

Formation DS-IML PROJET 4 : OLIST Customers Segmentation



olist
empowering commerce



olist
empowering commerce

MISSION

Conseil marketing

**Analyse et Segmentation de la
clientèle Olist.**

Résultats utilisables et opérationnels



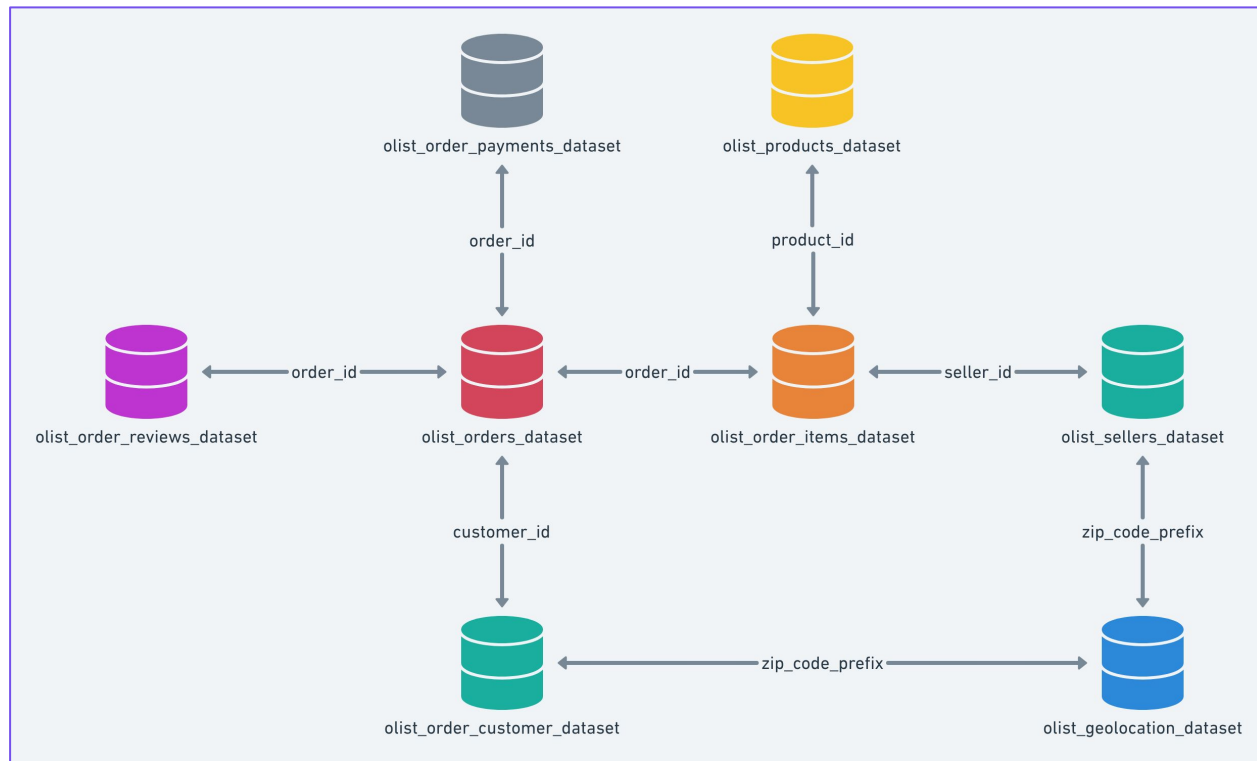
SOMMAIRE

- 1 - Data : Réunion - Exploration - Mise en forme
- 2 - Segmentation "RFM"
- 3 - Plus loin dans la segmentation...
- 4 - Conclusion et proposition de collaboration

Données à disposition

9 fichiers *.csv

- Clients
- Statut Com.
- Compo. Com.
- Paiements
- Reviews
- Produits
- Revendeurs
- Traduction
- Géo localisation



