

Segmentation des clients

olist

Présentation de la base de données

► orga. des tables

données

Cleaning

Feature engineering

RFM

Happy

Achat

Modélisation

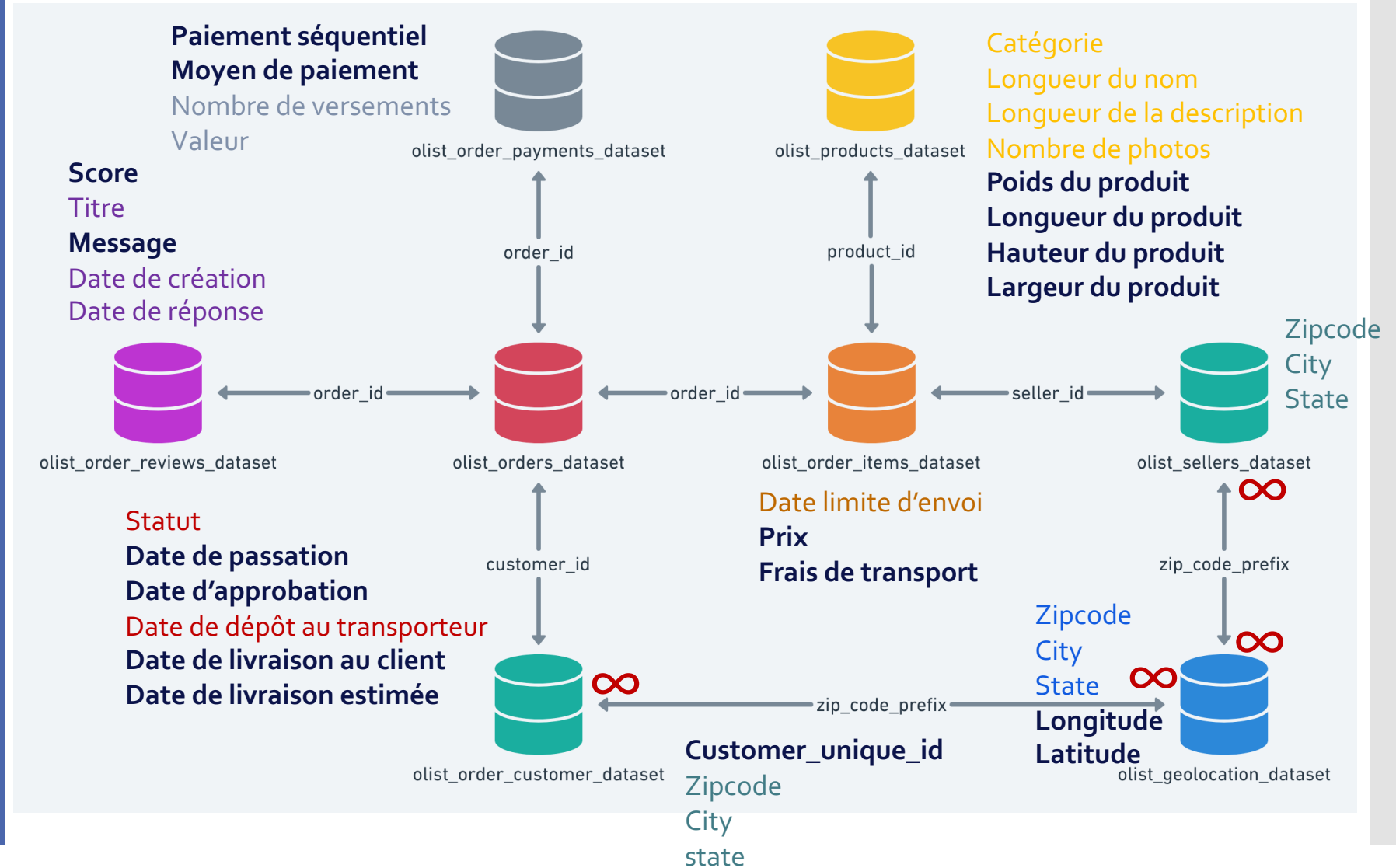
Outils de mesure

Les différents tests

Les résultats

Rendu client

Maintenance



Présentation de la base de données

organisation des tables

► **données**

Cleaning

Feature engineering

RFM

Happy

Achat

Modélisation

Outils de mesure

Les différents tests

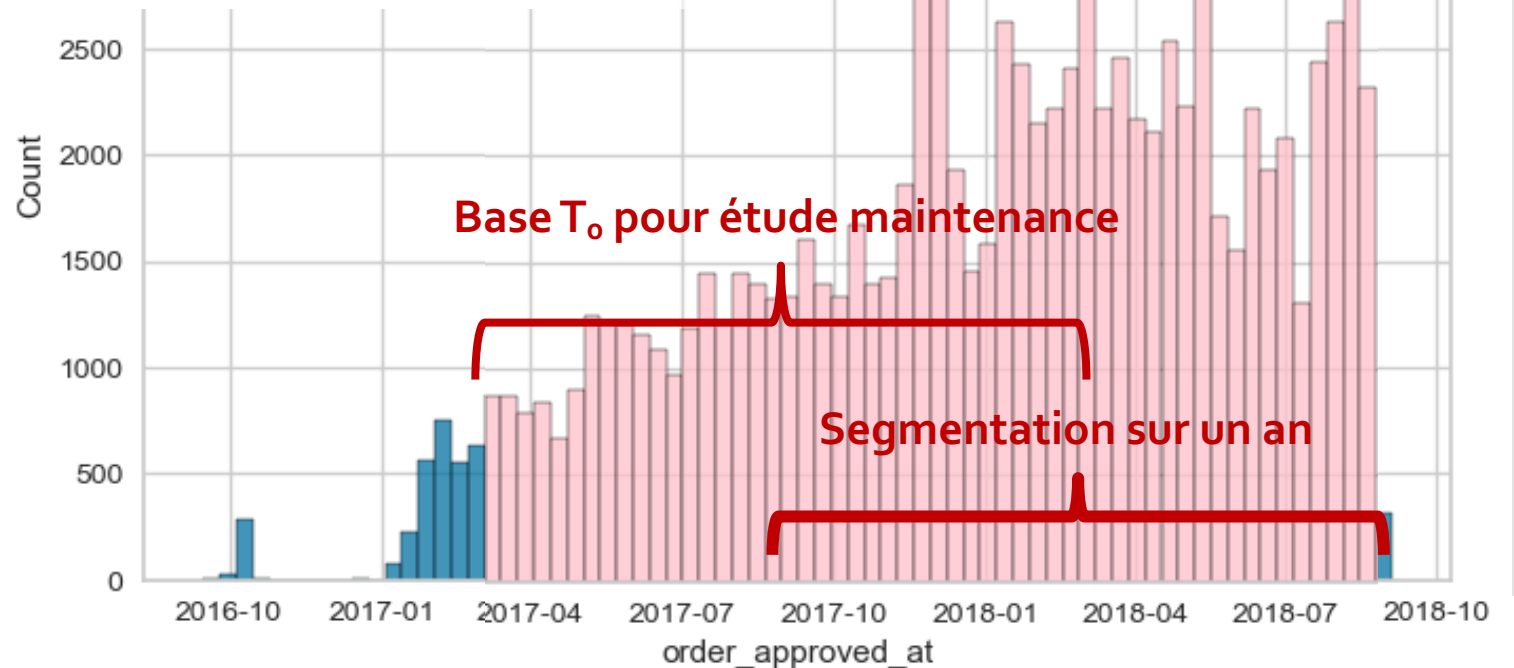
Les résultats

Rendu client

Maintenance

- Dataset à disposition :

