

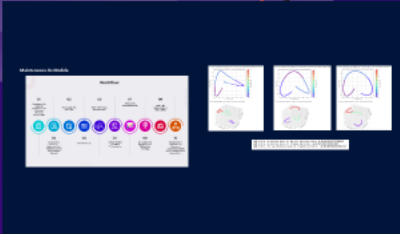
# Segmentez des clients d'un site e-commerce



## Introduction

Face à un marché en constante évolution, Olist souhaite transformer sa stratégie Data pour mieux comprendre et servir sa clientèle.

Pour cela nous devons développer une segmentation client dynamique, intuitive et évolutive.



## Introduction

Face à un marché en constante évolution, Olist souhaite transformer sa stratégie Data pour mieux comprendre et servir sa clientèle.

Pour cela nous devons développer une segmentation client dynamique, intuitive et évolutive.



# Table des matières

1

Feature Engineering

2

Exploration des Données

3

Nettoyage des Données

4

Modélisation

5

Maintenance du Modèle

# Comprendre Olist et Sa Base de Données

<table><tr><th>translation</th></tr><tr><td>123 index</td></tr><tr><td>AC product_category_name</td></tr><tr><td>AC product_category_name_english</td></tr></table>	translation	123 index	AC product_category_name	AC product_category_name_english	<table><tr><th>sellers</th></tr><tr><td>123 index</td></tr><tr><td>AC seller_id</td></tr><tr><td>123 seller_zip_code_prefix</td></tr><tr><td>AC seller_city</td></tr><tr><td>AC seller_state</td></tr></table>	sellers	123 index	AC seller_id	123 seller_zip_code_prefix	AC seller_city	AC seller_state	<table><tr><th>customers</th></tr><tr><td>123 index</td></tr><tr><td>AC customer_id</td></tr><tr><td>AC customer_unique_id</td></tr><tr><td>123 customer_zip_code_prefix</td></tr><tr><td>AC customer_city</td></tr><tr><td>AC customer_state</td></tr></table>	customers	123 index	AC customer_id	AC customer_unique_id	123 customer_zip_code_prefix	AC customer_city	AC customer_state	<table><tr><th>geoloc</th></tr><tr><td>123 index</td></tr><tr><td>123 geolocation_zip_code_prefix</td></tr><tr><td>123 geolocation_lat</td></tr><tr><td>123 geolocation_lng</td></tr><tr><td>AC geolocation_city</td></tr><tr><td>AC geolocation_state</td></tr></table>	geoloc	123 index	123 geolocation_zip_code_prefix	123 geolocation_lat	123 geolocation_lng	AC geolocation_city	AC geolocation_state	<table><tr><th>order_pymts</th></tr><tr><td>123 index</td></tr><tr><td>AC order_id</td></tr><tr><td>123 payment_sequential</td></tr><tr><td>AC payment_type</td></tr><tr><td>123 payment_installments</td></tr><tr><td>123 payment_value</td></tr></table>	order_pymts	123 index	AC order_id	123 payment_sequential	AC payment_type	123 payment_installments	123 payment_value	<table><tr><th>order_items</th></tr><tr><td>123 index</td></tr><tr><td>AC order_id</td></tr><tr><td>123 order_item_id</td></tr><tr><td>AC product_id</td></tr><tr><td>AC seller_id</td></tr><tr><td>AC shipping_limit_date</td></tr><tr><td>123 price</td></tr><tr><td>123 freight_value</td></tr></table>	order_items	123 index	AC order_id	123 order_item_id	AC product_id	AC seller_id	AC shipping_limit_date	123 price	123 freight_value	<table><tr><th>order_reviews</th></tr><tr><td>123 index</td></tr><tr><td>AC review_id</td></tr><tr><td>AC order_id</td></tr><tr><td>123 review_score</td></tr><tr><td>AC review_comment_title</td></tr><tr><td>AC review_comment_message</td></tr><tr><td>AC review_creation_date</td></tr><tr><td>AC review_answer_timestamp</td></tr></table>	order_reviews	123 index	AC review_id	AC order_id	123 review_score	AC review_comment_title	AC review_comment_message	AC review_creation_date	AC review_answer_timestamp	<table><tr><th>orders</th></tr><tr><td>123 index</td></tr><tr><td>AC order_id</td></tr><tr><td>AC customer_id</td></tr><tr><td>AC order_status</td></tr><tr><td>AC order_purchase_timestamp</td></tr><tr><td>AC order_approved_at</td></tr><tr><td>AC order_delivered_carrier_date</td></tr><tr><td>AC order_delivered_customer_date</td></tr><tr><td>AC order_estimated_delivery_date</td></tr></table>	orders	123 index	AC order_id	AC customer_id	AC order_status	AC order_purchase_timestamp	AC order_approved_at	AC order_delivered_carrier_date	AC order_delivered_customer_date	AC order_estimated_delivery_date
translation																																																																		
123 index																																																																		
AC product_category_name																																																																		
AC product_category_name_english																																																																		
sellers																																																																		
123 index																																																																		
AC seller_id																																																																		
123 seller_zip_code_prefix																																																																		
AC seller_city																																																																		
AC seller_state																																																																		
customers																																																																		
123 index																																																																		
AC customer_id																																																																		
AC customer_unique_id																																																																		
123 customer_zip_code_prefix																																																																		
AC customer_city																																																																		
AC customer_state																																																																		
geoloc																																																																		
