

A background network diagram featuring several clusters of nodes connected by thin lines. The nodes are colored in groups: light green, yellow, orange, and teal. The clusters are arranged in a circular pattern around the central text.

olist

SEGMENTEZ LES CLIENTS D'UN SITE E-COMMERCE

INTRODUCTION



estoque online

múltiplos canais de venda

coleta de pedidos

fluxo financeiro

omnicanalidade

PDV

loja oficial em marketplaces

frete barato

faturar pedidos

campanhas promocionais

atração de clientes

Introduction.

La problématique



**COMPRENDRE LES DIFFÉRENTS TYPES
D'UTILISATEURS GRÂCE À LEUR COMPORTEMENT
ET À LEURS DONNÉES PERSONNELLES.**

Introduction.

Les **livrables**

essais de modélisation

description actionnable de clusters

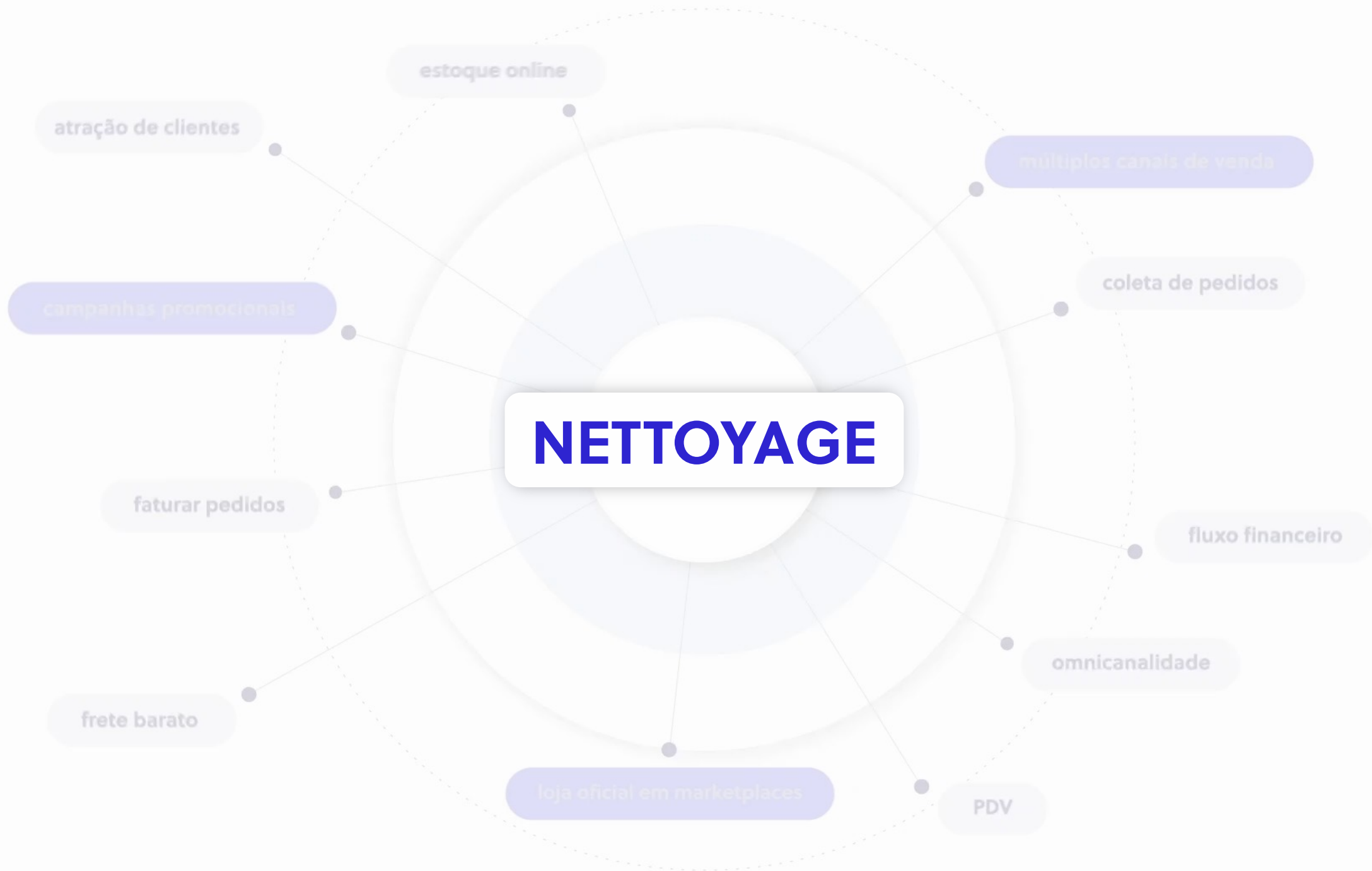
1

proposition de contrat de maintenance

analyse de la stabilité des segments au cours du temps

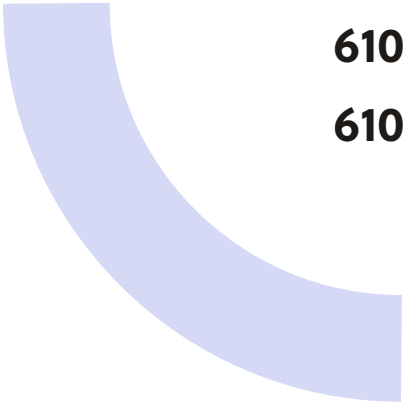
2










Nettoyage.

Les NaN



160		order_approved_at
1783		order_delivered_carrier_date
610		product_name_lenght
610		product_description_lenght
610		product_photos_qty

Nettoyage.

Les outliers

```
data.drop_duplicates(subset=['customer_id'],
                    inplace=True)

# Conservation des données utiles uniquement
data_orders = data_orders[data_orders.order_status == "delivered"]

# Modification au format datetime
data['order_purchase_timestamp'] = data[
    'order_purchase_timestamp'].astype('datetime64')

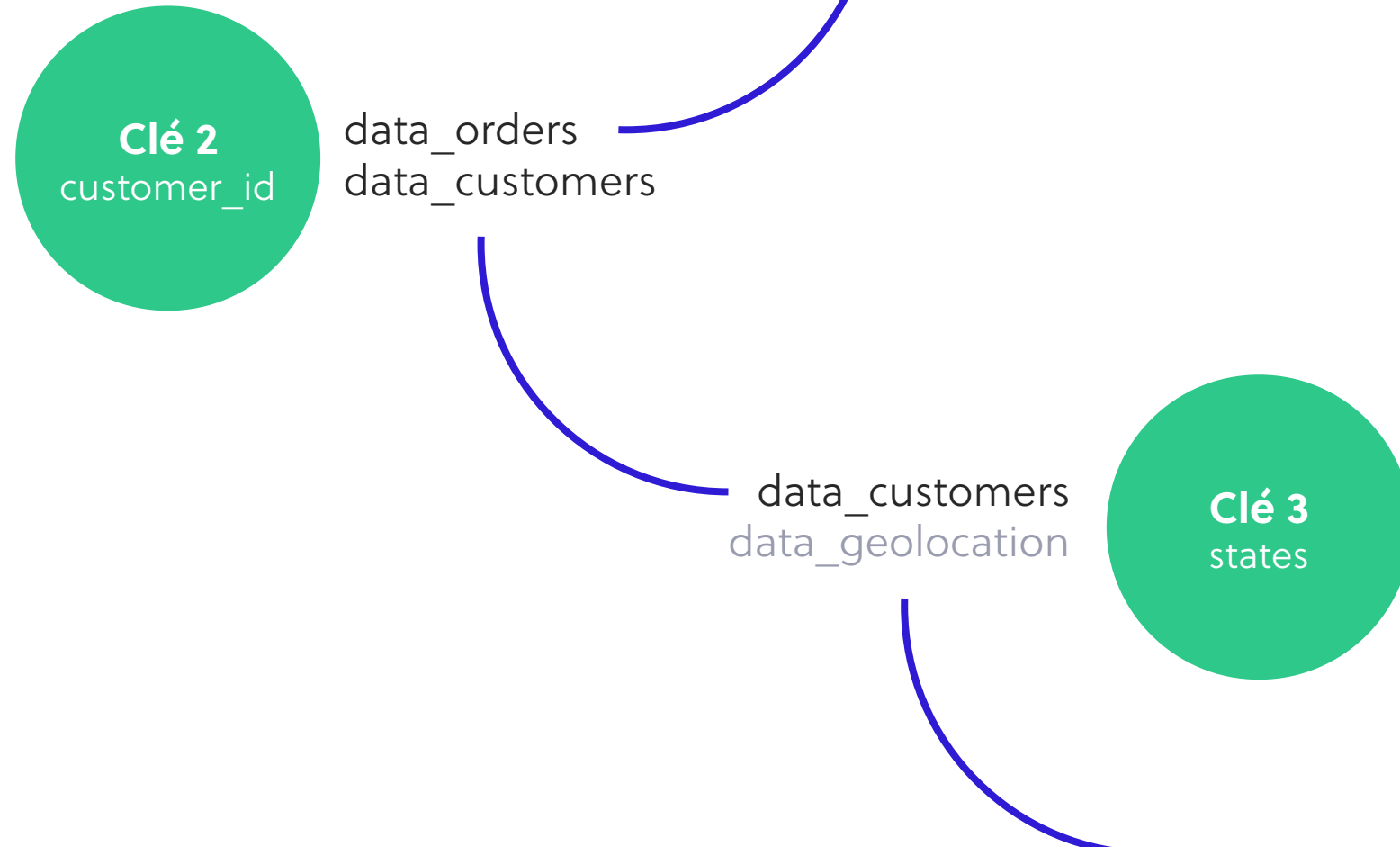
# Modification au format datetime
data['order_estimated_delivery_date'] = data[
    'order_estimated_delivery_date'].astype('datetime64')

# Modification au format datetime
data['order_delivered_customer_date'] = pd.to_datetime(
    data['order_delivered_customer_date'],
    errors='coerce')

# Traitement des valeurs abérrantes
final_data = final_data[final_data['mean_delay'] < 1000]
final_data = final_data[final_data['mean_delay'] > 0]
```

