



Projet 4

Segmentez des clients d'un site de e-commerce

SOMMAIRE

1

Le contexte

Problématique et jeux de données

2

EDA et Preprocessing

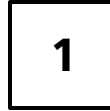
3

Modélisation

4

Analyse de la stabilité



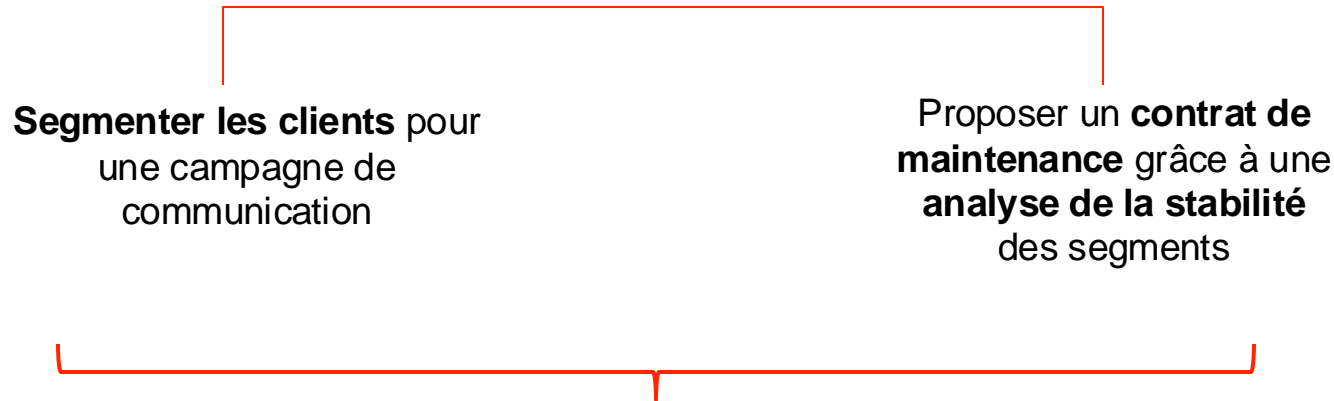


Le contexte

Problématique et jeux de données

Présentation de la problématique

Objectifs



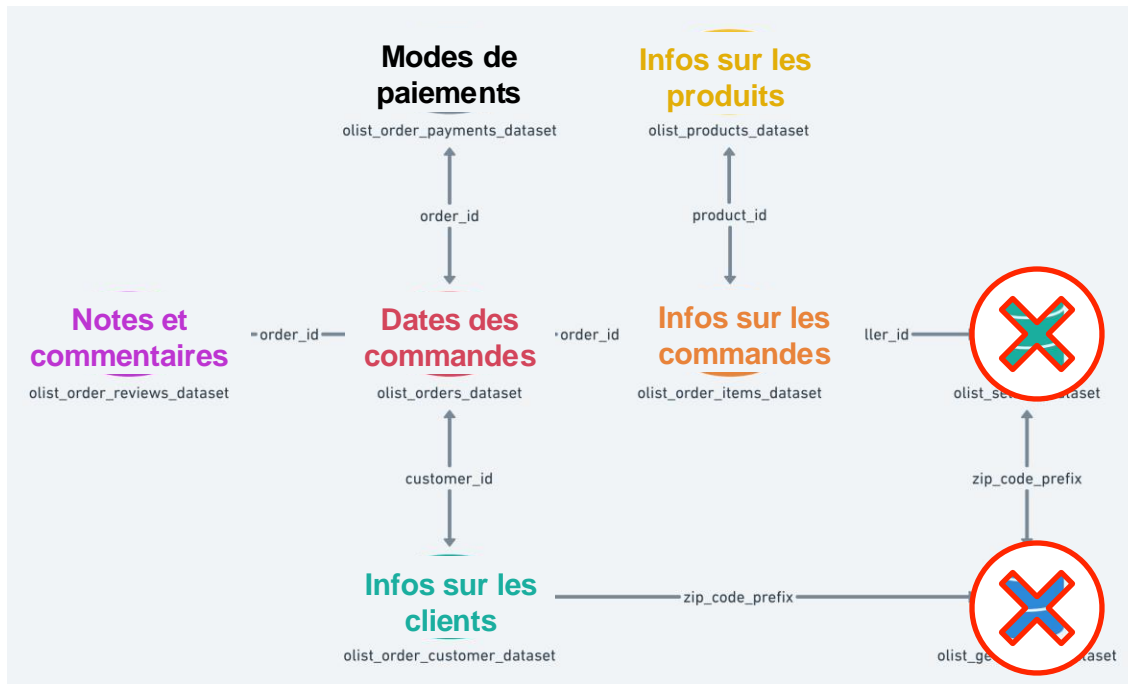
Segmenter les clients pour
une campagne de
communication

