Projet 4

Segmentez des clients d'un site e-commerce
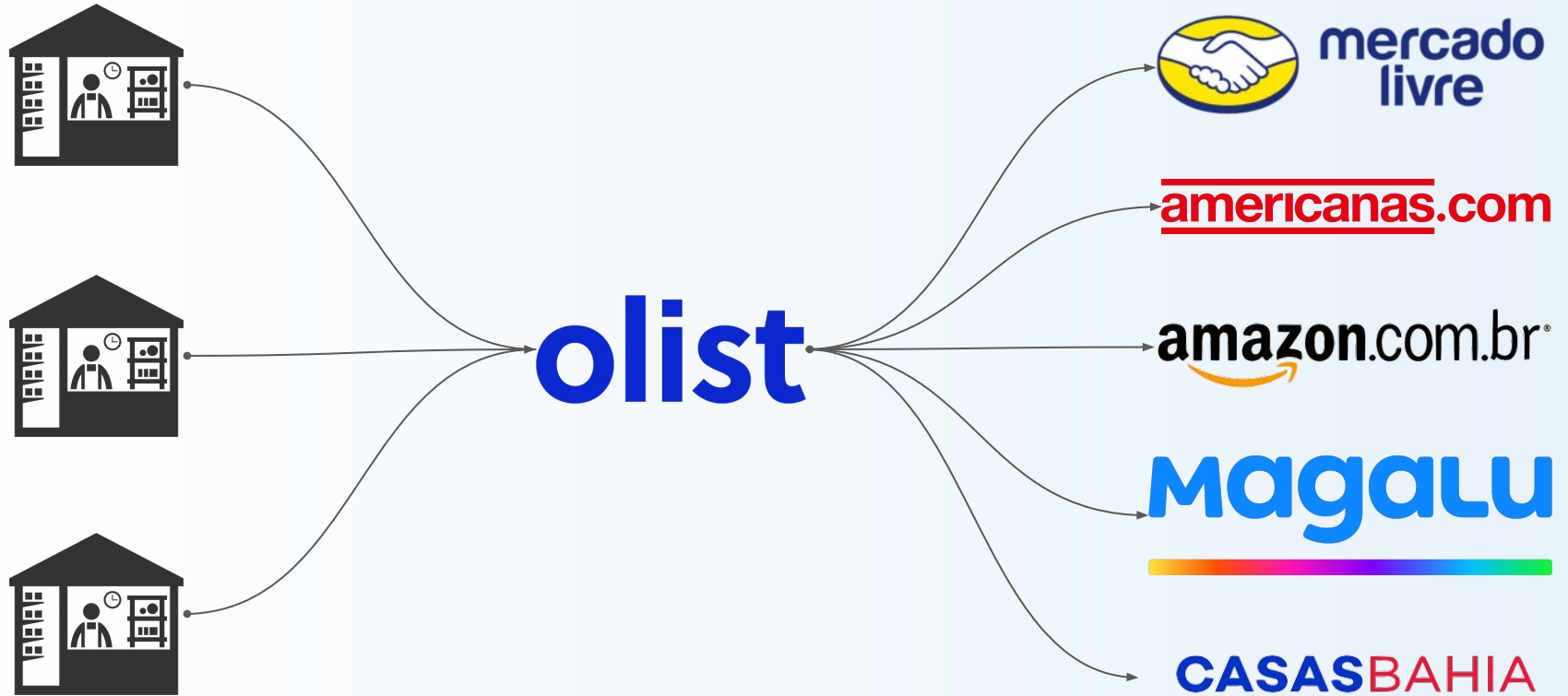
Adrian Rodriguez
Ingénieur Machine Learning

