

Projet 5

Segmentez des clients d'un site e-commerce

Fournir à Olist une segmentation de ses clients

Réaliser des campagnes marketing

- 1 - Sélection et nettoyage de variables
- 2 - Feature engineering
- 3 - Approche méthodologique
- 4 - Interprétation de la segmentation
- 5 - Définition d'un contrat de maintenance
- 6 - Conclusion

1- Sélection et nettoyage des variables

Base de données fournie par Olist

Customers

Données clients

Geolocation

Données localisations

- villes

- coordonnées

Reviews

Données satisfactions

- notes

- commentaires

Orders

Données commandes

- statuts

- dates de suivis

order_items

Données produits liés

à la commande

Sellers

Données vendeurs

Payments

Données paiements

- types

- montants

Products

Données produits

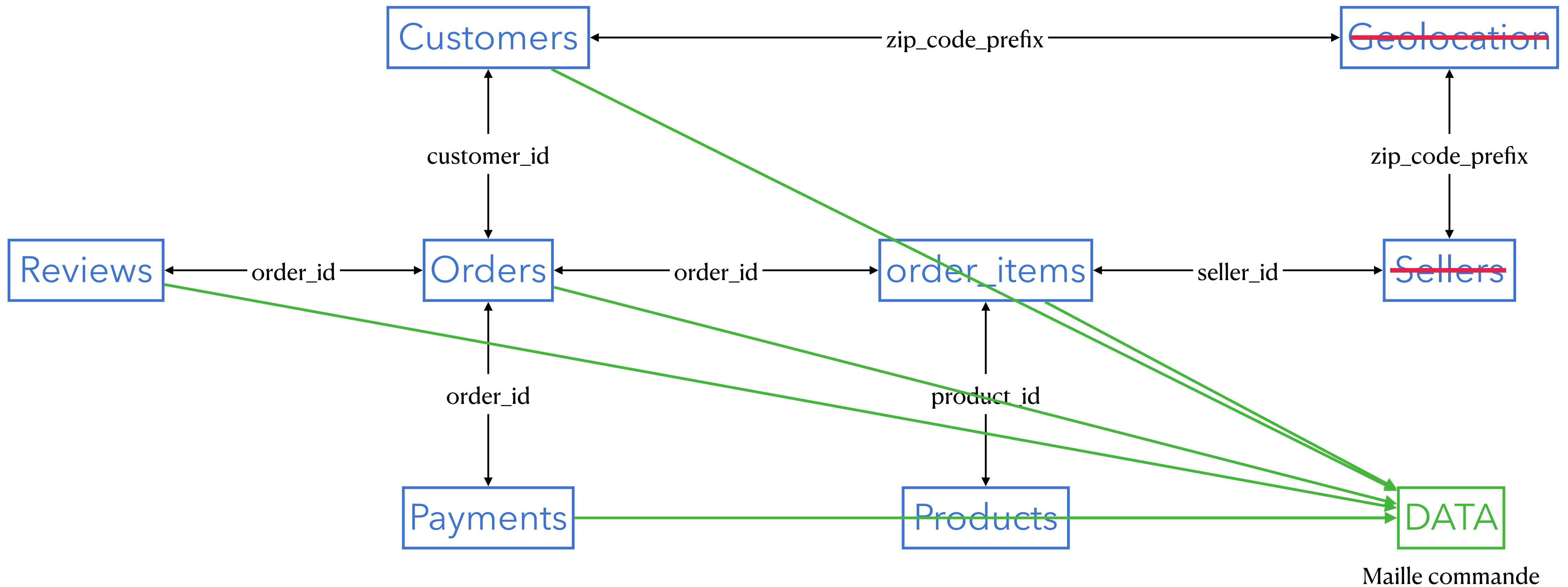
- catégorie

- dimensions

- description

1- Sélection et nettoyage des variables

Base de données fournie par Olist



1- Sélection et nettoyage des variables

DATA

Segmentation clients :
campagnes marketing

FE_1

R_{ecence}

F_{réquence}

M_{ontant}

S_{atisfaction}

Compréhension et mise en oeuvre
simple

Se concentre sur les aspects clés

Meilleure précision

Plus de personnalisation

FE_2

RFMS

+

Autres facteurs

1- Sélection et nettoyage des variables

DATA

Maille commande

Customers

Orders

Payments

Numéro client

Numéro de commande
Date et heure de commande

Type de paiement
+ valeur associée

Montant total

Indice d'échelonnement
de paiement
0 / 1

DATA_1

Complétion des
valeurs manquantes

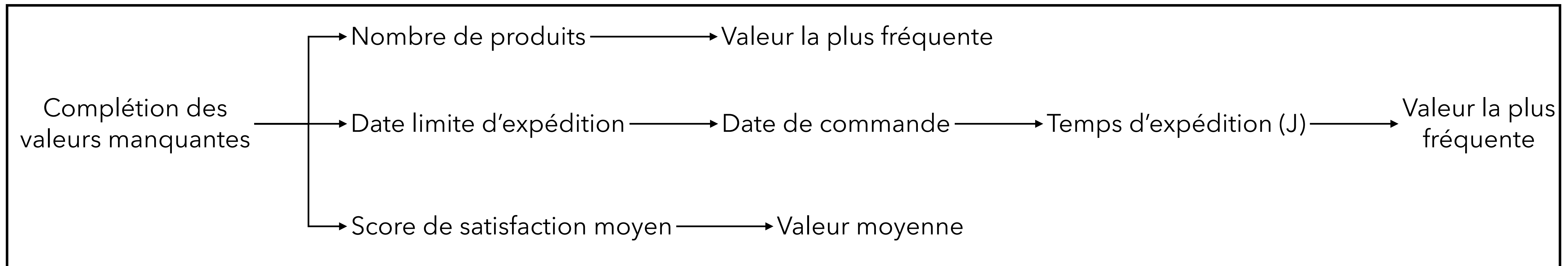
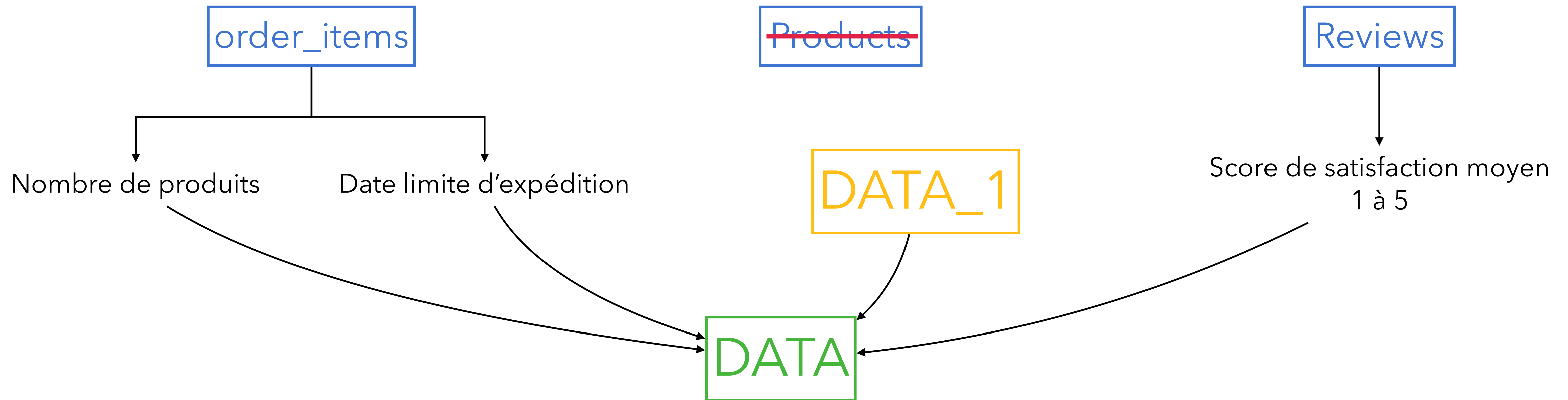
- Prix du produit
- Frais de port

Suppression de la commande

Montant total

Type de paiement
le plus fréquent

1- Sélection et nettoyage des variables



2- Feature engineering

DATA

- Numéro de client
- Numéro de commande
- Date et heure de commande
- Montant en espèce
- Montant en CB
- Montant en carte de débit
- Montant en voucher
- Montant total
- Indice d'échelonnement
- Nombre de produit
- Temps d'expédition (J)
- Score de satisfaction moyen

FE_1

R

- Recence
date de fin data - date de commande

F

- fréquence
Somme des commandes par client

M

- Montant
Somme totale dépensée par client

S

- Satisfaction
Moyenne des scores du client

FE_2

RFMS

+

Autres facteurs

- Différents types de paiement
- Nombre de produit total par client
- Temps d'expédition moyen
- Jour le plus fréquent de commande
- Moment de la journée le plus fréquent

