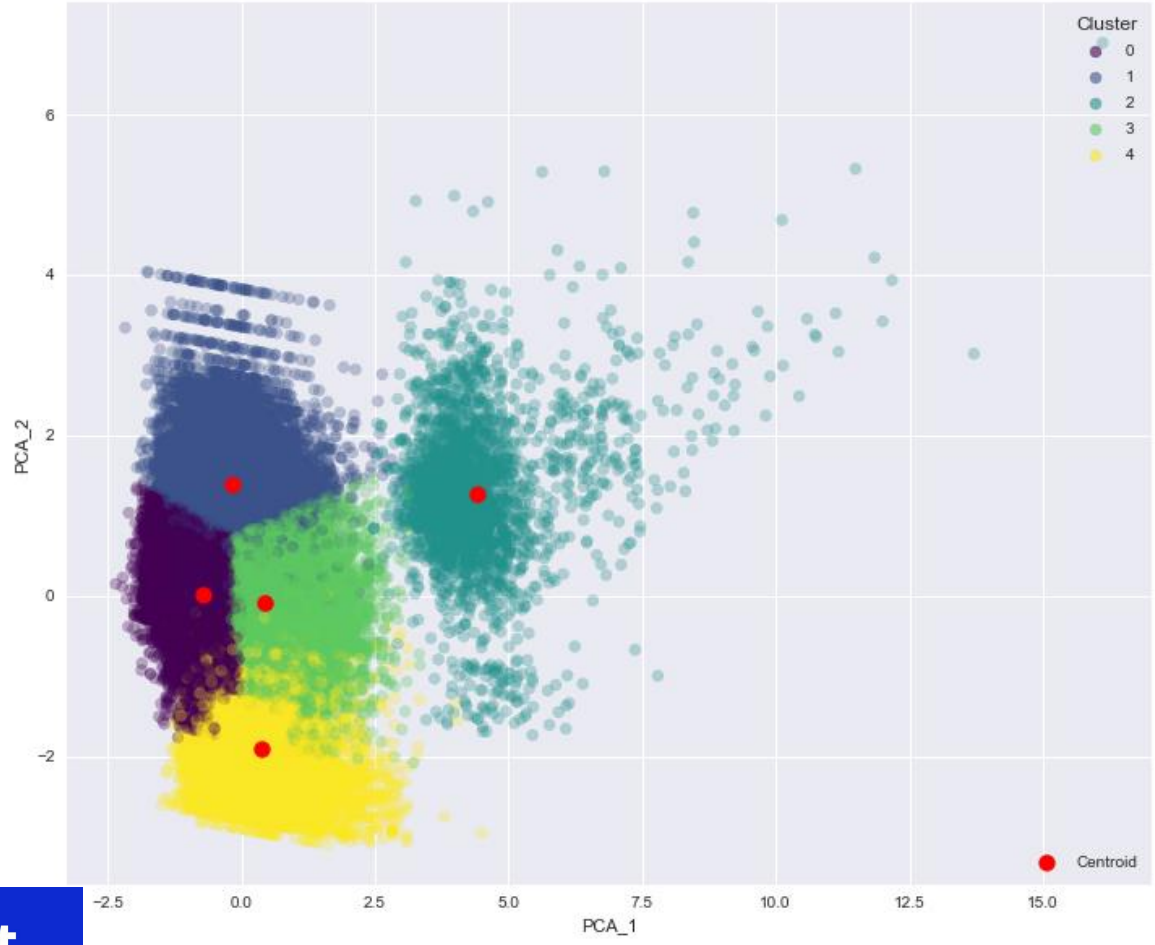
olist

# Représentation des données



# Représentation des données




- 0 : L'économe certainement parti
- 1 : Nouveau client neutre
- 2 : Meilleur client
- 3 :Dépensier occasionnel
- 4 :L'insatisfait