Après réunion des données**116 581 individus**

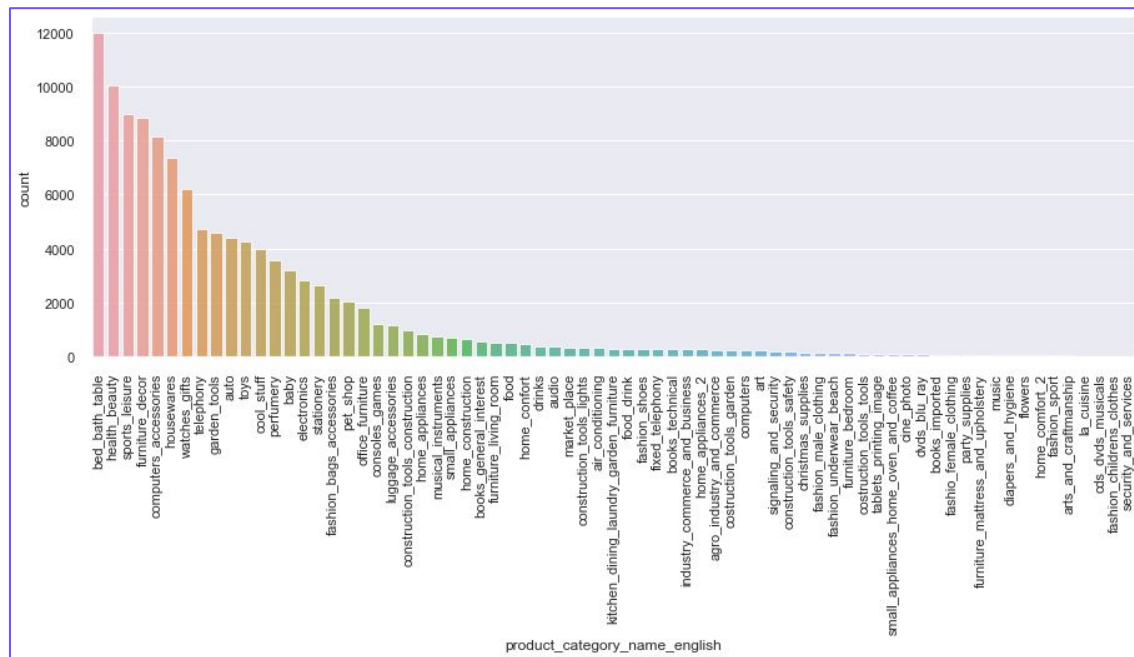
	customer_unique_id	customer_city	customer_state	order_status	order_purchase_timestamp
0	861eff4711a542e4b93843c6dd7febb0	franca	SP	delivered	2017-05-16 15:05:35
1	9eae34bbd3a474ec5d07949ca7de67c0	santarem	PA	delivered	2017-11-09 00:50:13
2	9eae34bbd3a474ec5d07949ca7de67c0	santarem	PA	delivered	2017-11-09 00:50:13

Client**Commande****Produits Commandés****FOCUS
CLIENT****40 à 9 colonnes**

0	customer_unique_id	116581 object
1	customer_state	116581 object
2	order_id	116581 object
3	order_purchase_timestamp	116581 object
4	payment_type	116581 object
5	payment_installments	116581 int64
6	payment_value	116581 float64
7	review_score	116581 int64
8	product_category_name_english	116581 object

Date d'achat

- Mois
- Semaine du mois
- Jour de la semaine
- Créneau Horaire

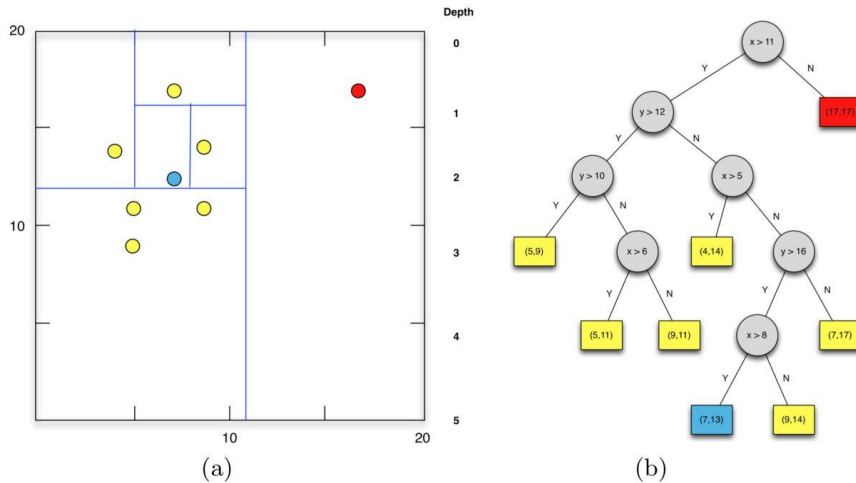


Réduction catégorielles

Catégories de Produits : xx à 18

Province des clients

Algorithme Isolation Forest



Hyper-paramètre de
CONTAMINATION

Proportion d'individus
les plus “anormaux”
devant être considérés
comme “outliers”



SOMMAIRE

- 1 - Data : Réunion - Exploration - Mise en forme
- 2 - Segmentation "RFM"
- 3 - Plus loin dans la segmentation...
- 4 - Conclusion et proposition de collaboration

Pourquoi segmenter des clients ?

Segmentation **RFM** ?

Une clientèle est variée, identifier des groupes permet de cibler ses actions commerciales.

Procédé classique en marketing. 1er niveau de segmentation sur des critères fondamentaux.

