

Analyse

Olist

Segmentez des clients d'un site e-commerce

Sofiane Mouhab
10 juin 2021

1 - Généralités

1.1 - Problématique

Notre client **Olist** souhaite fournir à ses équipes une segmentation des clients afin de les utiliser au quotidien pour des campagnes de communication.

1.2 - Objectif

- Comprendre les différents types d'utilisateurs grâce à leur comportement et à leurs données personnelles.
- Fournir à l'équipe marketing une description actionable des segmentations et des logiques sous-jacente pour une utilisation optimale
- Proposition de contrat de maintenance basée sur une analyse de la stabilité des segments au cours du temps.

1.3 - Condition de mise en oeuvre

Pour pouvoir sereinement réaliser ses trois objectifs, il nous faut donc diverses informations qui pourrait se trouver dans notre base de données.

À nous donc, d'examiner celle-ci, de déterminer à quel point les informations sont viables, ou perfectible.

Il y a donc 3 grandes interrogations :

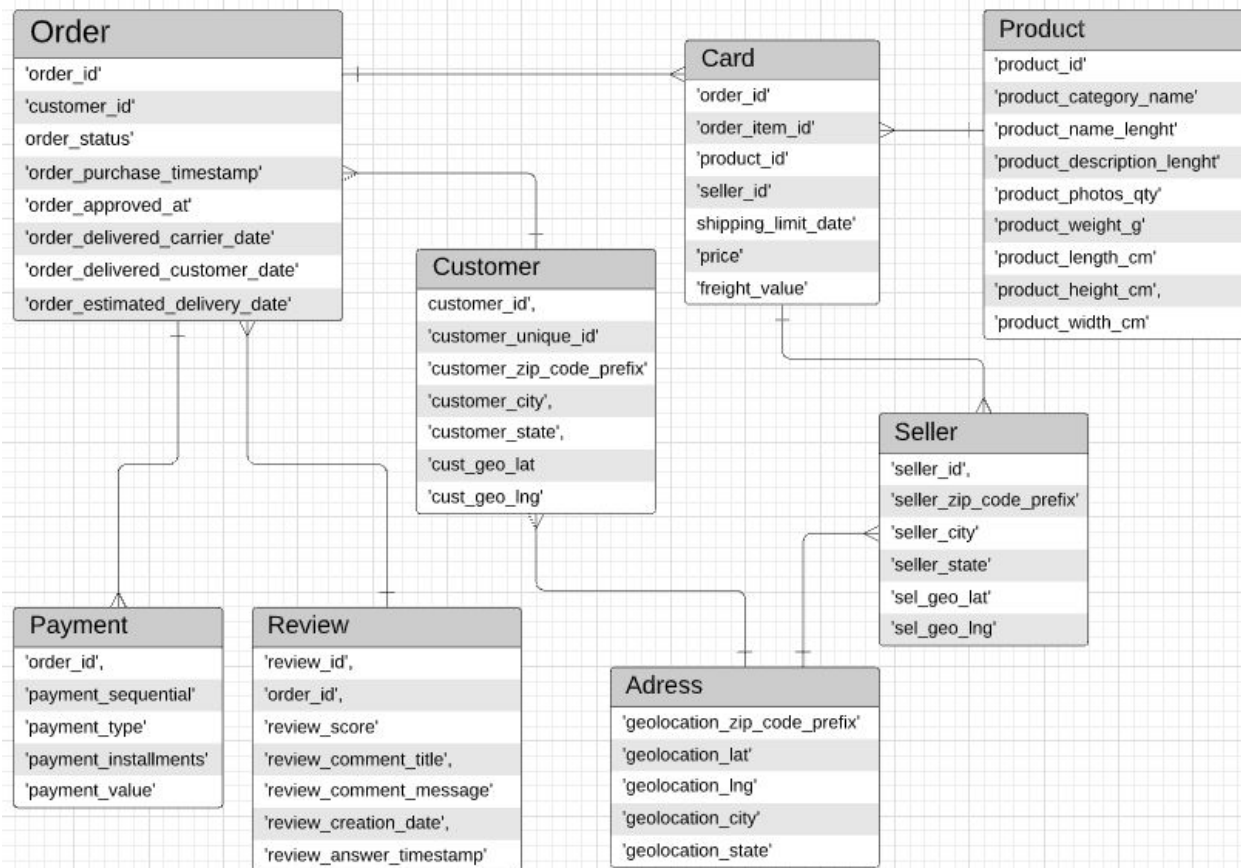
- A-t-on assez de données ?
- Peut-on faire des clusters cohérente ?
- Peut-on faire des clusters durables et adaptables ?

Passons de suite à ce travail, en commençant par rapidement prendre connaissance des données en présence...

2 - Les données

2.1 - Description

Dans ce fichier volumineux nous comptons 8 Fichiers (typique du langage SQL) :

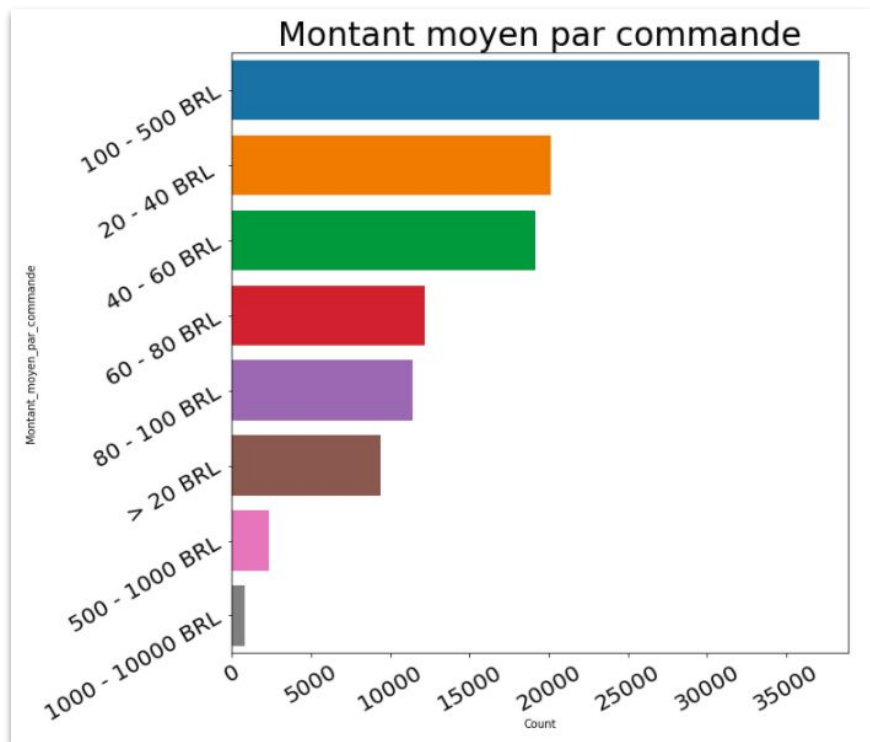


Chiffres clés :