- 95 827 clients dont 3% avec plusieurs commandes
 - ⚠ La fréquence risque d'être peu exploitable pour l'analyse RFM
- 99 441 commandes, dont 87% avec un seul article
 - ⚠ L'étude des catégories d'articles sera sans doute inexploitable
- un an et demi de données exploitables



Présentation de la base de données

organisation des tables
données

► Cleaning

Feature engineering

RFM

Happy

Achat

Modélisation

Outils de mesure

Les différents tests

Les résultats

Rendu client

Maintenance

Table	Suppression	Nombre de lignes supprimées	Commentaires
Geolocation	Latitude > 10	4 zip codes	Pas de suppression d'individus. Des positions médianes seront données pour les zip codes manquants
toutes	Sur merge how = 'inner'	Environ 0,7% de clients	Perte d'individus ayant une commande sans article par exemple
Dataframe de regroupement	Frequence > 15	1 individu	
	Distance acheteur vendeur > 4000 km	0 individu	Il n'y a plus de cas avec l'ajout de la suppression de la latitude > 10 cependant nous laissons ce point pour d'éventuels nouveaux clients
	Price_total_sum > 10 000	1 individu	
	Freight_value_sum > 1 500	1 individu	

Présentation de la base de données

organisation des tables
données

Cleaning

► Feature engineering

RFM

Happy

Achat

Modélisation

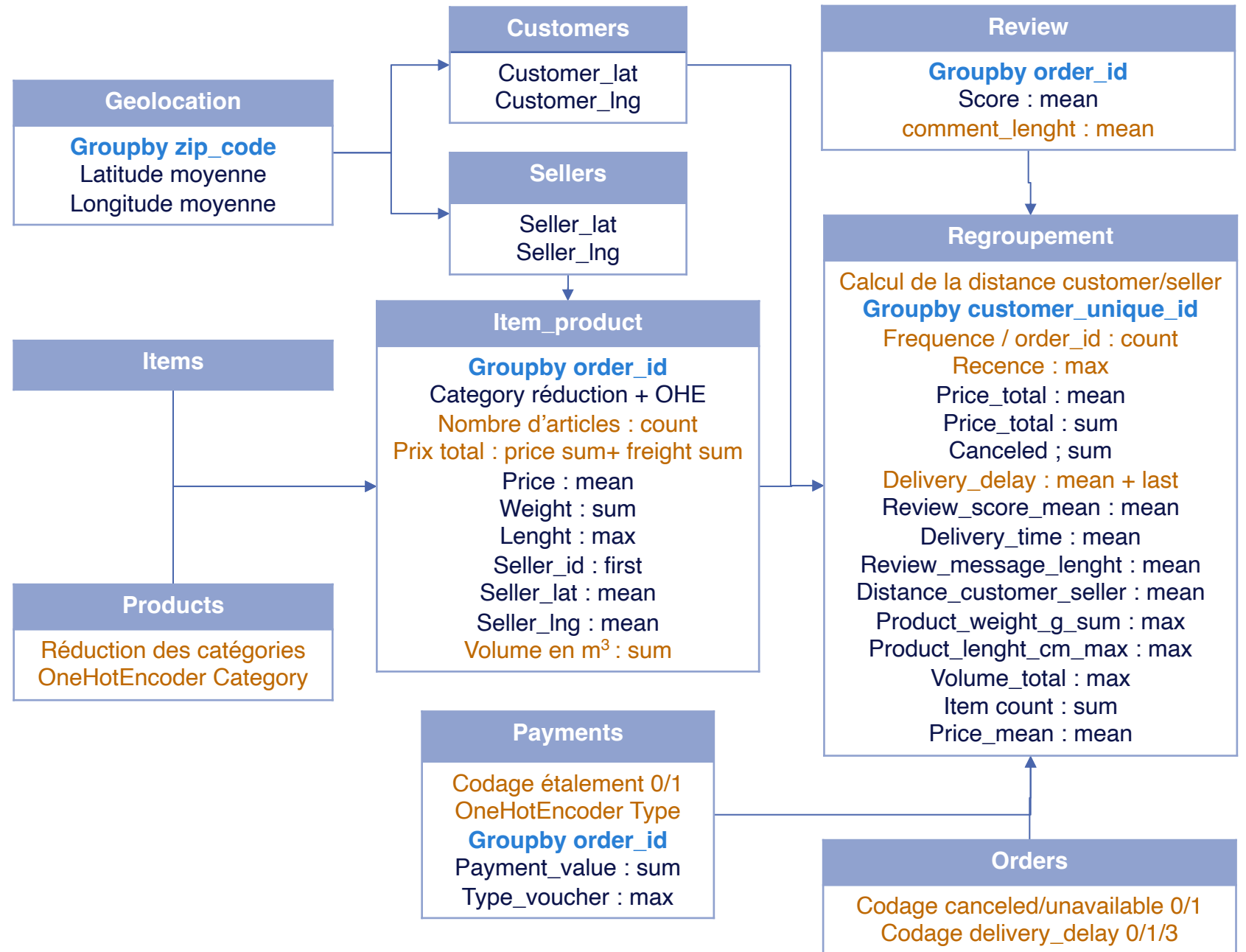
Outils de mesure

Les différents tests

Les résultats

Rendu client

Maintenance



Présentation de la base de données

organisation des tables données

Cleaning

Feature engineering

► **RFM**

Happy

Achat

Modélisation

Outils de mesure

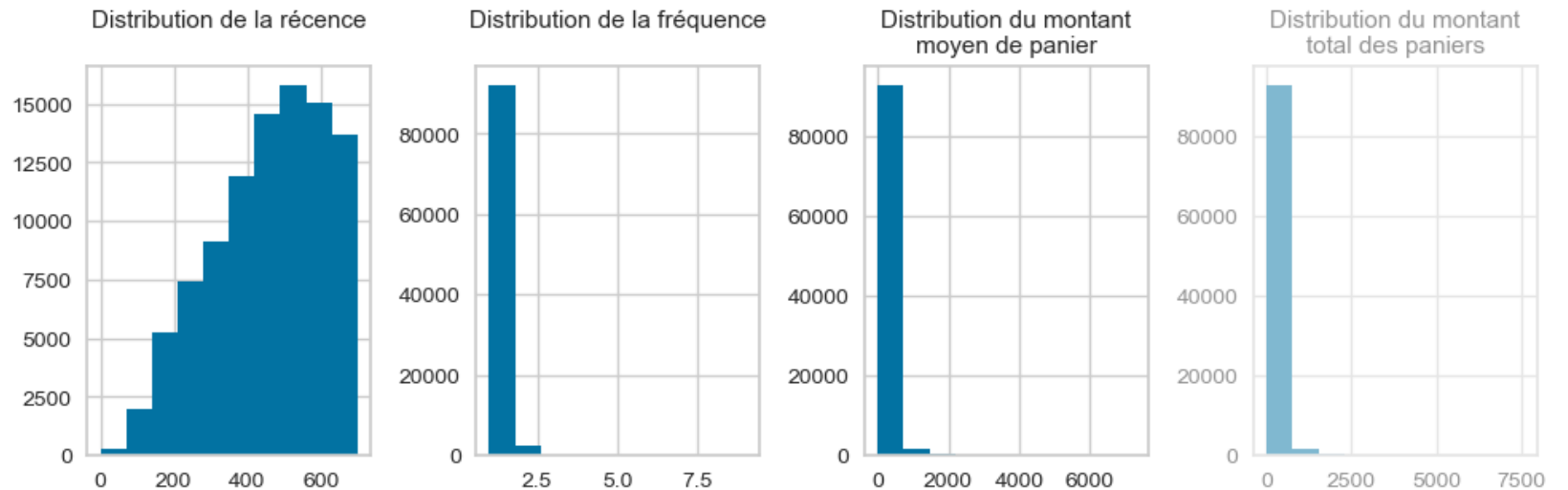
Les différents tests

Les résultats

Rendu client

Maintenance

- Variables RFM et leur distribution



Présentation de la base de données

organisation des tables données

Cleaning

Feature engineering

RFM

► **Happy**

Achat

Modélisation