On note les clients sur les critères suivants :

- **R**ECENCY - Date du dernier achat
- **F**REQUENCY - Nombre de commandes passées
- **M**ONETARY VALUE - Montant total dépensé

Classe RFM

111

333

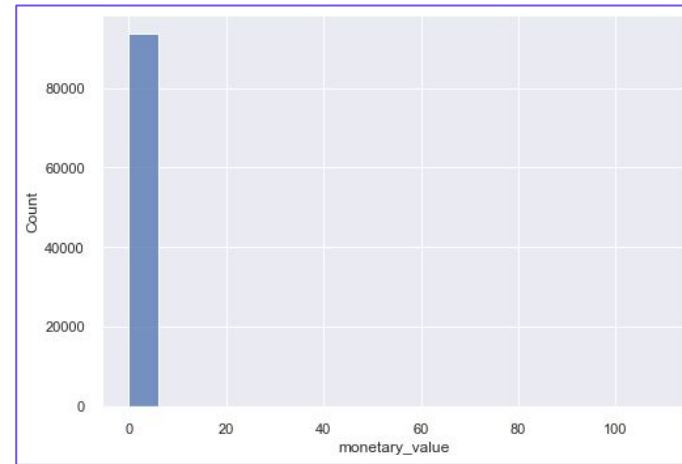
etc...

On part d'une **Table RFM**

*Opération sur
le dataset*



customer_unique_id	recency	frequency	monetary_value
8d50f5eadf50201ccdcedfb9e2ac8455	14	15	820.15
3e43e6105506432c953e165fb2acf44c	188	9	1963.58
ca77025e7201e3b30c44b472ff346268	94	7	2126.44



Chantier prioritaire : FIDÉLISATION CLIENT

Changement de plan

Montant dépensé :

Fréquence :

Récence :

RFM "ad-hoc" sur 3 groupes

Quantiles 2/3 et 1/3.