olist

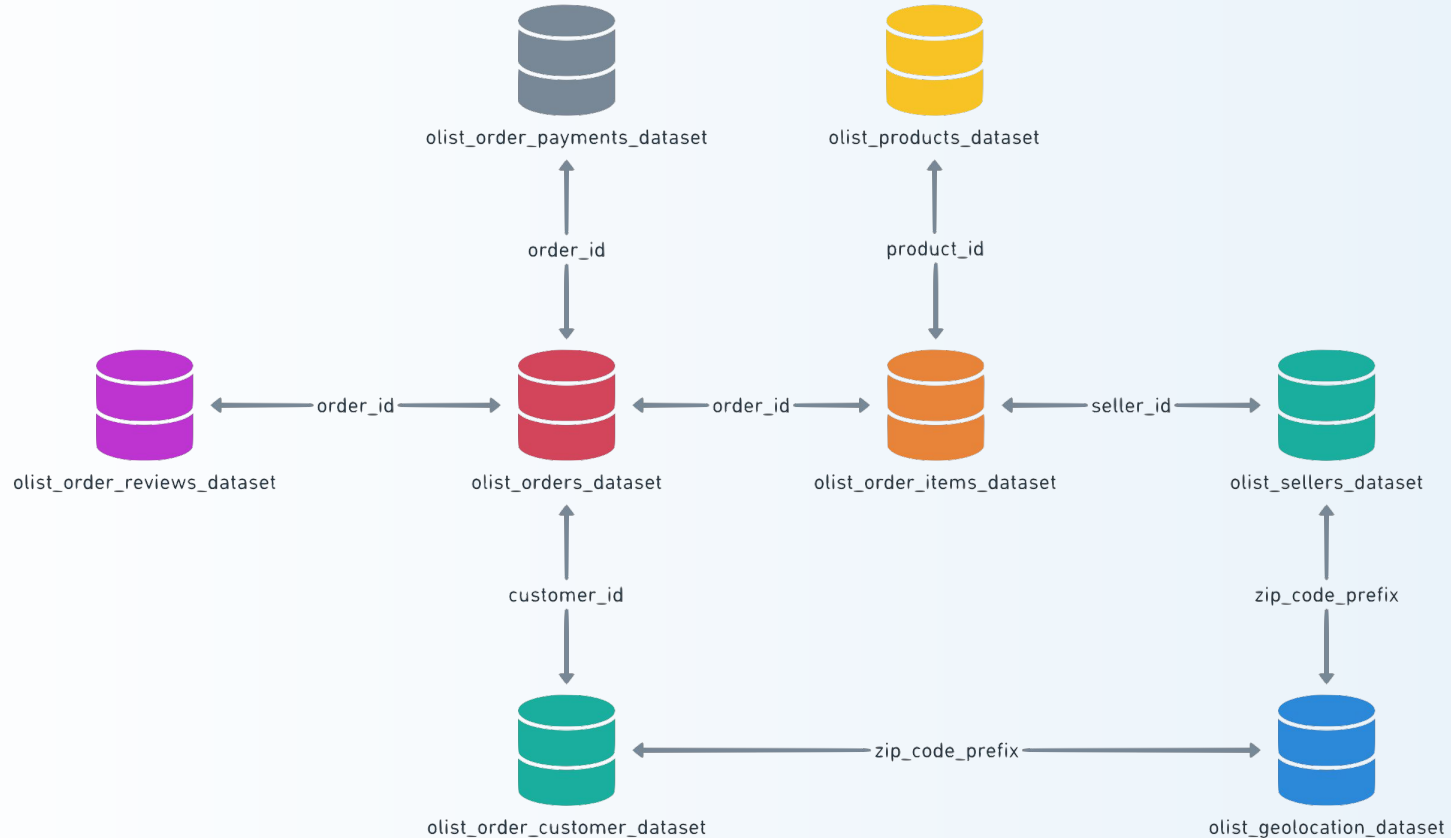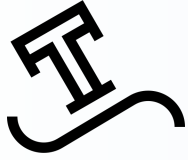CentraleSupélec

OPENCLASSROOMS

Olist : regroupement de marchand

Système d'information Client/Marchand Olist
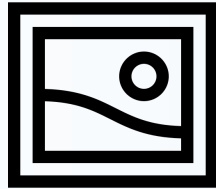
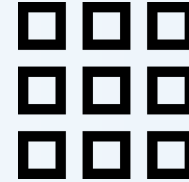# Création / transformation des variables
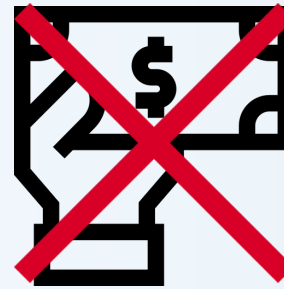
Description