3- Approche méthodologique

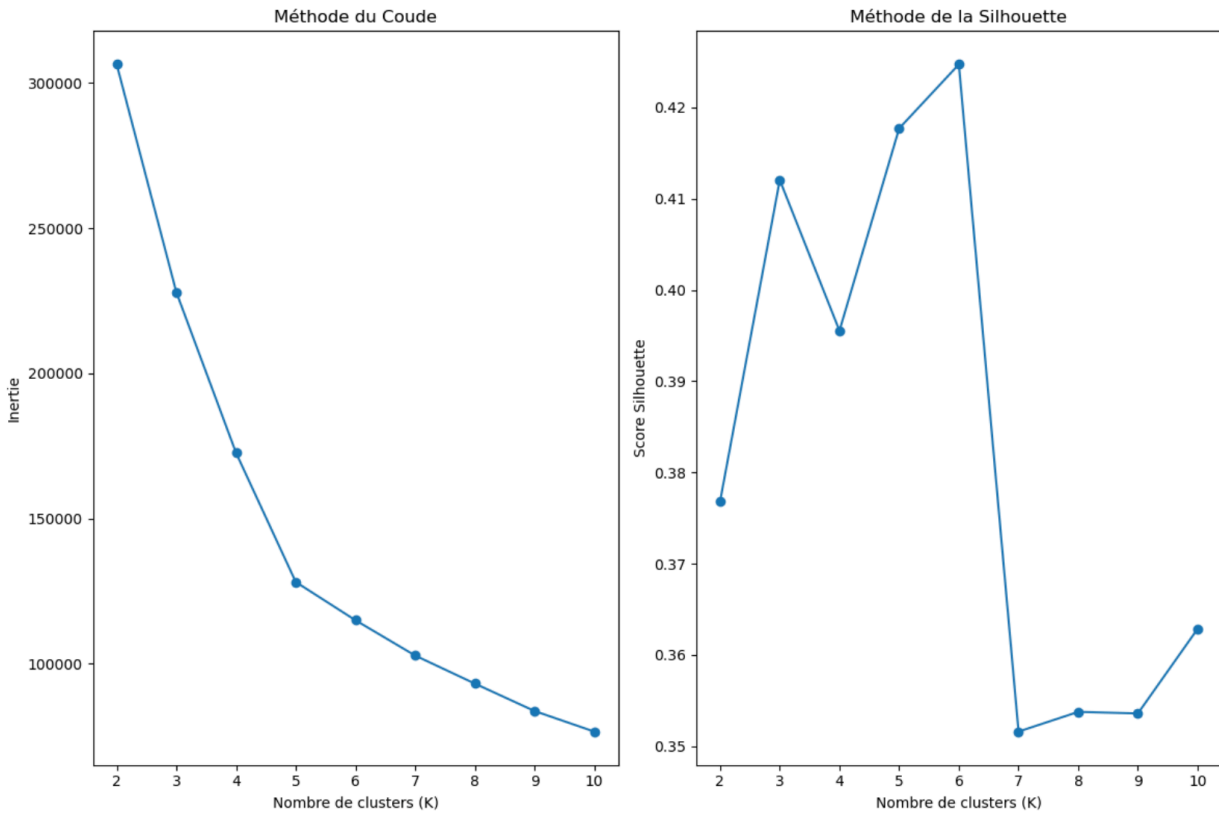
Choix et mise en place des modèles			
	KMeans	AgglomerativeClustering	DBSCAN
Choix	Populaire Facilité et simplicité d'interprétation	Regroupement hiérarchique	Regroupement par densité Capacité de détection de clusters complexes
Entrainement	Méthode du coude et de la silhouette	Dendrogramme	Tests de combinaisons d'hyperparamètres
Evaluation	Indice de Davies-Bouldin Qualité des clusters	Coefficient de silhouette Séparation des clusters	Score de stabilité Evaluation de la stabilité du modèle