Nettoyage.

Le **merge**



Nettoyage.

Le merge

Merged

114 859 x 20

order_id	object
order_item_id	float64
product_id	object
seller_id	object
price	float64
freight_value	float64
payment_type	object
payment_installments	float64
payment_value	float64
review_score	float64
review_comment_message	object
customer_id	object
order_purchase_timestamp	datetime64[ns]
order_delivered_customer_date	datetime64[ns]
order_estimated_delivery_date	datetime64[ns]
customer_unique_id	object
customer_state	object
geolocation_lat	float64
geolocation_lng	float64
seller_state	object

Clé 4
seller_id

data_order_items
data_sellers

FEATURE ENGINEERING



The diagram consists of a central white circle with the text 'FEATURE ENGINEERING' in bold blue letters. Surrounding this is a light blue ring. Further out is a dashed circle with 11 points. Each point is connected to the central ring by a thin line and has a corresponding feature label in a rounded rectangle. The labels are: 'estoque online', 'múltiplos canais de venda', 'coleta de pedidos', 'fluxo financeiro', 'omnicanalidade', 'PDV', 'loja oficial em marketplaces', 'frete barato', 'faturar pedidos', 'campanhas promocionais', and 'atração de clientes'.

estoque online

múltiplos canais de venda

coleta de pedidos

fluxo financeiro

omnicanalidade

PDV

loja oficial em marketplaces

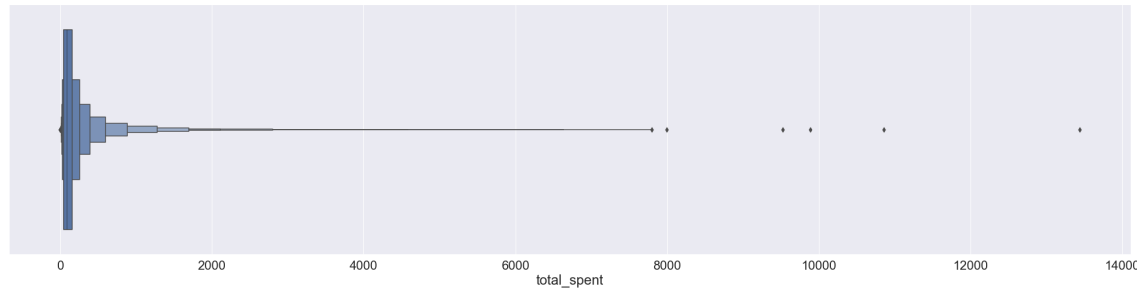
frete barato

faturar pedidos

campanhas promocionais

atração de clientes

Feature engineering.



total_spent

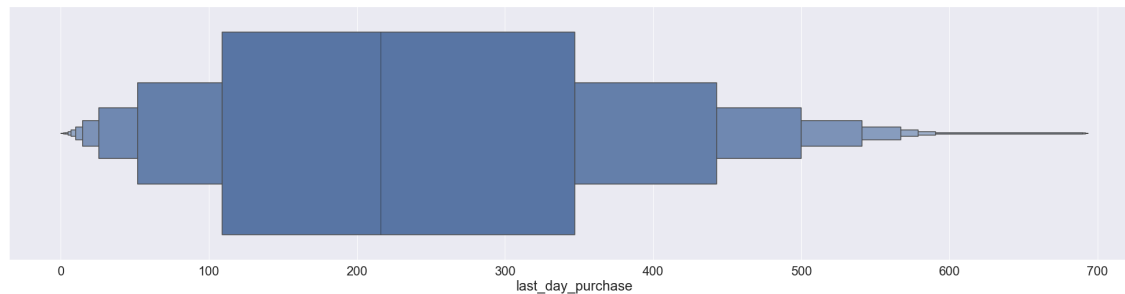
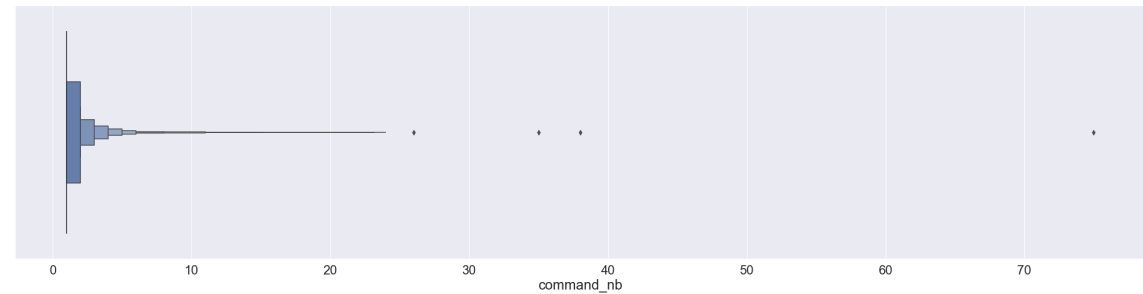
customer_unique_id

.sum()

command_nb

customer_unique_id

.count()

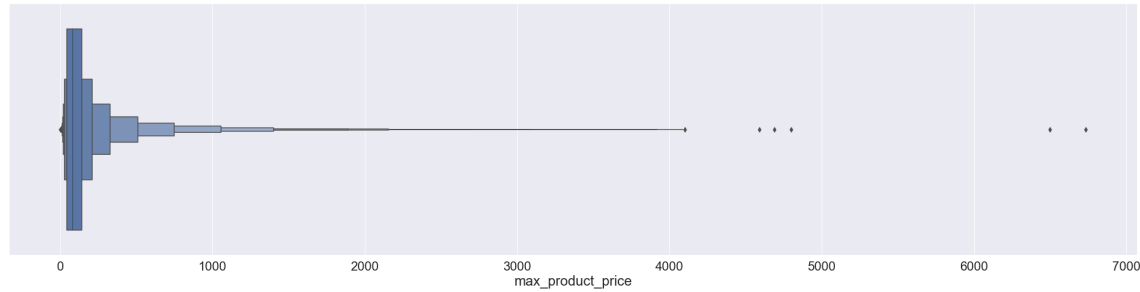


last_day_purchase

order_purchase_timestamp

= .max() - apply(lambda x : x.days())

Feature engineering.



