FORMATION DATA SCIENTIST : PROJET 5

Segmentez des clients d'un site e-commerce

Sommaire

1. Introduction

- **Contexte :**

Olist, le pont entre vendeurs et marketplaces.

Objectif : Innover avec une équipe Data dédiée

Problématique :

Segmentation client : Clé d'une communication marketing sur-mesure

- **But :**

Décrypter les comportements pour des actions ciblées. Une segmentation client précise et utile.

2. Données et Sources

- **Base de Données Riches**

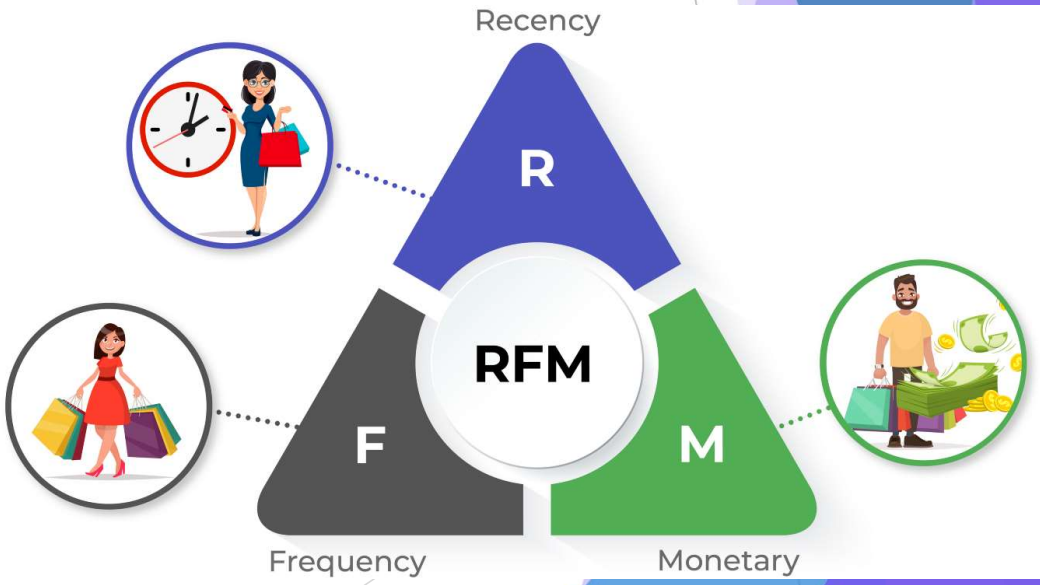
  Accès à un trésor d'informations : commandes, satisfaction, localisation etc.
  De janvier 2017 à octobre 2018.

- **Diversité des Données**

  Un panorama complet : Récence, Fréquence, Montant et bien plus.

- **Analyse en Profondeur**

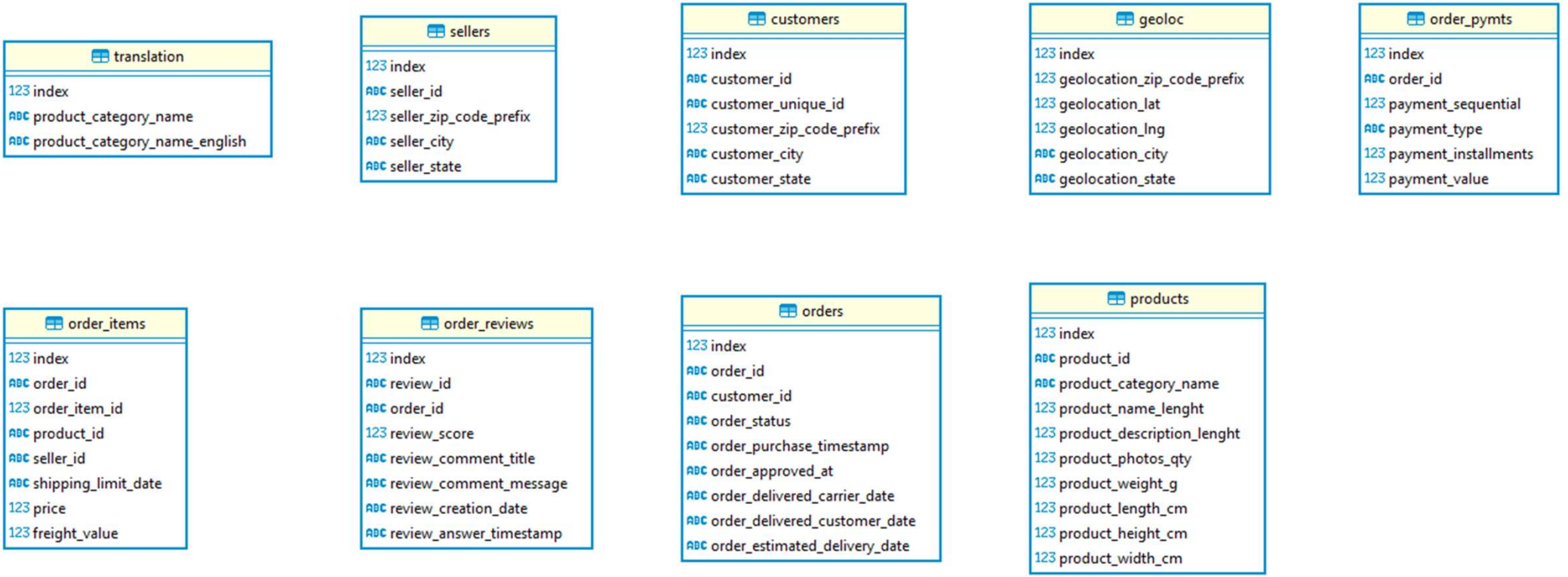
  Des données prêtes pour l'exploration : Identifier les tendances, comportements et préférences.



3. Extraction des données

- Base de données SQL

9 tables



3. Extraction des données

- Sélection et pré traitement

3 requêtes pour sortir les données pertinentes de la base

Extract .csv	Variables brutes	Features engineering
Customers	customer_unique_id customer_city customer_state geolocation_lat geolocation_lng	FirstOrderDate, LastOrderDate, TotalOrders, TotalSpent, TotalFreight, AvgItems, nb_item, ActualDeliveryTime, DeliveryDelay, AverageReviewScore, NumberOfReviews, NumberOfCommentTitles, NumberOfComments, DifferentCategories, AvgWeight, AvgVolume
Category	customer_unique_id product_category_name_english	CategoryCount TotalSpentPerCategory
Payment	customer_unique_id payment_type	PaymentCount TotalInstallments, TotalPaymentValue

- Merge, Nettoyage et features engineering

92411 clients, 1 dataframe

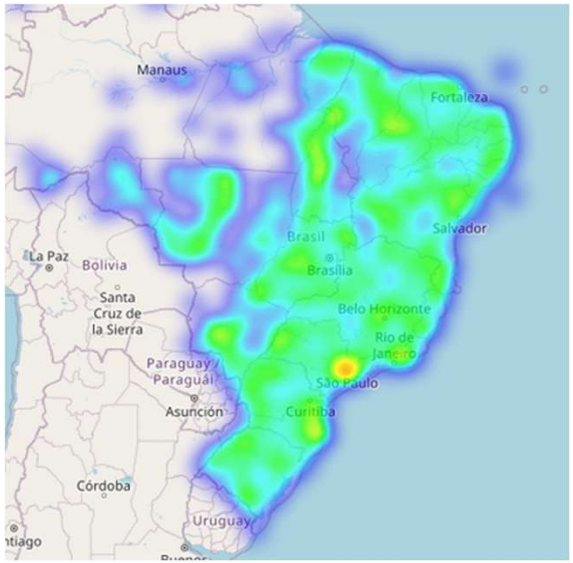
TotalOrders
nb_item
ActualDeliveryTime
DeliveryDelay
AverageReviewScore
DifferentCategories
AvgWeight
AvgVolume
TotalInstallments

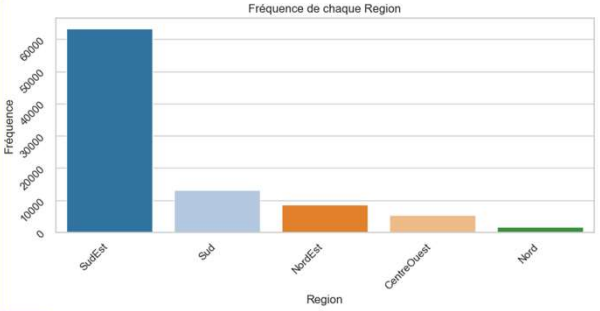
boleto_pct
credit_card_pct
debit_card_pct
voucher_pct
Fashion_Beauty_pct
Home_pct
Food_drink_pct
Miscellaneous_pct
Books_pct
Office_pct
Construction_pct
Sports_Leisure_pct
Technology_Electronics_pct

Recency
Frequency
TotalFreightPct
AvgBasket
EngagementIndex
Region

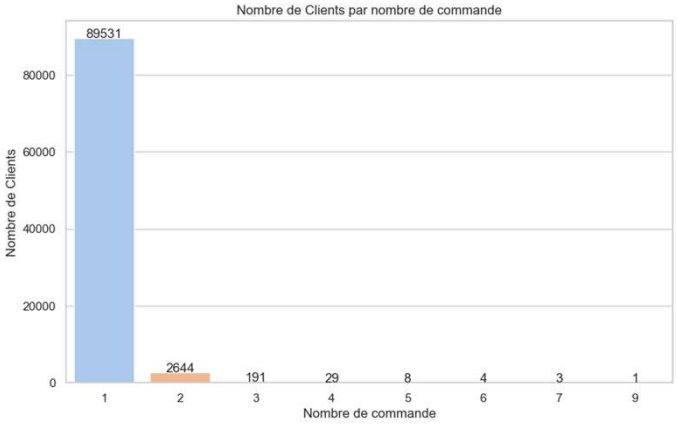
4. Analyse exploratoire et prétraitement des données

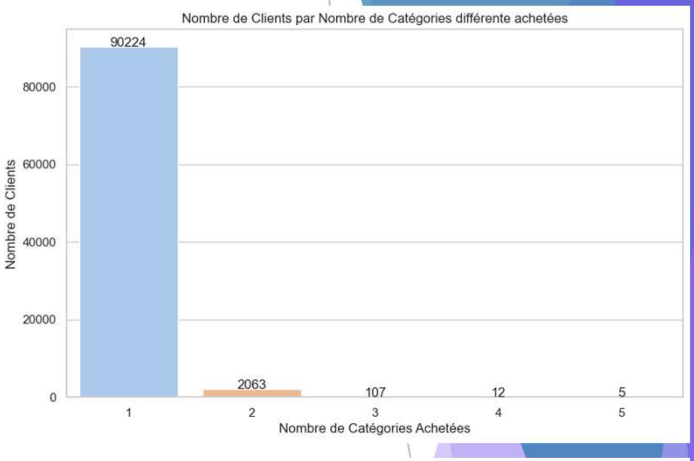
- Position géographique des clients d'olist