3- Approche méthodologique

Entrainement des modèles

KMeans

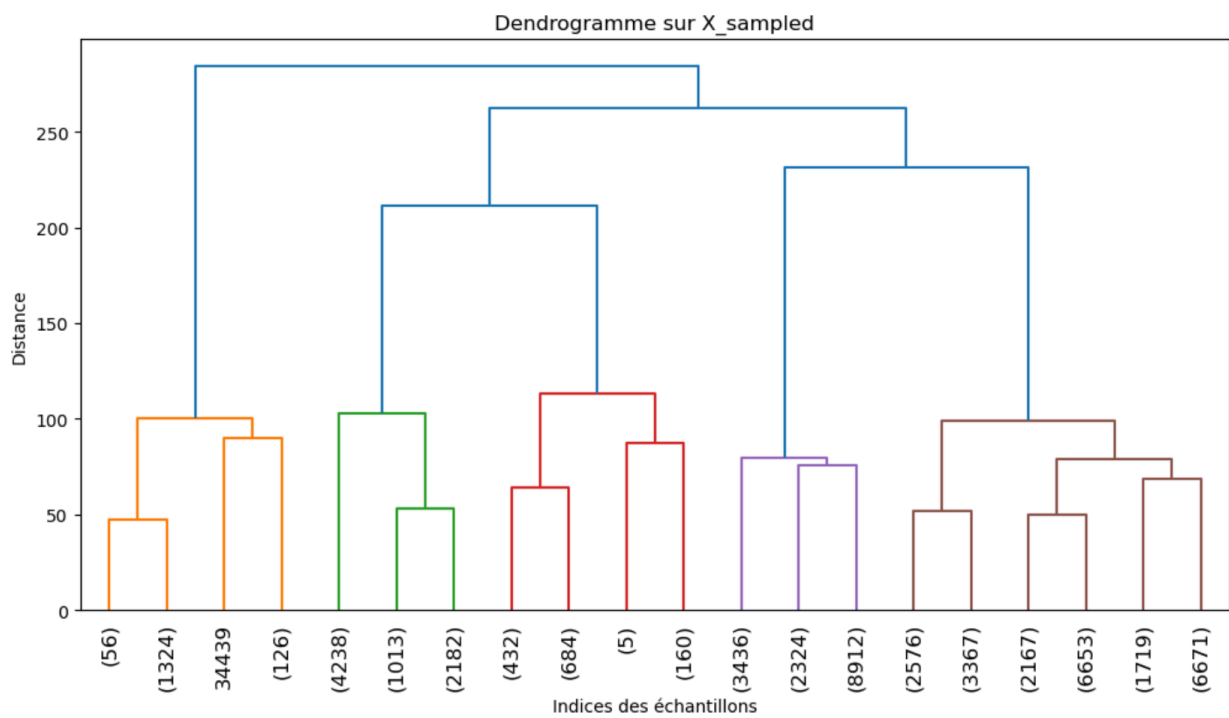
Méthode du coude
et de la silhouette



5 clusters

AgglomerativeClusteting

Dendrogramme



5 clusters

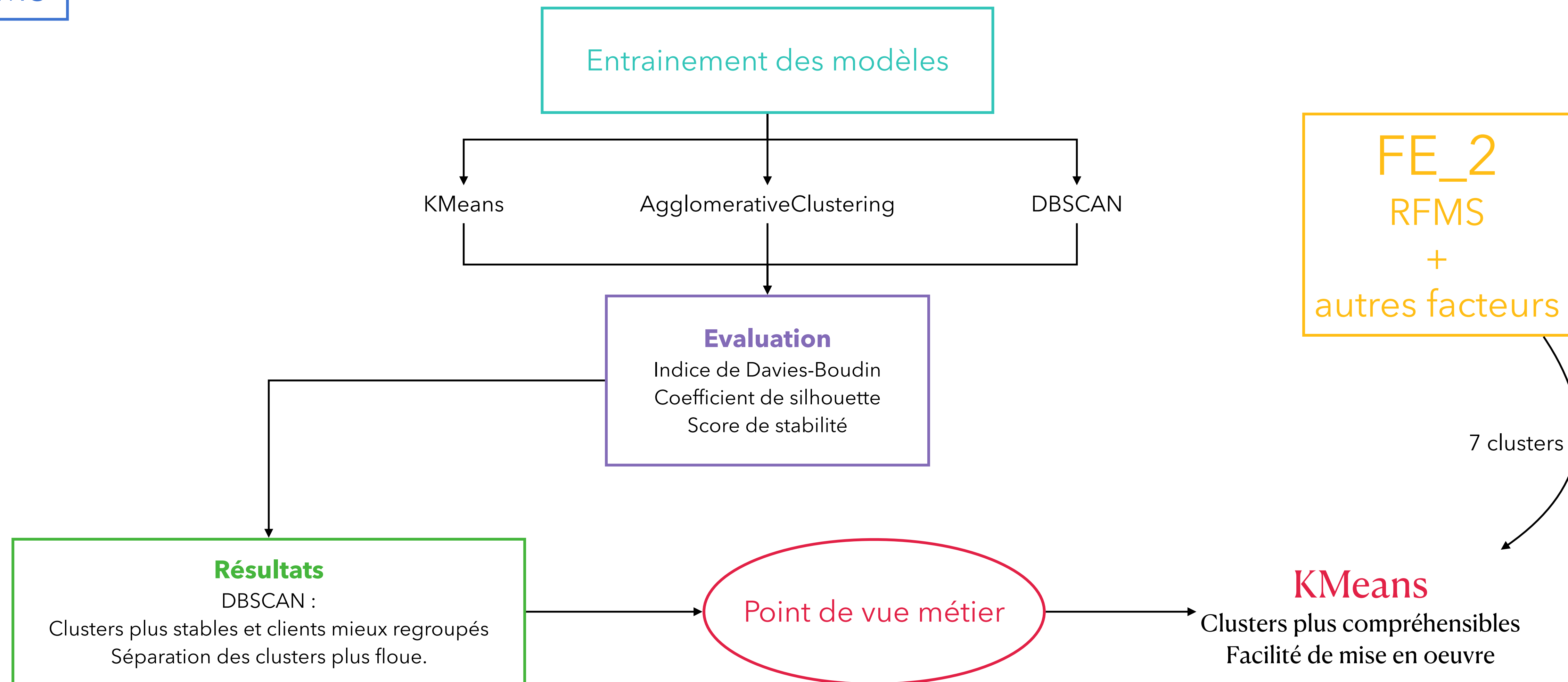
DBSCAN

Tests d'hyperparamètres

	Eps	Min_samples	Nbre_clusters	Nbre_outliers	Taille_moy_clusters	Davies_Bouldin_score
0	1.0	5	6	100	15998.833333	1.229297
1	1.0	6	6	107	15997.666667	1.222519
2	1.0	7	5	122	19194.200000	1.210245
3	1.3	5	7	61	13718.857143	1.217567
4	1.3	6	5	76	19203.400000	1.237751
5	1.3	7	4	84	24002.250000	1.197702
6	1.5	5	6	47	16007.666667	1.103372
7	1.5	6	6	55	16006.333333	1.214615
8	1.5	7	4	74	24004.750000	1.218712

6 clusters

3- Approche méthodologique



4- Interprétation de la segmentation

FE_1
RFMS

- Cluster 0 : « Inactifs »
- Cluster 1 : « Actifs »
- Cluster 2 : « Fidèles »
- Cluster 3 : « Dépensiers ».
- Cluster 4 : « Insatisfaits »