max_product_price

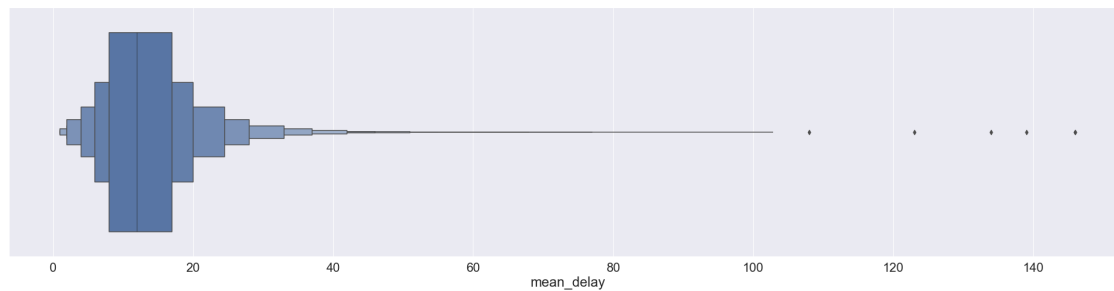
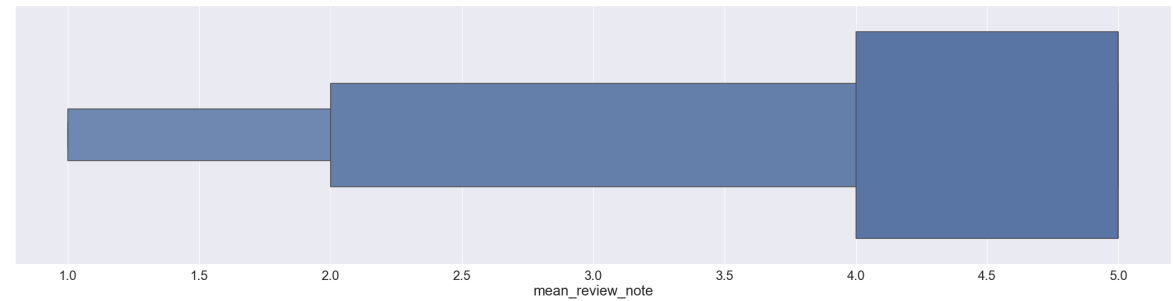
customer_unique_id

.max()

mean_review_note

customer_unique_id

.mean()

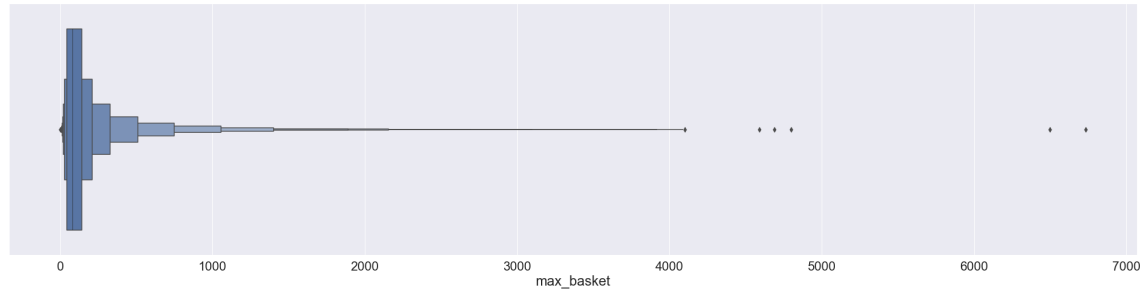


mean_delay

customer_unique_id

= estimated – effective delivery

Feature engineering.



mean_basket

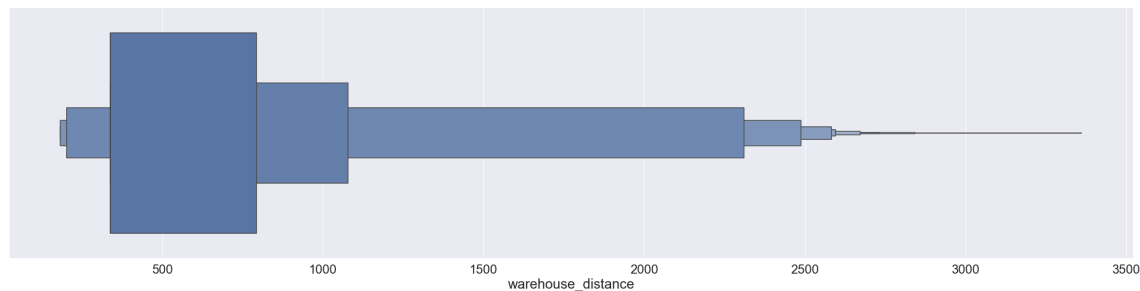
customer_unique_id, order_id

.mean()

last_weekday_purchase

customer_unique_id

= .max(). apply(lambda x : x.weekday())



warehouse_distance

customer_unique_id

= .haversine(geolocation) - olist(lat,lon)

EXPLORATION



The diagram features a central white circle with the word 'EXPLORATION' in bold blue letters. Surrounding this is a light blue ring. Twelve points are arranged in a circle around the ring, each connected to the ring by a thin line. These points are: 'estoque online', 'múltiplos canais de venda', 'coleta de pedidos', 'fluxo financeiro', 'omnicanalidade', 'PDV', 'loja oficial em marketplaces', 'frete barato', 'faturar pedidos', 'campanhas promocionais', 'atração de clientes', and 'estoque online' (repeated at the top). The points are contained within light blue rounded rectangular boxes.

estoque online

múltiplos canais de venda

coleta de pedidos

fluxo financeiro

omnicanalidade

PDV

loja oficial em marketplaces

frete barato

faturar pedidos

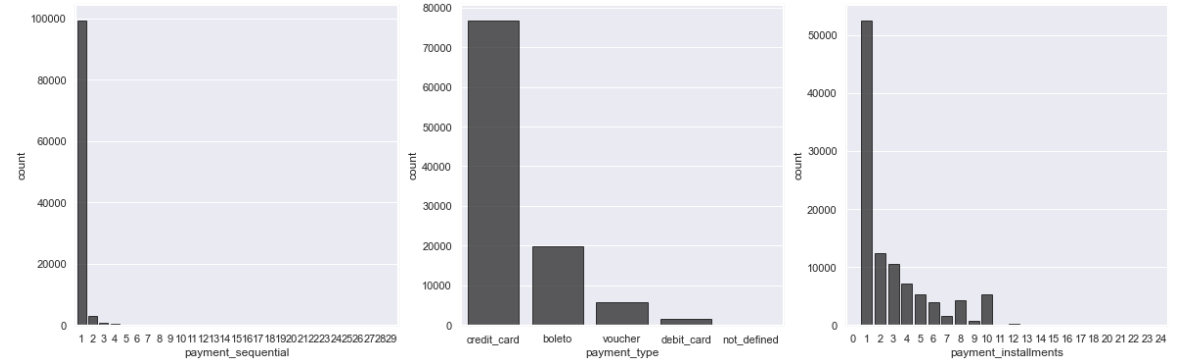
campanhas promocionais

atração de clientes

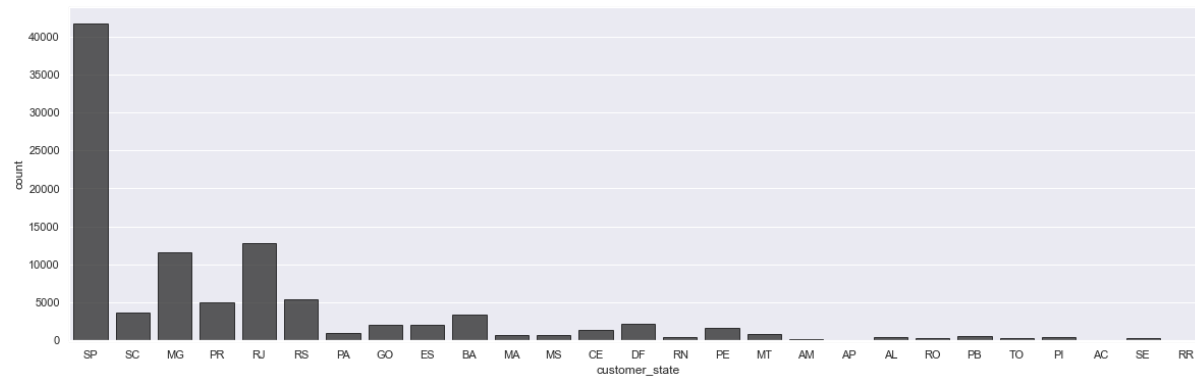
Exploration.

univariée

Moyens de paiement



Répartition par État



SEGMENTATION RFM



The diagram features a central white circle with the text 'SEGMENTATION RFM' in bold blue letters. Surrounding this is a light blue ring, followed by a larger white ring, and an outermost dotted circle. Twelve points are distributed around the dotted circle, each connected to the center by a thin line. These points are: 'estoque online' (top), 'múltiplos canais de venda' (top-right), 'coleta de pedidos' (right), 'fluxo financeiro' (bottom-right), 'omnicanalidade' (bottom), 'PDV' (bottom), 'loja oficial em marketplaces' (bottom-left), 'frete barato' (left), 'faturar pedidos' (top-left), 'campanhas promocionais' (top-left), 'atração de clientes' (top), and 'estoque online' (top).