3+ commandes

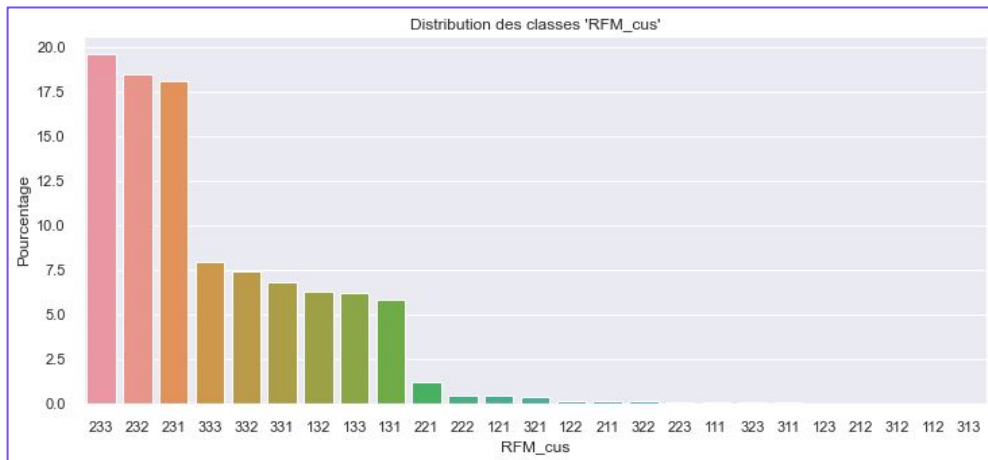
2 commandes

1 commande

- de 90 jours

90 à 365 jours

+ de 365 jours



27 classes en tout.
La majorité **peu**
représentée

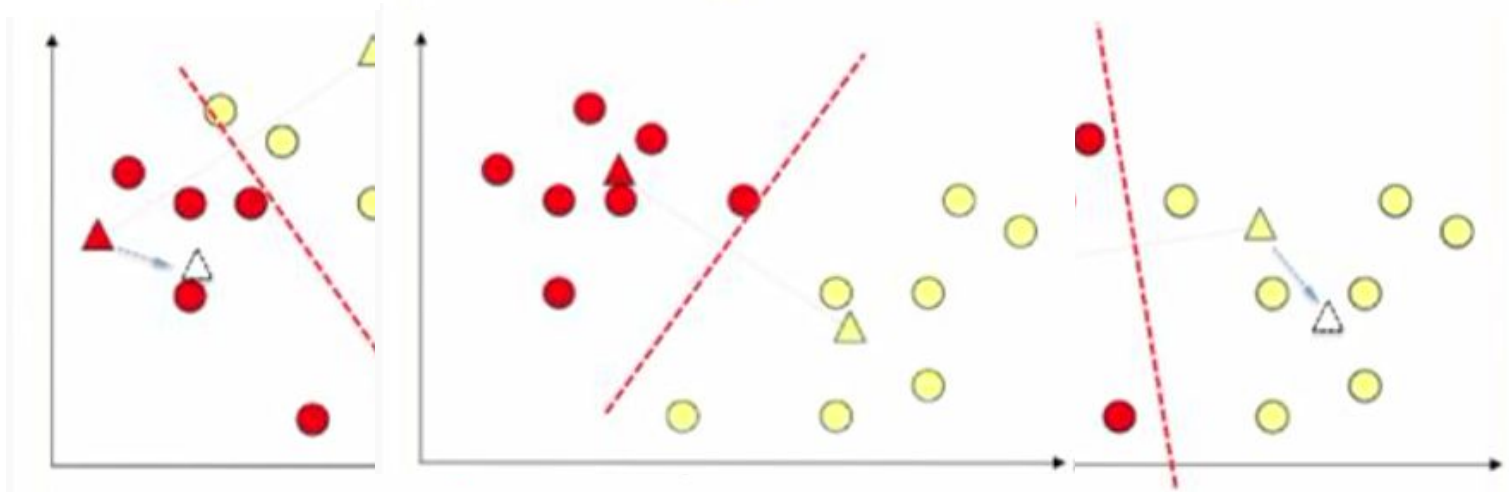
Machine Learning

Supervisé

Renforcé

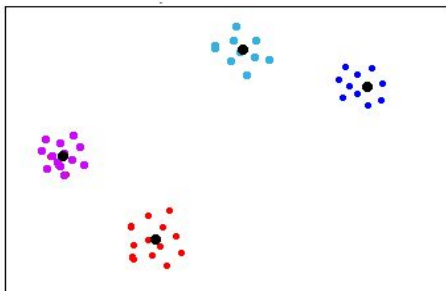
Non Supervisé

K-Means

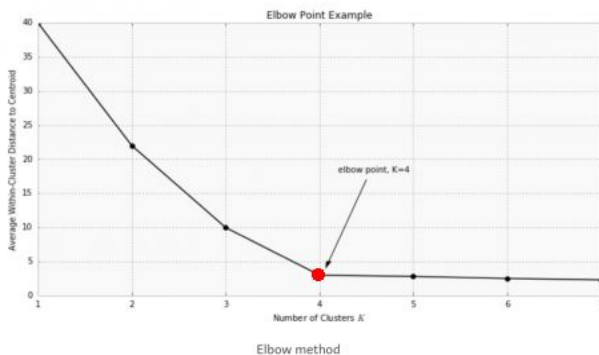
Segmente des données numériques en **K** groupes

Choix basé que plusieurs éléments...

Méthode du coude



cost = somme des variances des clusters

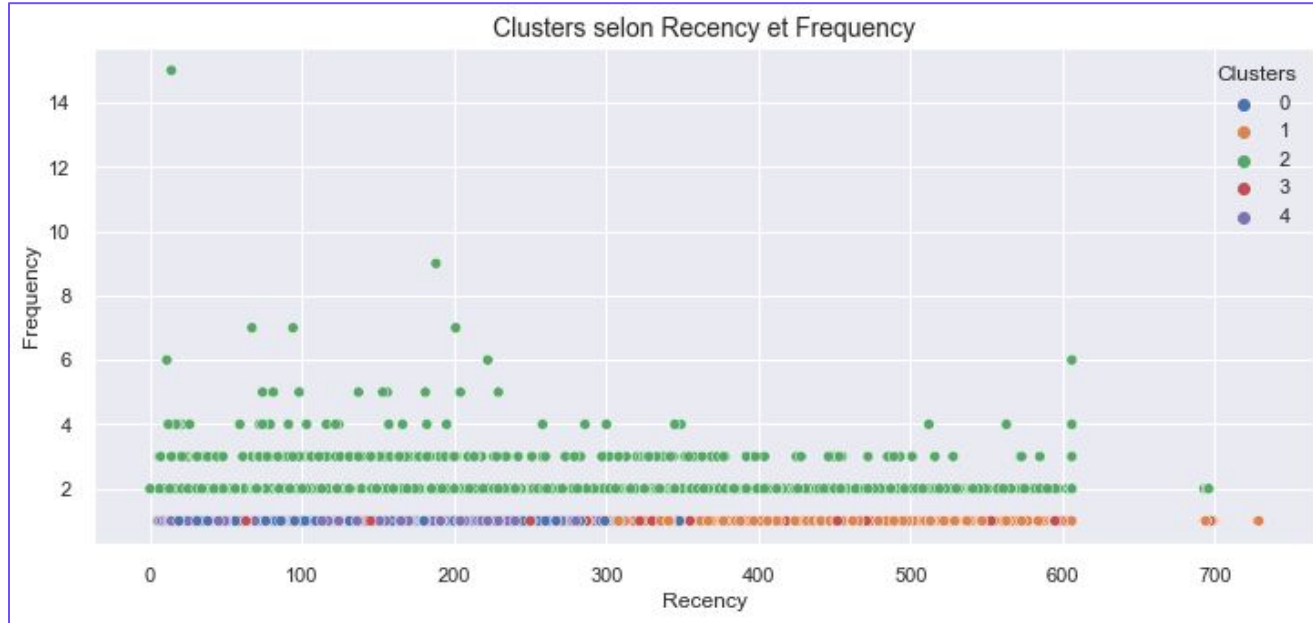


Silhouette Score

Note de “qualité” de la segmentation

Observation des clusters

Sens, interprétation...