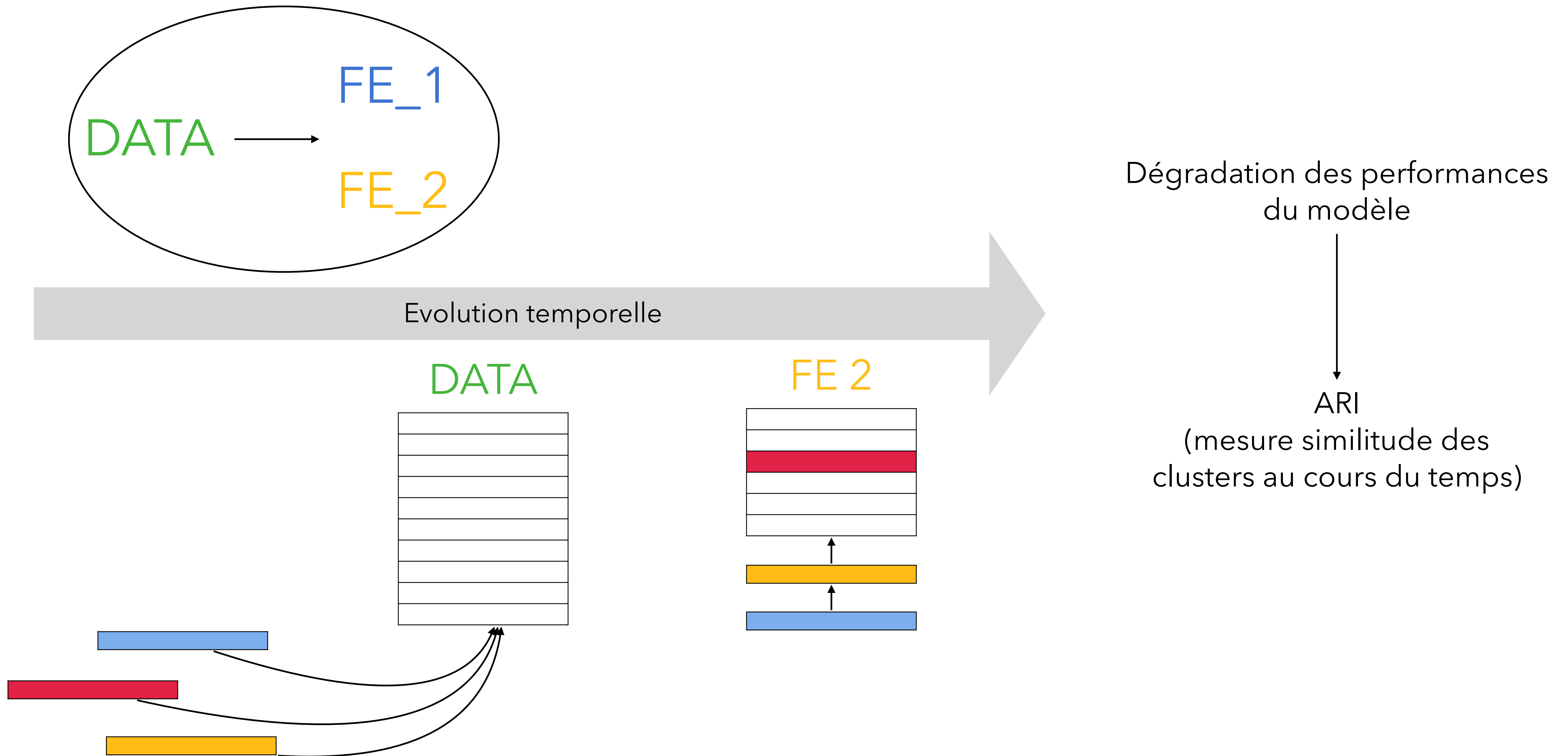
FE_2
RFMS + autres facteurs

- Cluster 0 : « Satisfaits avec Paiement Échelonné »
- Cluster 1 : « Satisfaits avec Faibles Dépenses et Paiement Échelonné »
- Cluster 2 : « Clients Inactifs et Insatisfaits »
- Cluster 3 : « Satisfaits avec Moyennes Dépenses et Fidélité »
- Cluster 4 : « Insatisfaits avec Faibles Dépenses »
- Cluster 5 : « Satisfaits avec Fortes Dépenses et Paiement Échelonné »
- Cluster 6 : « Satisfaits avec Fortes Dépenses en Espèces »

Réserve :

Interprétations subjectives basées sur la répartition des variables
—> Avis d'experts marketing plus approprié