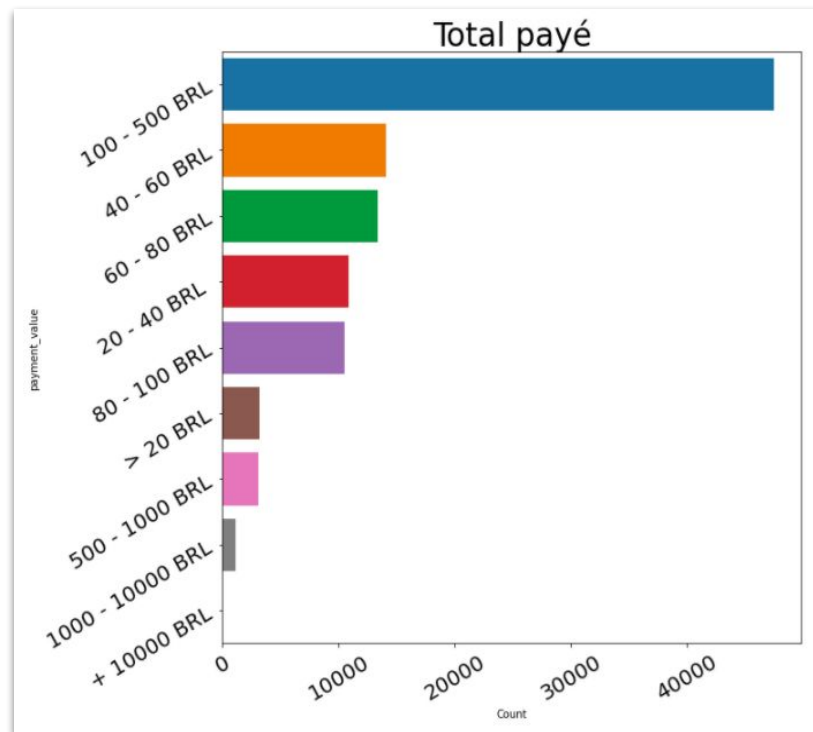
- 96 000 clients
- 3 000 vendeurs
- 99 000 commandes
- 113 000 ventes réussies
- 104 000 paiements
- 33 000 produits vendus
- 8 000 villes et leurs coordonnées

3 - Analyse

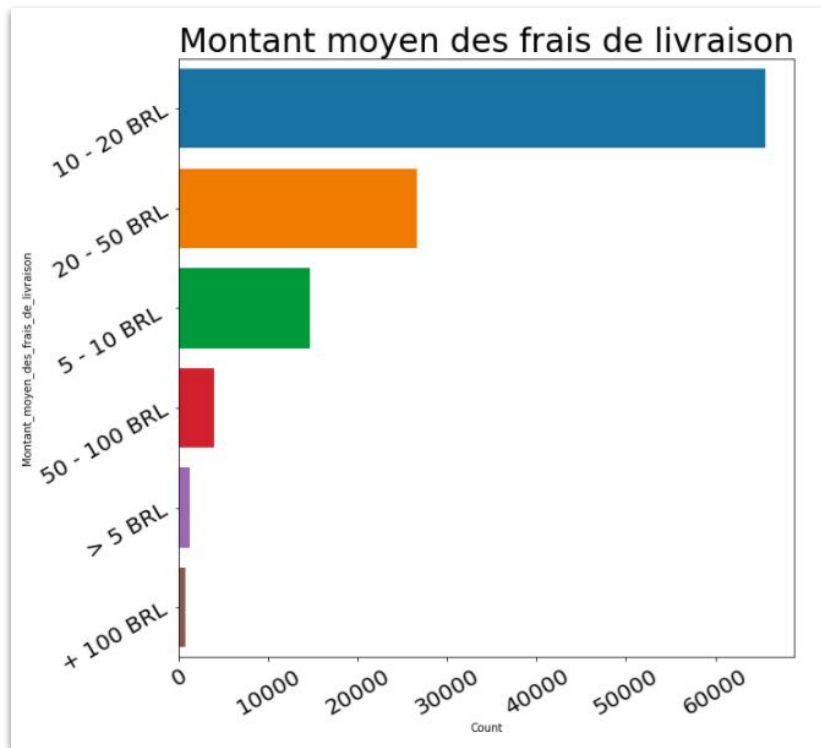
3.1 - Montant moyen par commande



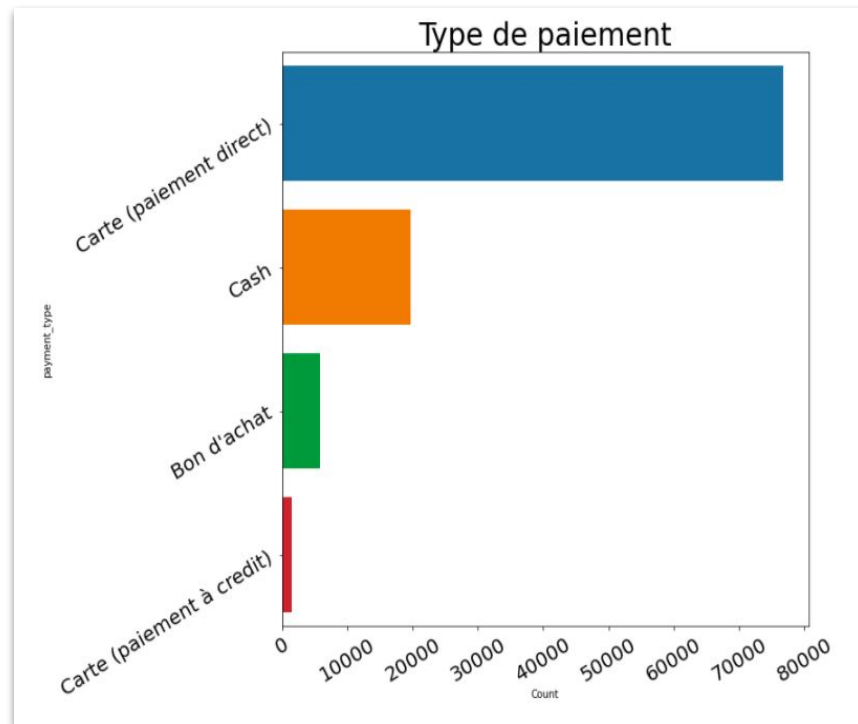
3.2 - Dépense totale par client



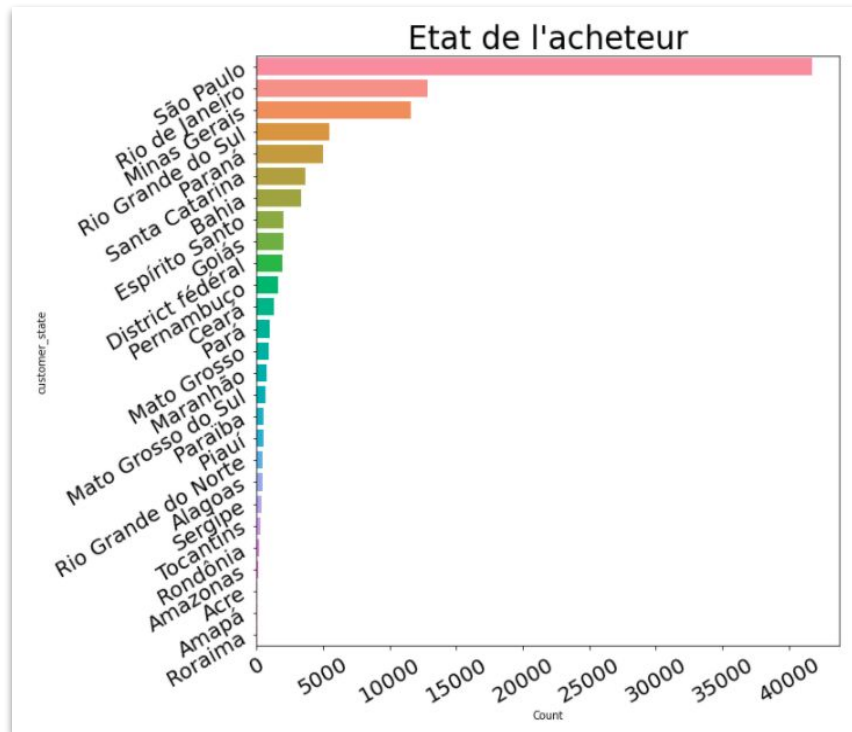
3.3 - Montant moyen des frais de livraison