Livraison

Avis

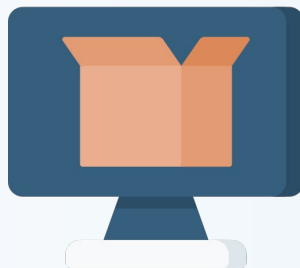Image

Catégories

Vérification de la qualité des données

Chiffres-clés Olist

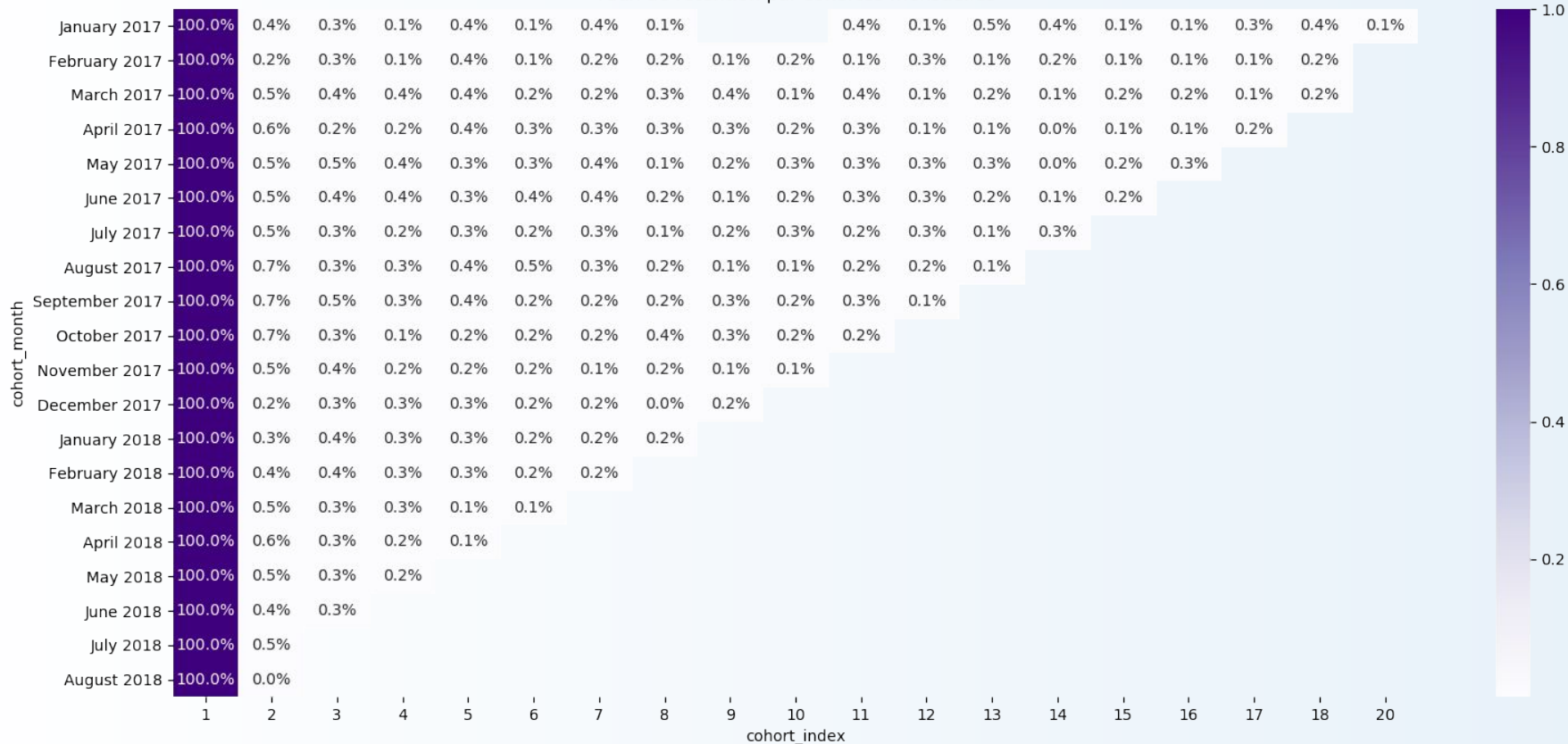## Période du 01/01/2017 au 30/09/2018

Près de 94 000 clients

Presque 97 000 commandes reçues

Plus de 110 000 articles concernés

Taux de rétention



Taux de rétention par cohortes mensuelles

cohort_month	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	20
January 2017	100.0%	0.4%	0.3%	0.1%	0.4%	0.1%	0.4%	0.1%		0.4%	0.1%	0.5%	0.4%	0.1%	0.1%	0.3%	0.4%	0.1%	
February 2017	100.0%	0.2%	0.3%	0.1%	0.4%	0.1%	0.2%	0.2%	0.1%	0.2%	0.1%	0.3%	0.1%	0.2%	0.1%	0.1%	0.1%	0.2%	
March 2017	100.0%	0.5%	0.4%	0.4%	0.4%	0.2%	0.2%	0.3%	0.4%	0.1%	0.4%	0.1%	0.2%	0.1%	0.2%	0.2%	0.1%	0.2%	
April 2017	100.0%	0.6%	0.2%	0.2%	0.4%	0.3%	0.3%	0.3%	0.3%	0.2%	0.3%	0.1%	0.1%	0.0%	0.1%	0.1%	0.2%		
May 2017	100.0%	0.5%	0.5%	0.4%	0.3%	0.3%	0.4%	0.1%	0.2%	0.3%	0.3%	0.3%	0.3%	0.0%	0.2%	0.3%			
June 2017	100.0%	0.5%	0.4%	0.4%	0.3%	0.4%	0.4%	0.2%	0.1%	0.2%	0.3%	0.3%	0.2%	0.1%	0.2%				
July 2017	100.0%	0.5%	0.3%	0.2%	0.3%	0.2%	0.3%	0.1%	0.2%	0.3%	0.2%	0.3%	0.1%	0.3%					