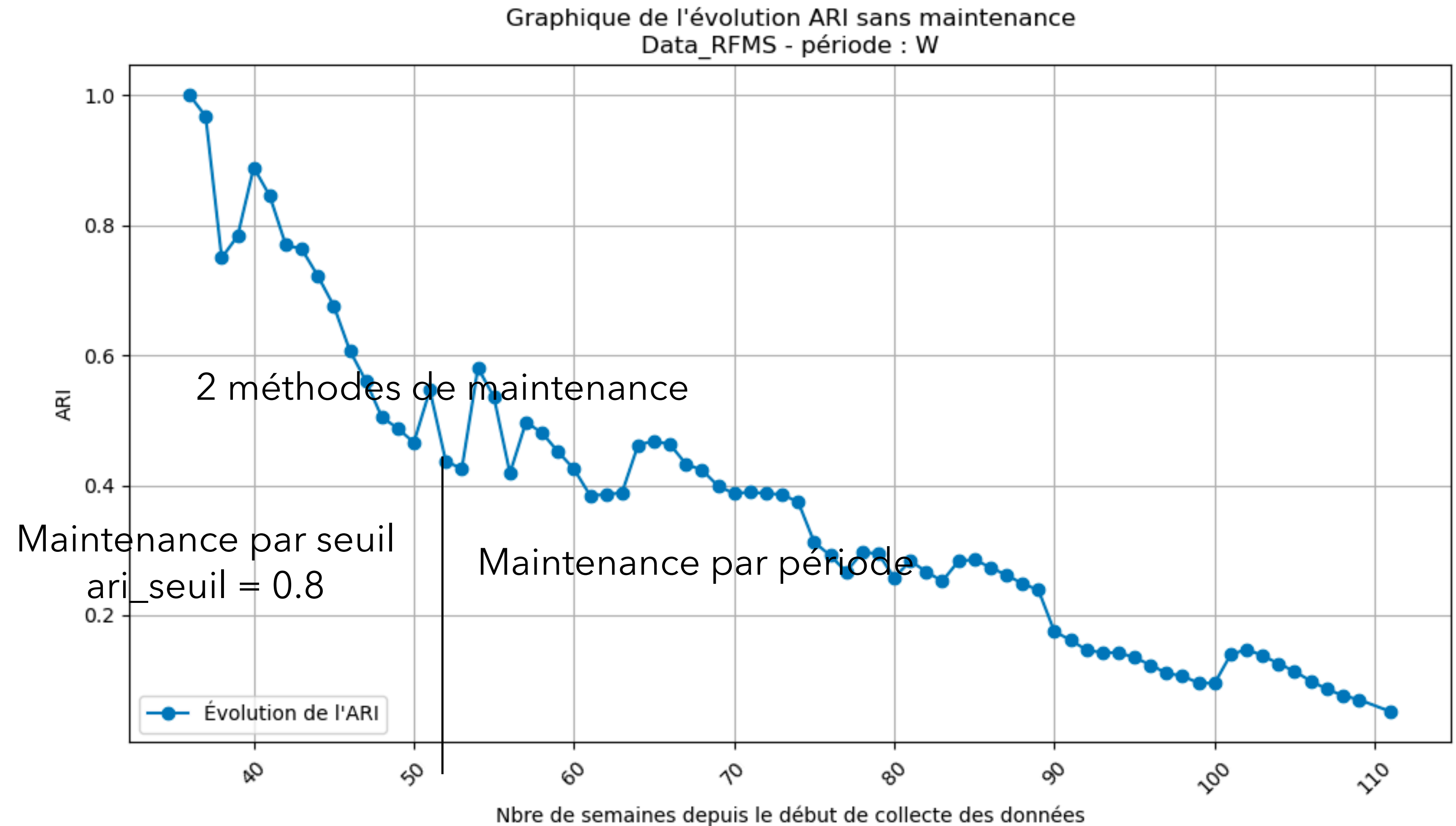
5- Définition d'un contrat de maintenance



5- Définition d'un contrat de maintenance

DATA

Période initiale	10 %
Période 1	
Période 2	
.	
.	
.	