Proposer un **contrat de
maintenance** grâce à une
analyse de la stabilité
des segments

Base de données Olist comportant des informations sur les commandes
depuis Octobre 2016

Présentation des données

Base de données Olist comportant des informations sur les commandes depuis Octobre 2016



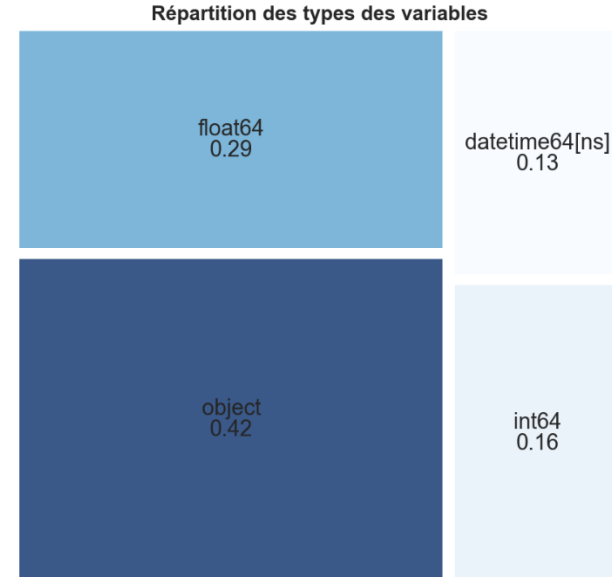
Présentation des données brutes

115k lignes

39 colonnes

| | |
|------------|--|
| | |
| Commande 1 | |
| Commande 2 | |

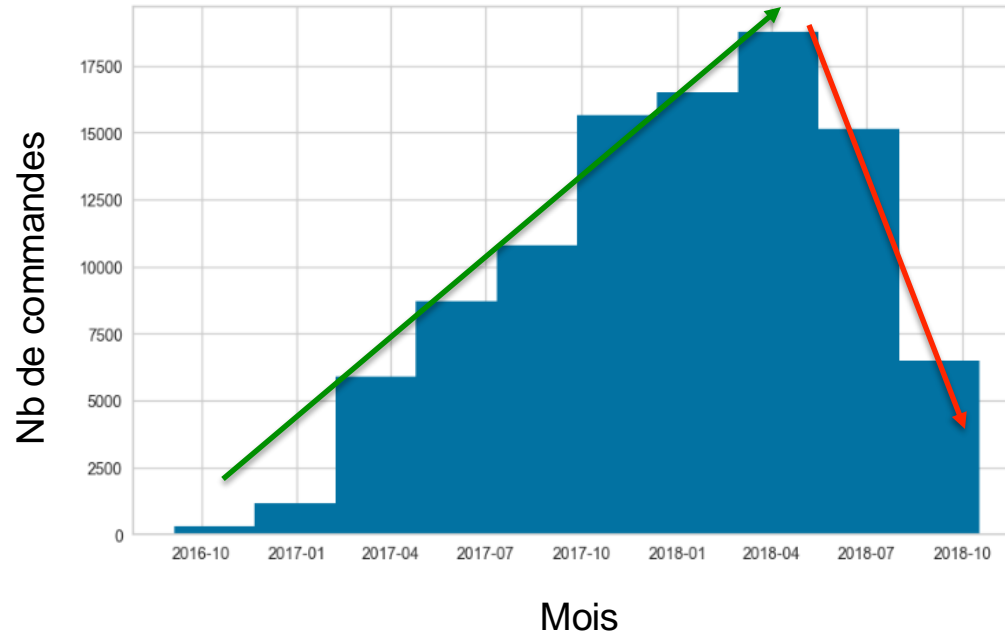
2 colonnes avec 60% de NaNs ou plus



2

EDA et Preprocessing

Nb de commandes dans le temps



EDA



- Toutes les commandes ont reçu une note
- 75% de 4-5
- 40% a reçu un commentaire
- Nans liés à cette colonnes