estoque online

múltiplos canais de venda

coleta de pedidos

fluxo financeiro

omnicanalidade

PDV

loja oficial em marketplaces

frete barato

faturar pedidos

campanhas promocionais

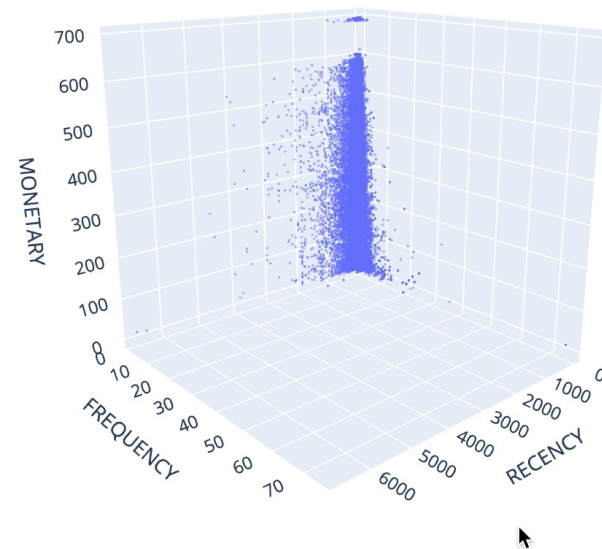
atração de clientes

Segmentation RFM.

Données **brutes**

RFM

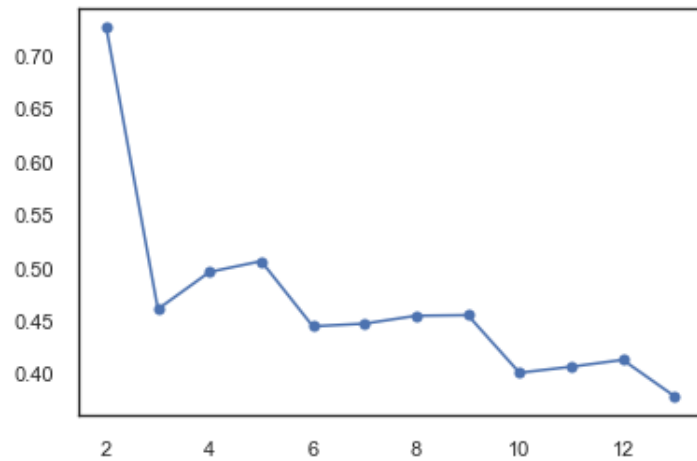
last_day_purchase,	récence
command_nb,	fréquence,
mean_basket	montant



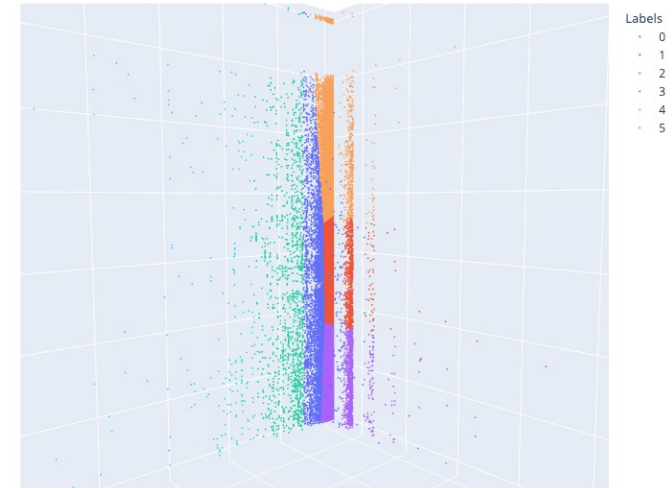
Segmentation RFM.

test **KMeans**

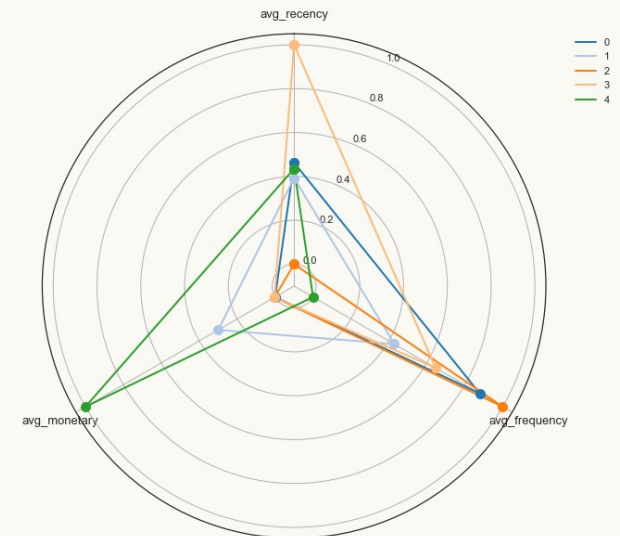
▼ KMeans
KMeans(n_clusters=5)



Silhouette score



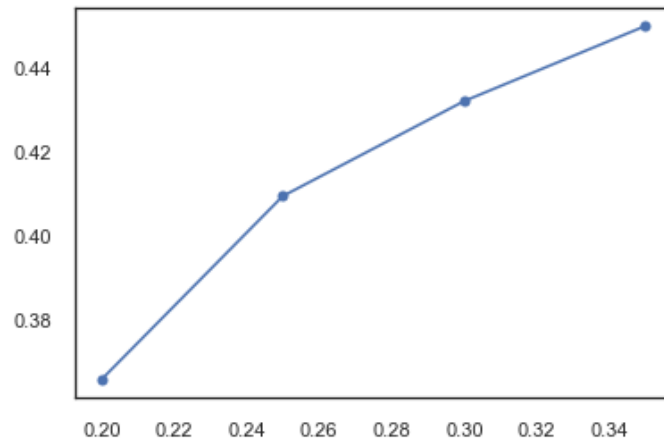
Labels plot



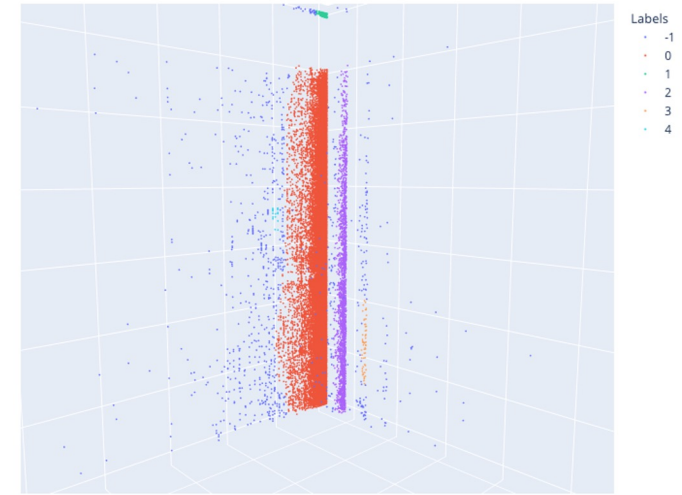
Segmentation RFM.

test **DBSCAN**

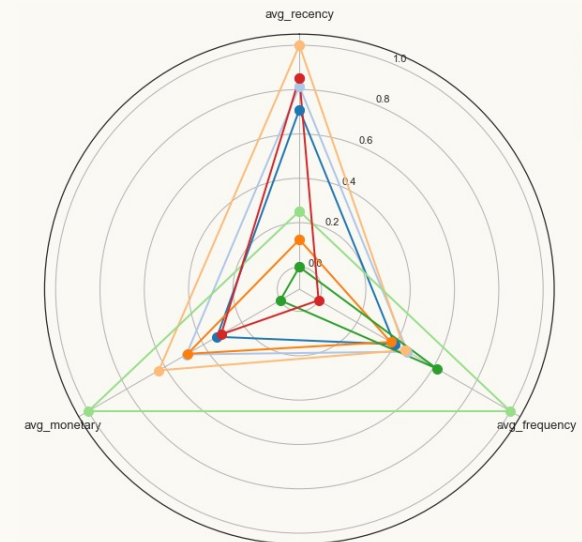
DBSCAN
`DBSCAN(eps=0.35, min_samples=35, n_jobs=-1)`