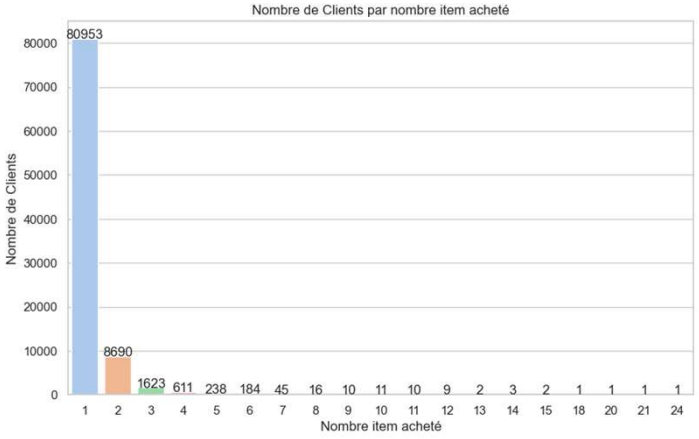


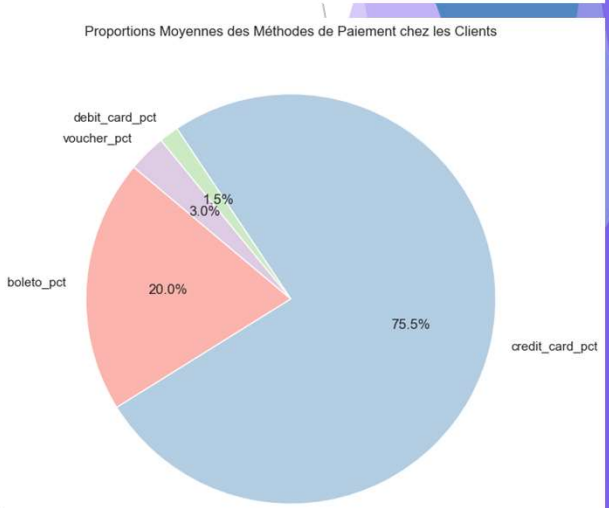


- Statistiques notables



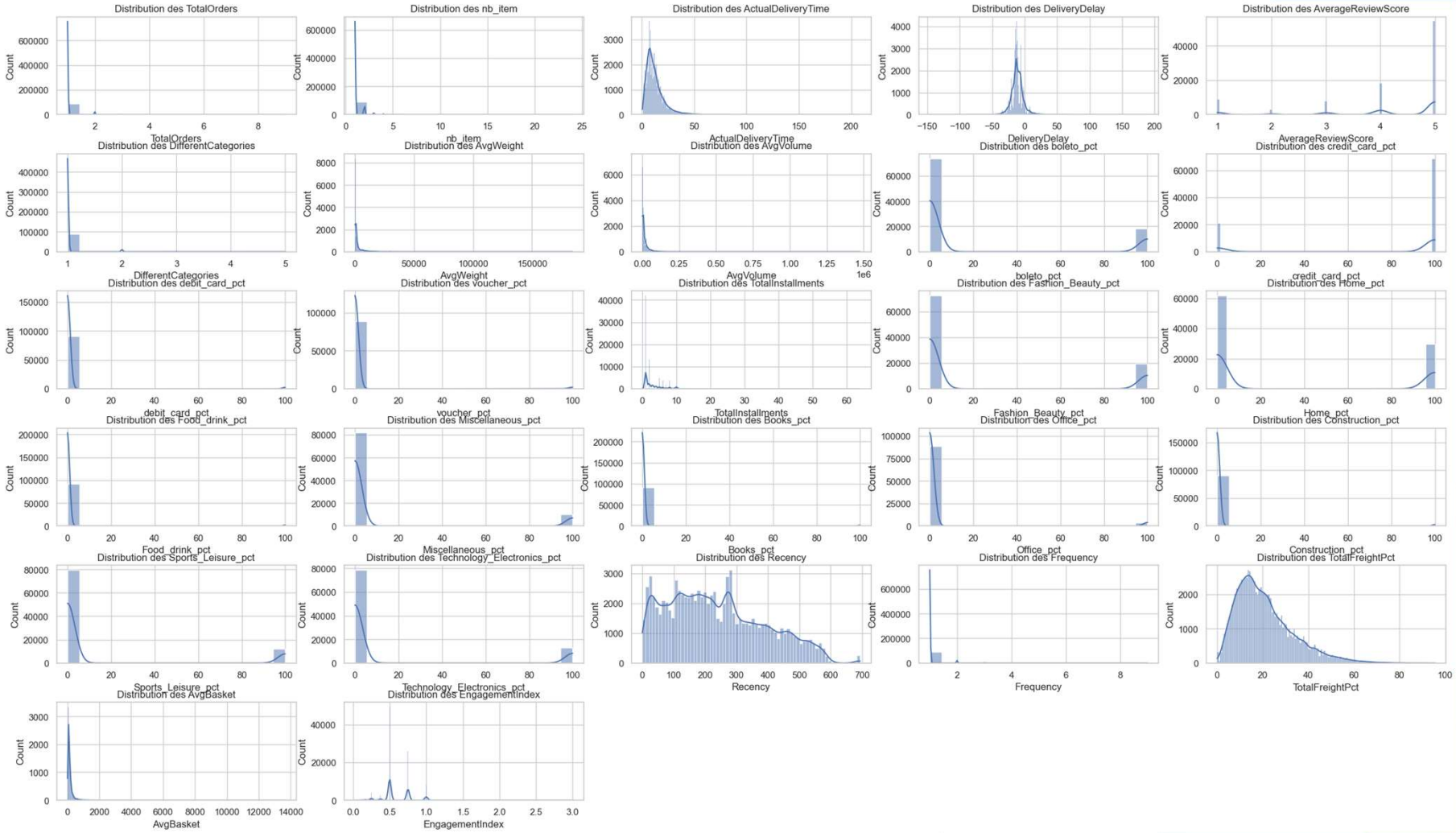






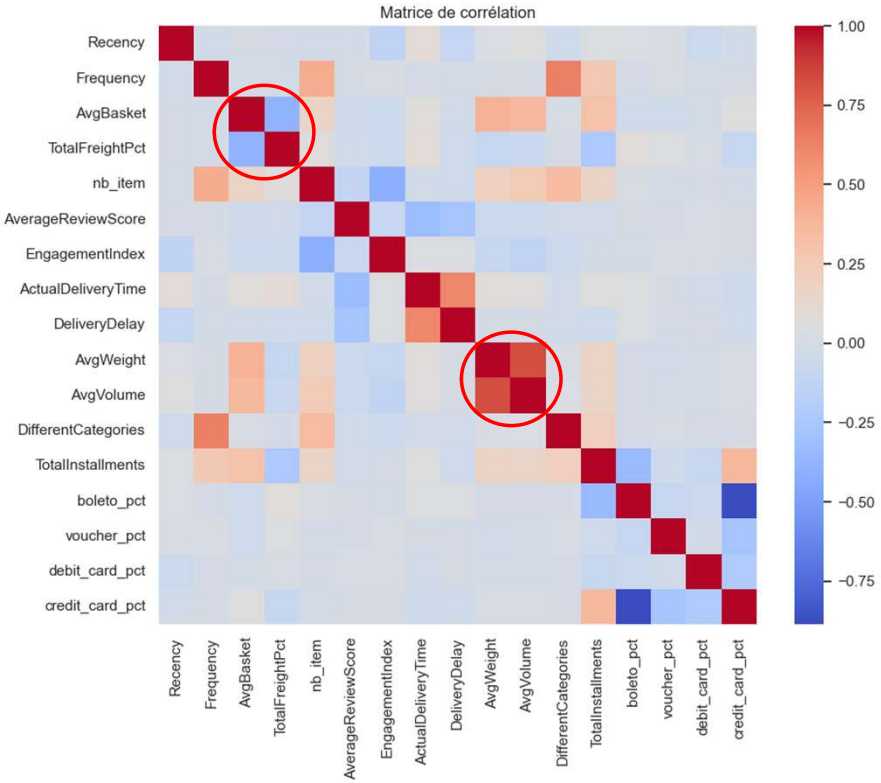
4. Analyse exploratoire et prétraitement des données

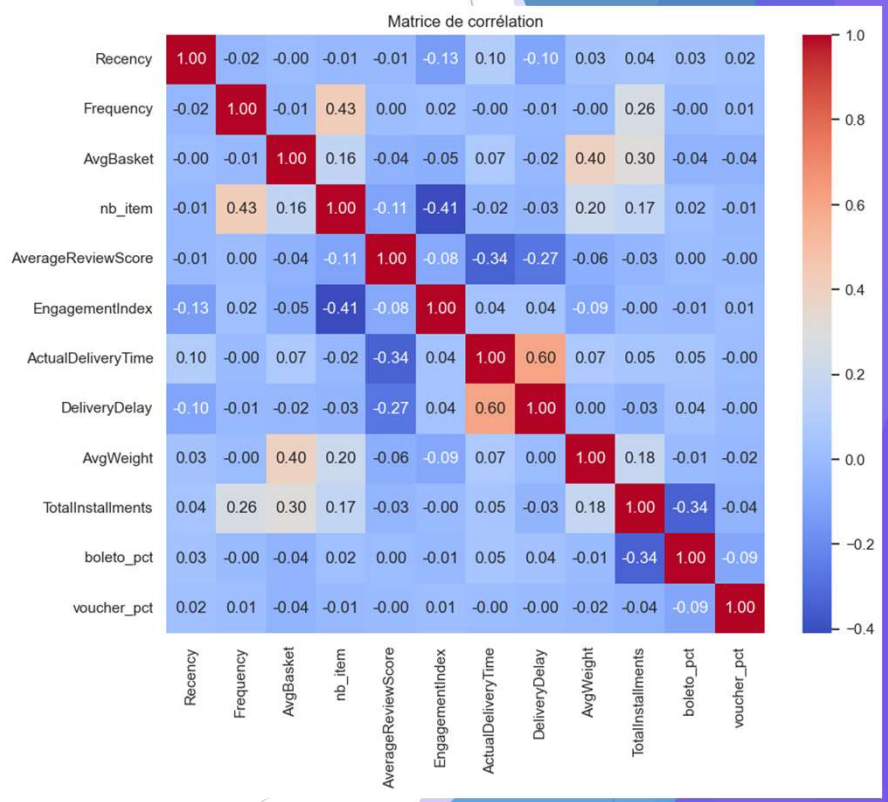
- Distribution des variables numériques



4. Analyse exploratoire et prétraitement des données

- Analyse des corrélations, ACP et sélection finale des indicateurs





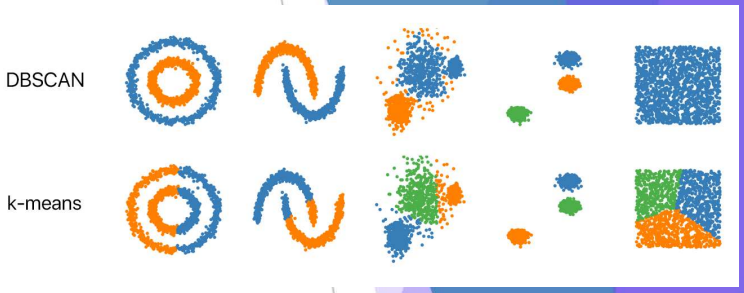
5. Segmentation : Approches testées

- 7 indicateurs
  - 'Recency',
  - 'Frequency',
  - 'AvgBasket',
  - 'nb_item',
  - 'AverageReviewScore',
  - 'EngagementIndex',
  - 'DeliveryDelay'