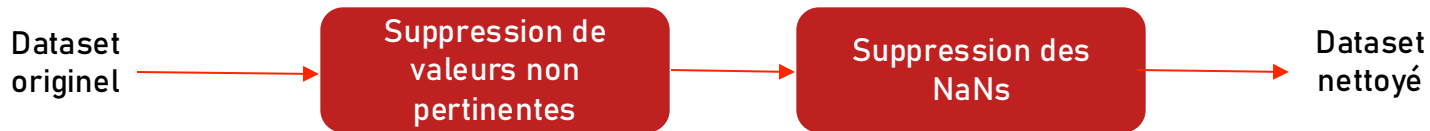


- 3% de clients qui font plusieurs commandes

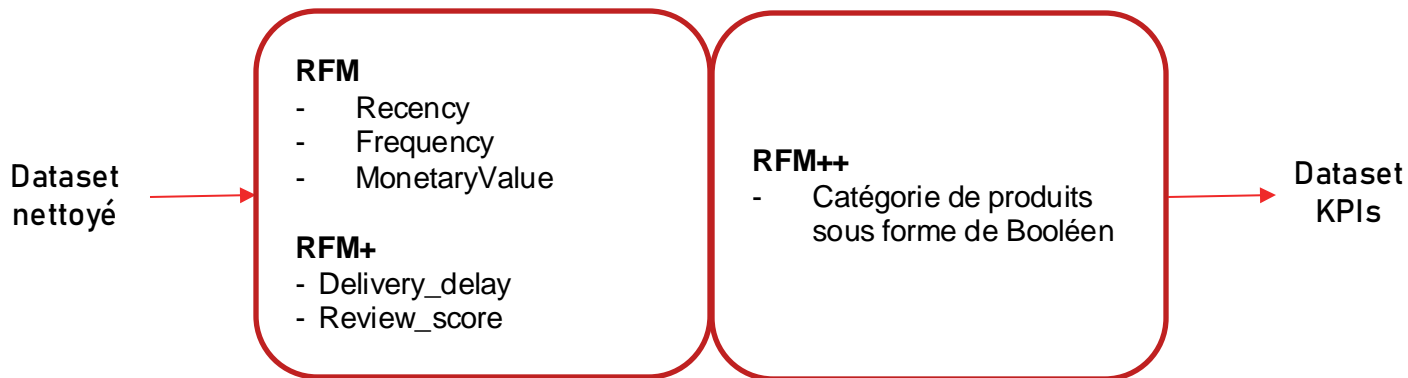


- 75% des achats faits en CB