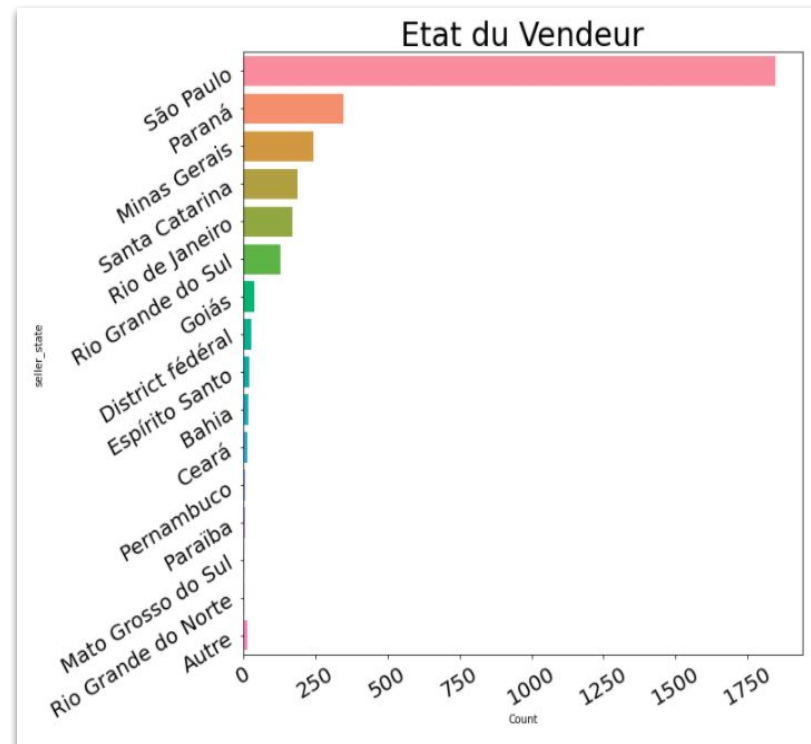
3.4 - Type de paiement



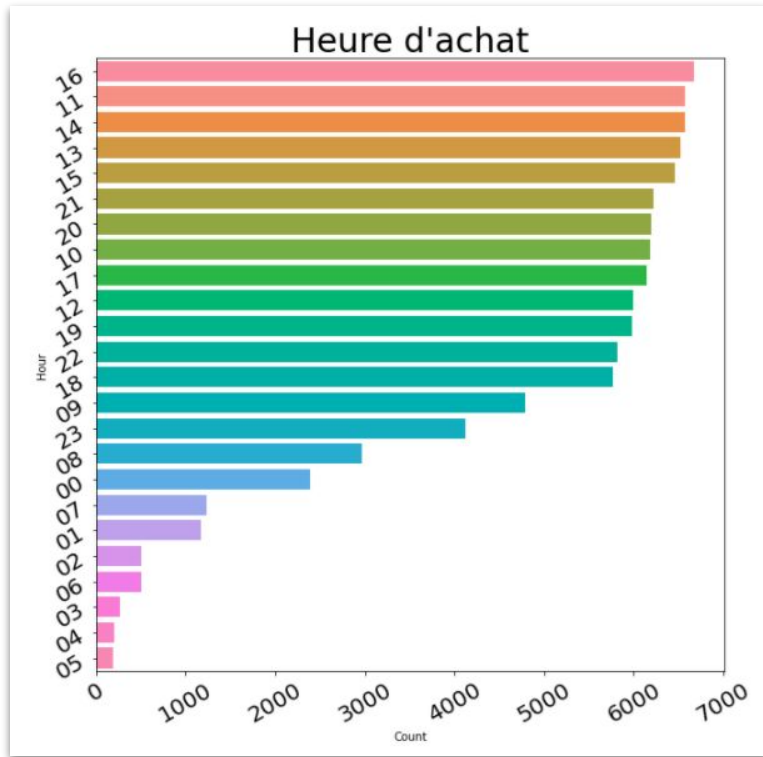
3.5 - Localisation du client



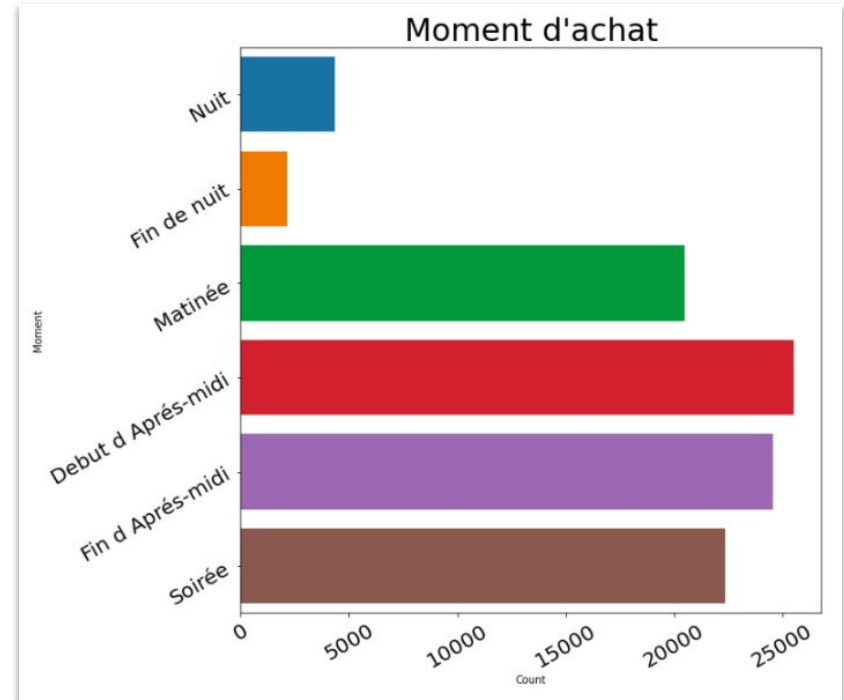
3.6 - Localisation du vendeur



3.7 - Heure d'achat



3.8 - Moment d'achat



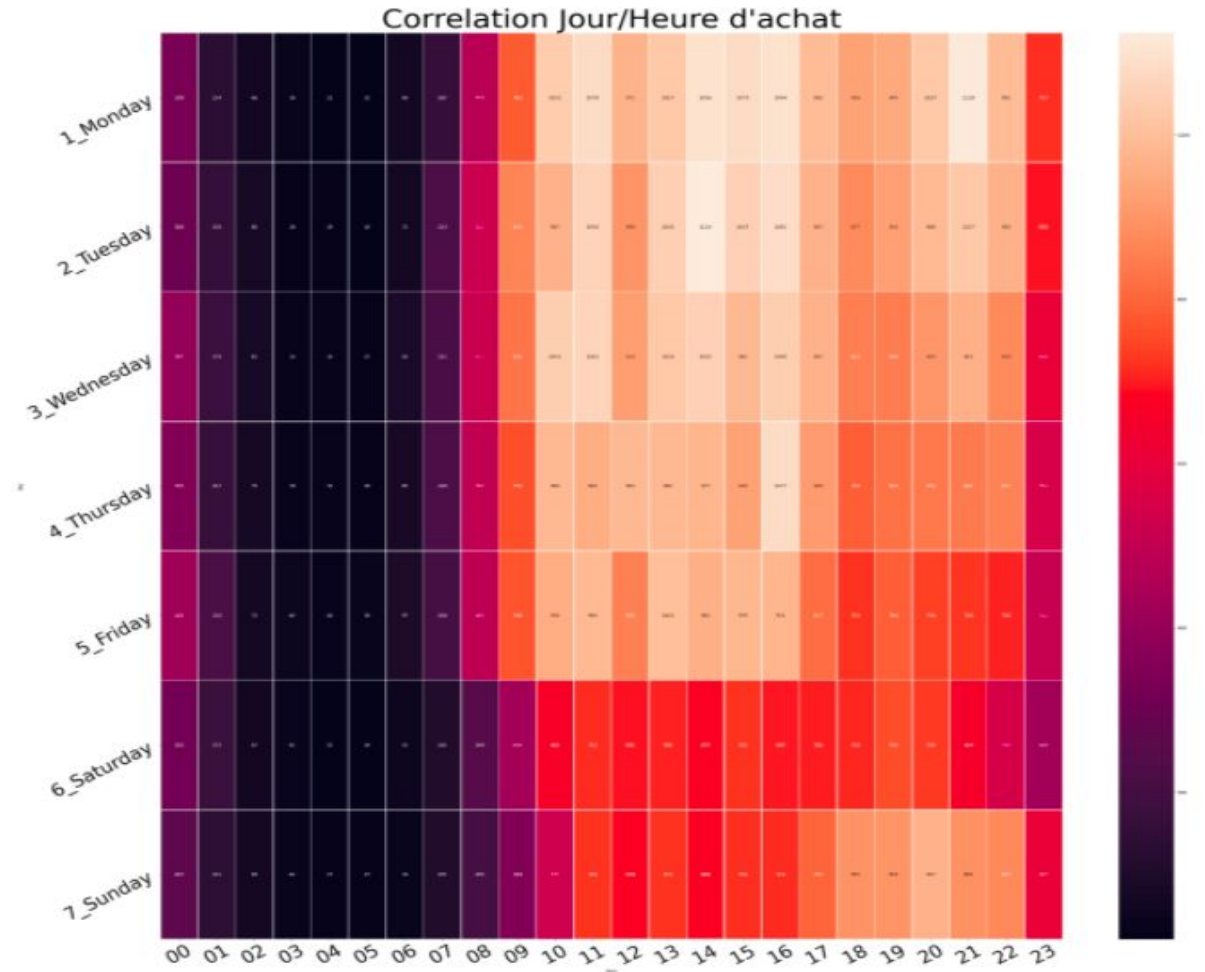
3.14

-

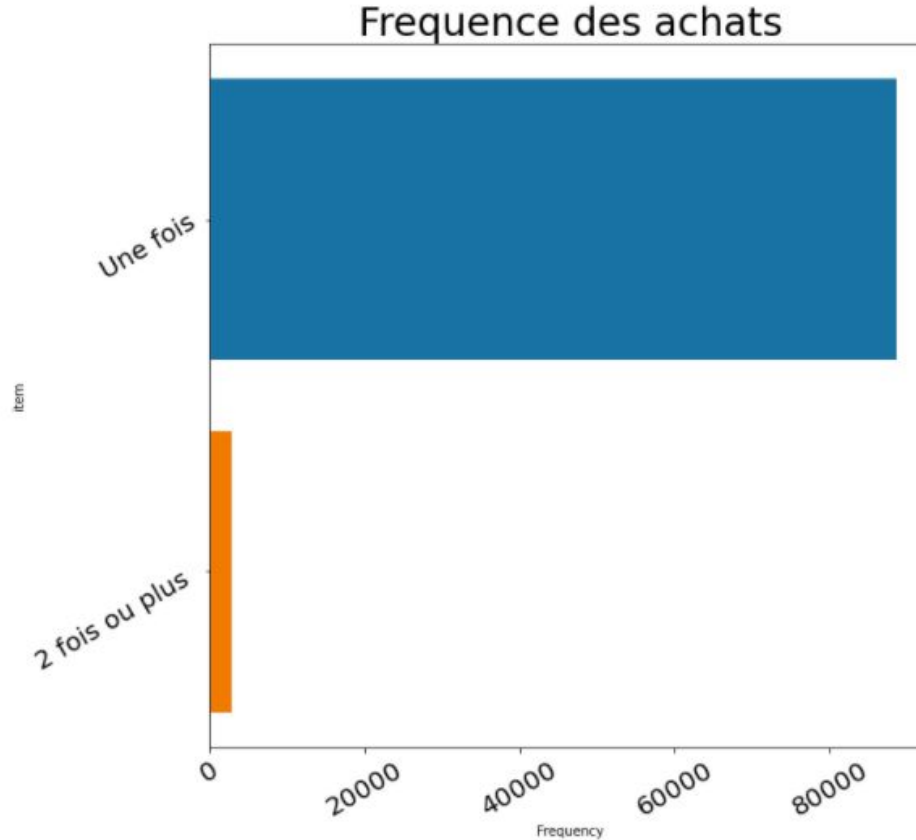
HeatMap

-