Silhouette score



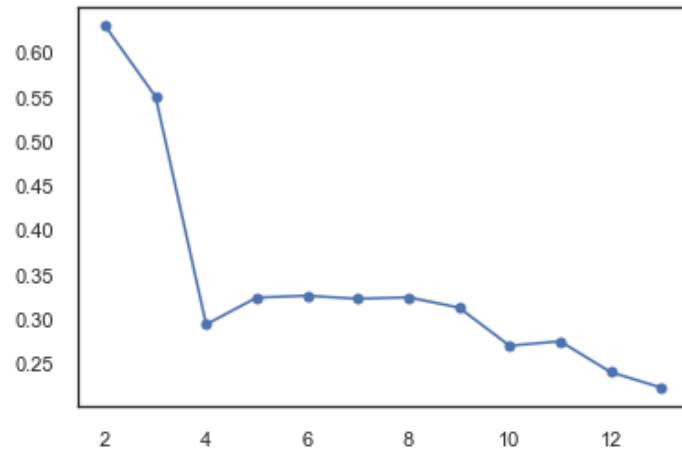
Labels plot



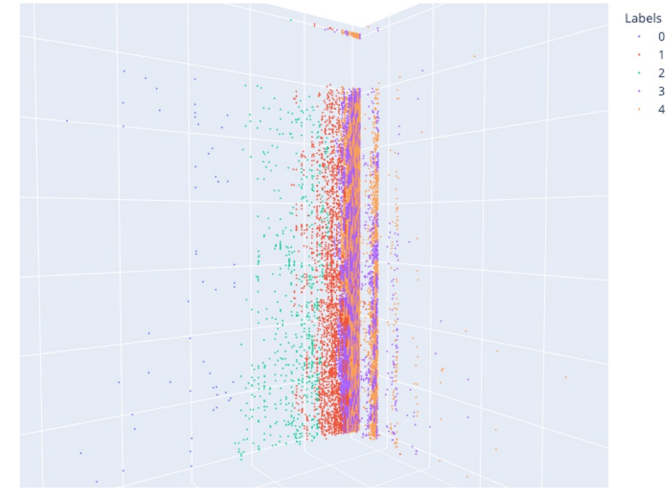
Segmentation RFM.

test **Clustering Agglomératif (CAH)**

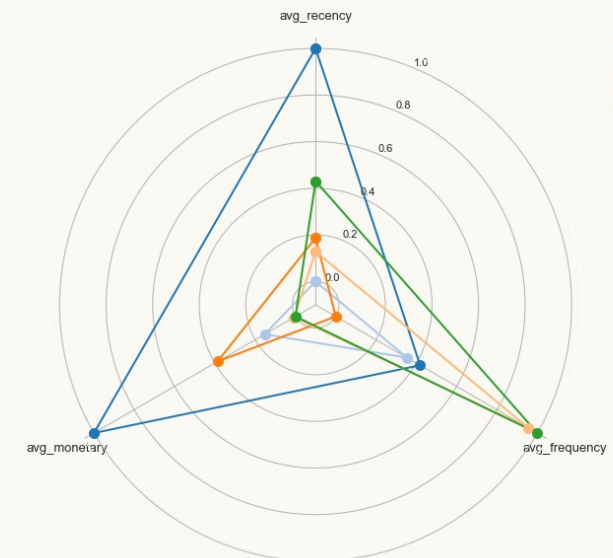
▼ AgglomerativeClustering
AgglomerativeClustering(n_clusters=5)



Silhouette score



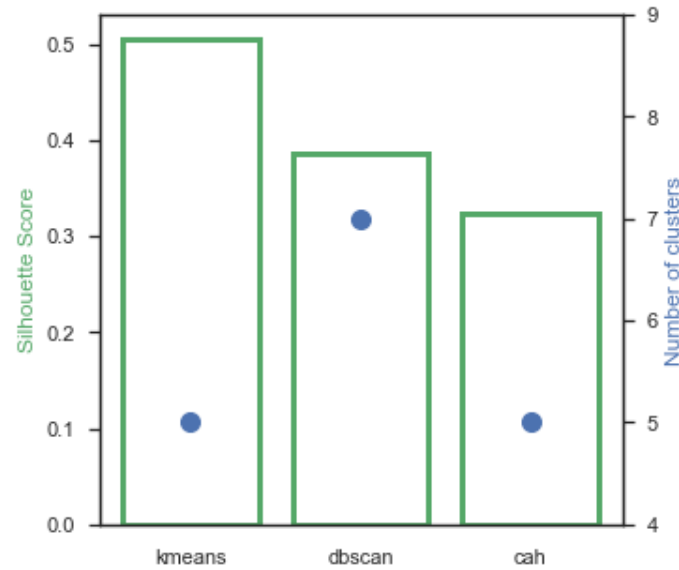
Labels plot



Segmentation RFM.

Choix du **modèle**

Comparatif des modèles



+ Importance de l'interprétabilité et des ressources utilisées

CLUSTERING FINAL



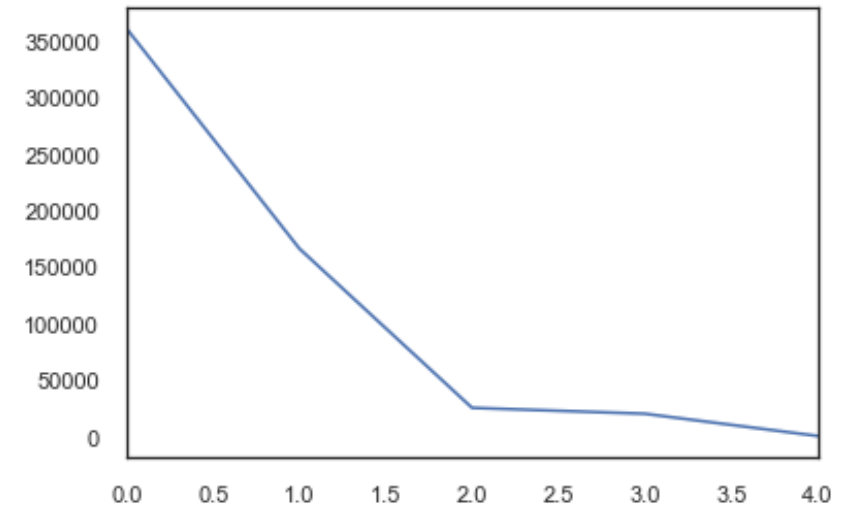
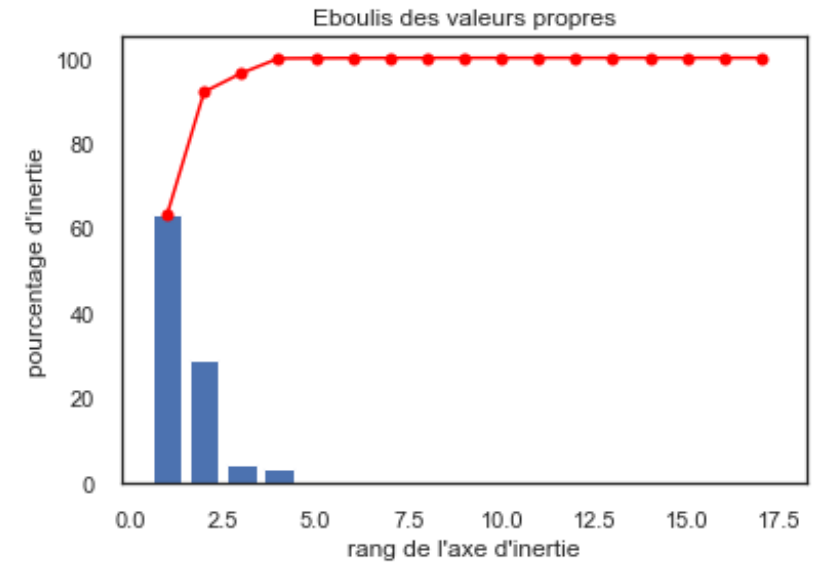
Clustering final.