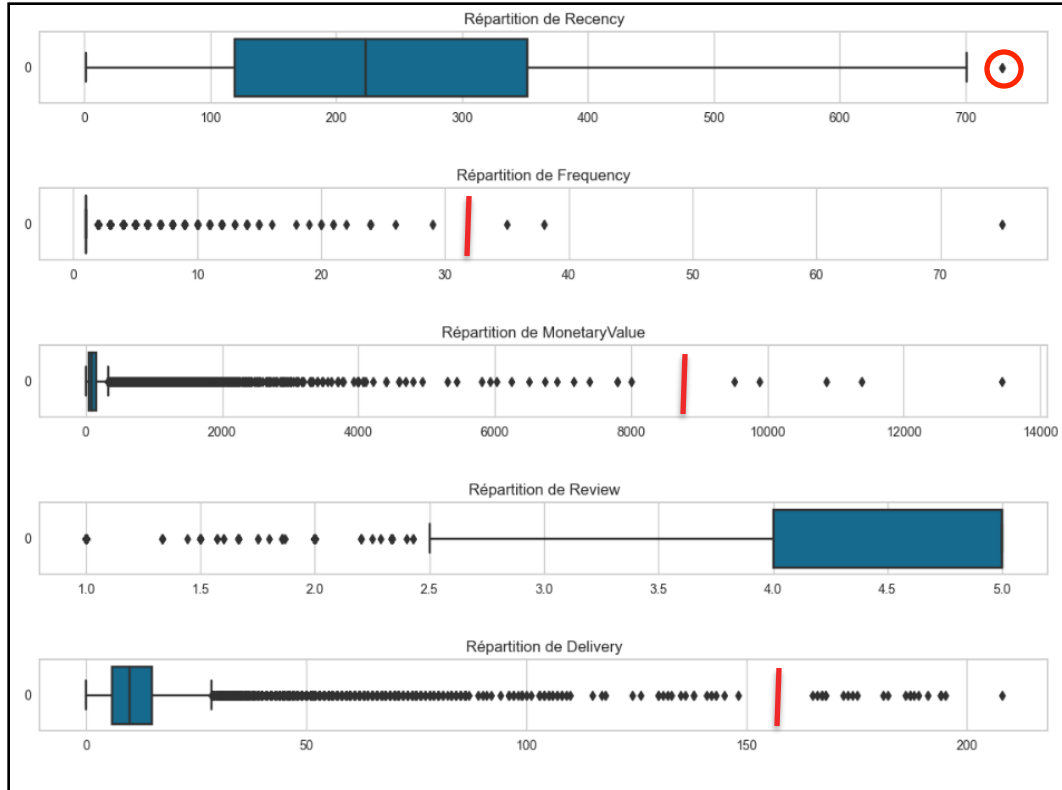
Nettoyage et Feature Engineering



GROUPBY



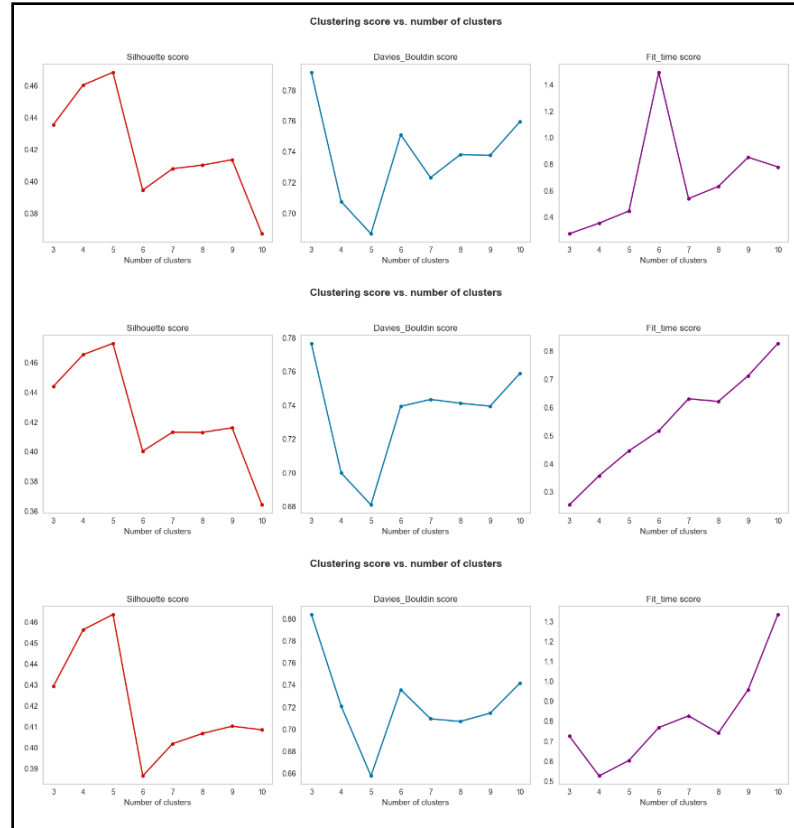
Outliers



3

Modélisation

Quel dataset ?