August 2017	100.0%	0.7%	0.3%	0.3%	0.4%	0.5%	0.3%	0.2%	0.1%	0.1%	0.2%	0.2%	0.1%						
September 2017	100.0%	0.7%	0.5%	0.3%	0.4%	0.2%	0.2%	0.2%	0.3%	0.2%	0.3%	0.1%							
October 2017	100.0%	0.7%	0.3%	0.1%	0.2%	0.2%	0.2%	0.4%	0.3%	0.2%	0.2%								
November 2017	100.0%	0.5%	0.4%	0.2%	0.2%	0.2%	0.1%	0.2%	0.1%	0.1%									
December 2017	100.0%	0.2%	0.3%	0.3%	0.3%	0.2%	0.2%	0.0%	0.2%										
January 2018	100.0%	0.3%	0.4%	0.3%	0.3%	0.2%	0.2%	0.2%											
February 2018	100.0%	0.4%	0.4%	0.3%	0.3%	0.2%	0.2%												
March 2018	100.0%	0.5%	0.3%	0.3%	0.1%	0.1%													
April 2018	100.0%	0.6%	0.3%	0.2%	0.1%														
May 2018	100.0%	0.5%	0.3%	0.2%															
June 2018	100.0%	0.4%	0.3%																
July 2018	100.0%	0.5%																	
August 2018	100.0%	0.0%																	

cohort_index

Résultat de la segmentation RFM

RFM_Score	recency	frequency	monetary_value	customer_unique_id
3.0	457.7	1.0	43.7	5930.0
4.0	367.8	1.0	64.1	11791.0
5.0	300.7	1.0	87.9	17294.0
6.0	240.6	1.0	163.7	22728.0
7.0	176.7	1.0	203.7	17229.0
8.0	121.8	1.1	258.2	12035.0
9.0	73.7	1.1	395.5	6191.0
10.0	75.7	2.1	399.2	507.0
11.0	87.0	3.5	567.8	63.0
12.0	58.6	5.0	945.6	23.0