Outils de mesure

Les différents tests

Les résultats

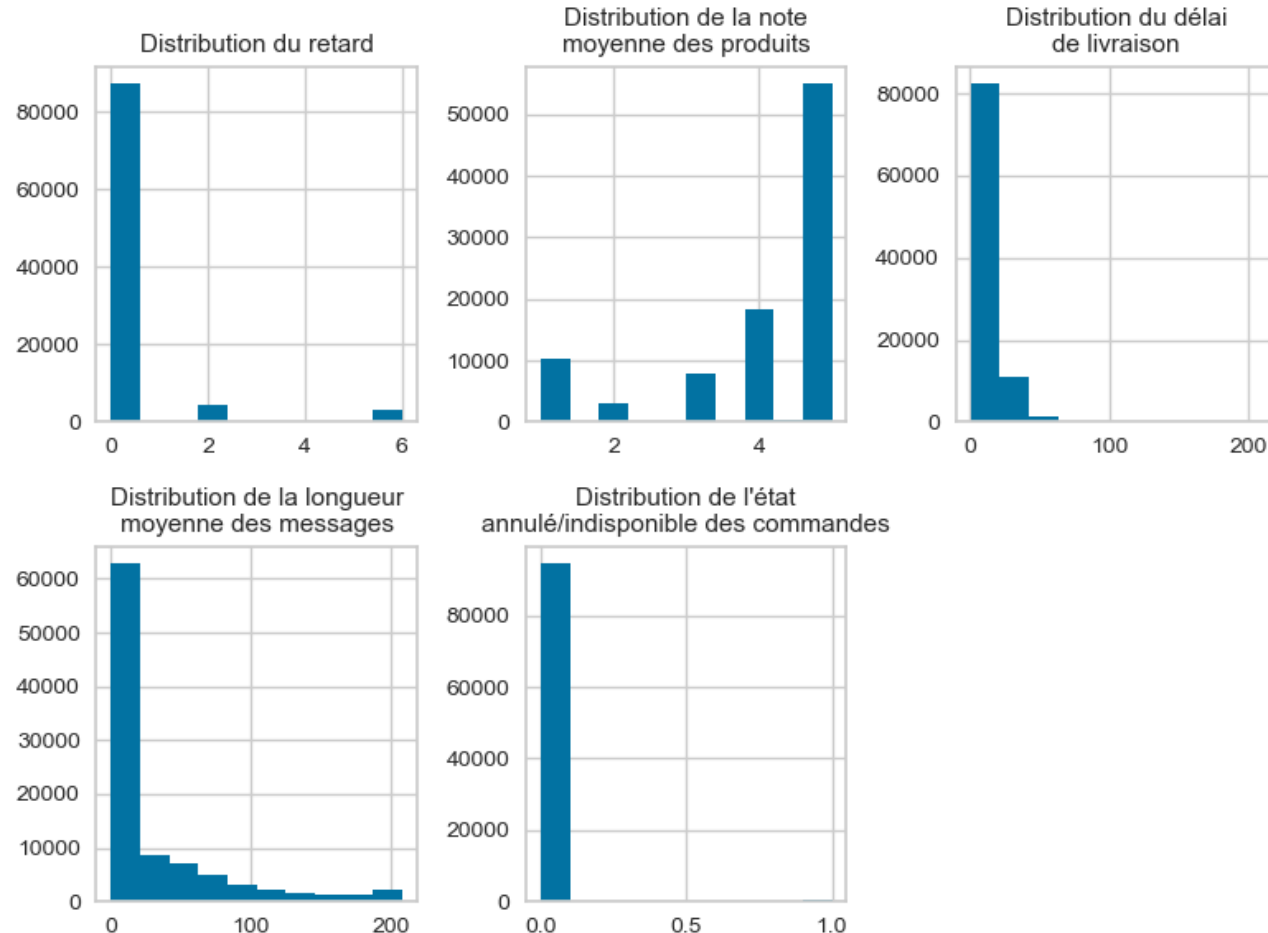
Rendu client

Maintenance

- Variables segmentation Happy

Avant regroupement : retard = 0 si nul, 1 si < 7 jours, sinon 3

Après regroupement : retard = retard.last + retard.mean



Présentation de la base de données

organisation des tables données

Cleaning

Feature engineering

RFM

Happy

► **Achat**

Modélisation

Outils de mesure

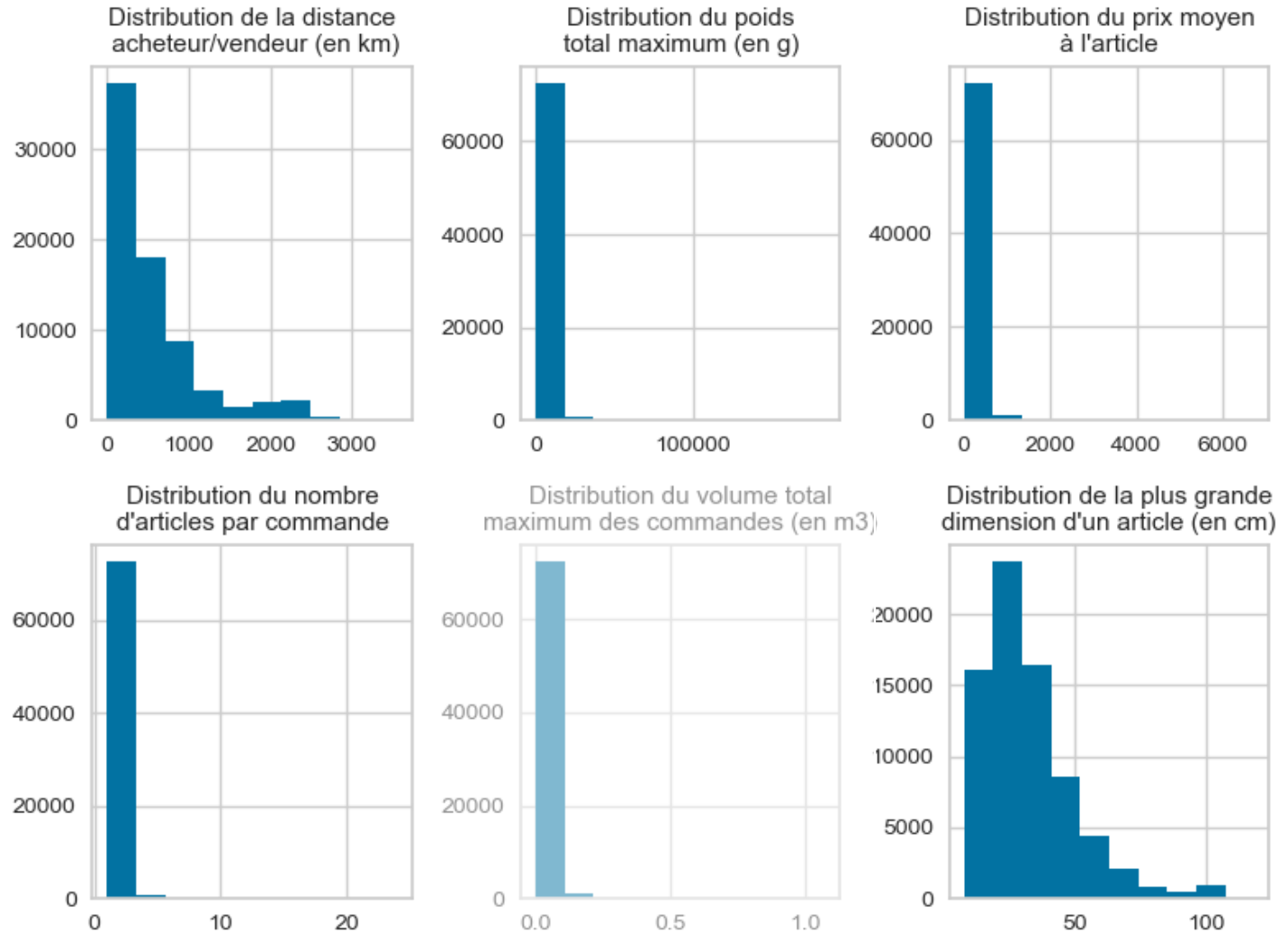
Les différents tests

Les résultats

Rendu client

Maintenance

- Variables segmentation Achat



Présentation de la base de données

organisation des tables
données

Cleaning

Feature engineering

RFM

Happy

Achat

Modélisation

► **Outils de mesure**

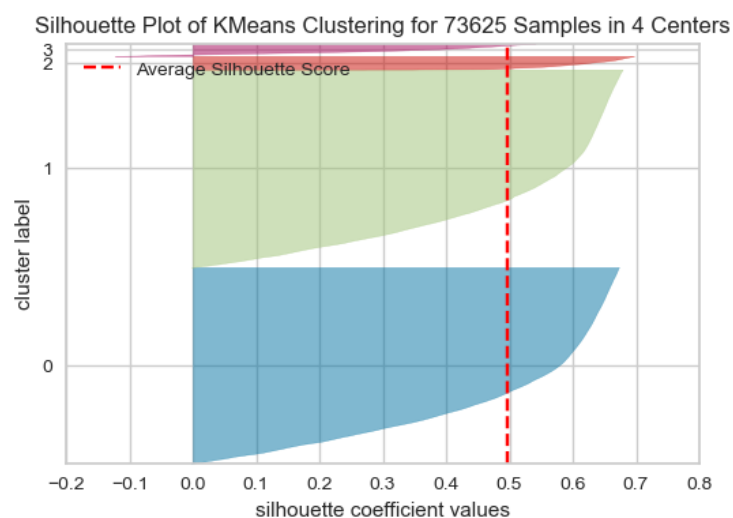
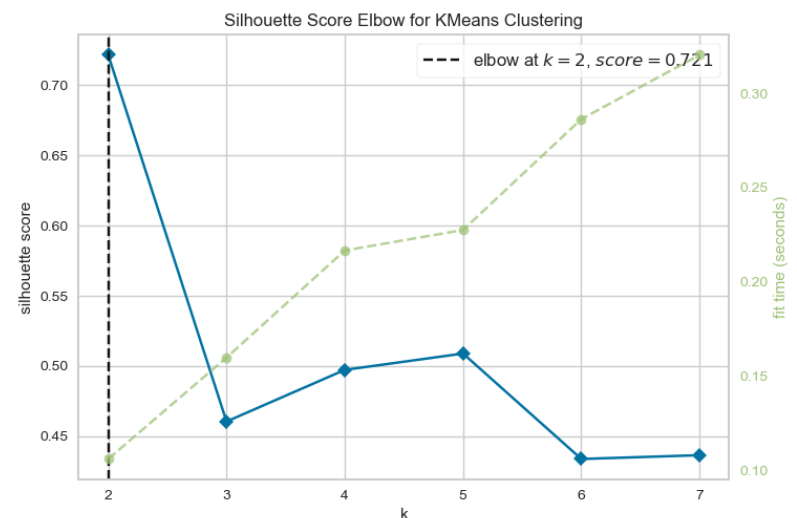
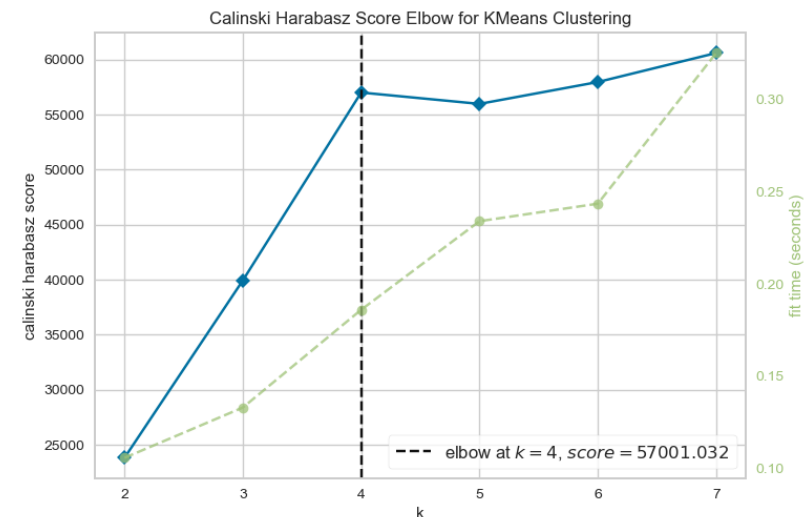
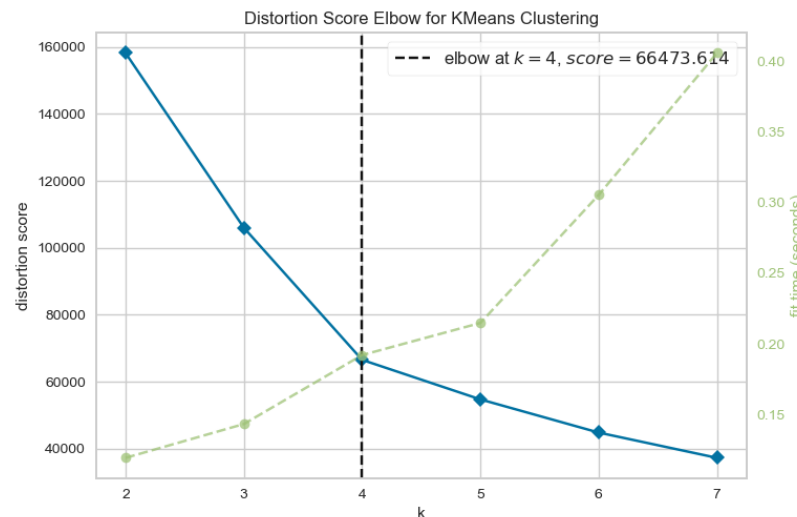
Les différents tests