K = 5**2ème essai avec F et M “passées au log”****Cela nous aide-t-il ?****NON****1 - Un cluster regroupe les clients à au moins deux achats.****2 - Les autres sont partagés sur R et M**

Objectif : augmentez fréquence/fidélité

95% de clients à 1 commande

Segmentation

Objectif - Fréquence

Récence pour les clients “one shot”

Gold

$F > 2$

Silver

$F = 2$

B new

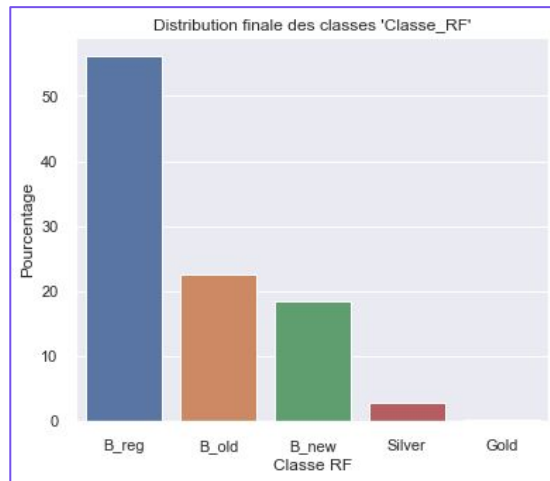
$F = 1 \text{ \& } R < 90$

B reg

$F = 1 \text{ \& } 90 < R < 365$

B old

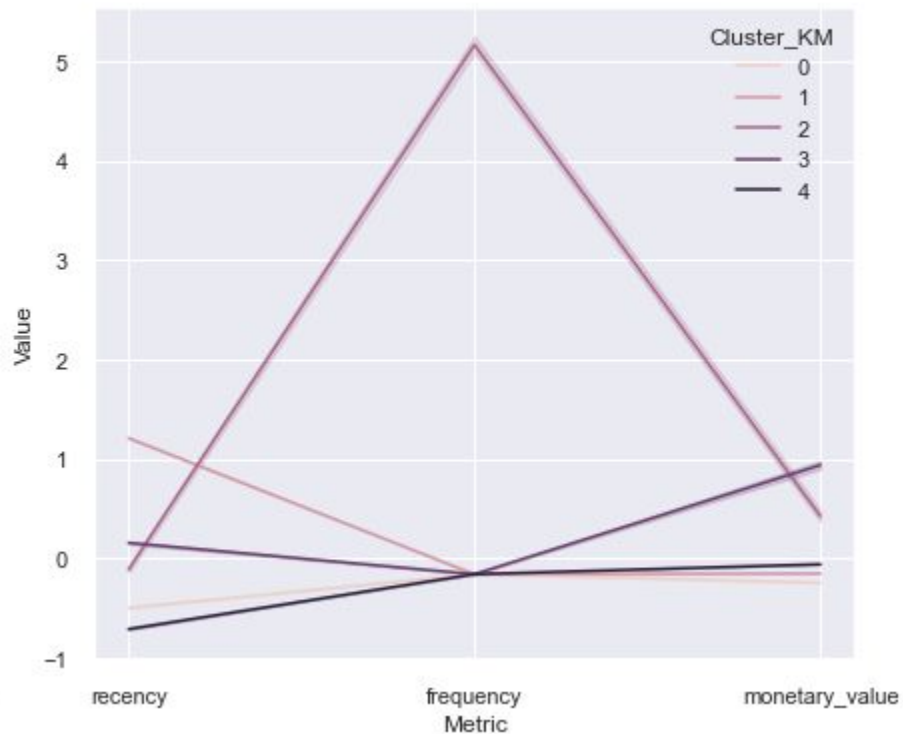
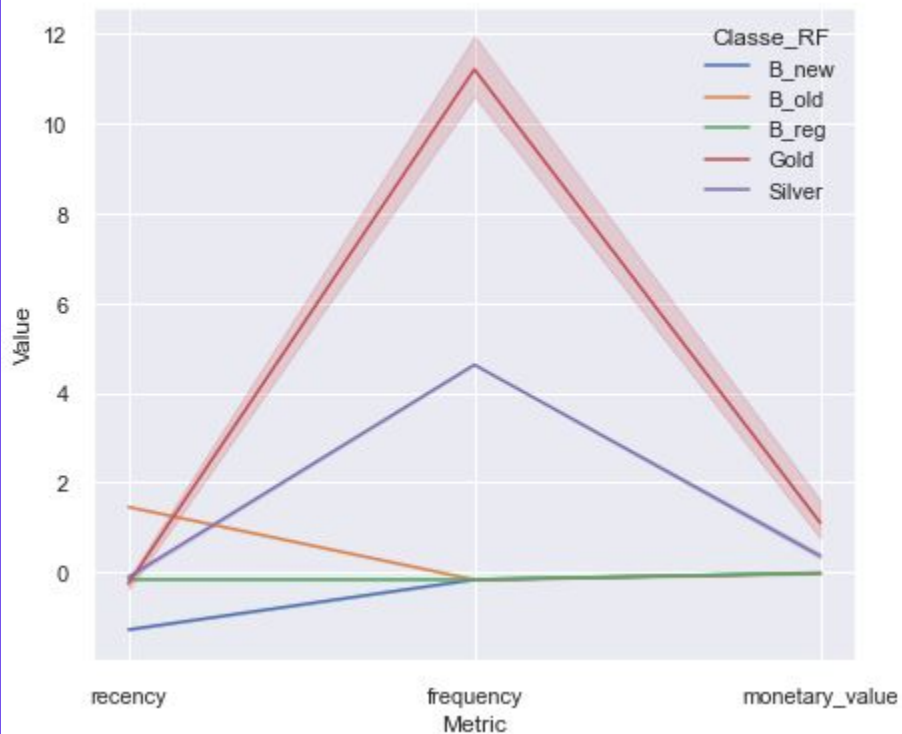
$F = 1 \text{ \& } 365 < R$