- 9 indicateurs : 7 indicateurs +
  - 'boleto_pct',
  - 'voucher_pct'

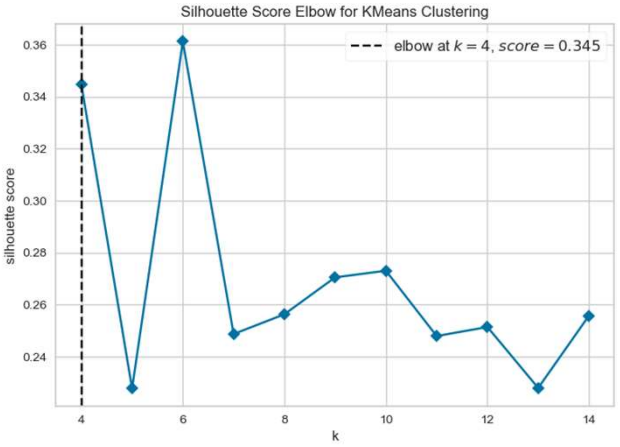
- 12 indicateurs : 9 indicateurs +
  - 'AvgWeight',
  - 'ActualDeliveryTime',
  - 'TotalInstallments'

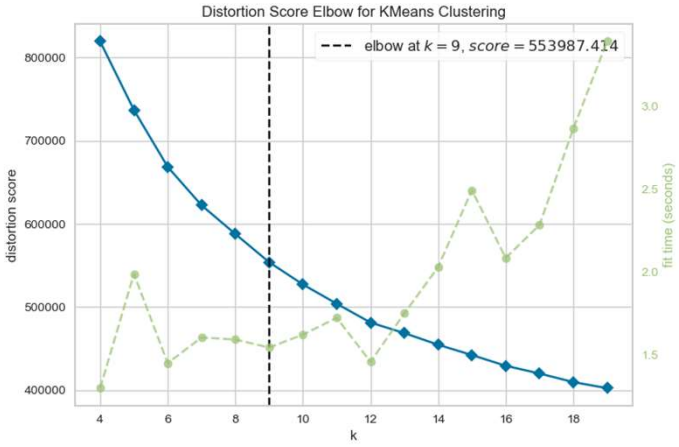
- 3 algorithmes testés
  - K-means++
  - DBScan
  - K-means + Hiérarchique

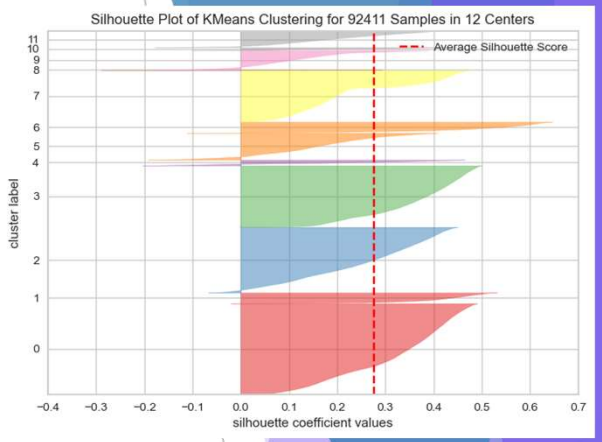


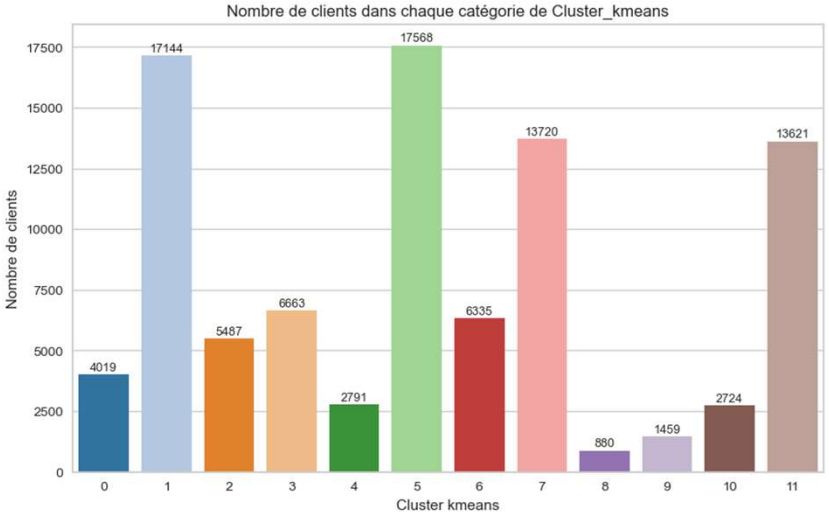
- Métriques
  - Silhouette score
  - Distortion score
  - Score de Davies-Bouldin
  - Taille des clusters
  - **Pertinence des profils**

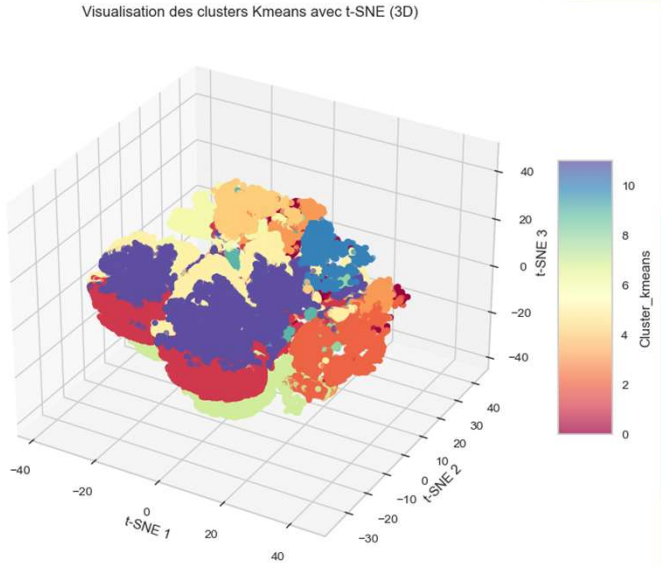
6. Modèle final



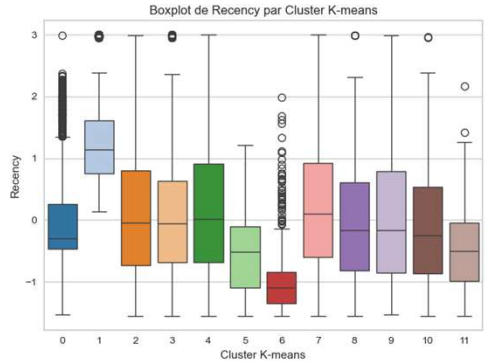


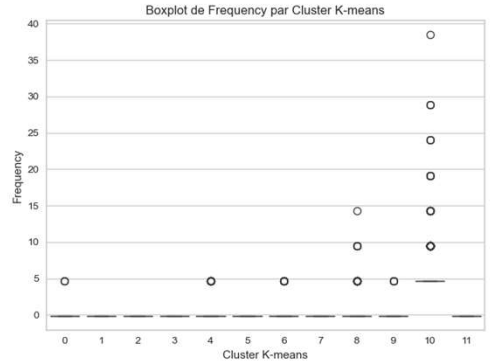


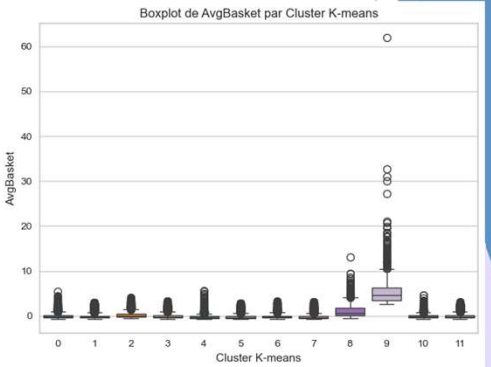


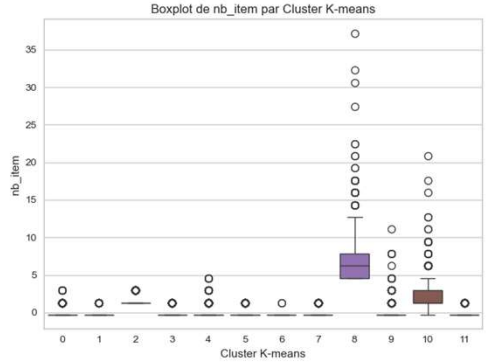


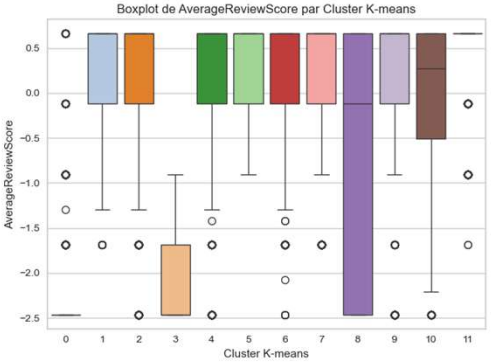
7. Profils

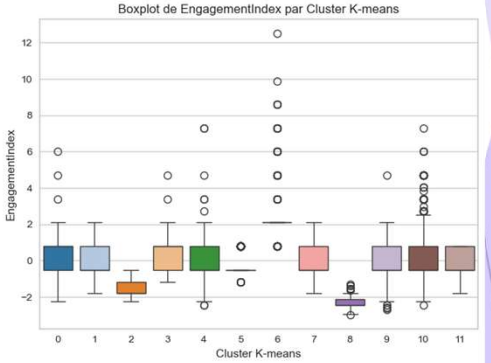

Boxplot de Recency par Cluster K-means
Recency
Cluster K-means

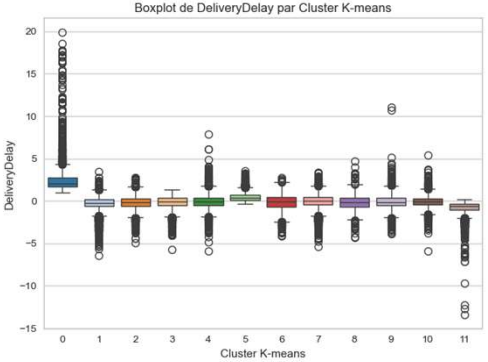

Boxplot de Frequency par Cluster K-means
Frequency
Cluster K-means

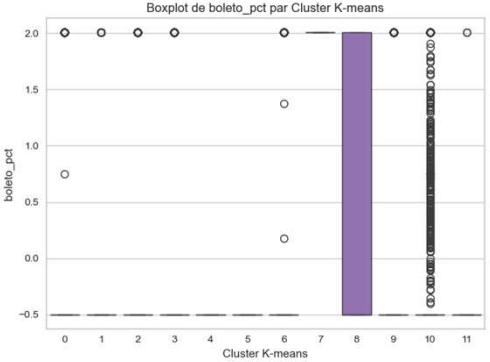

Boxplot de AvgBasket par Cluster K-means
AvgBasket
Cluster K-means

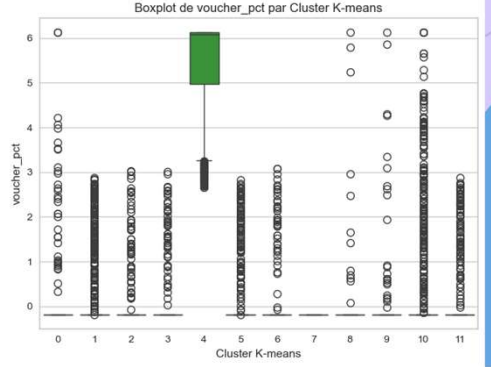

Boxplot de nb_item par Cluster K-means
nb_item
Cluster K-means


Boxplot de AverageReviewScore par Cluster K-means
AverageReviewScore
Cluster K-means