Les résultats

Rendu client

Maintenance

• Détermination du nombre de clusters



Présentation de la base de données

organisation des tables données

Cleaning

Feature engineering

RFM

Happy

Achat

Modélisation

► Outils de mesure

Les différents tests

Les résultats

Rendu client

Maintenance

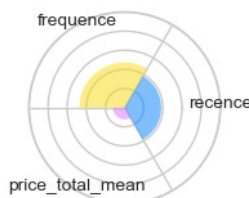
• Validation des clusters générés

Rfm_mean, StandardScaler

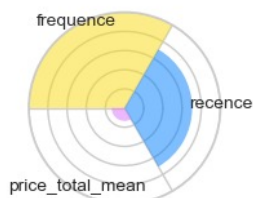
Cluster 0
47.1% des individus
38.0% du chiffre d'affaires



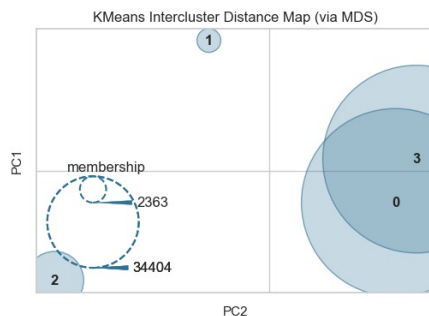
Cluster 1
46.7% des individus
37.0% du chiffre d'affaires



Cluster 2
3.2% des individus
6.0% du chiffre d'affaires

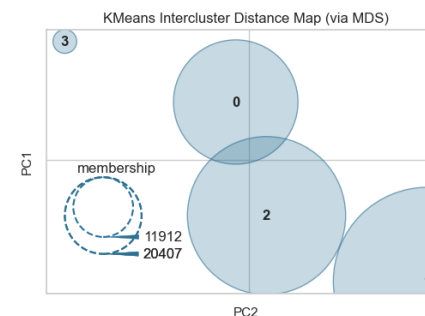
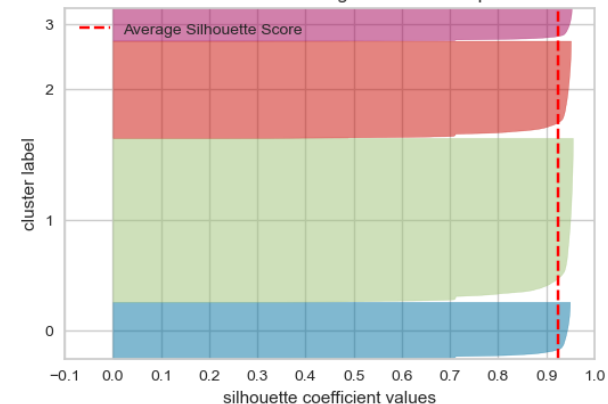


Cluster 3
2.9% des individus
19.0% du chiffre d'affaires

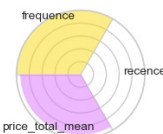


Rfm_simple, MinMaxScaler

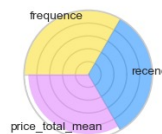
Silhouette Plot of KMeans Clustering for 73627 Samples in 4 Centers



Cluster 0
46.8% des individus
45.5% du chiffre d'affaires



Cluster 1
9.3% des individus
9.1% du chiffre d'affaires



Cluster 2
27.7% des individus
28.5% du chiffre d'affaires



Cluster 3
16.2% des individus
16.9% du chiffre d'affaires



Présentation de la base de données

organisation des tables données

Cleaning

Feature engineering

RFM

Happy

Achat

Modélisation

Outils de mesure

► Les différents tests

Les résultats

Rendu client

Maintenance

Paramètre testé	Scaling	Transf. log	Méthode	Nombre de variables	Silhouette _score	Nombre de clusters	Séparation des clusters	Exploitation métier
Nombre de variables	StandardScaler		KMeans	52	0,296	2	mauvaise	impossible
Base	StandardScaler		KMeans	3 - base	0,498	4	chevauchement de 2 clusters	Aisée
Transf. log	StandardScaler	oui	KMeans	3 - log sur price_total_mécan	0,368	4	clusters presque bien séparés	Aisée
Feature scaling			KMeans	3 - codage recence	0,641	4	superposition de 3 clusters	discrimination sur price uniquement
Feature scaling			KMeans	3 - base	0,422	4	chevauchement de 3 clusters	discrimination sur recence et price
Feature scaling	MinMaxScaler		KMeans	3 - codage recence	0,925	4	bonne séparation	discrimination sur recence uniquement
Feature scaling	MinMaxScaler		KMeans	3 - base	0,538	4	bon	discrimination sur recence uniquement
Feature scaling	RobustScaler		KMeans	3 - codage recence	0,483	4	chevauchement de 3 clusters	discrimination sur price uniquement
Feature scaling	RobustScaler		KMeans	3 - base	0,448	4	superposition de 2 clusters	discrimination sur price uniquement
Méthode	StandardScaler		DBSCAN	3 - base	0,729	6/7		inexploitable
Complexité des variables	StandardScaler		KMeans	3 - codage recence	0,591	4	chevauchement de 2 clusters	Aisé

Présentation de la base de données

organisation des tables données

Cleaning

Feature engineering

RFM

Happy

Achat

Modélisation

Outils de mesure

Les différents tests

► **Les résultats**

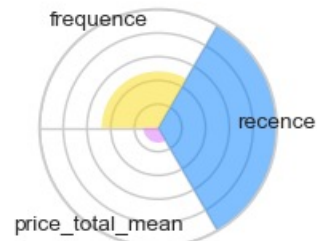
Rendu client

Maintenance

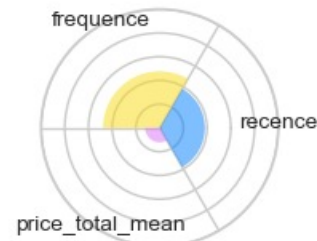
• Segmentation RFM

Cluster	% de clients	% de CA	Type de clients
0	47,4 %	38,5 %	Nouveaux clients
1	47,1 %	37,6 %	Clients inactifs, pas de commande récente, une seule commande sur la dernière année, et montant du panier faible
2	2,5 %	4,7 %	Clients réguliers, montant moyen du panier faible
3	2,9 %	19,2 %	Très bons clients sur le montant des paniers

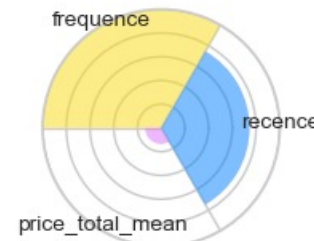
Cluster 0
47.4% des individus
38.5% du chiffre d'affaires



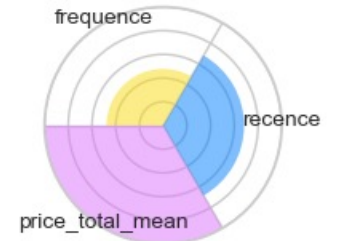
Cluster 1
47.1% des individus
37.6% du chiffre d'affaires