Segmentation non supervisée

Features engineering
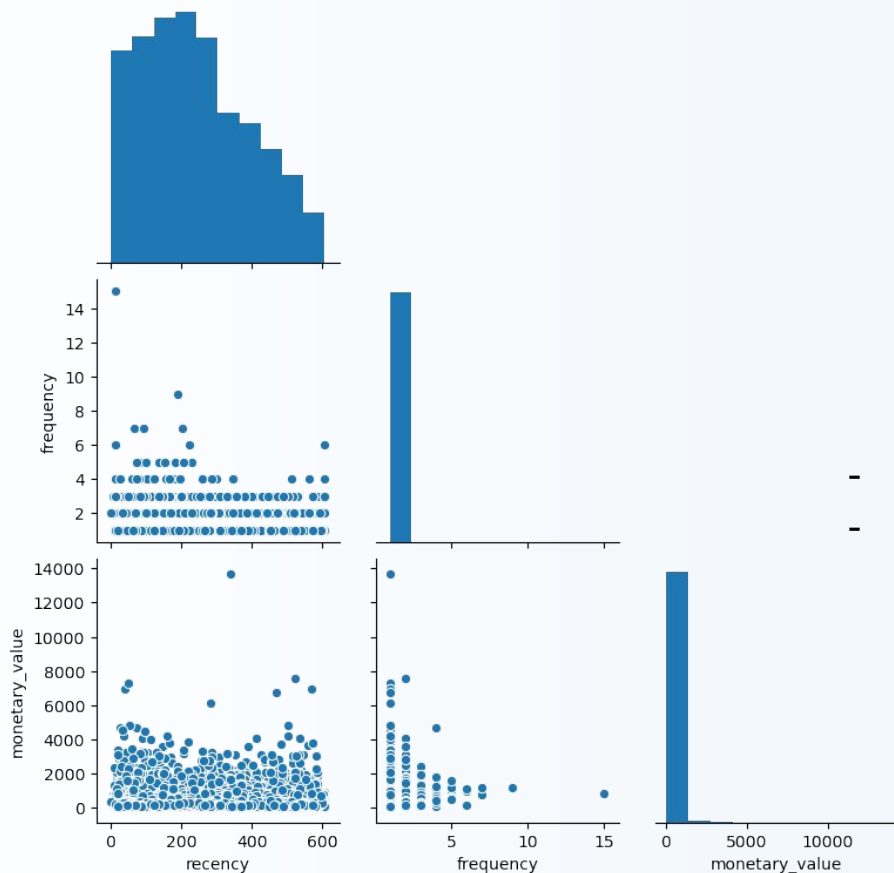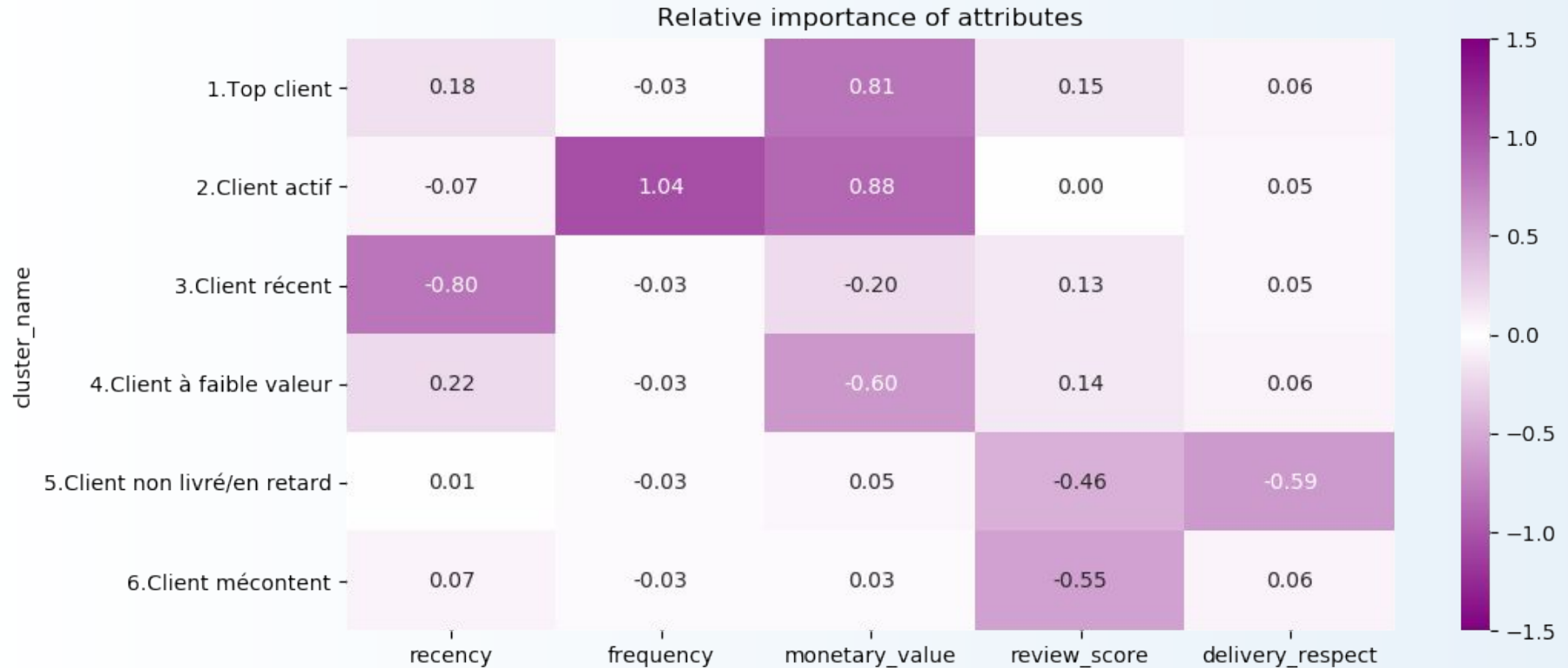