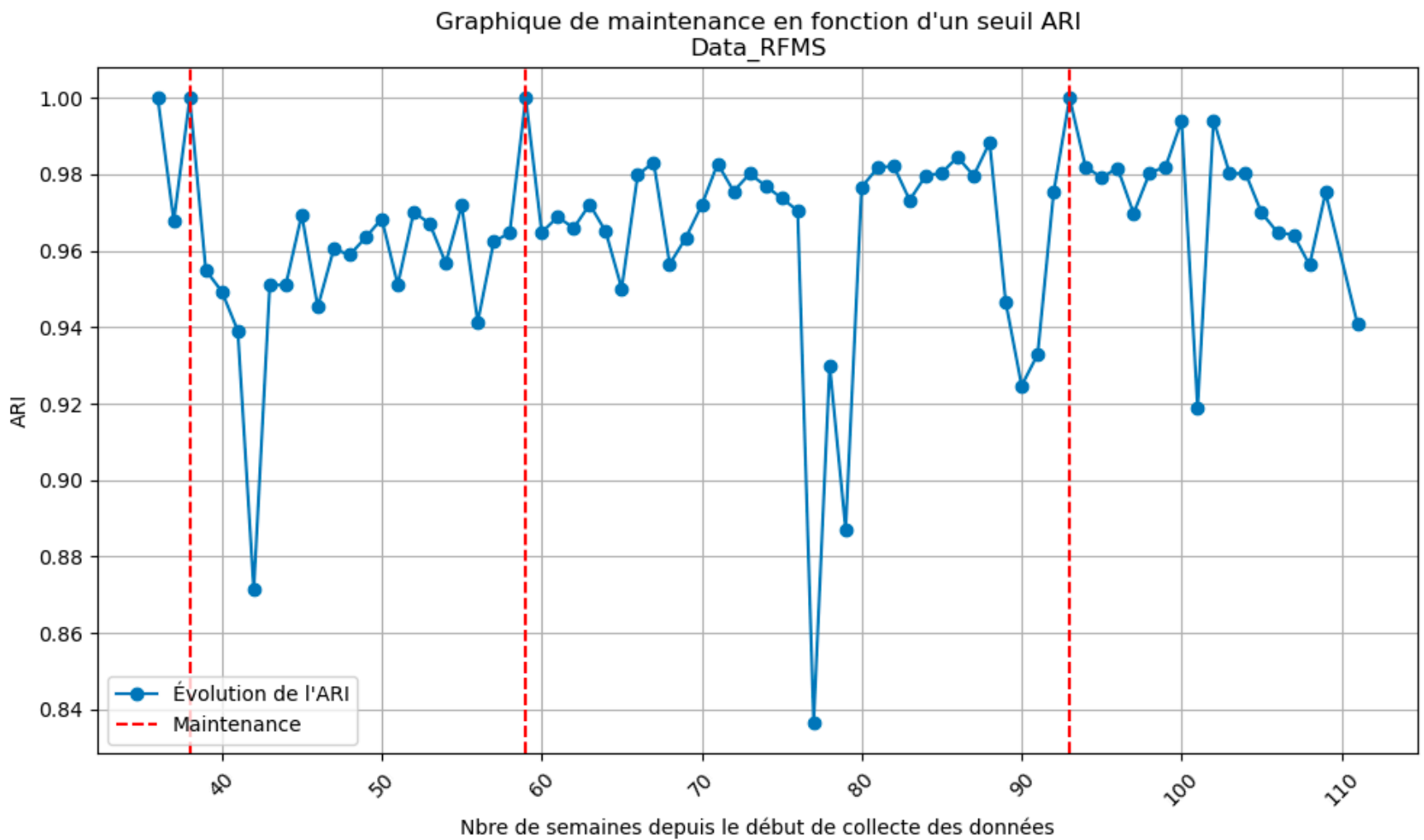
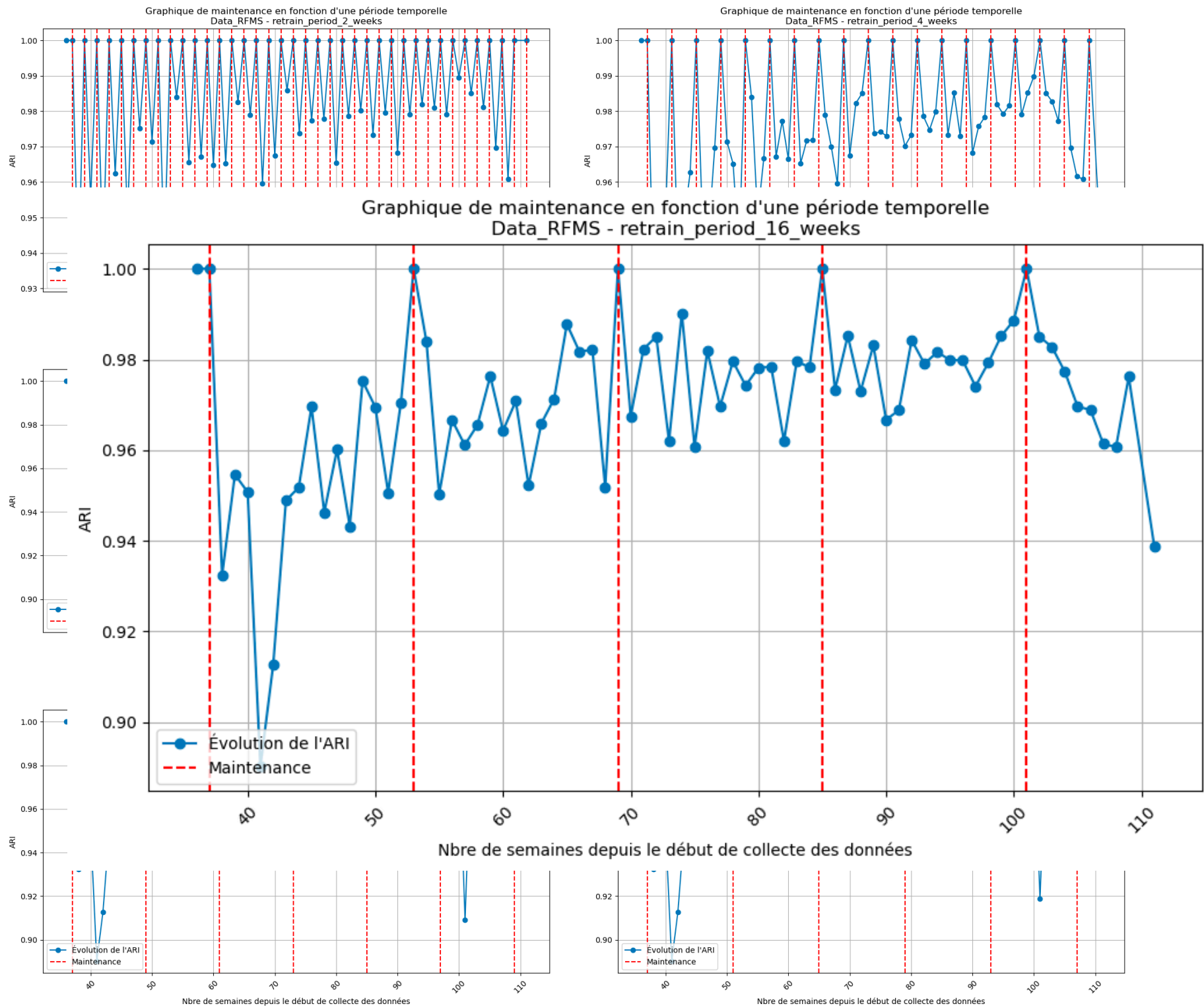