Intensité des achats sur une
semaine



3.15 - Nombre de commandes par client



4 - Classification

4.1.1 - Segmentation RFM

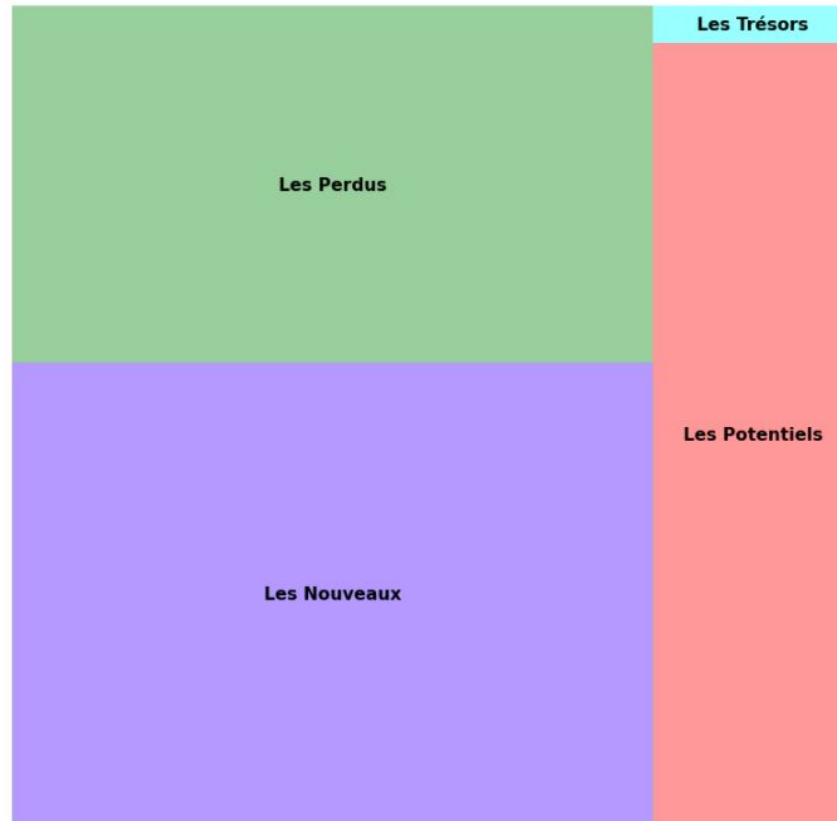
La segmentation **RFM** est une méthode très utilisée, elle permet un classement des clients autour de 3 paramètres. On peut ainsi optimiser une stratégie marketing en fonction de nos clients.