- Distributions non gaussienne
- Variances élevées

Choix du nombre de cluster
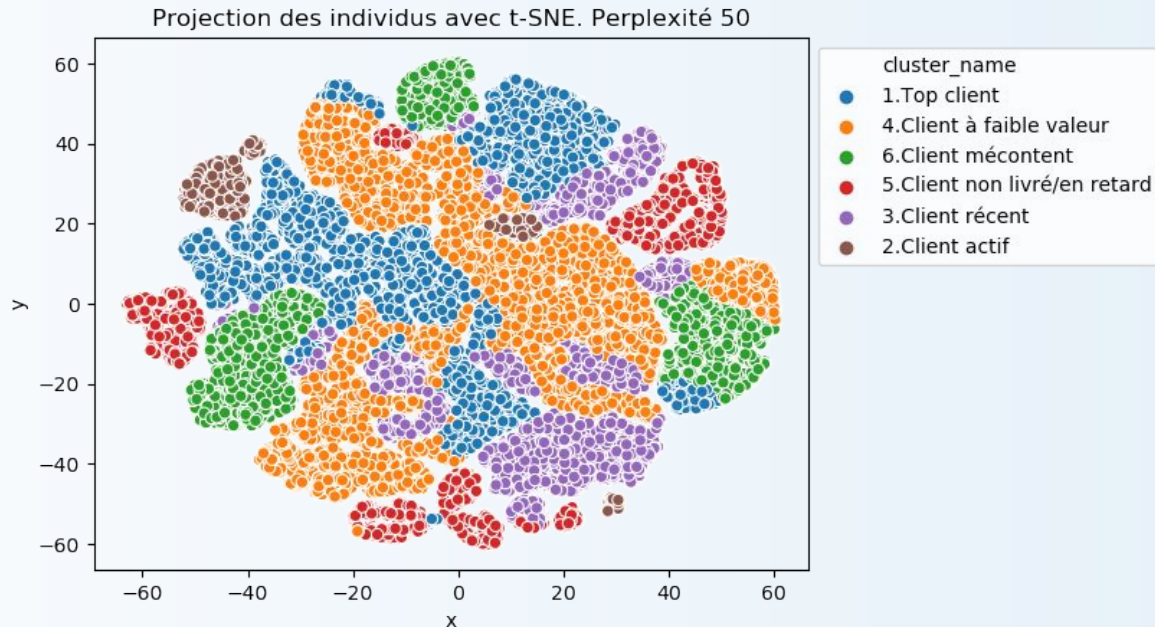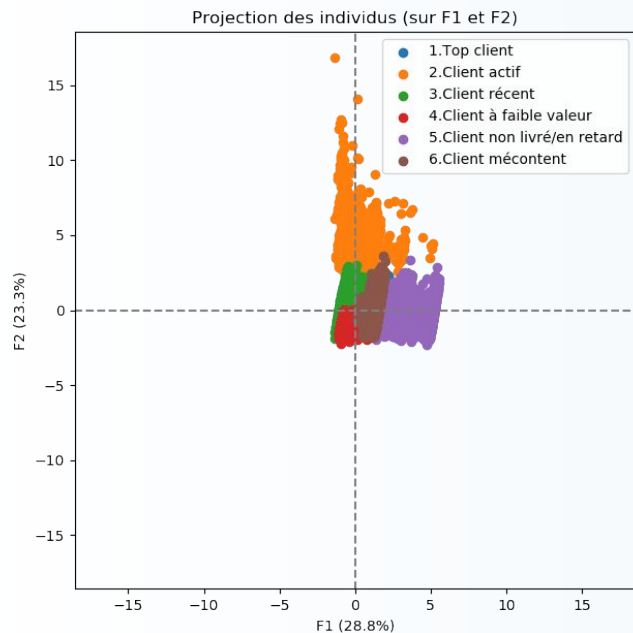