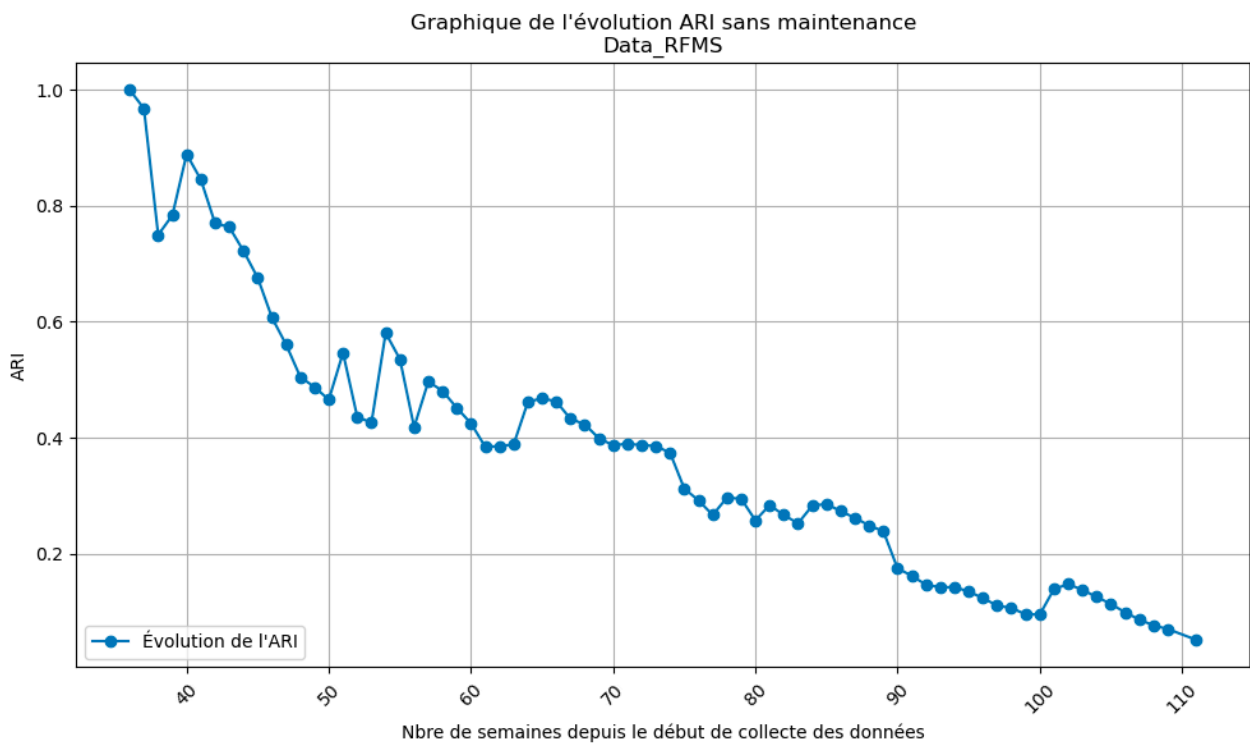