Boxplot de EngagementIndex par Cluster K-means
EngagementIndex
Cluster K-means

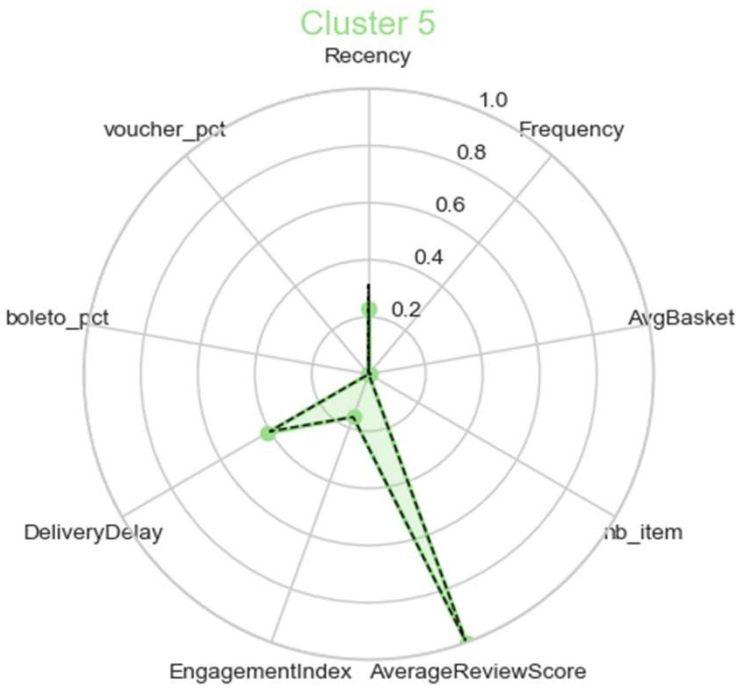

Boxplot de DeliveryDelay par Cluster K-means
DeliveryDelay
Cluster K-means


Boxplot de boleto_pct par Cluster K-means
boleto_pct
Cluster K-means

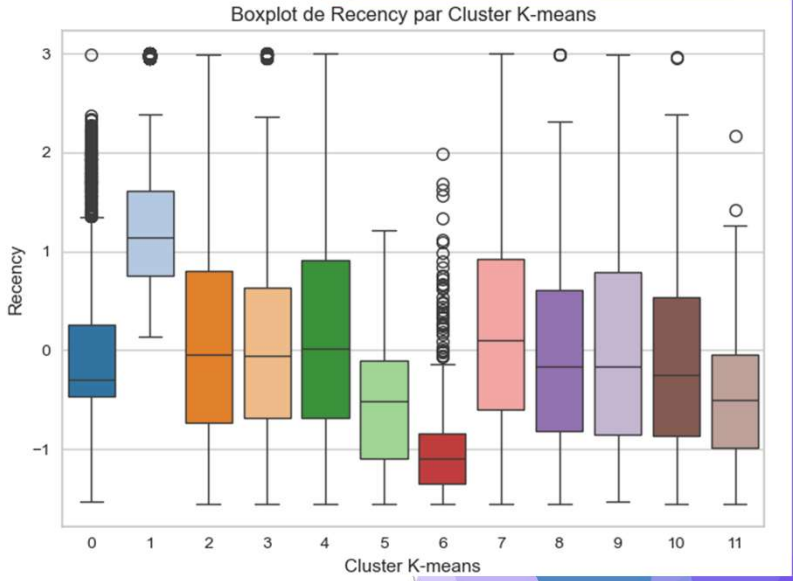

Boxplot de voucher_pct par Cluster K-means
voucher_pct
Cluster K-means

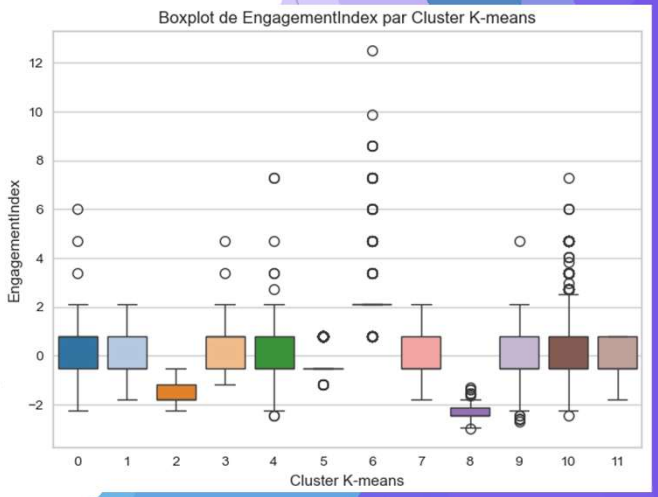
7. Profils

Clients Actifs Récemment



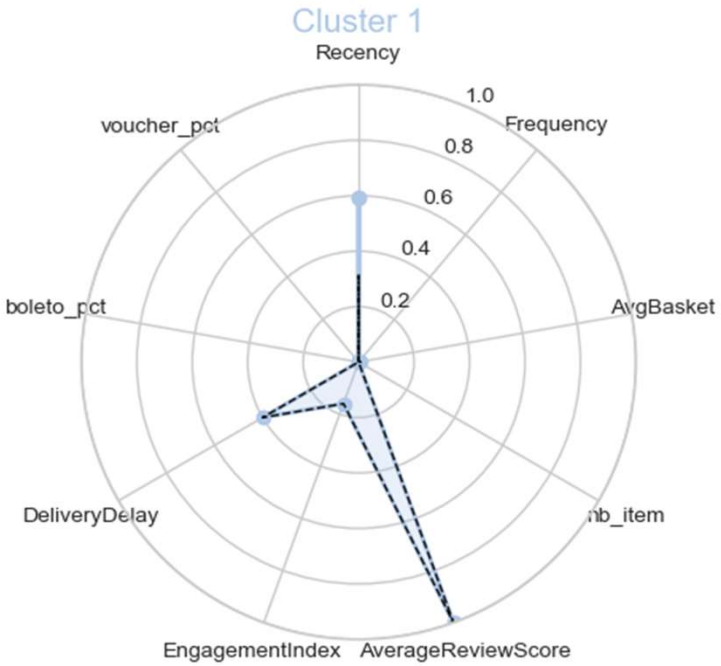
- Particularités :
  - Commande récente
- Proportion :
  - 17568 clients
  - 19 %



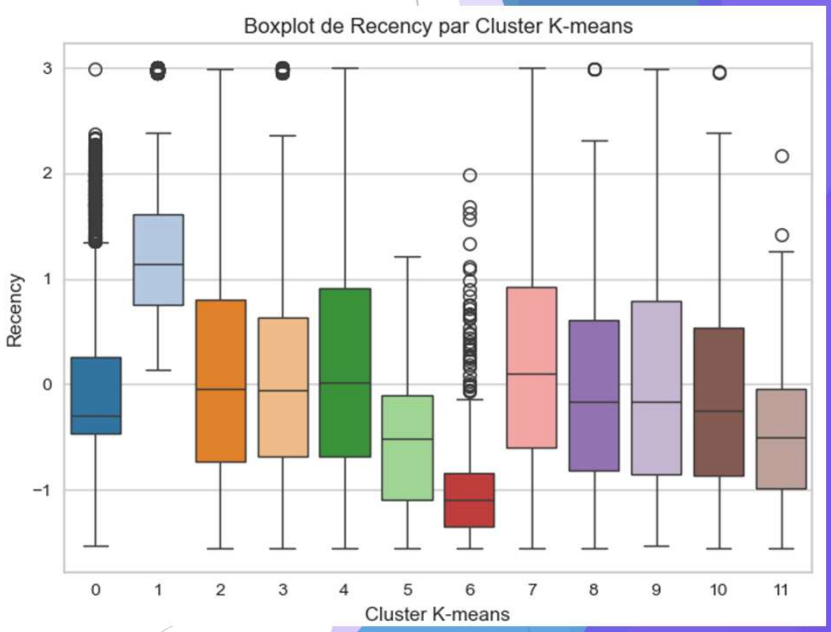


7. Profils

Clients Inactifs

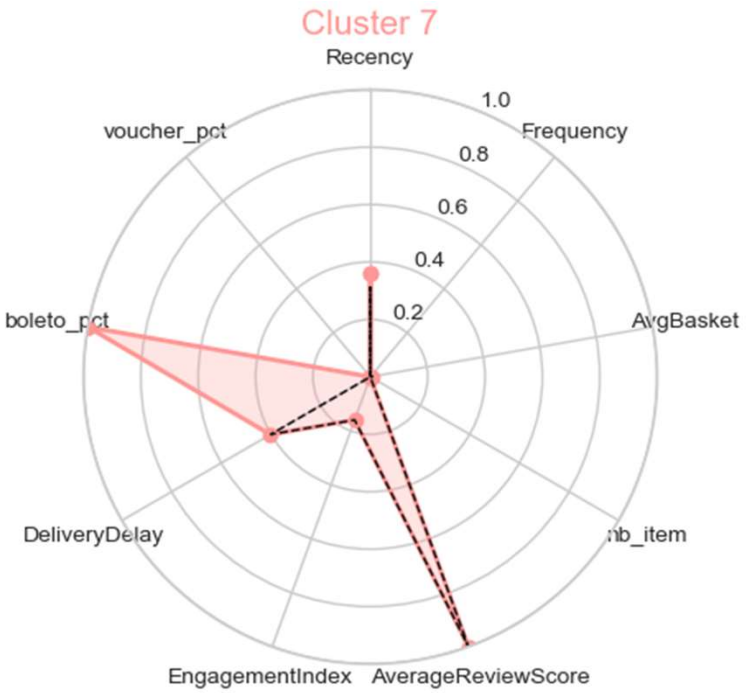


- Particularités :
  - Pas de commandes depuis longtemps
- Proportion :
  - 17144 clients
  - 18,6 %