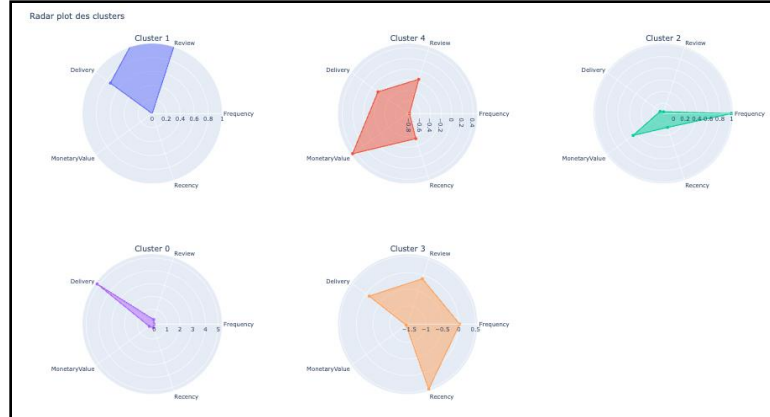


Quel dataset ?

RFM

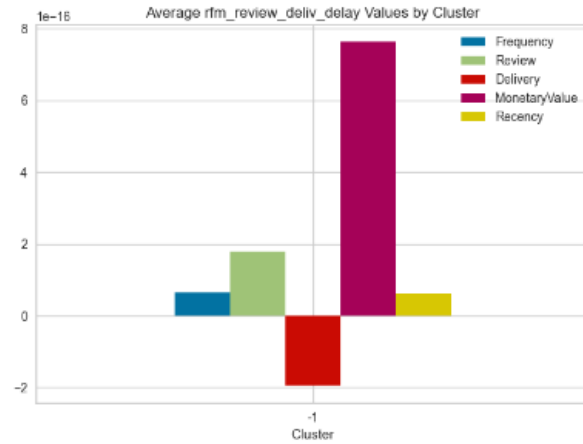


RFM+

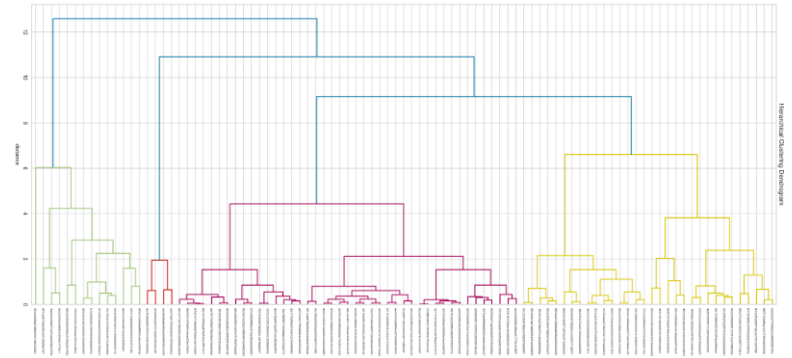


Quel modèle ?