5- Définition d'un contrat de maintenance

FE_1

Maintenance par période

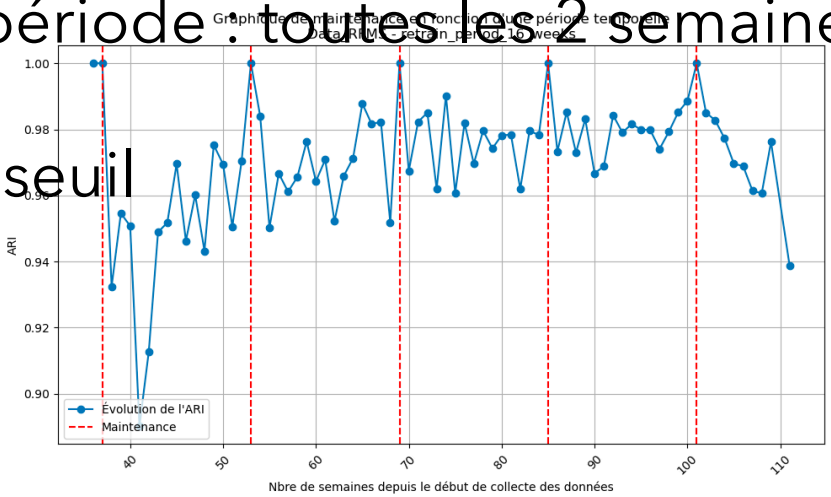
Maintenance par seuil
ari_seuil = 0.8



FE_2

Maintenance par période : toutes les 2 semaines

Maintenance par seuil



6- Conclusion

Réalisation de deux segmentations différentes

Basée sur les principes RFMS

Segmentation clients facilement actionnable
et simple en compréhension

Préconisation de maintenance :

- toutes les 16 semaines
- surveillance constante par seuil

Basée sur des variables plus spécifiques

Segmentation clients plus détaillées
permettant une personnalisation plus poussée
des campagnes de marketing

Préconisation de maintenance :

- toutes les 2 semaines
- surveillance constante par seuil

Réserve : Ces résultats et interprétations devraient être soumis à la validation du service marketing

Question ?