- **Récence** : date du dernier achat
- **Fréquence** : fréquence des achats
- **Montant** : somme des achats cumulés

Etant donné la faible fréquence d'achat sur votre site, cette méthode ne s'adapte pas à notre étude.

Nous pouvons néanmoins établir un rapide diagnostic sous le prisme “ RFM”

4.1.2 - Segmentation RFM



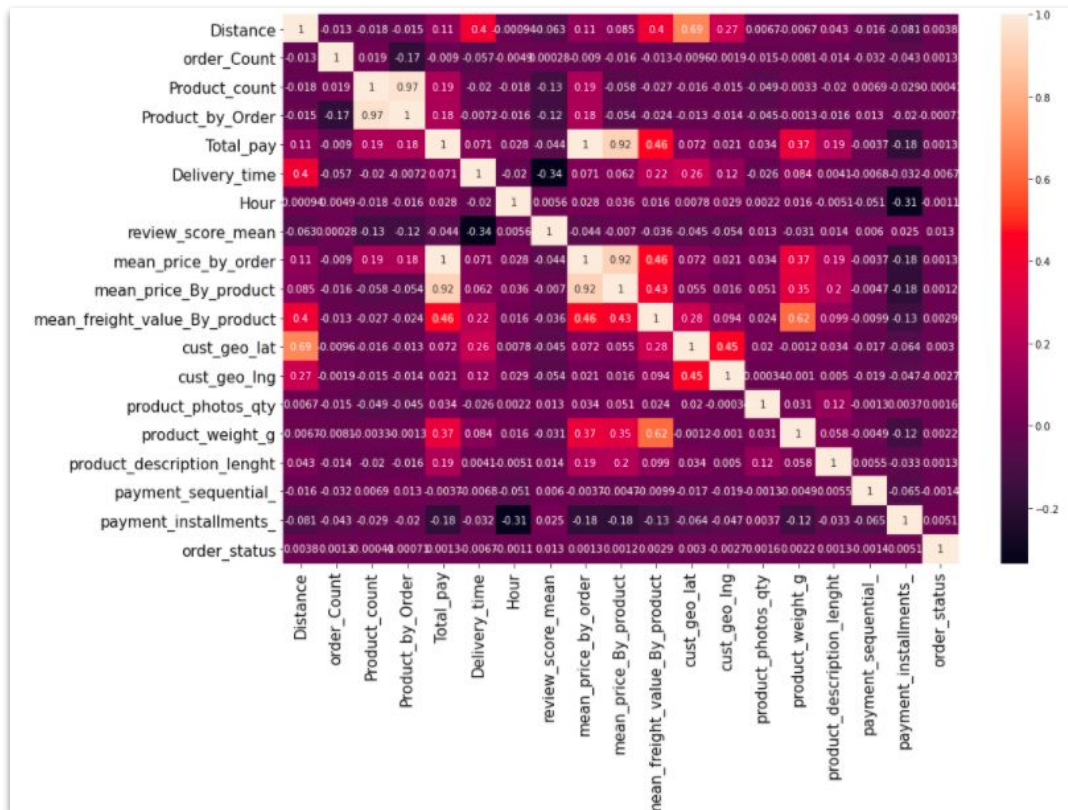
4.2 - K Means

4.2.1 - Variable en présence

- Distance Client/Vendeur
- Délais de Livraison
- Nombre de produit par Commande
- Catégorie du produit
- Type de paiement
- Heure d'achat
- Jour d'achat
- Avis du produit
- Prix moyen d'un panier créé
- Prix moyen d'une commande achetée
- Poids moyen d'un produit acheté
- Latitude
- Longitude
- Ville du client
- Nombre de photo pour les produits achetés
- Longueur de la description les produits achetés
- Statuts de la commande à date
- Mois d'achat
- Nombre de paiement à crédit
- Nombre de paiement échelonné

4.2.2 - Variable retenue

- **Nombre de produit par Commande**
- **Catégorie du produit**
- **Note de l'utilisateur**
- **Somme dépensé**
- **Geolocalisation**
- **Date du dernier achat**
- **Fréquence d'achat**

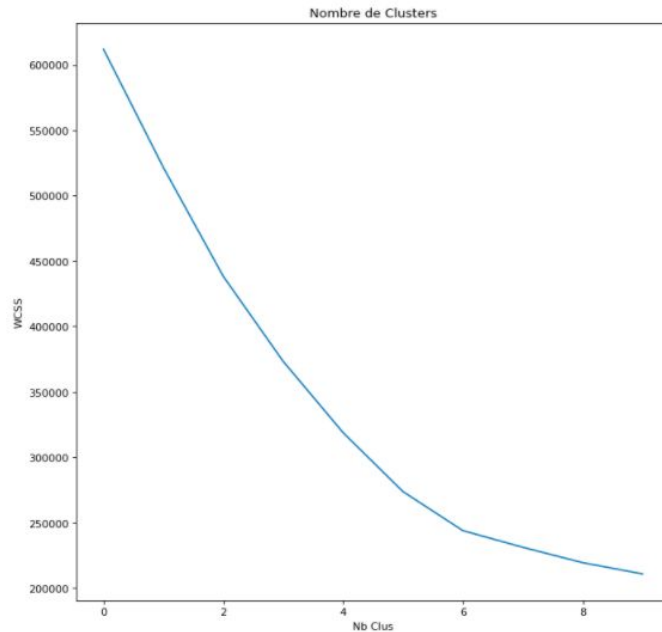


4.2.2.1 - Pré-Processing

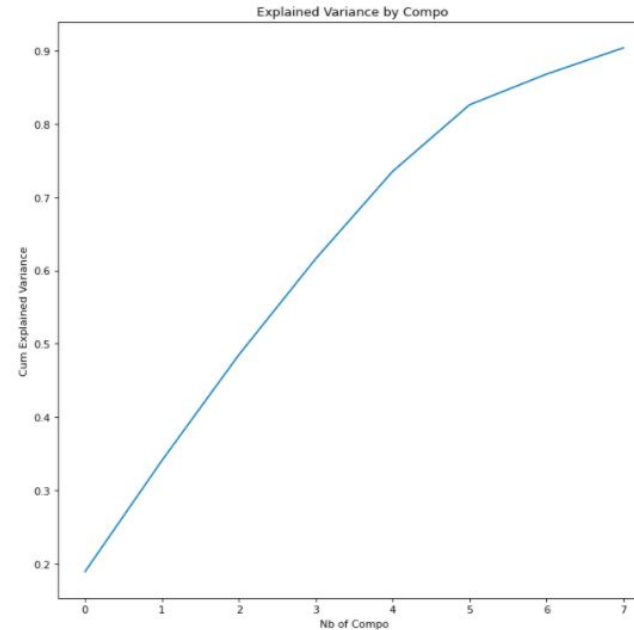
Avant la mise en place de notre étude de segmentation, nous avons procédé à un processing de notre sélection de variable, via différents outils