**Segmentation
“RF”**

Snake Plot

Différence de classification : 'RF' vs K-Means



Objectif : augmentez fréquence/fidélité

95% de clients à 1 commande

Gold

$F > 2$

Globale :

Silver

$F = 2$

Mécanisme de fidélisation

B new

$F = 1 \text{ \& } R < 90$

Ciblée :

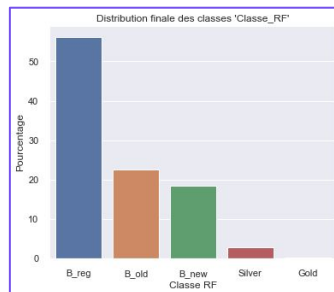
B reg

$F = 1 \text{ \& } 90 < R < 365$

Communiquer avec le client en fonction de leur classe “RF”.

B old

$F = 1 \text{ \& } 365 < R$



Stacked Area Plot

Percentage Stacked Area Plot

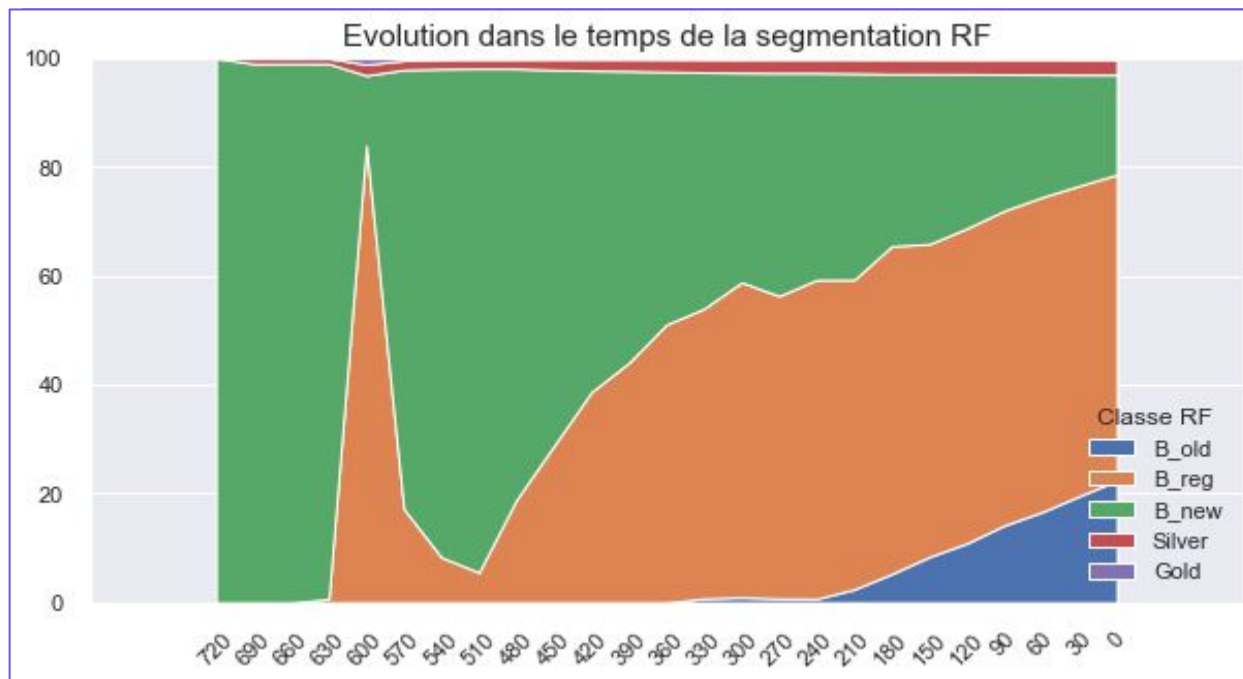
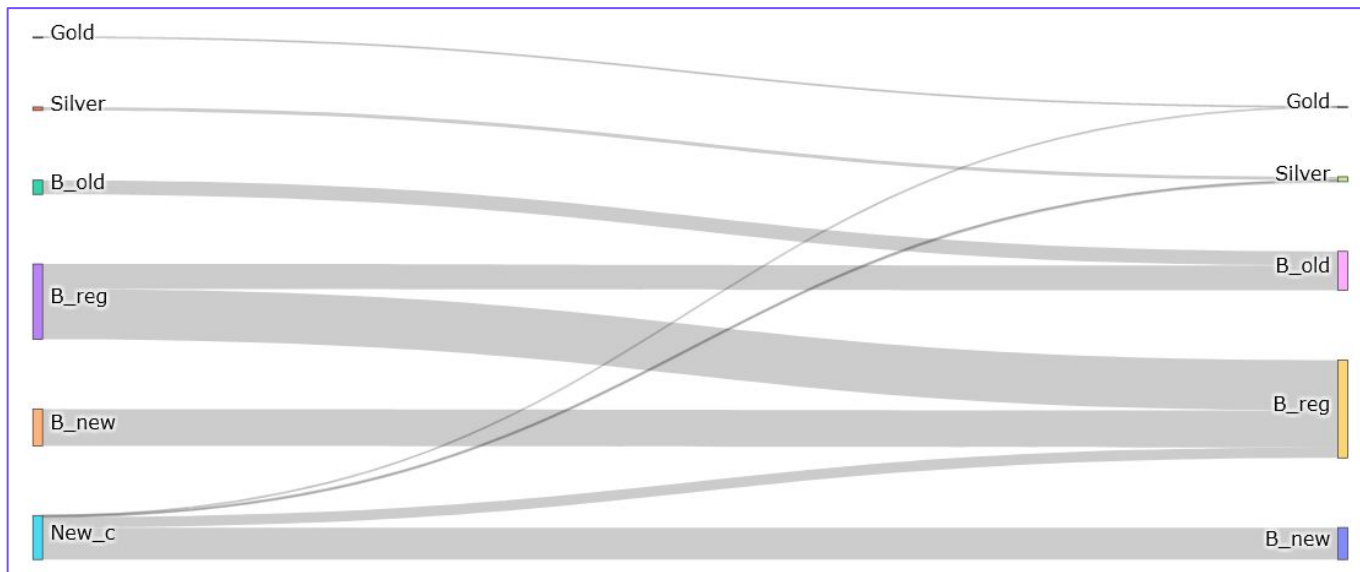
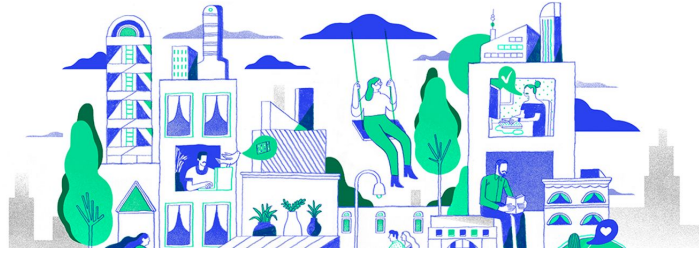