Choix des **variables**

100%

De la variance

mean_delivery_time,
mean_delay,
last_day_purchase,
mean_review_note,
total_of_products,
recidivist,
delay_between_purchases,
command_nb,
freight_max_value,
mean_basket,
max_basket,
max_product_price,
total_spent,
mean_product_nb_per_basket,
last_weekday_purchase,
last_hour_purchase,
warehouse_distance



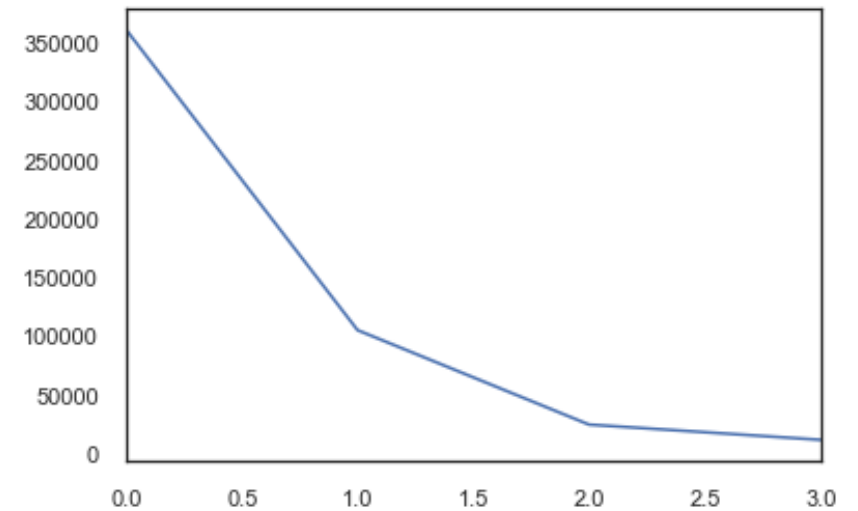
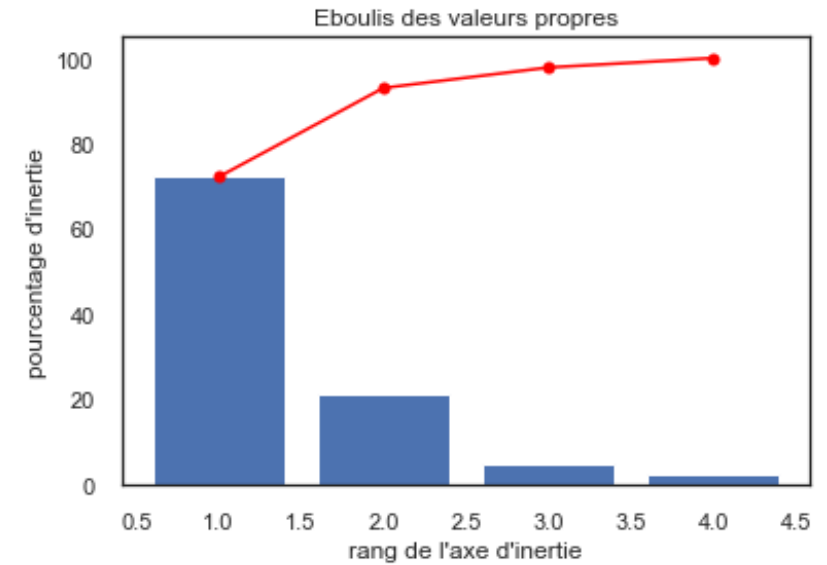
Clustering final.

Choix des **variables**

87%

De la variance

last_day_purchase,
max_product_price,
total_spent,
warehouse_distance



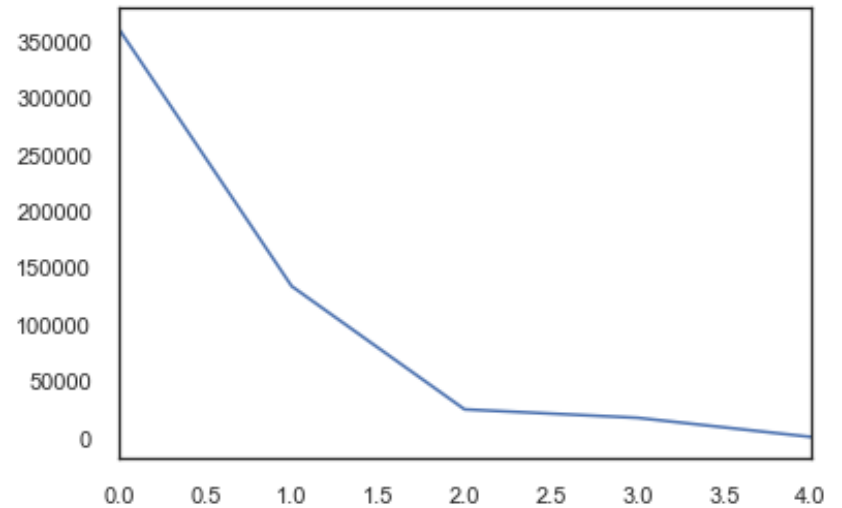
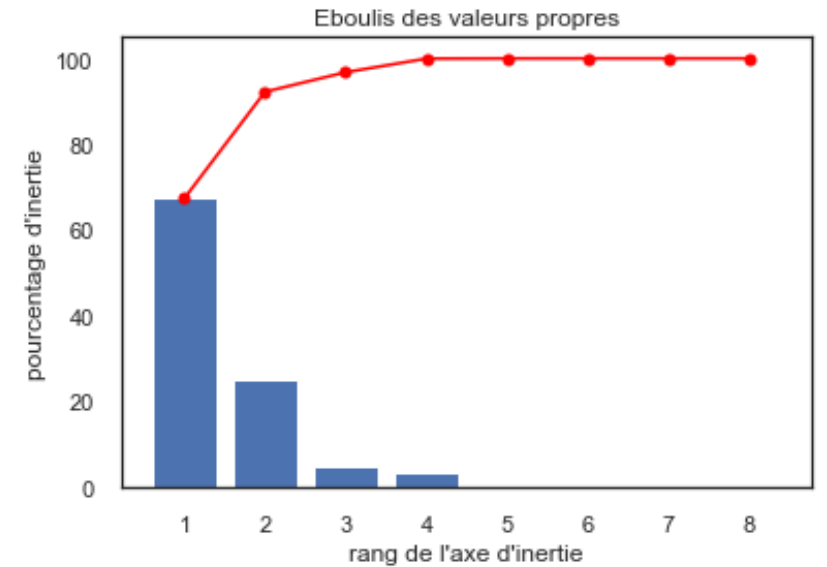
Clustering final.

Choix des **variables**

94%

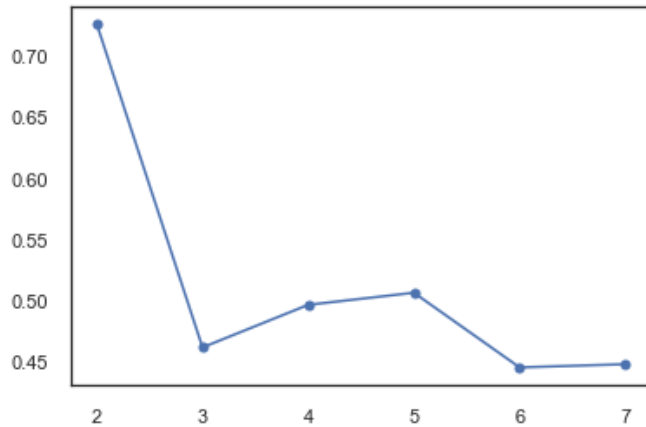
De la variance

last_day_purchase,
max_product_price,
total_spent,
warehouse_distance,
command_nb,
mean_basket,
mean_review_note,
mean_delay

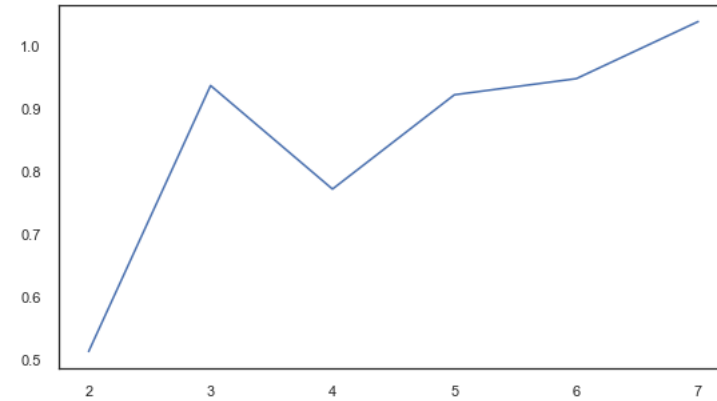


Clustering final.

Entrainement **du modèle**



Silhouette score



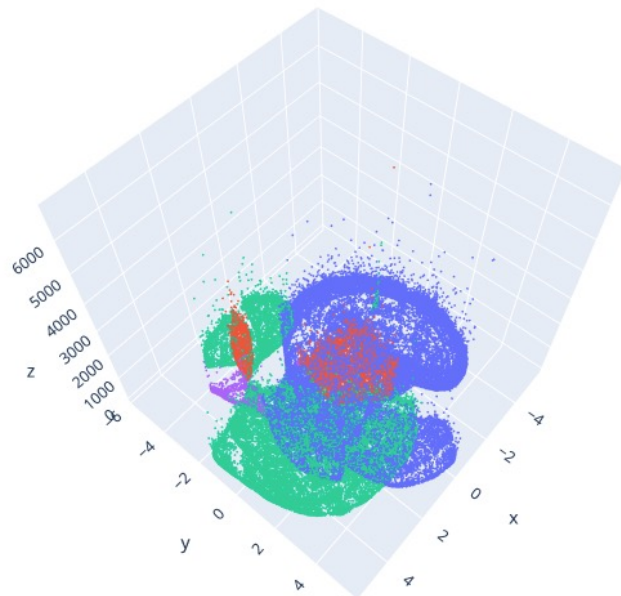
Davies_Bouldin score

Clustering final.