- Passage au Log
- Standard Scaler
- Dummy Encoding
- PCA

4.2.3 - K Means - PCA - Nombre idéal de Cluster

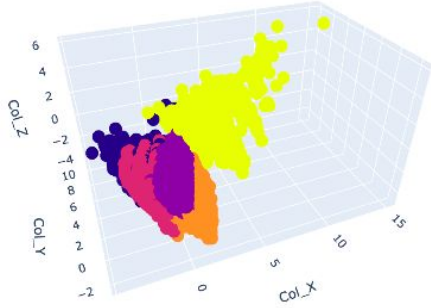


4.2.4 - K Means - PCA - Nombre idéal de Composantes principale

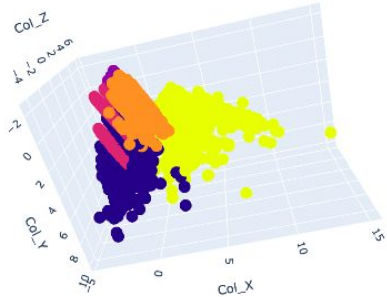


Nombre de cluster retenu : **5** / Nombre de composante principale: **6** / Silhouette score : **0.32**

4.2.5 - K Means - Visualisation avec PCA



Visualisation
des 5 clusters
en 3D



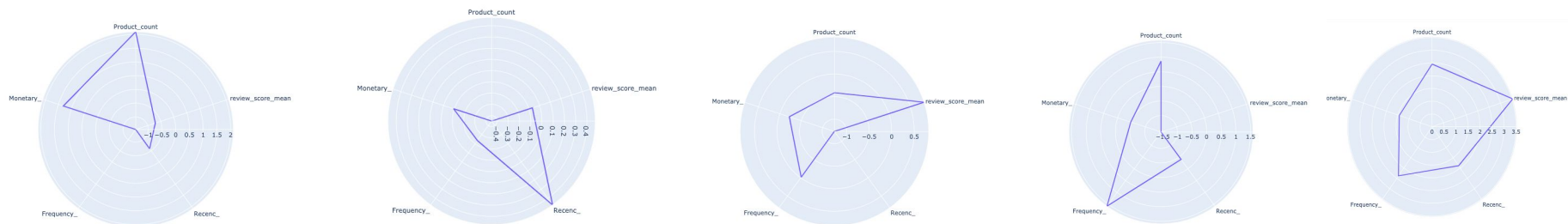
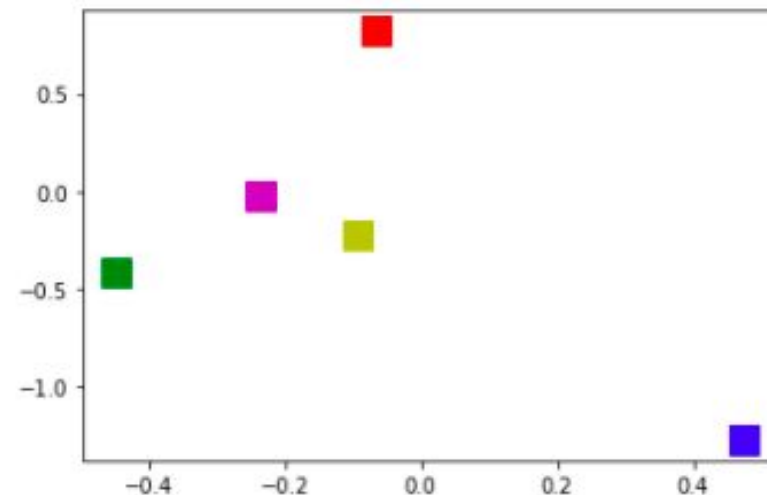
Visualisation des 5 clusters en 2D