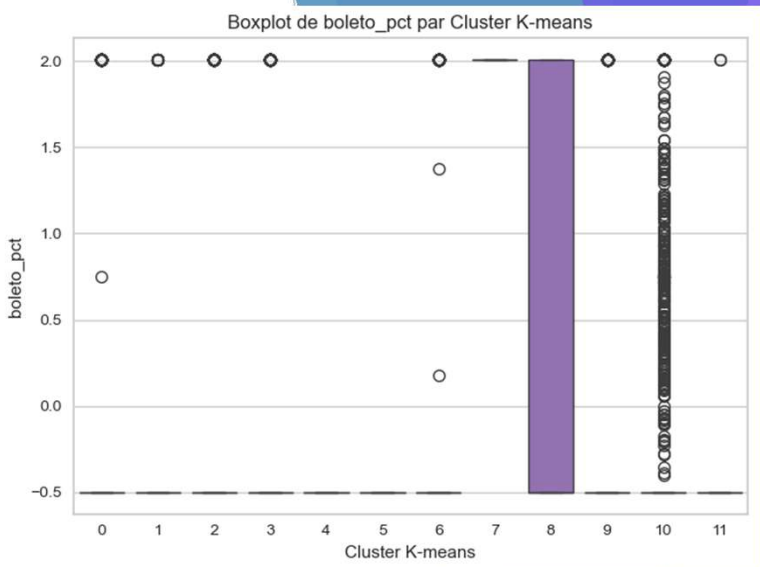
7. Profils

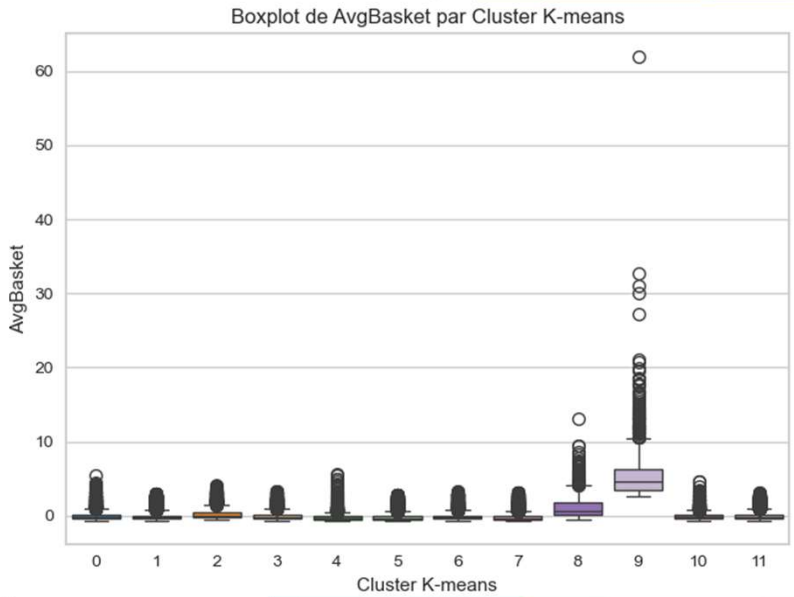
Les économiques



- Particularités :
  - Paye en Boletos

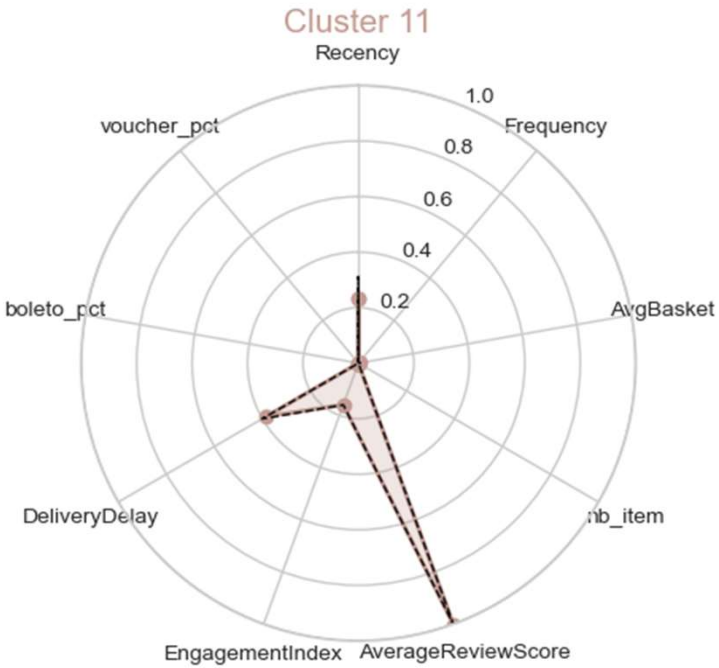
- Proportion :
  - 13720 clients
  - 14,8 %





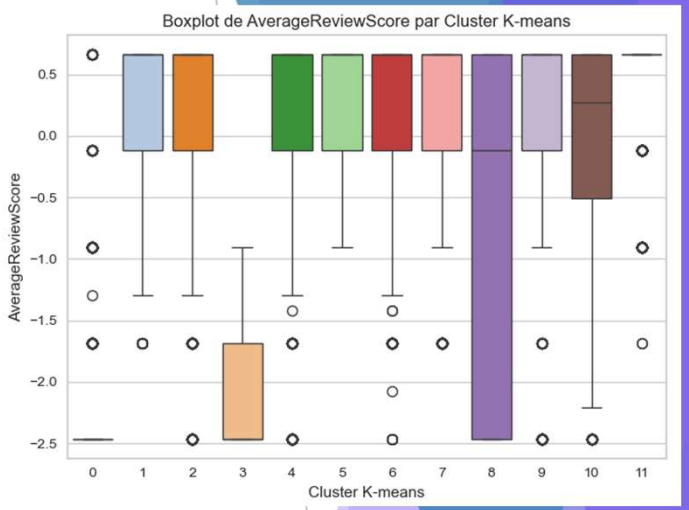
7. Profils

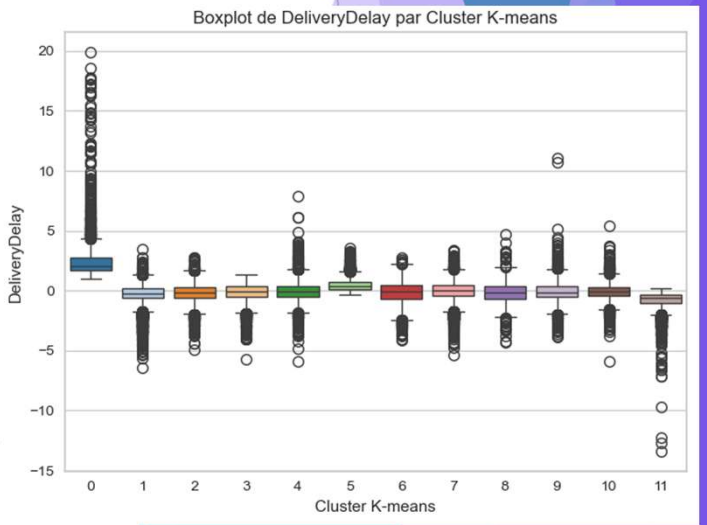
Les Satisfaits Privilégiés



- Particularités :
  - Satisfaction très élevée
  - Livraisons en avance

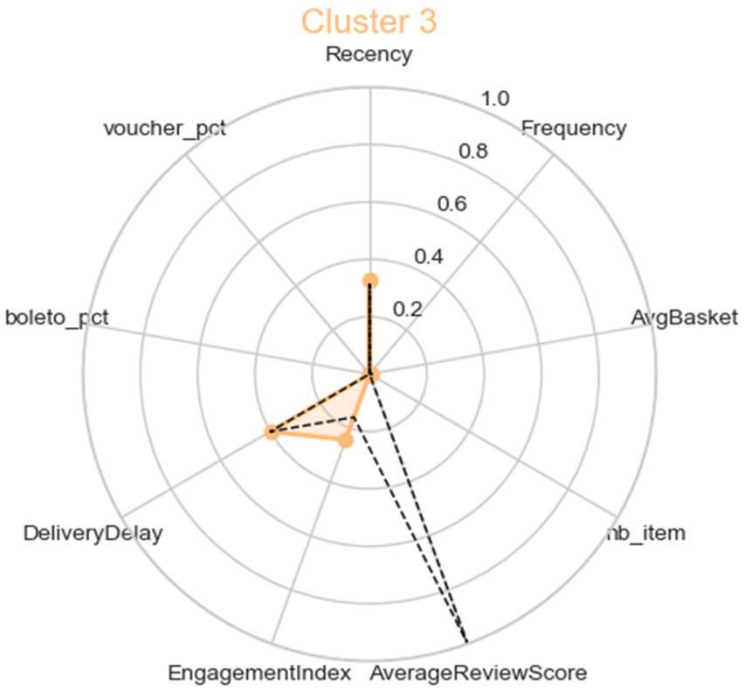
- Proportion :
  - 13621 clients
  - 14,7 %





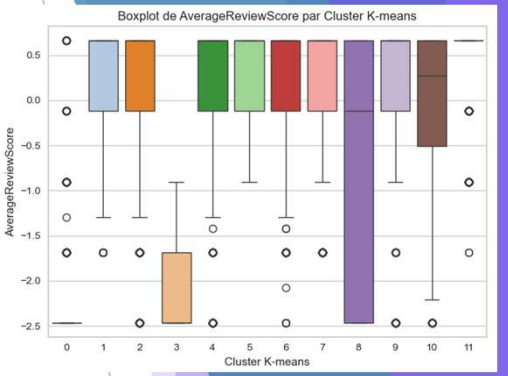
7. Profils

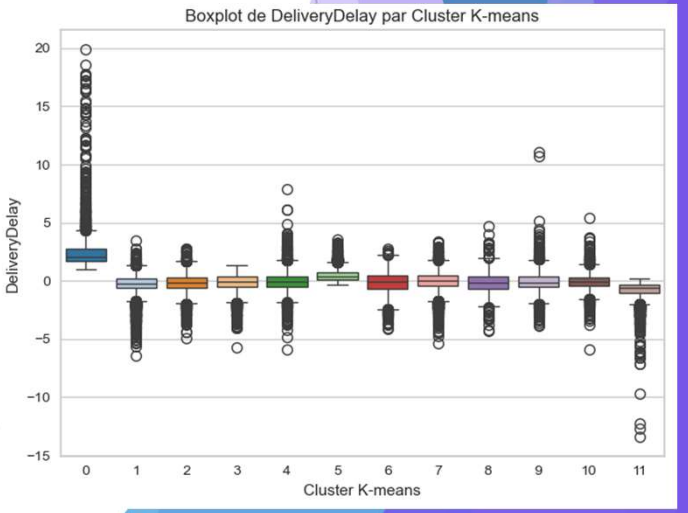
Commentateurs Actifs Non Satisfaits



- Particularités :
  - Clients mécontents
  - Pas de retards de livraison
  - Laisse plus de commentaires que la médiane