Elbow Method For Optimal k
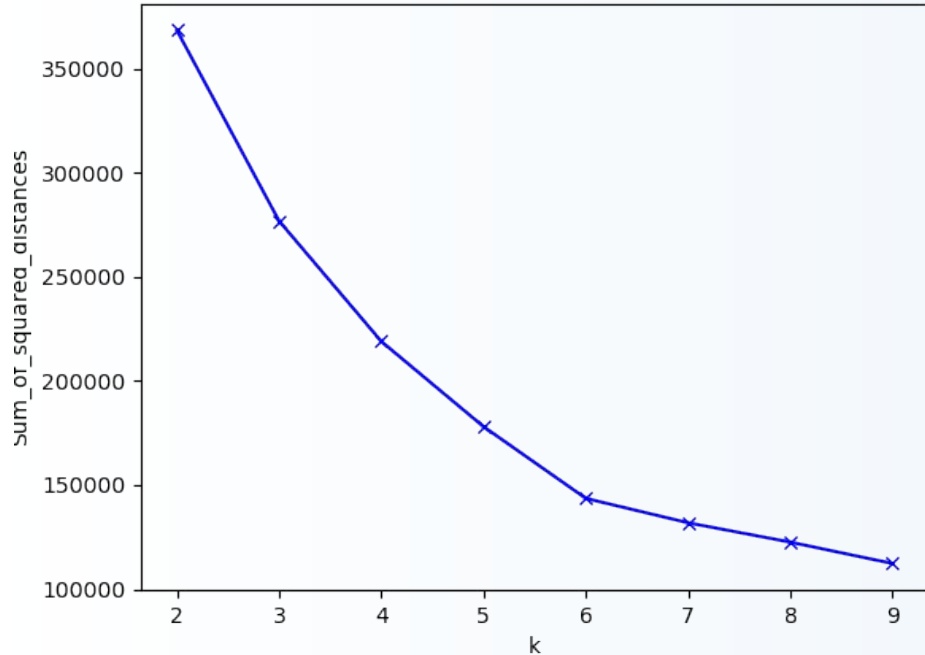
Nombre de cluster	Score de silhouette
2.0	0.45
3.0	0.34
4.0	0.32
5.0	0.3
6.0	0.32
7.0	0.29
8.0	0.3
9.0	0.31

Importance des attributs

Relative importance of attributes

cluster_name	recency	frequency	monetary_value	review_score	delivery_respect
1.Top client	0.18	-0.03	0.81	0.15	0.06
2.Client actif	-0.07	1.04	0.88	0.00	0.05
3.Client récent	-0.80	-0.03	-0.20	0.13	0.05
4.Client à faible valeur	0.22	-0.03	-0.60	0.14	0.06
5.Client non livré/en retard	0.01	-0.03	0.05	-0.46	-0.59
6.Client mécontent	0.07	-0.03	0.03	-0.55	0.06

Projection des clusters

Résultats de la segmentation non supervisée

cluster_name	recency	frequency	monetary_value	review_score	delivery_respect	volume
1.Top client	285.2	1.0	300.7	4.7	2.0	23828
2.Client actif	225.7	2.1	311.5	4.1	2.0	2837
3.Client récent	49.3	1.0	133.1	4.6	2.0	15197
4.Client à faible valeur	296.8	1.0	65.9	4.7	2.0	32350
5.Client non livré/en retard	243.9	1.0	175.0	2.2	0.8	8664
6.Client mécontent	259.6	1.0	170.5	1.8	2.0	10915

Persona : Fiche produit

Creme Facial Clareador Teatrical 100g

(Cód.41304206) ★★★★☆ (81)

🏬 **pegue ou receba** hoje!

Compre **por R$ 14,99** e pegue na loja sem pagar frete ou **receba em casa em 3 horas**
confira as regras de entrega

pegar ou receber hoje

Escolha uma loja abaixo e compre

○ LOJAS AMERICANAS
R$ 14,99
Frete: -

● Olist
R$ 15,10
Frete: -

○ FarmaLua
R$ 31,90
Frete: -

Mais opções deste produto a partir de **R$ 14,80** ›

vendido e entregue por **Olist**

R$ 15,10

🛒 **comprar**

comprar com ♥ame

◆+▭ R$ 15,10 em 1x no cartão de crédito

◆+▭ R$ 15,10 em até 3x de R$ 5,03 s/ juros no cartão Americanas

:) Este produto é vendido por uma loja parceira. A **Americanas** garante sua compra, do pedido à entrega.

A linha **Teatrical** se nutre de células-tronco provenientes da flor Buddleja davidii, popularmente conhecida como Árvore das Borboletas. O cultivo das células-tronco em plantas é uma tecnologia inovadora e sustentável, que permite proporcionar à pele as melhores propriedades da natureza.
Buddleja davidii é uma espécie pertencente a família das escrofulariáceas. Originária do noroeste da China e Japão, seu uso se estendeu por todo mundo como decoração de jardins.