# Personae

## Jean-Bernard

**Meilleur client**

Dépense moyenne : 308 réal
Fréquence : 2 achats minimum
Satisfaction : 4.2

3%

## Mathilde

**Dépensière occasionnelle**

Dépense moyenne : 290 réal
Fréquence : 1 achat
Satisfaction : 4.55

30%

## Léo

**Client neutre (nouveau)**

Dépense moyenne : 133 réal
Fréquence : 1 achat
Satisfaction : 4.57

15%

## Annette

**L'économe (partie)**

Dépense moyenne : 66 réal
Fréquence : 1 achat
Satisfaction : 4.53

40%

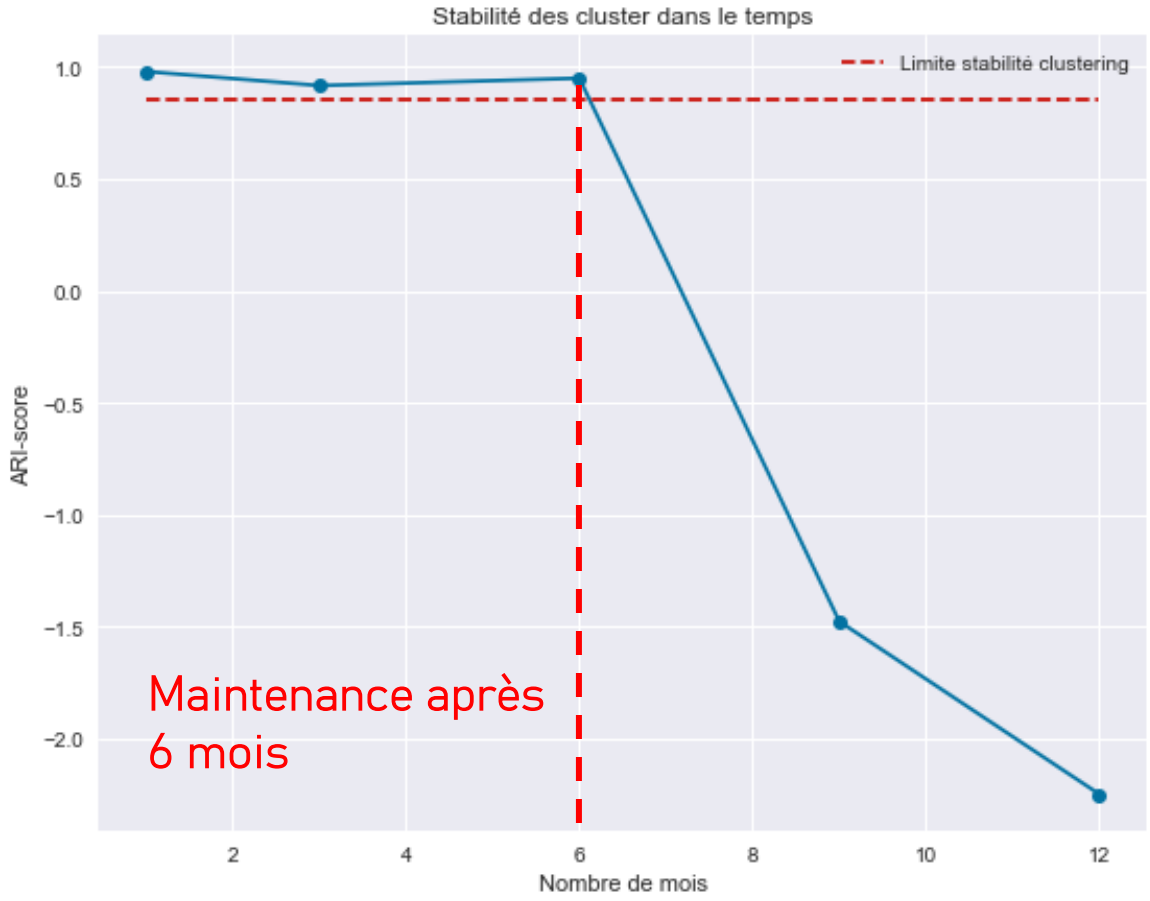
## Luna(tique)