Diagramme Sankey

Flux entre classes sur une période données



Nous fournirons les listes des clients changeant de classe...



SOMMAIRE

- 1 - Data : Réunion - Exploration - Mise en forme
- 2 - Segmentation "RFM"
- 3 - Plus loin dans la segmentation...
- 4 - Conclusion et proposition de collaboration

Utilisation du ML non-supervisé

Infos, tendances, "persona"

1 - Cadre de recherche**2** - Algo.**3** - Observation

Algorithmes

K-Means

Data numériques

K-Modes

Data catégorielles

K-Prototypes

Mixed data

DBSCAN

Densité de points

Principe Simple

Pas d'assurance de résultat

DBSCAN



k-means



K-modes : M_class / états / paiement / month / product_cat

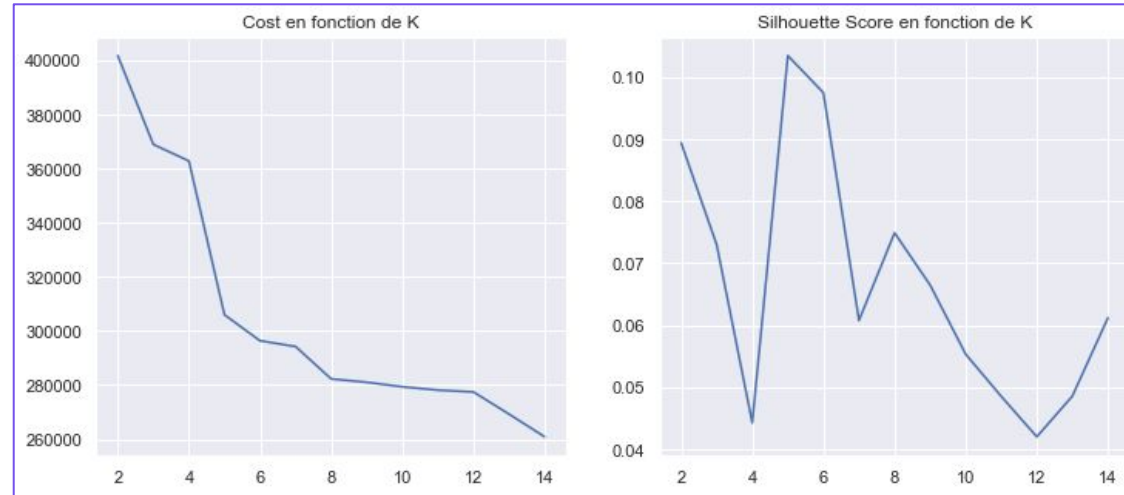
que des variables catégorielles !