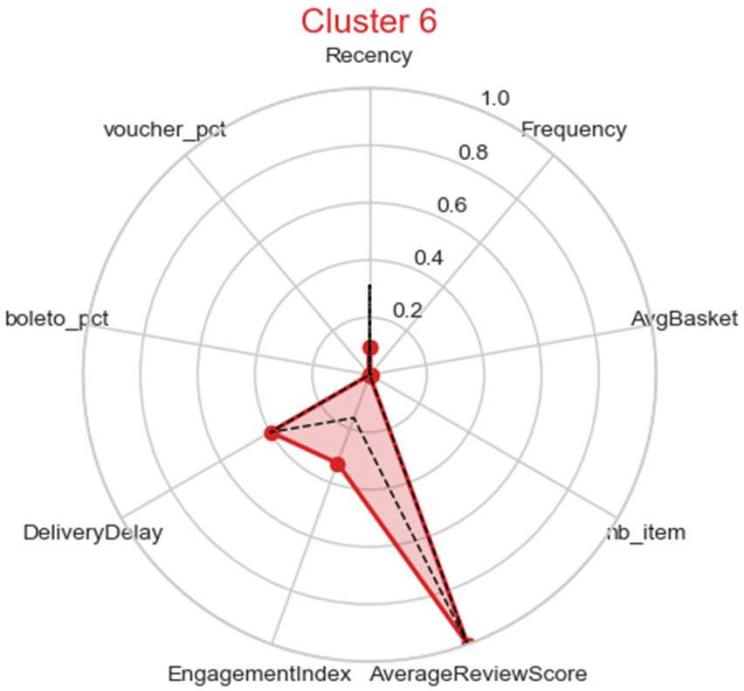
- Proportion :
  - 6663 clients
  - 7,2 %





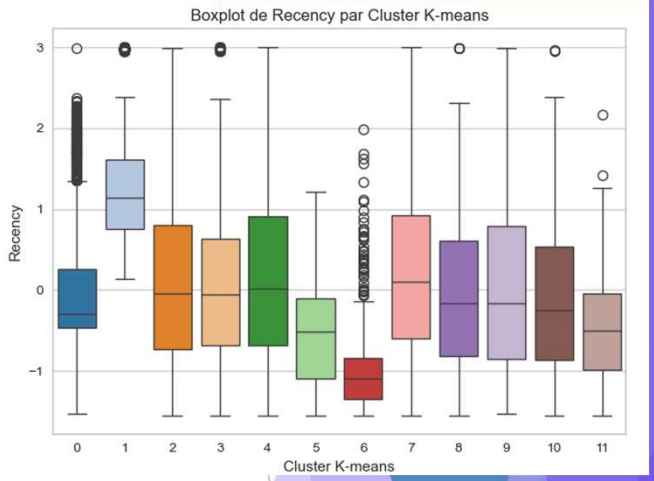
7. Profils

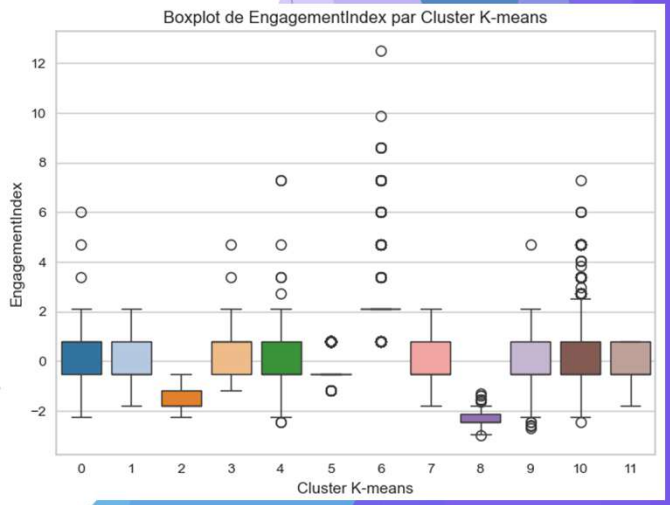
Commentateurs Actifs Récents



- Particularités :
  - Commande très récente
  - Laisse beaucoup de commentaires

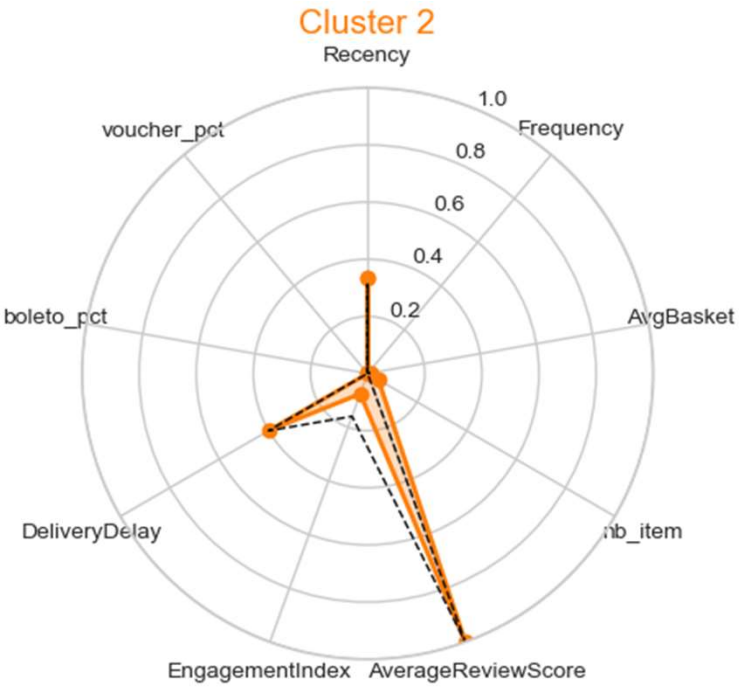
- Proportion :
  - 6335 clients
  - 6,9 %





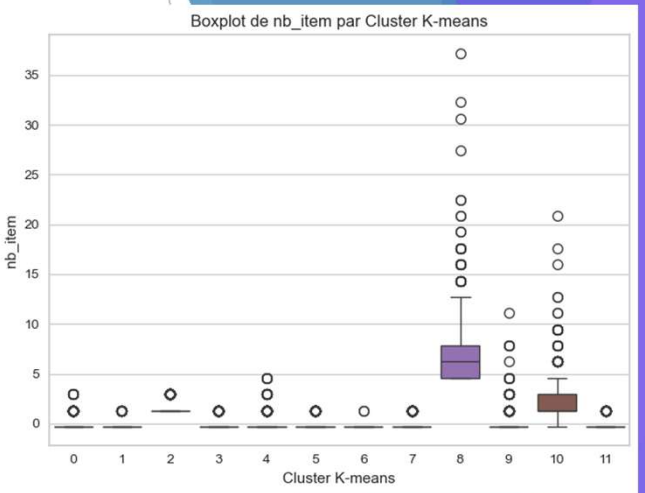
7. Profils

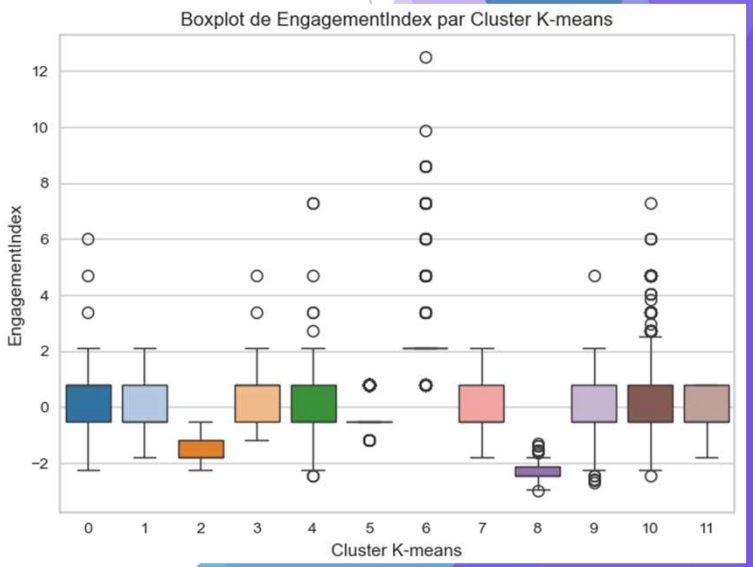
Acheteurs Silencieux



- Particularités :
  - Plus d'items que la médiane
  - Très peu de commentaires laissés

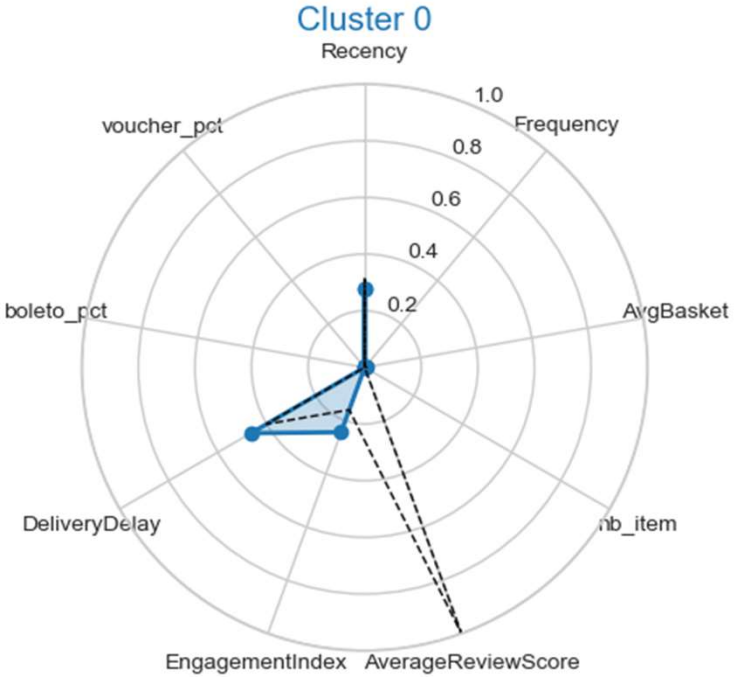
- Proportion :
  - 5487 clients
  - 5,9 %





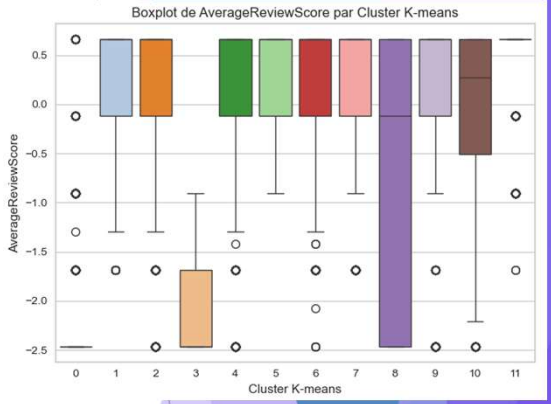
7. Profils

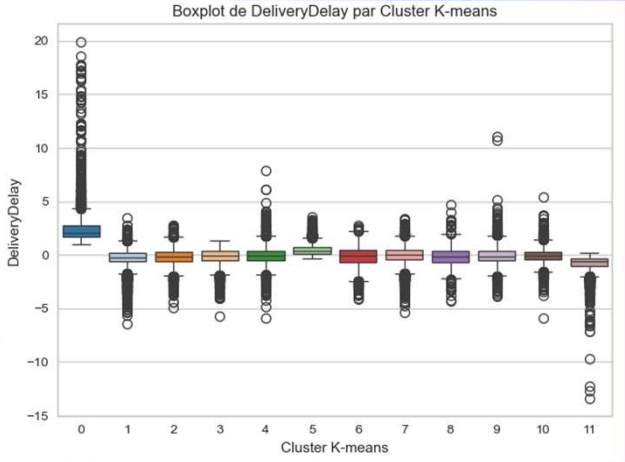
Clients Insatisfaits et Expressifs



- Particularités :
  - Satisfaction très faible
  - Gros retards de livraison
  - Laisse plus de commentaires que la médiane