Cluster 2
2.5% des individus
4.7% du chiffre d'affaires



Cluster 3
2.9% des individus
19.2% du chiffre d'affaires



Présentation de la base de données

organisation des tables
données

Cleaning

Feature engineering

RFM

Happy

Achat

Modélisation

Outils de mesure

Les différents tests

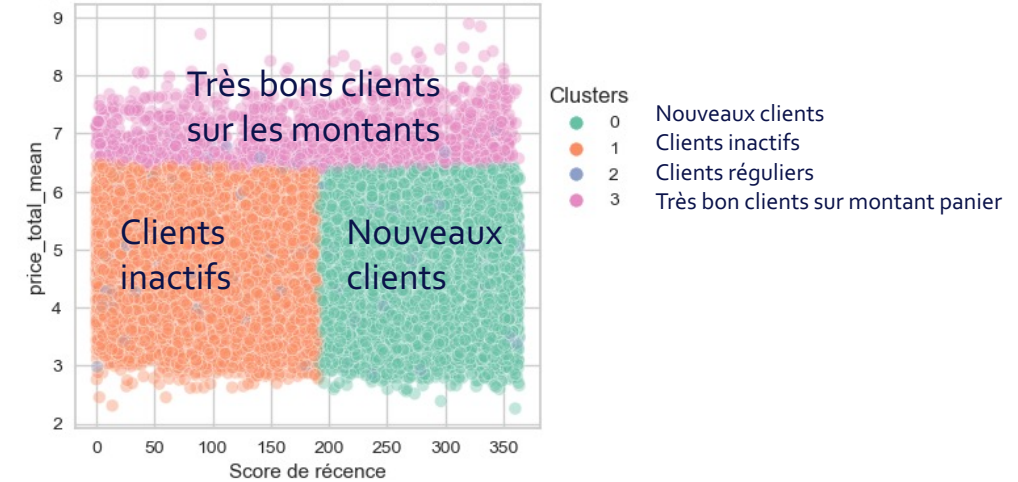
► **Les résultats**

Rendu client

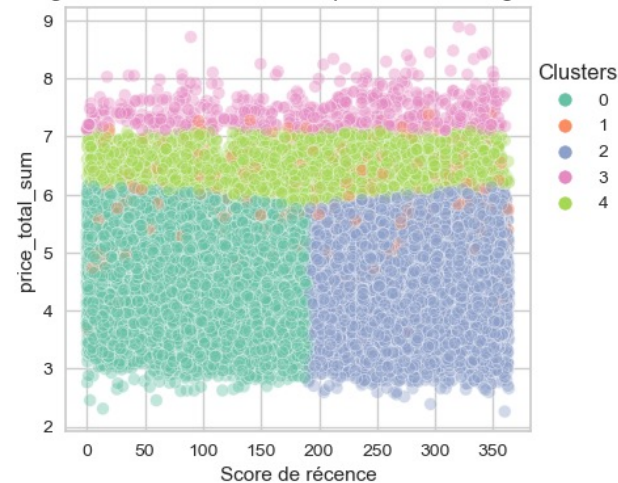
Maintenance

- Variation du nombre de clusters sur segmentation RFM

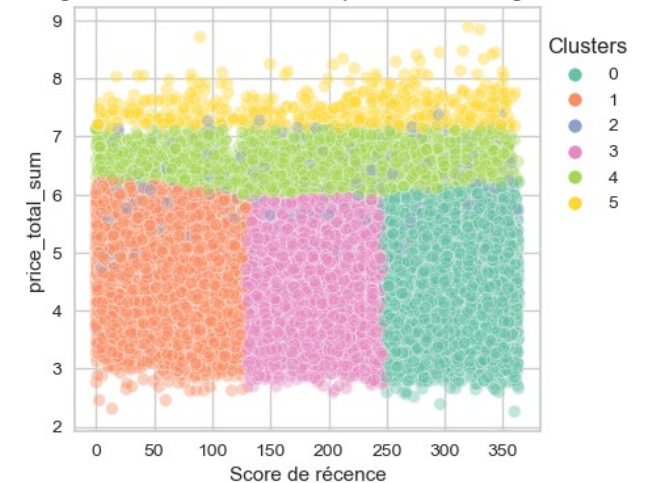
Segmentation RFM Montant du panier moyen - 4 segments



Segmentation RFM Montant du panier total - 5 segments



Segmentation RFM Montant du panier total - 6 segments



Présentation de la base de données

organisation des tables
données

Cleaning

Feature engineering

RFM

Happy

Achat

Modélisation

Outils de mesure

Les différents tests

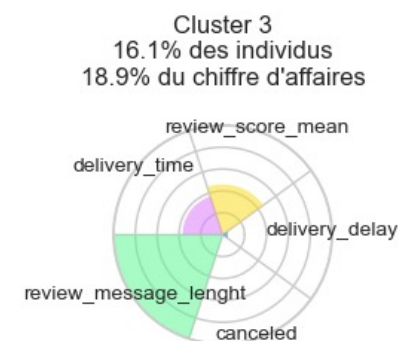
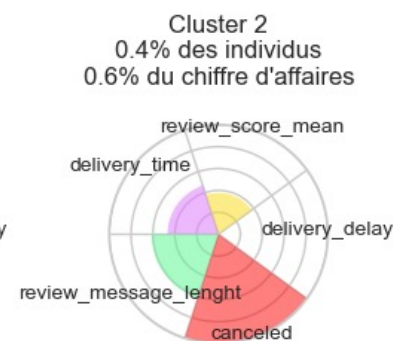
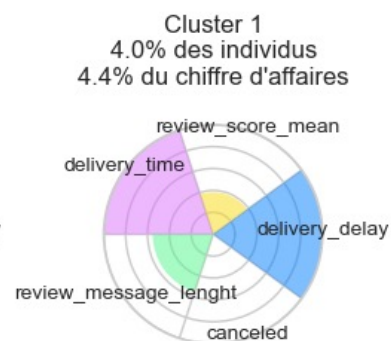
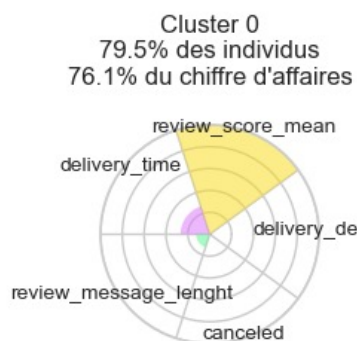
► Les résultats

Rendu client

Maintenance

• Segmentation Happyness

Cluster	% de clients	% de CA	Type de clients
0	79,5 %	76,0 %	Clients très satisfaits
3	16,1 %	18,9 %	Clients globalement peu satisfaits des produits, ayant laissé de longs commentaires avec la note, clients qui ont des choses à dire et veulent s'exprimer
1	4,0 %	4,5 %	Clients très insatisfaits à cause de longs délais de livraison et de gros retards
2	0,4 %	0,6 %	Clients très insatisfaits, ayant annulé leur commande ou ayant vu leur commande annulée pour indisponibilité des produits



Présentation de la base de données

organisation des tables données

Cleaning

Feature engineering

RFM

Happy

Achat

Modélisation

Outils de mesure

Les différents tests

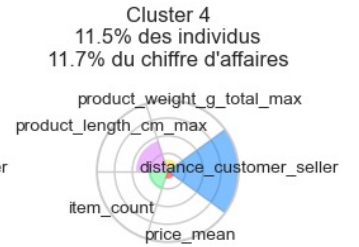
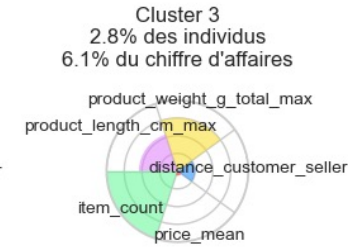
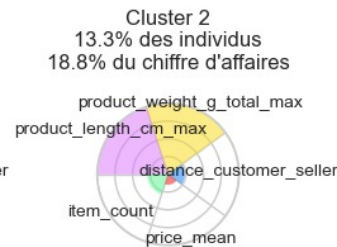
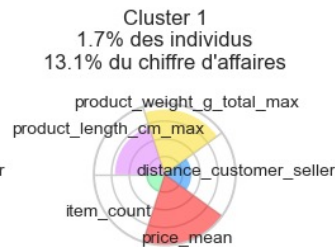
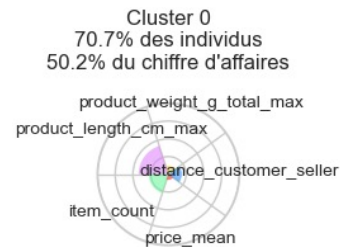
► **Les résultats**