Choix de K

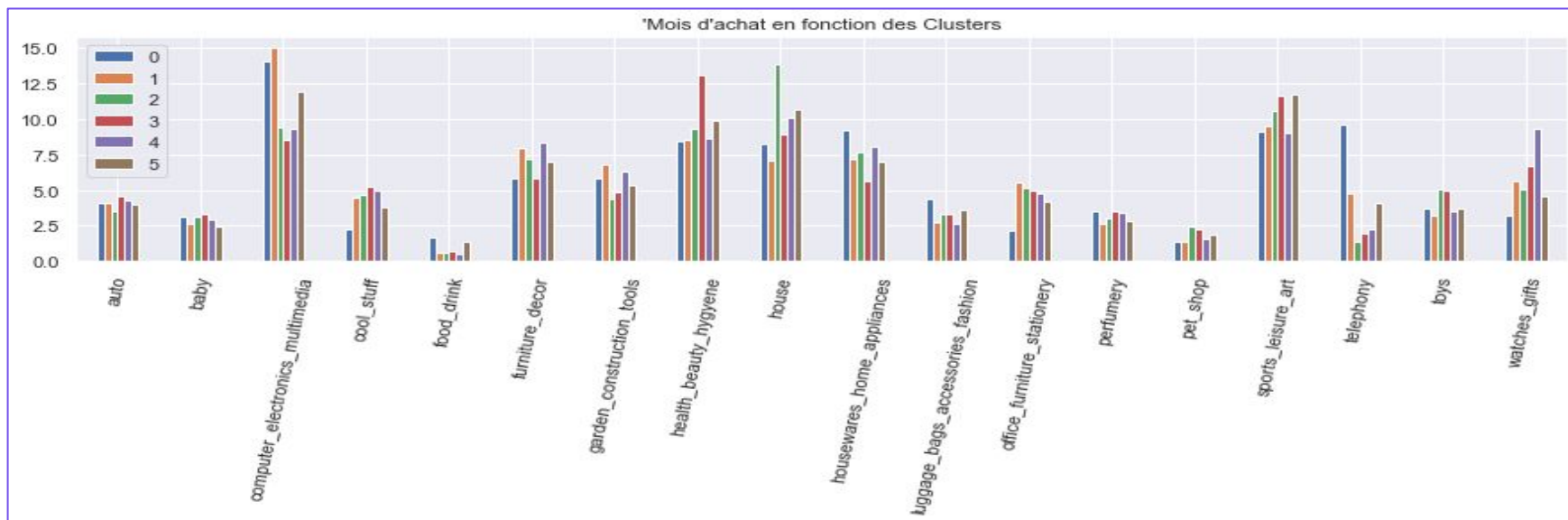
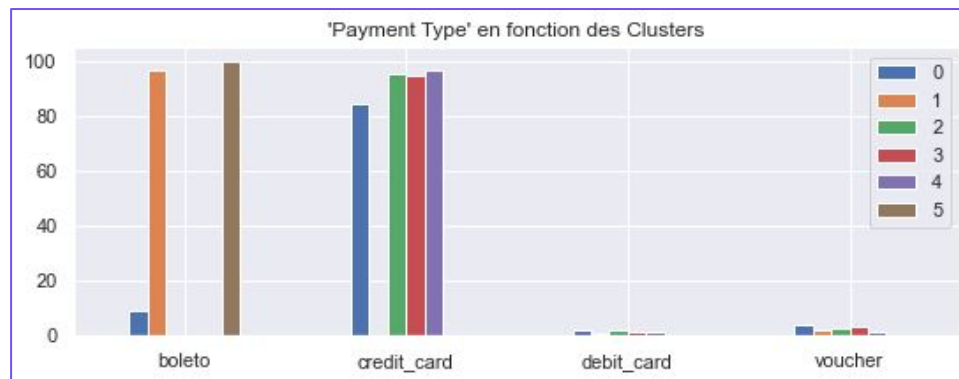
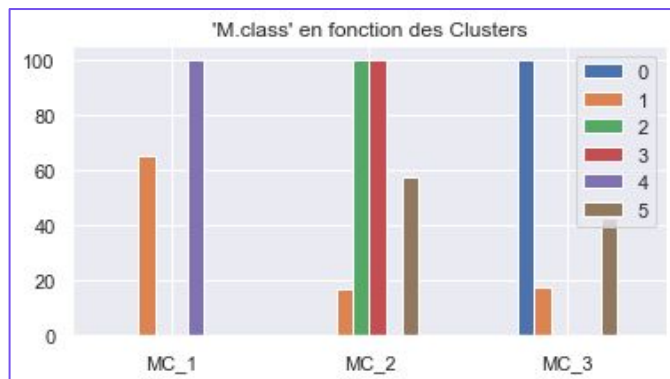
45 minutes

Qualité
médiocre

K = 6



0	27079
4	25475
2	19385
1	8595
5	7905
3	5648



K-modes : M_class

K-modes : M_class

--