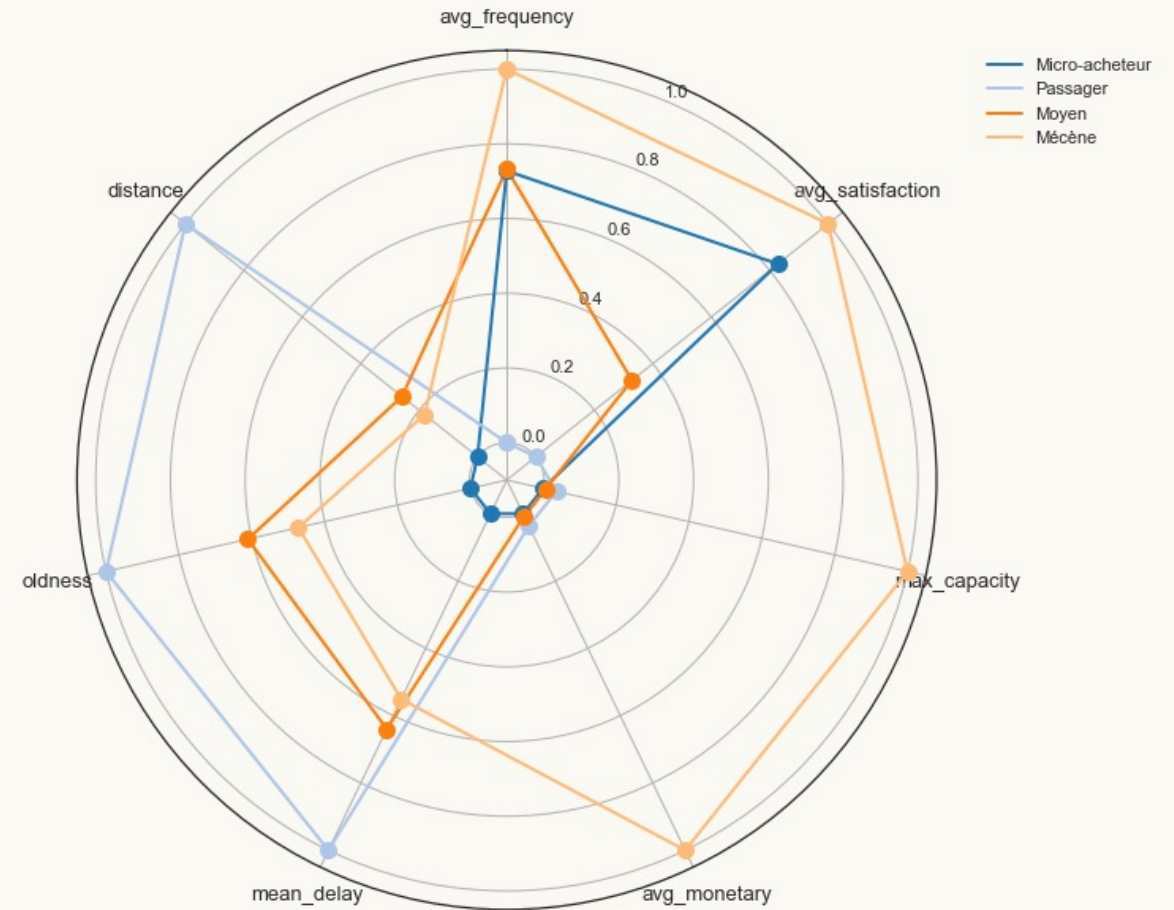
Représentation **du modèle**

Labels

- Micro-acheteur
- Passager
- Moyen
- Mécène



Clusters



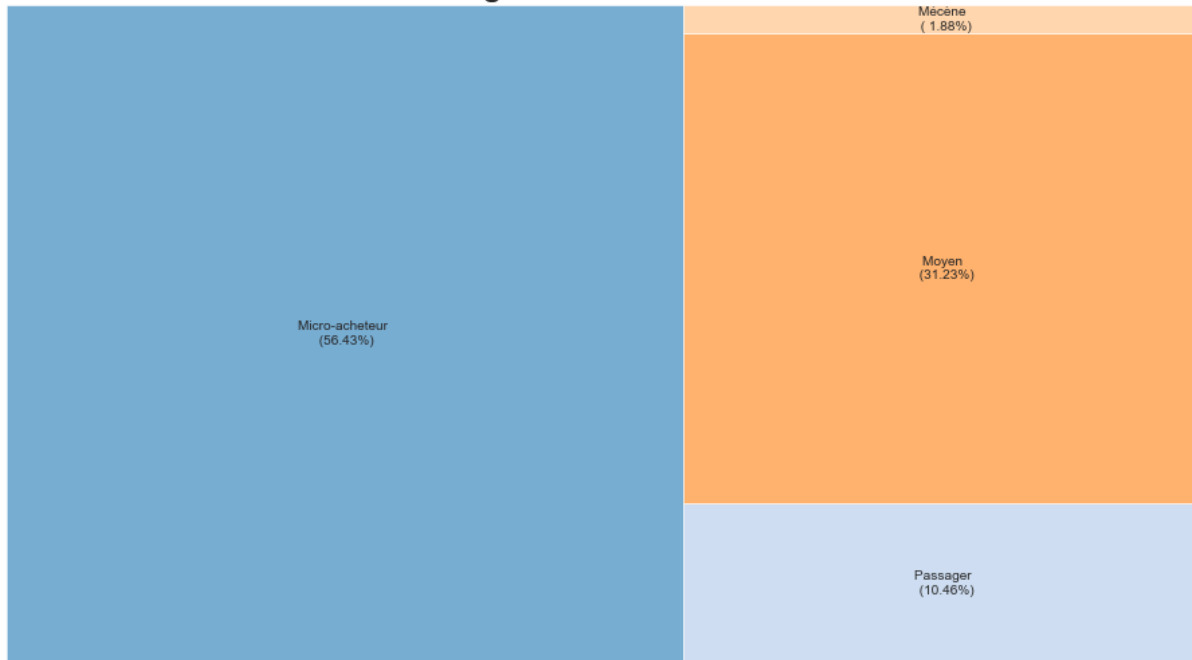
Clustering final.

Qualité du modèle

Labels

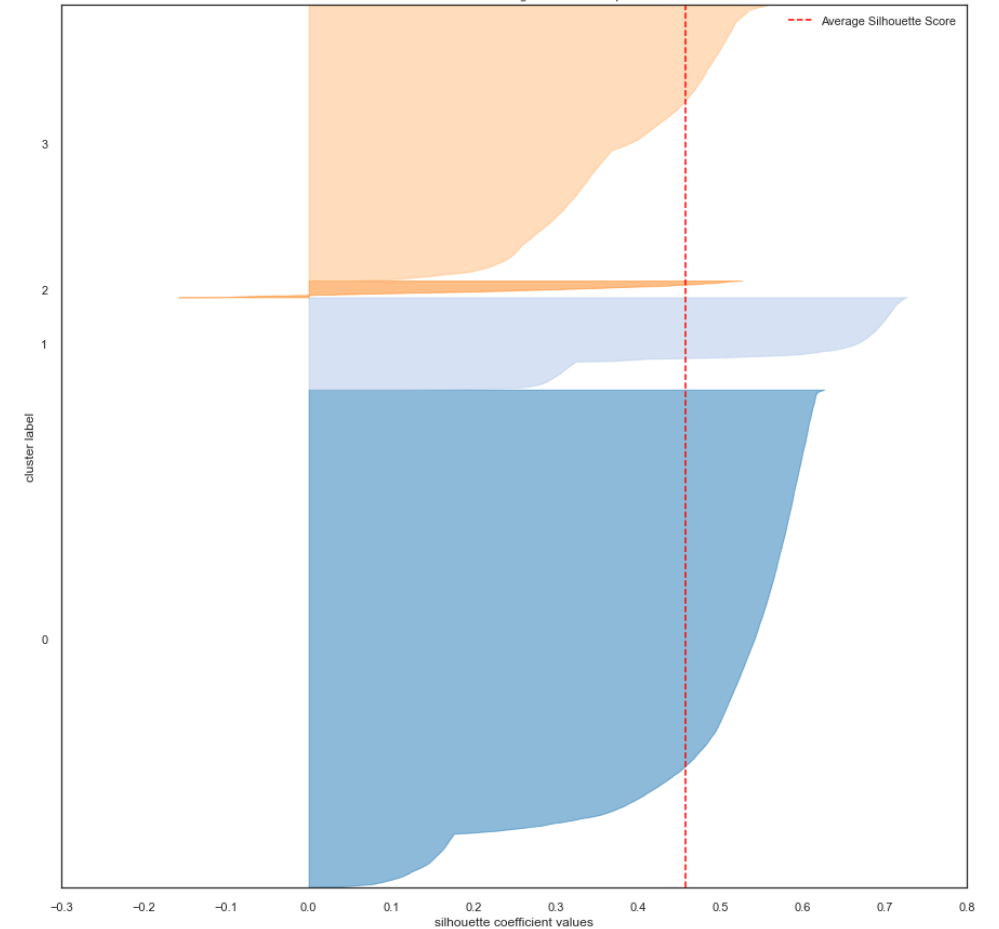
- Micro-acheteur
- Passager
- Moyen
- Mécène