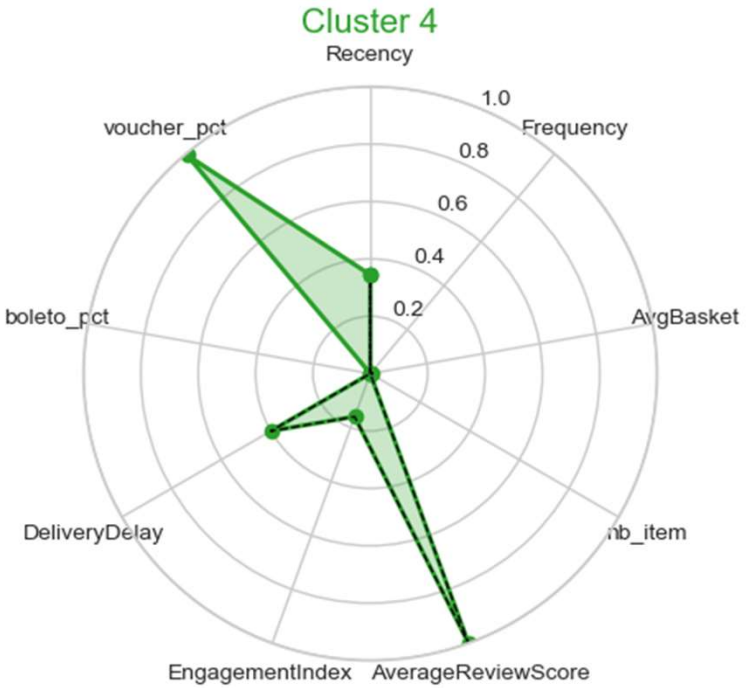
- Proportion :
  - 4019 clients
  - 4,3 %





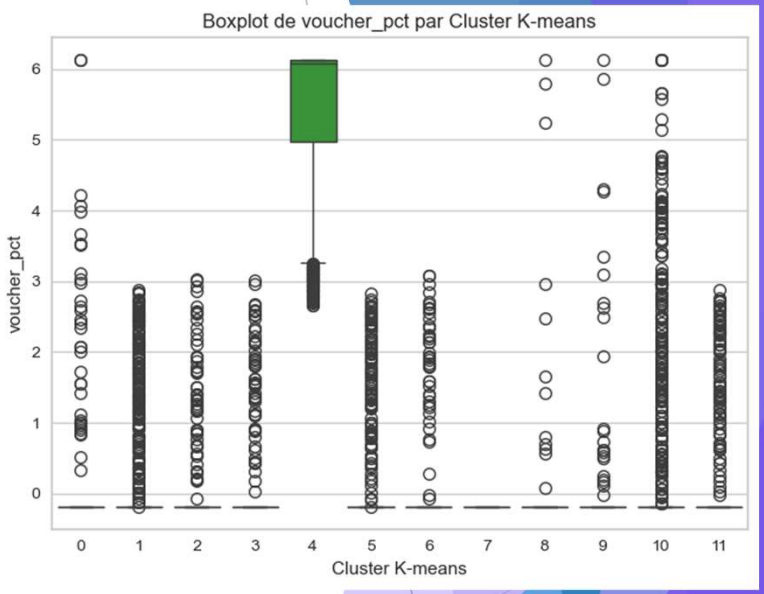
7. Profils

Voucher



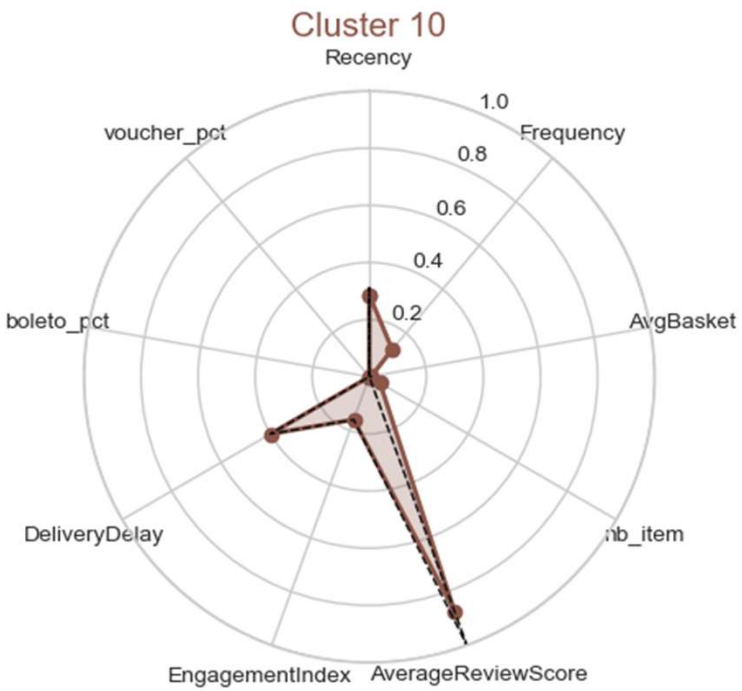
- Particularités :
  - Payent en vouchers

- Proportion :
  - 2791 clients
  - 3 %



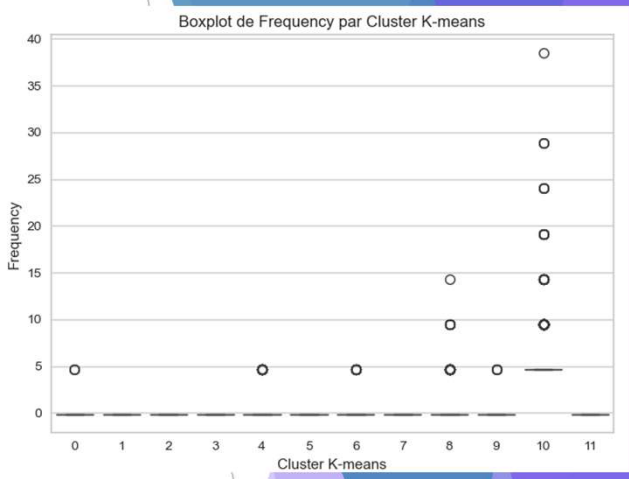
7. Profils

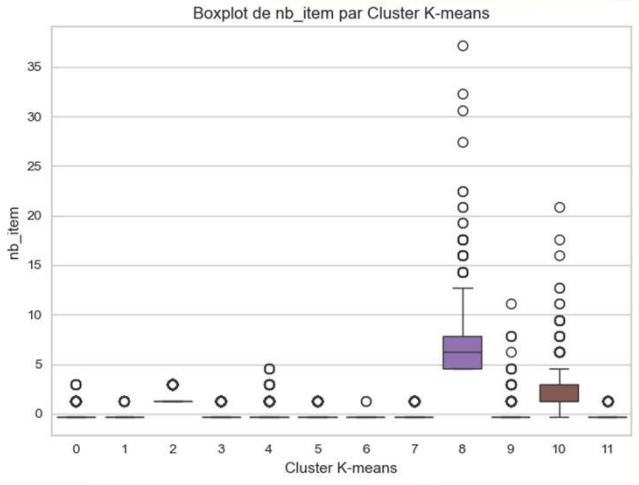
Clients Fidèles



- Particularités :
  - Achats fréquents
  - Plus d'items que la médiane

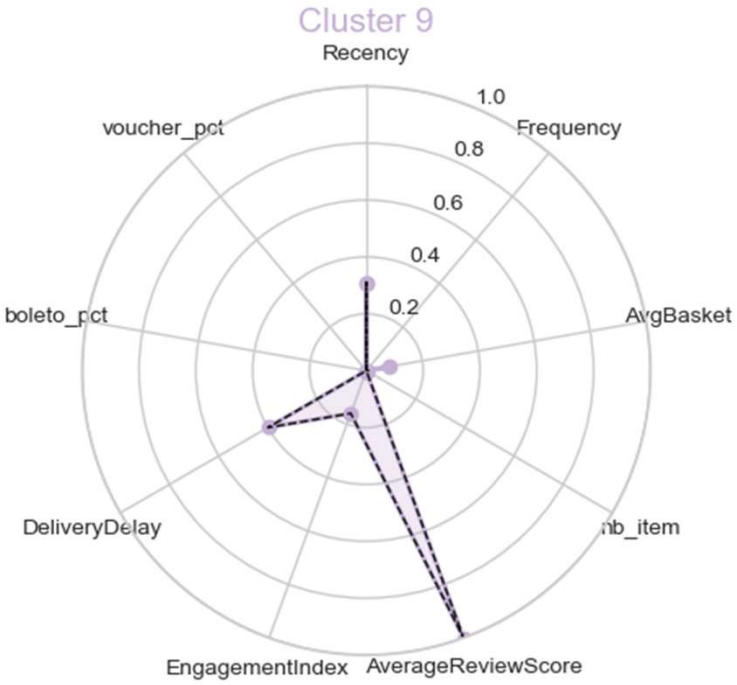
- Proportion :
  - 2724 clients
  - 2,9 %