4.2.9 - K Means

- PCA -

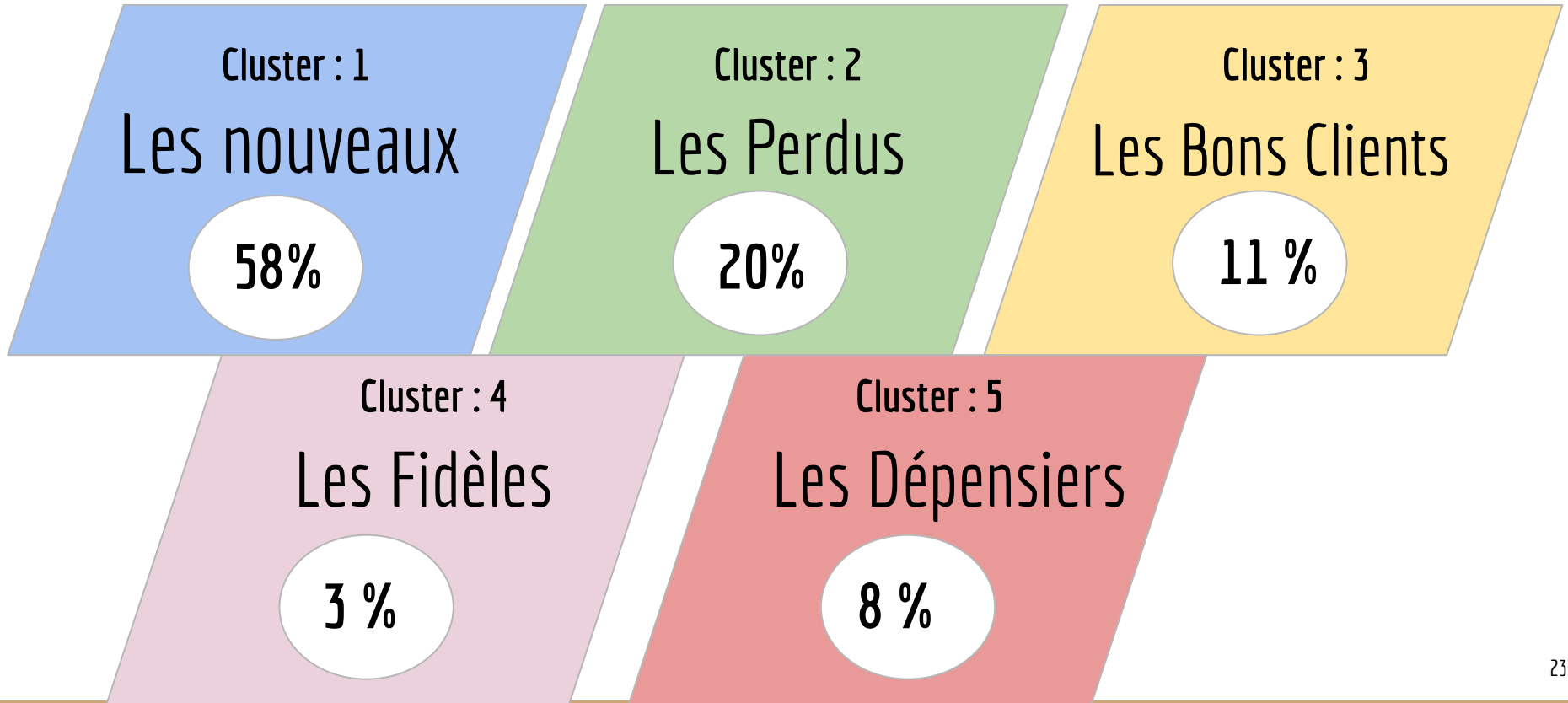
Centroïde

Visualisation des Centroïdes en 2D



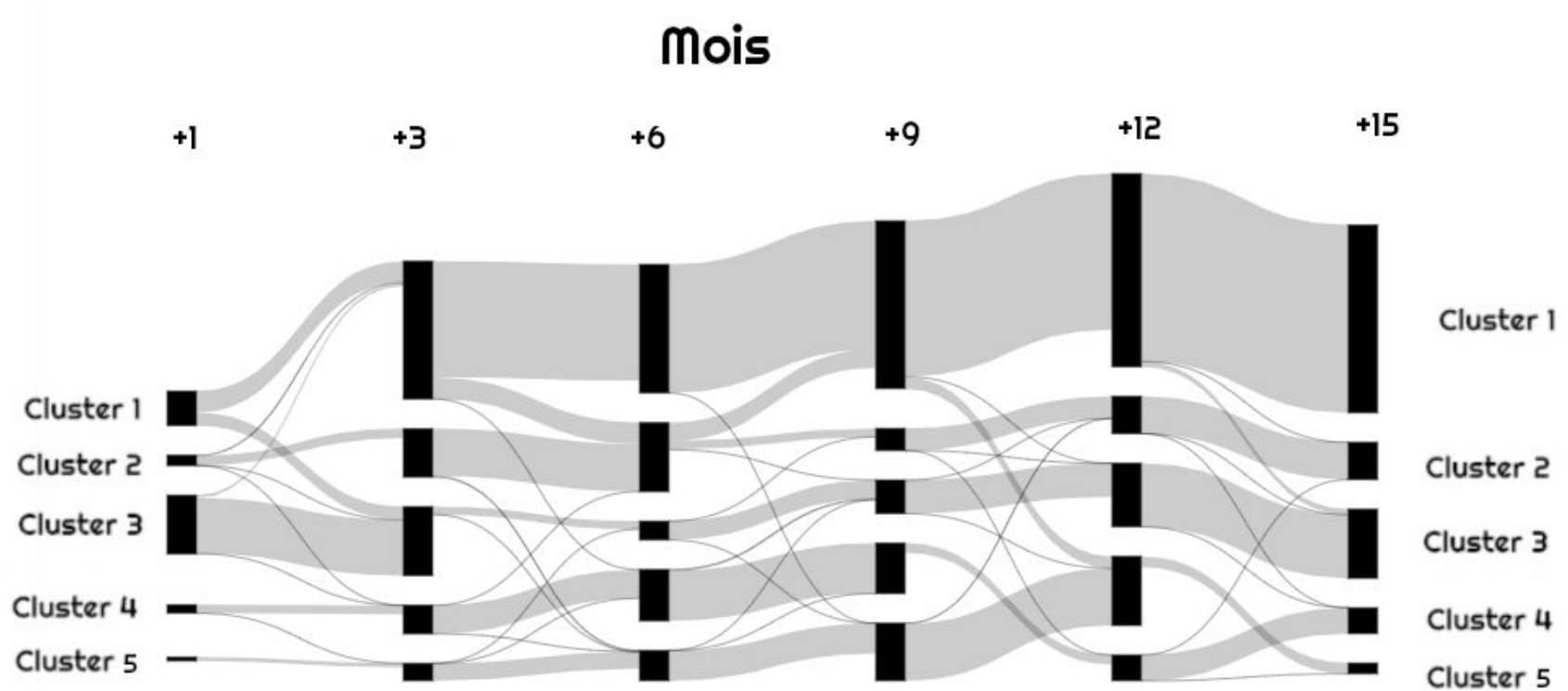
Caractéristique de chaque Cluster

4.2.10 - K Means - Conclusion



5 - Evaluation des performances

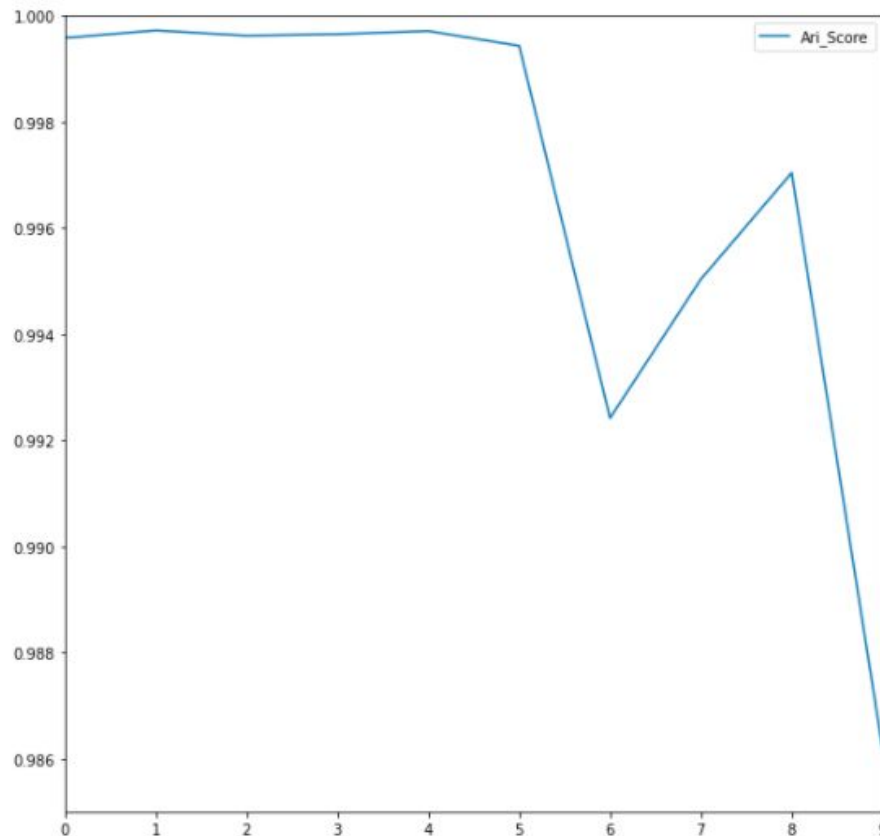
5.1 - Diagramme de Sankey



5.2 - Stabilité dans le temps

On observe une bonne stabilité du modèle dans le temps.

Néanmoins on constate que les données sont moins précises à partir du 10eme mois.



6 - Conclusion

Nous avons donc établi sur la base des données en présence, la segmentation de la clientèle en 5 segments utiles pour vos équipes marketings.

Les Nouveaux / Les Perdus / Les Bons Clients / Les Fidèles / Les Dépensiers

Néanmoins, nous constatons que les données fournis sont assez pauvre pour mener une étude complète. 97% des clients n'ont qu'un achat enregistré. Ce qui pose de nombreux soucis de fiabilité dans l'analyse d'un site de e-commerce.

Nous préconisons de même une maintenance quasi-annuel pour faire à nouveau ce travail

De même nous tenons à rappeler que cet étude est une piste pour de nouveaux axes de travail, et reste très évolutif sur le moyen terme. Il y a tout de même quelques clusters intéressants pour vos équipes

Voici quelques pistes qui nous permettront à l'avenir une étude plus précise :