**L'insatisfait(e)**

Dépense moyenne: 186 réal
Fréquence : 1 achat
Satisfaction : 1.17

12%

# Maintenance du clustering



# Stratégie d'ajout des nouveaux clients

## Calcul de distance

Calcul distance centroïde

## Classification multi-classe

Apprentissage supervisé sur les données labélisées par le clustering

olist

# Conclusion


olist

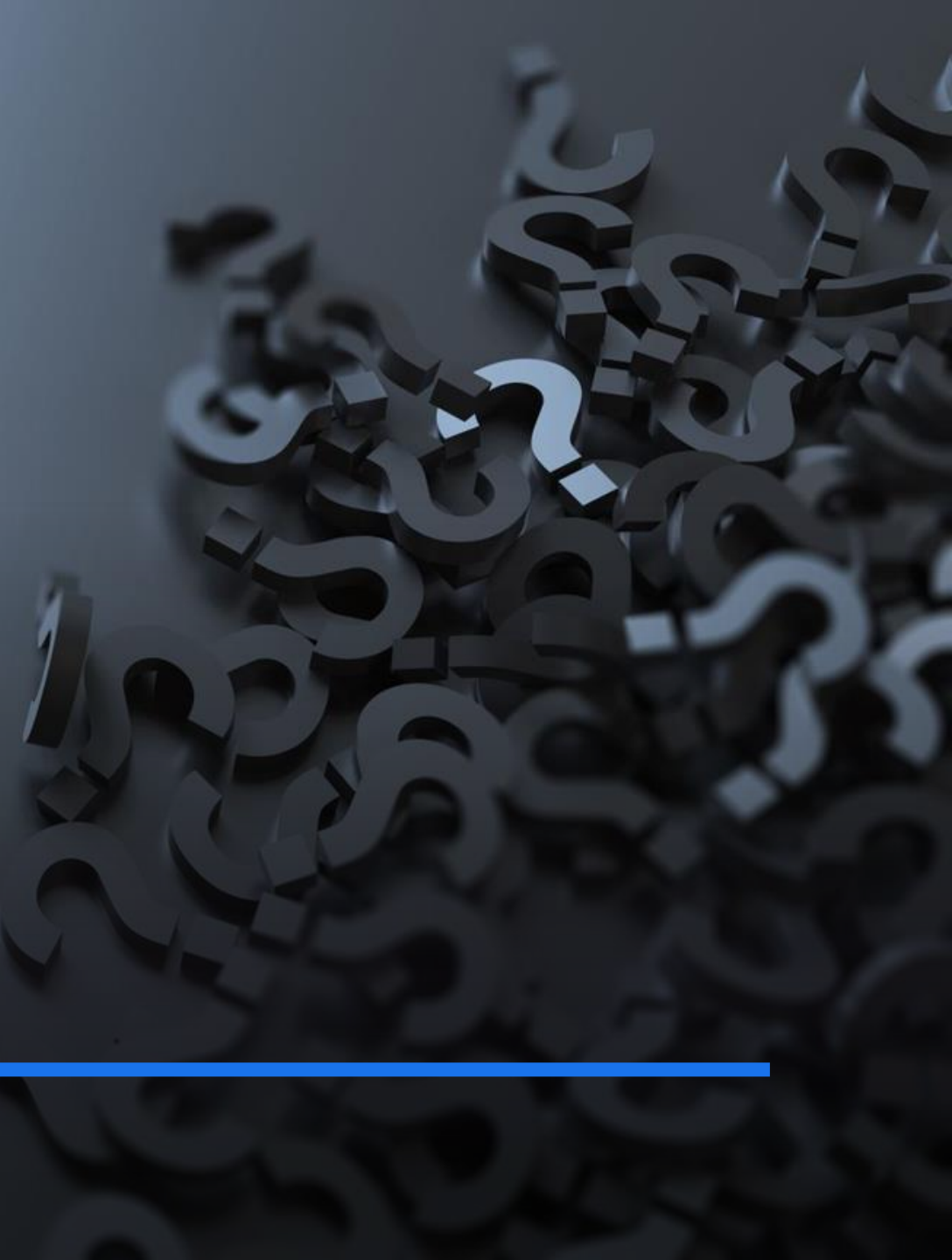
# Conclusion

- Ajout dimension du type d'achat voir NLP pour les commentaires
- Prédiction de l'évolution des clients dans le temps

- Interprétation des résultats difficiles (équipe marketing inexistante)

olist

Merci de votre attention,
avez-vous des questions ?



olist