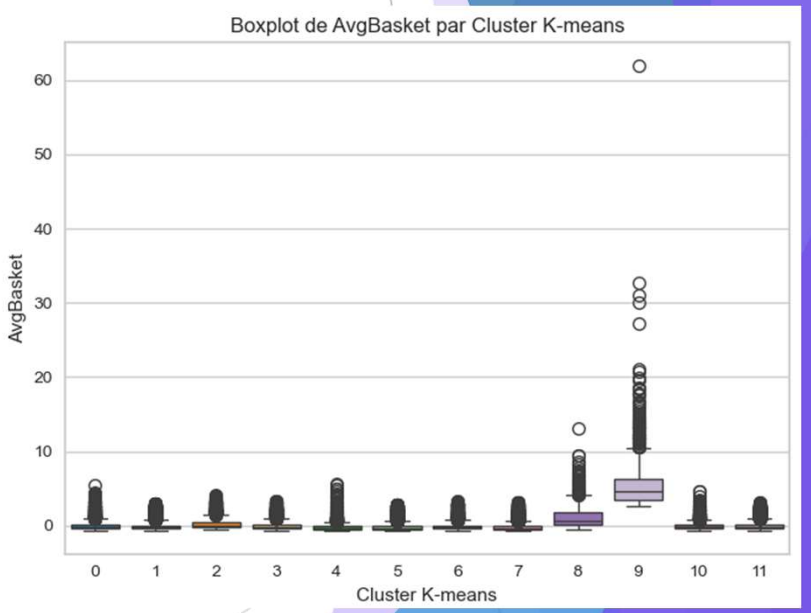


7. Profils

Clients à Forte Valeur

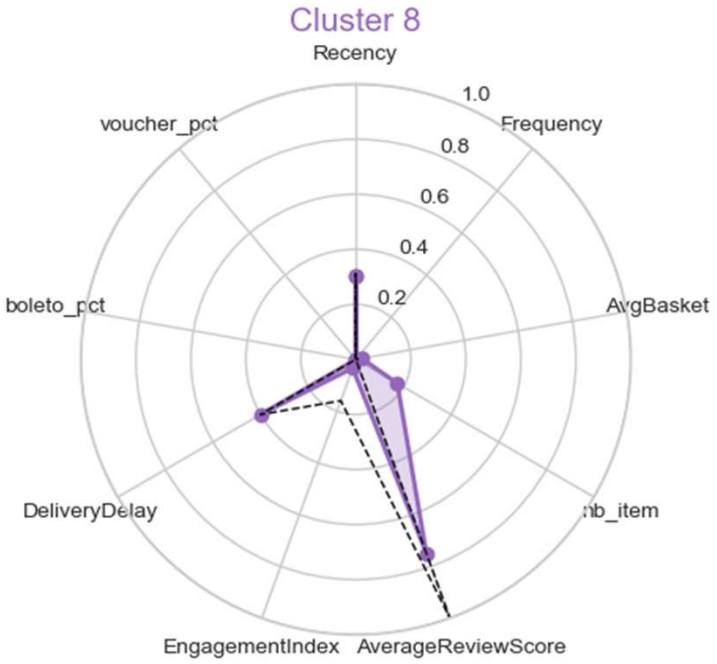


- Particularités :
  - Panier moyen beaucoup plus élevé
- Proportion :
  - 1459 clients
  - 1,6 %



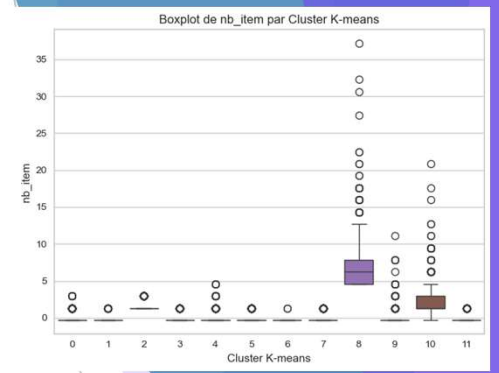
7. Profils

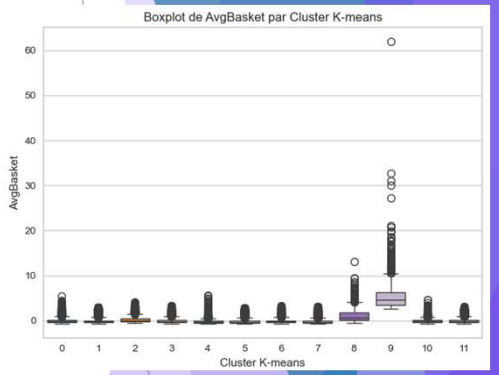
Grands Acheteurs Discrets

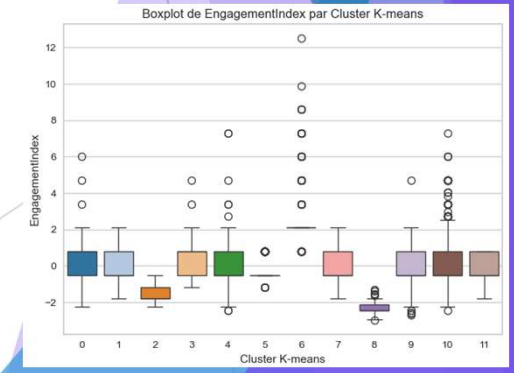


- Particularités :
  - Achète beaucoup d'item
  - Panier moyen plus élevé
  - Laisse très peu de commentaires

- Proportion :
  - 880 clients
  - 1 %



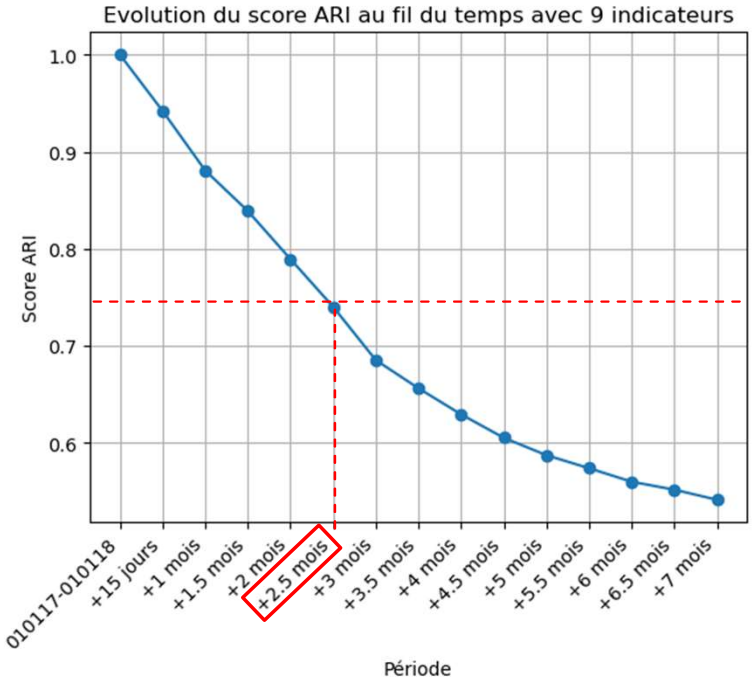


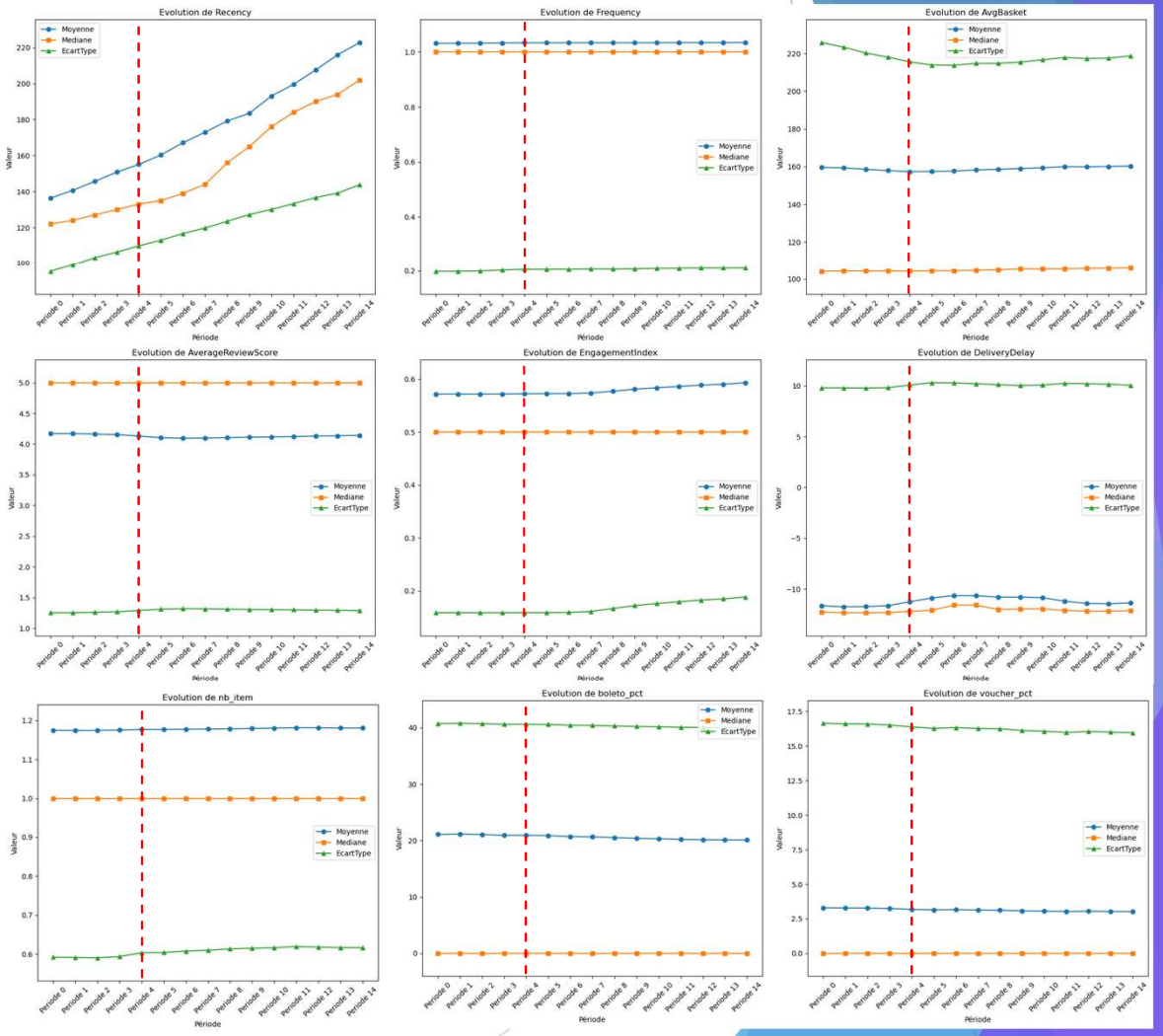


8. Stratégie de mise à jour et maintenance

- Ajout de nouveaux clients
  - Pipeline de traitement des données
  - Application du modèle

- Mise à jour tous les 2 mois et demi recommandée





Conclusion et Recommendations

- Synthèse
  - **Segmentation Éclairée** : Identification de 12 clusters significatifs via le modèle K-means++, combinant pertinence statistique et insights métier.
  - **Intelligence Client Accrue** : Profilage détaillé pour une approche marketing affinée et personnalisée.
- Recommandations
  - **Personnalisation Marketing** : Adapter les communications en fonction des clusters pour augmenter l'engagement.
  - **Réévaluation Régulière** : Revisiter le modèle tous les 2.5 mois maximums pour maintenir la pertinence des segments.
  - **Intégration Continue** : Intégrer systématiquement les nouveaux clients dans les segments existants.
  - **Surveillance des Métriques** : Monitorer les indicateurs clés pour une optimisation continue.
  - **Analyse Approfondie** : Étudier les segments exceptionnels pour des actions ciblées.
  - **Réaction Proactive** : Investiguer et répondre activement au feedback des segments critiques.
- Prochaines Étapes
  - Utilisation des insights pour des campagnes marketing ciblées.
  - Exploration d'opportunités de croissance basées sur la segmentation client.

Merci pour votre attention

Des questions ?