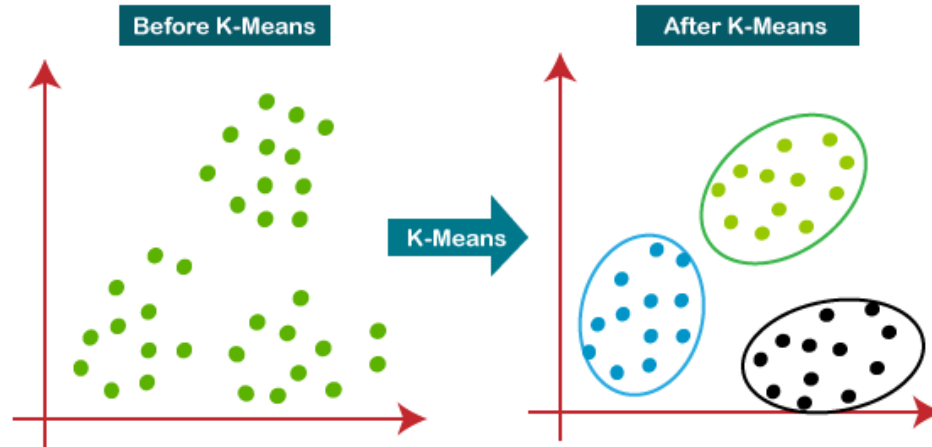
DBScan



Agglomerative clustering



Quel modèle ?



$K = 5$

Segmentation



Le régulier

A déjà acheté plusieurs produits
—> Newsletter



Le gros acheteur

Commente chacun des produits reçus.
—> Carte de fidélité



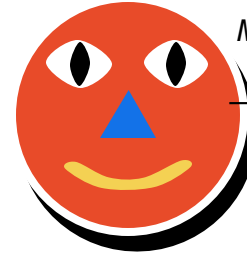
Le mécontent

Mauvaise note et/car fort délai de livraison
—> Contact service client



Le nouveau client

A acheté il y a moins de 6 mois
—> Promotion



L'ancien client

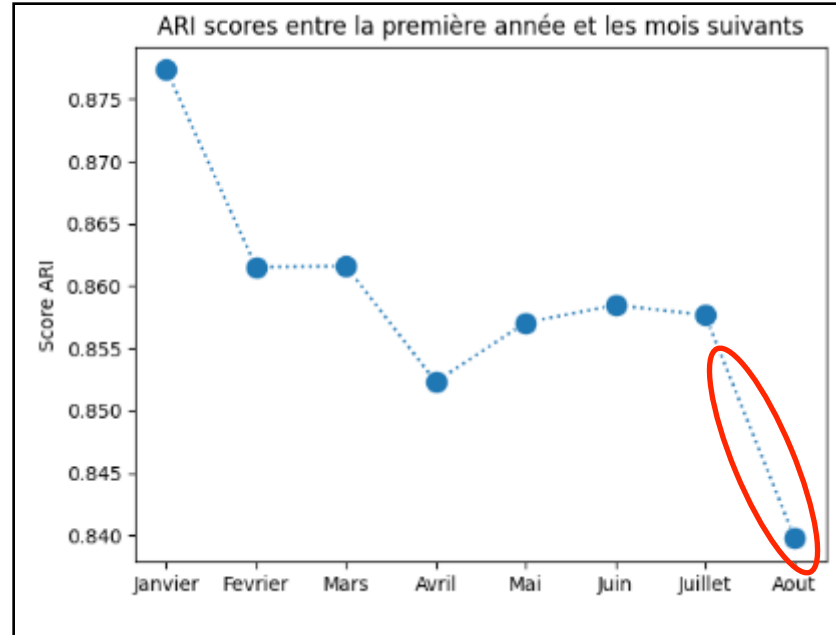
A acheté il y a plus d'un an
—> Promotion

4

Analyse de la stabilité

Analyse de la stabilité

Entraînement sur
2016-2017



→ Actualisation
tous les 8 mois

CONCLUSION

- 1) **Segmentation** en **5** clusters —> 5 profils de client —> 5 actions proposées
- 2) **Contrat de maintenance** : actualisation tous les 8 mois