Rendu client

Maintenance

• Segmentation Type d'achat

Cluster	% de clients	% de CA	Type de clients
0	70,7%	50,2 %	Clients sans critère particulier quant à la livraison
1	1,7 %	13,1 %	Clients se faisant livrer des articles lourds et à un prix élevé (hors port ici), donc sensibles à l'aspect pratique qu'offre une livraison pour ne pas avoir à porter des articles lourds et de qualité, qu'il ne faut donc pas abîmer pendant le transport
2	13,3 %	18,8 %	Clients ayant commandés un article encombrant et plutôt lourd, mais à un prix bas, donc sensibles aux avantages qu'offre la livraison mais pas à n'importe quel prix
3	2,8 %	6,1 %	Clients ayant commandé de nombreux articles, plutôt lourds au total, à un prix bas, profiteraient de devoir faire un achat pour ajouter des articles et "ne pas se faire livrer pour rien"
4	11,5 %	11,7 %	Clients très éloignés des points de vente, donc sensibles à l'aspect pratique qu'offre une livraison pour ne pas avoir à faire le trajet



Présentation de la base de données

organisation des tables
données

Cleaning

Feature engineering

RFM

Happy

Achat

Modélisation

Outils de mesure

Les différents tests

Les résultats

► **Rendu client**

Maintenance

- Fichier csv de segmentation

	RFM	happyness	usefull_delivery
customer_unique_id			
0000366f3b9a7992bf8c76cfd3221e2	new	happy	none
0000b849f77a49e4a4ce2b2a4ca5be3f	new	happy	none
0000f6ccb0745a6a4b88665a16c9f078	idle	happy	long_distance
0004aac84e0df4da2b147fca70cf8255	idle	happy	none
0004bd2a26a76fe21f786e4fbd80607f	new	happy	bulky_heavy_cheap
...
fffb09418989a0dbff854a28163e47c6	idle	happy	none
fffbf87b7a1a6fa8b03f081c5f51a201	idle	happy	long_distance
fffcc512b7dfecaffd80f13614af1d16	excellent	unhappy_about_products	none
fffea47cd6d3cc0a88bd621562a9d061	idle	happy	long_distance
ffff5962728ec6157033ef9805bacc48	new	unhappy_about_products	none