Persona : Commandes
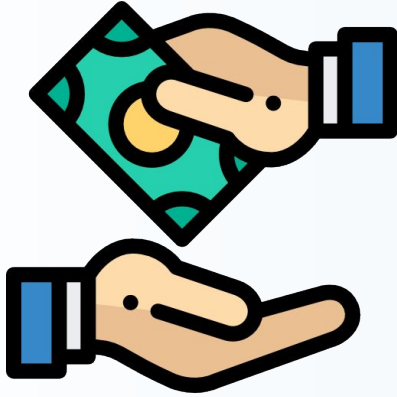
Score RFM et segmentation non supervisée

Score RFM moyen par clusters
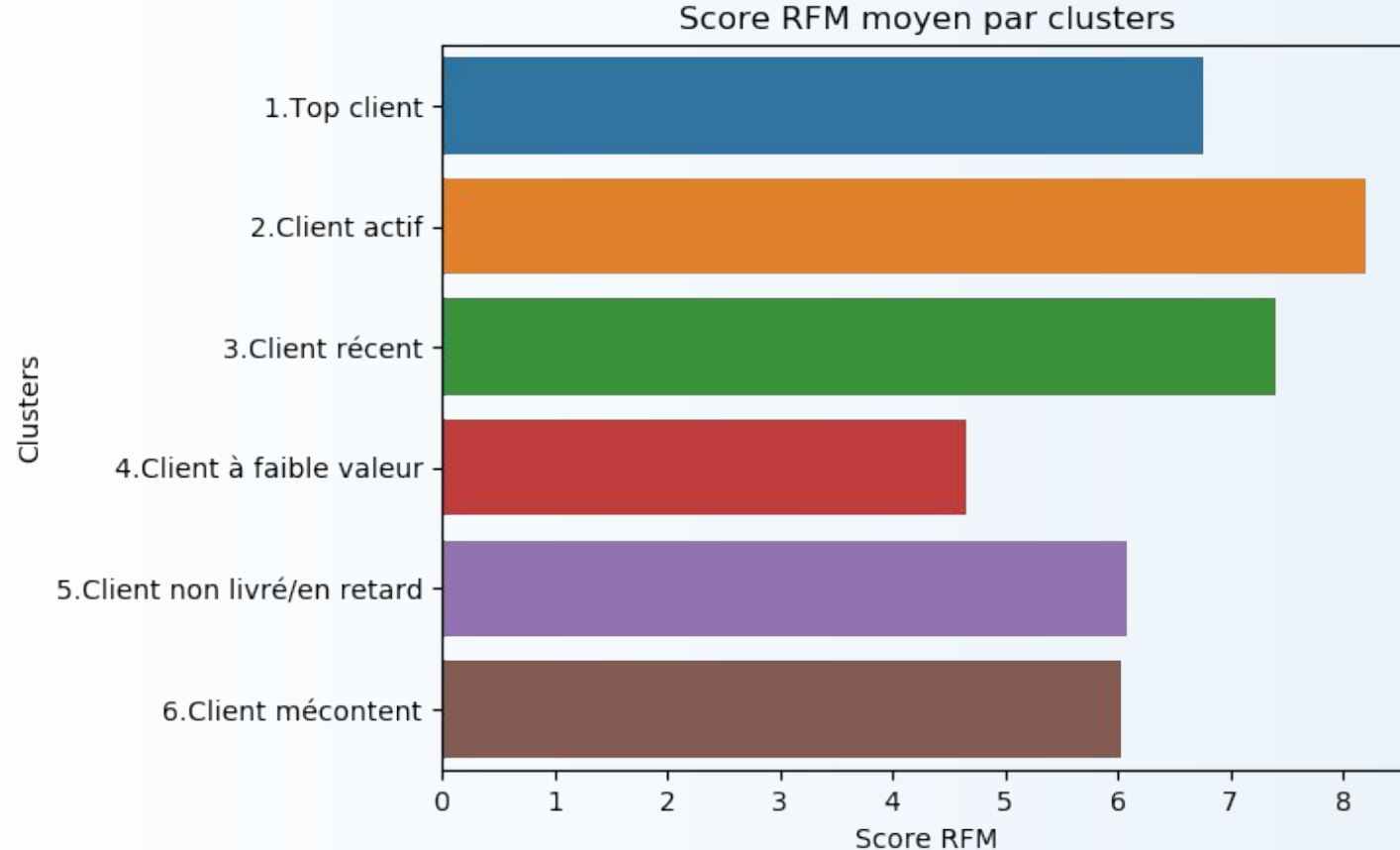
Etude de la stabilité des clusters

Stabilité trimestrielle des segments du 01/10/2017 au 31/08/2018