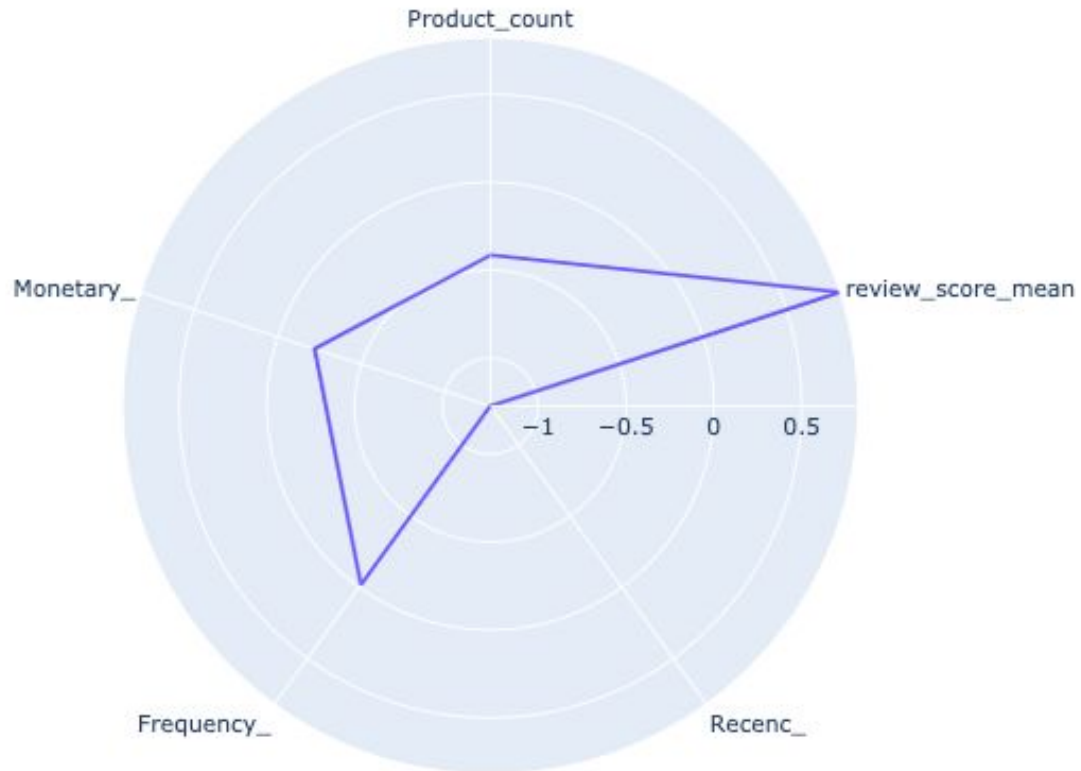
- Ajouter aux données l'âge des client
- Ajouter aux données le sexe des clients
- Créer une récurrence d'achat

Annexes

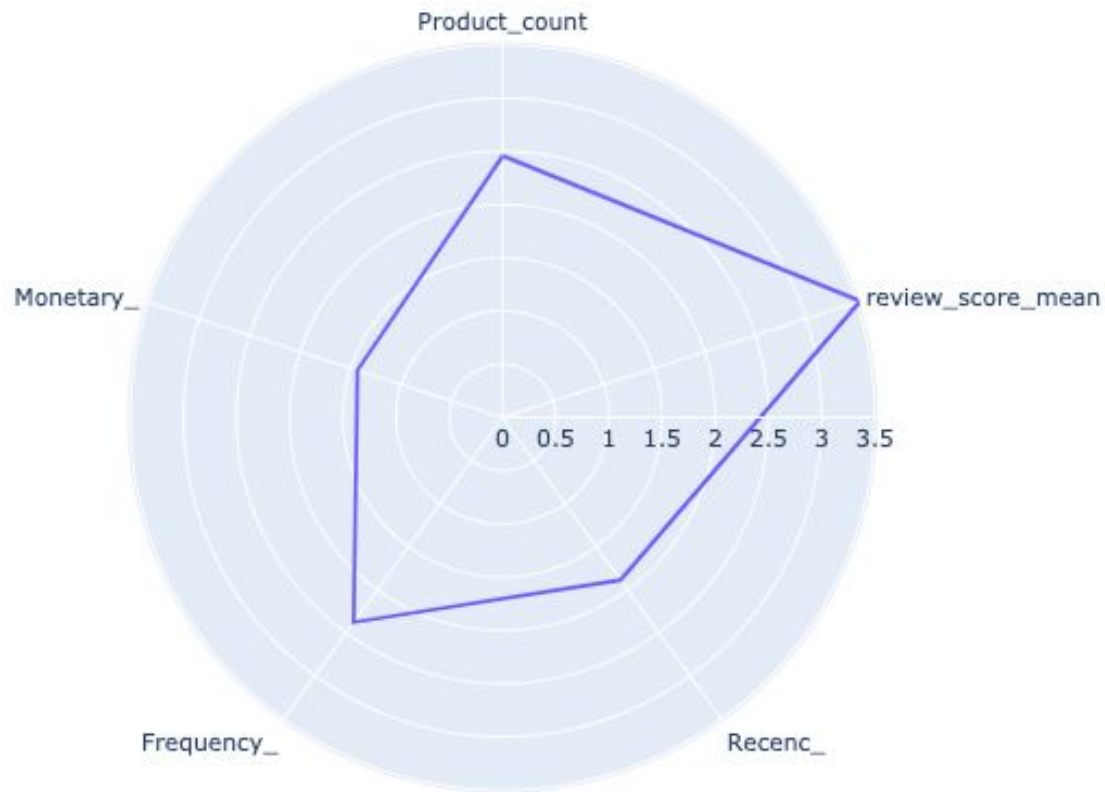
Cluster 1



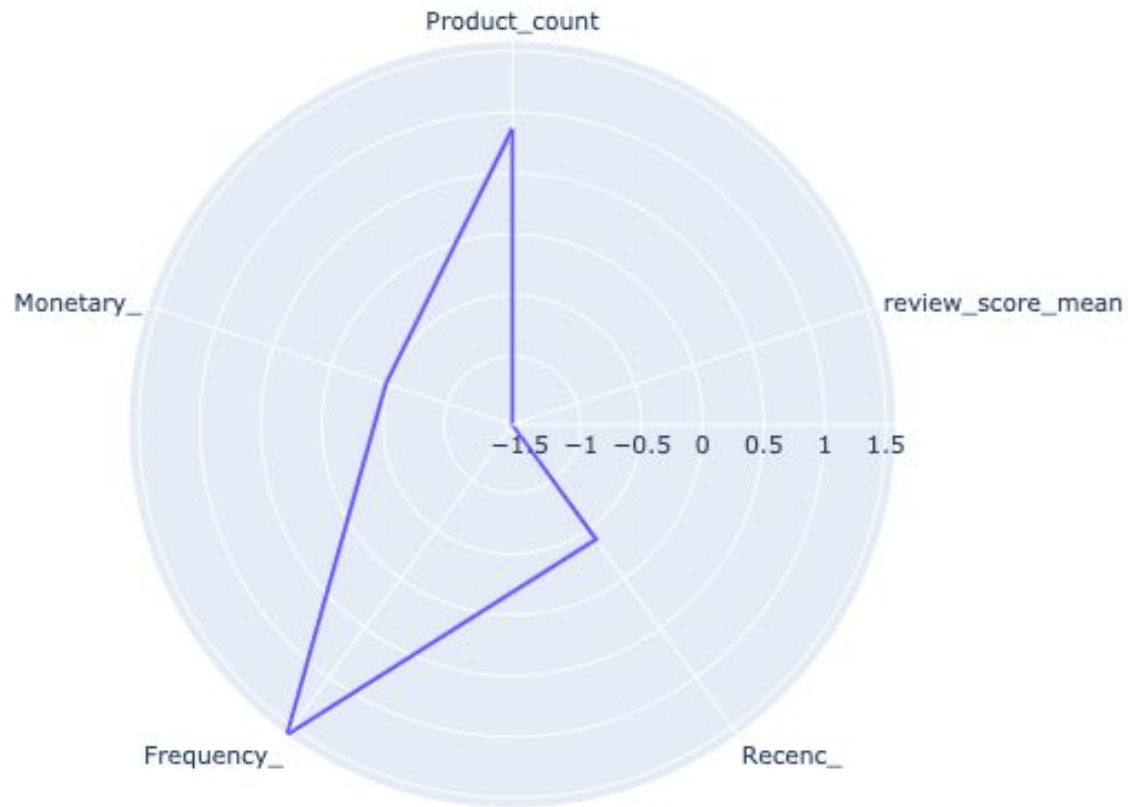
Cluster 2



Cluster 3



Cluster 4



Cluster 5