73627 rows × 3 columns

- Notebooks formatés avec Autopep8

Présentation de la base de données

organisation des tables données

Cleaning

Feature engineering

RFM

Happy

Achat

Modélisation

Outils de mesure

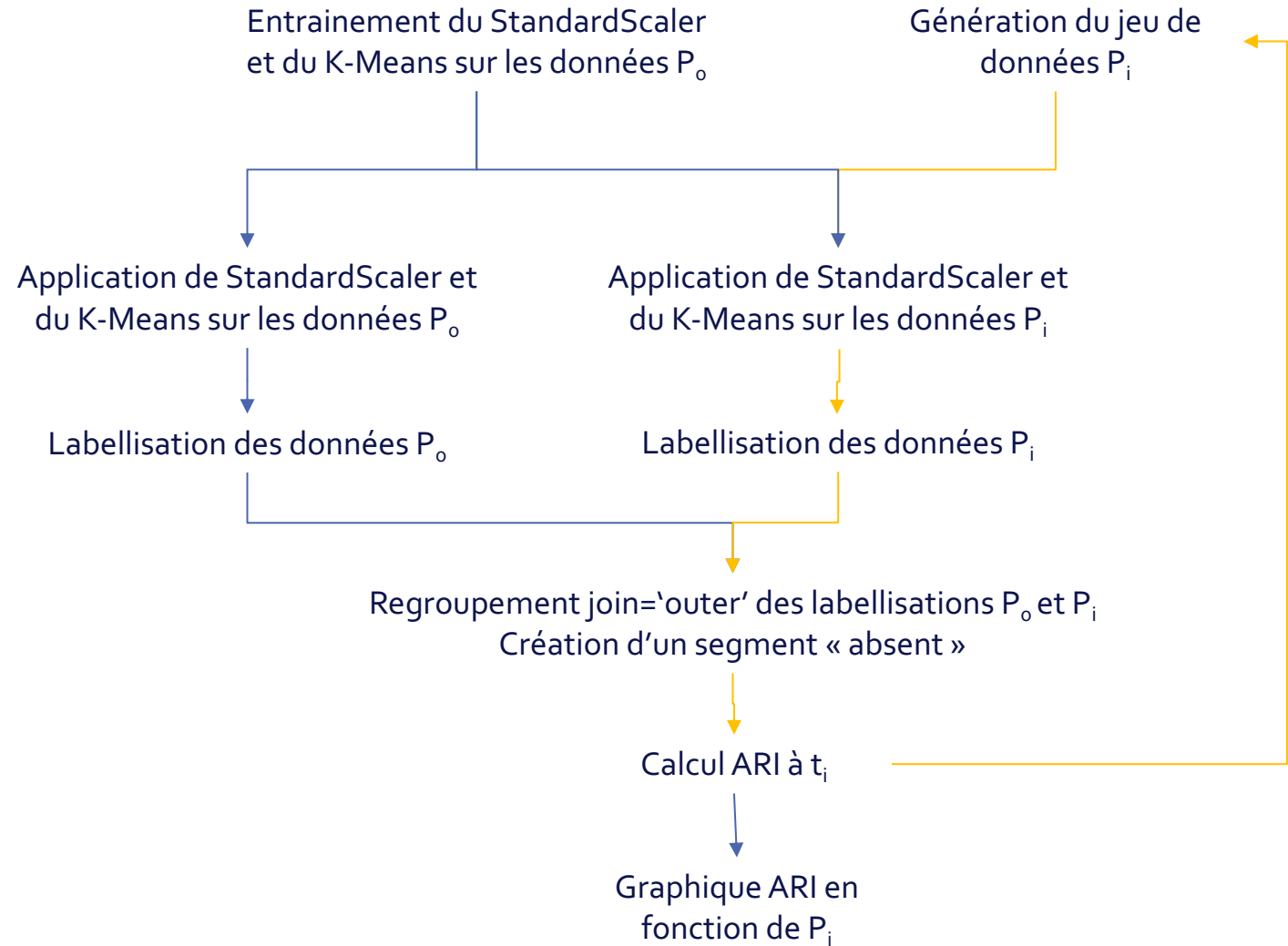
Les différents tests

Les résultats

Rendu client

► **Maintenance**

- Stratégie de calcul du temps de maintenance



Présentation de la base de données

organisation des tables données

Cleaning

Feature engineering

RFM

Happy

Achat

Modélisation

Outils de mesure

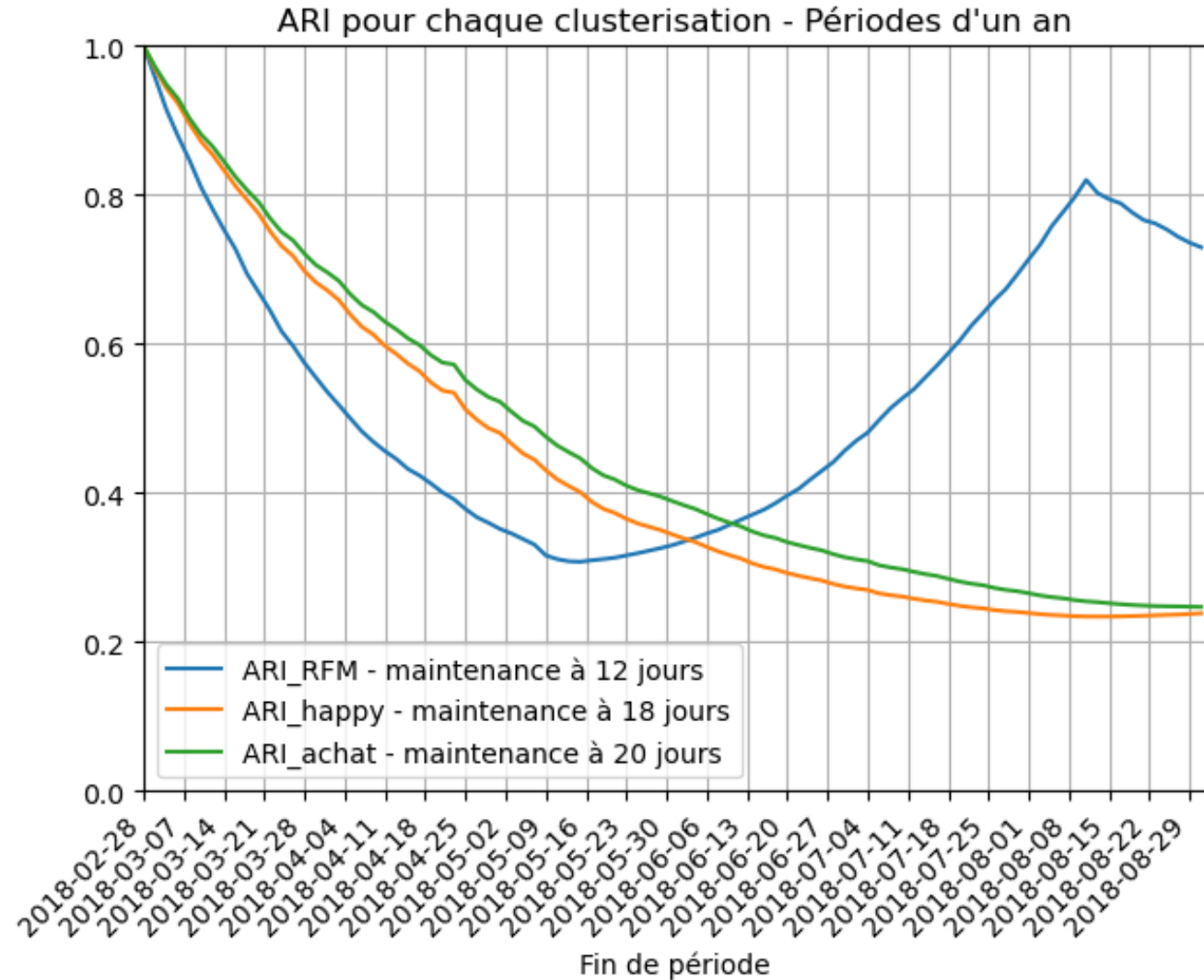
Les différents tests

Les résultats

Rendu client

► **Maintenance**

- Maintenance légère, d'une heure : tous les 12 jours



Présentation de la base de données

organisation des tables données

Cleaning

Feature engineering

RFM

Happy

Achat

Modélisation

Outils de mesure

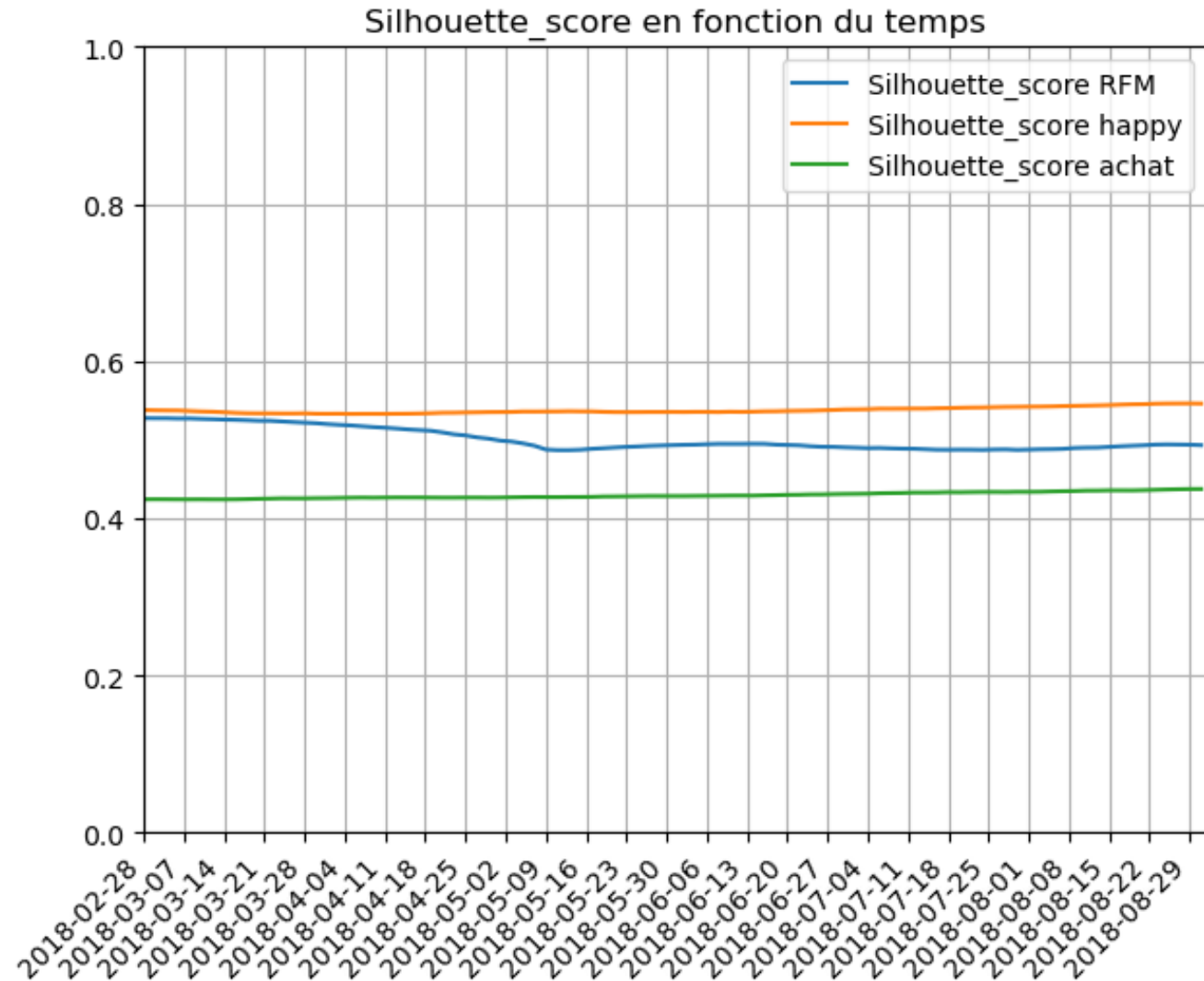
Les différents tests

Les résultats

Rendu client

► **Maintenance**

- Pas de maintenance lourde à prévoir, segmentation stable



Merci pour votre attention