Segmentation KMeans



Micro-acheteur : 47381
Passager : 8783
Moyen : 26218
Mécène : 1580

Silhouette Plot of KMeans Clustering for 83962 Samples in 4 Centers



MAINTENANCE



The diagram features a central white circle with the word 'MAINTENANCE' in bold blue letters. Surrounding this is a light blue ring. Eleven nodes, each in a rounded rectangle, are arranged in a circle around the ring. Lines connect each node to the central ring. The nodes are: 'estoque online' (top), 'múltiplos canais de venda' (top-right), 'coleta de pedidos' (right), 'fluxo financeiro' (bottom-right), 'omnicanalidade' (bottom), 'PDV' (bottom), 'loja oficial em marketplaces' (bottom-left), 'frete barato' (left), 'faturar pedidos' (top-left), 'campanhas promocionais' (top-left), and 'atração de clientes' (top-left).

estoque online

múltiplos canais de venda

coleta de pedidos

fluxo financeiro

omnicanalidade

PDV

loja oficial em marketplaces

frete barato

faturar pedidos

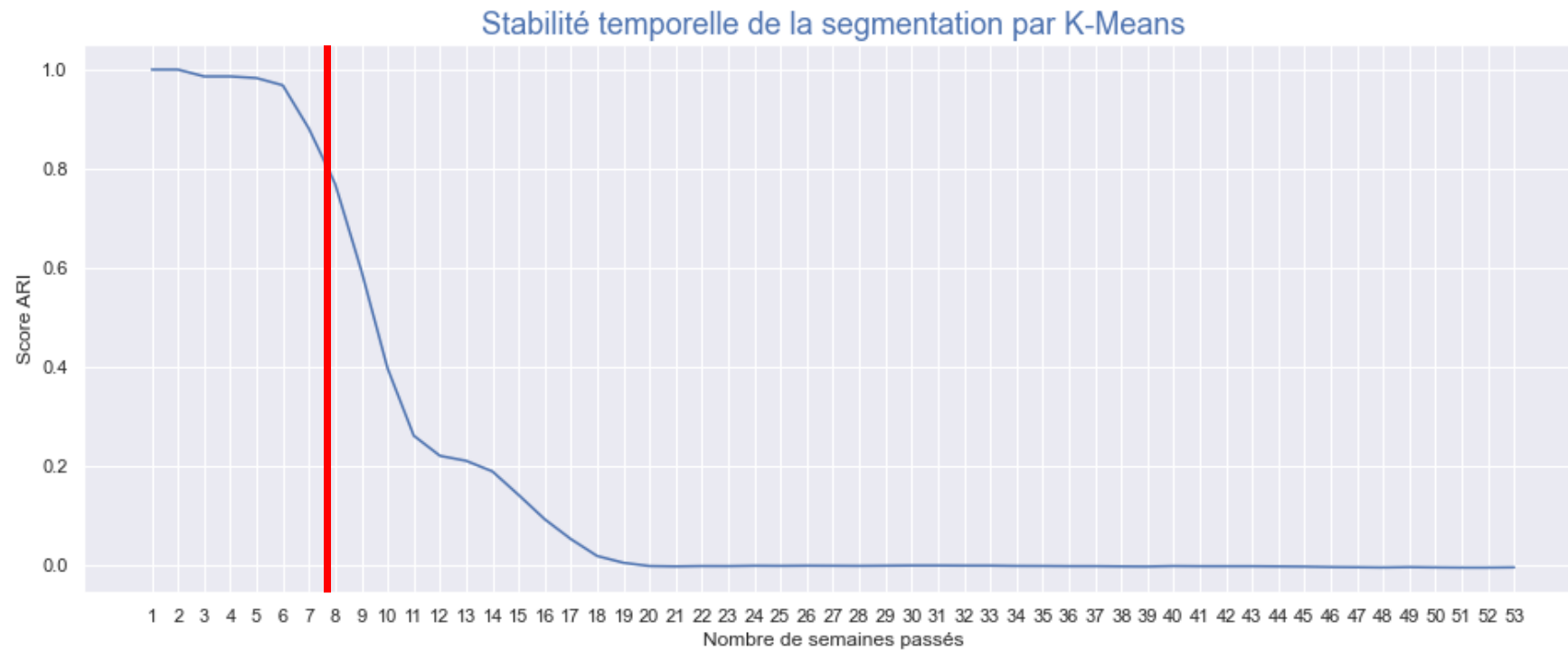
campanhas promocionais

atração de clientes

Maintenance.

Stabilité **sur 1 an**

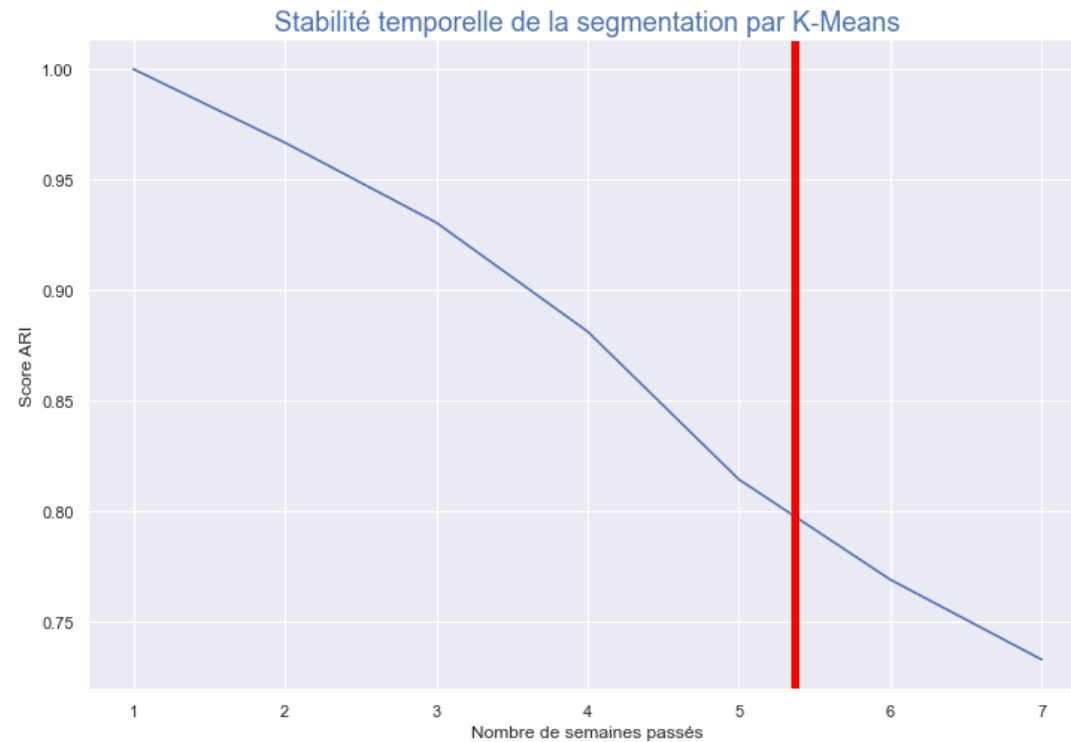
46 jours



Maintenance.

Stabilité sur les 8 **dernières semaines**

37 jours



Merci

Alexandre Delaguillaumie