123 index																																																																		
123 geolocation_zip_code_prefix																																																																		
123 geolocation_lat																																																																		
123 geolocation_lng																																																																		
AC geolocation_city																																																																		
AC geolocation_state																																																																		
order_pymts																																																																		
123 index																																																																		
AC order_id																																																																		
123 payment_sequential																																																																		
AC payment_type																																																																		
123 payment_installments																																																																		
123 payment_value																																																																		
order_items																																																																		
123 index																																																																		
AC order_id																																																																		
123 order_item_id																																																																		
AC product_id																																																																		
AC seller_id																																																																		
AC shipping_limit_date																																																																		
123 price																																																																		
123 freight_value																																																																		
order_reviews																																																																		
123 index																																																																		
AC review_id																																																																		
AC order_id																																																																		
123 review_score																																																																		
AC review_comment_title																																																																		
AC review_comment_message																																																																		
AC review_creation_date																																																																		
AC review_answer_timestamp																																																																		
orders																																																																		
123 index																																																																		
AC order_id																																																																		
AC customer_id																																																																		
AC order_status																																																																		
AC order_purchase_timestamp																																																																		
AC order_approved_at																																																																		
AC order_delivered_carrier_date																																																																		
AC order_delivered_customer_date																																																																		
AC order_estimated_delivery_date																																																																		
<table><tr><th>products</th></tr><tr><td>123 index</td></tr><tr><td>AC product_id</td></tr><tr><td>AC product_category_name</td></tr><tr><td>123 product_name_lenght</td></tr><tr><td>123 product_description_lenght</td></tr><tr><td>123 product_photos_qty</td></tr><tr><td>123 product_weight_g</td></tr><tr><td>123 product_length_cm</td></tr><tr><td>123 product_height_cm</td></tr><tr><td>123 product_width_cm</td></tr></table>	products	123 index	AC product_id	AC product_category_name	123 product_name_lenght	123 product_description_lenght	123 product_photos_qty	123 product_weight_g	123 product_length_cm	123 product_height_cm	123 product_width_cm																																																							
products																																																																		
123 index																																																																		
AC product_id																																																																		
AC product_category_name																																																																		
123 product_name_lenght																																																																		
123 product_description_lenght																																																																		
123 product_photos_qty																																																																		
123 product_weight_g																																																																		
123 product_length_cm																																																																		
123 product_height_cm																																																																		
123 product_width_cm																																																																		

# Segmentez des clients d'un site e-commerce



**Introduction**

Face à un marché en constante évolution, Olist souhaite transformer sa stratégie Data pour mieux comprendre et servir sa clientèle.

Pour cela nous devons développer une segmentation client dynamique, intuitive et évolutive.

# Feature Engineering

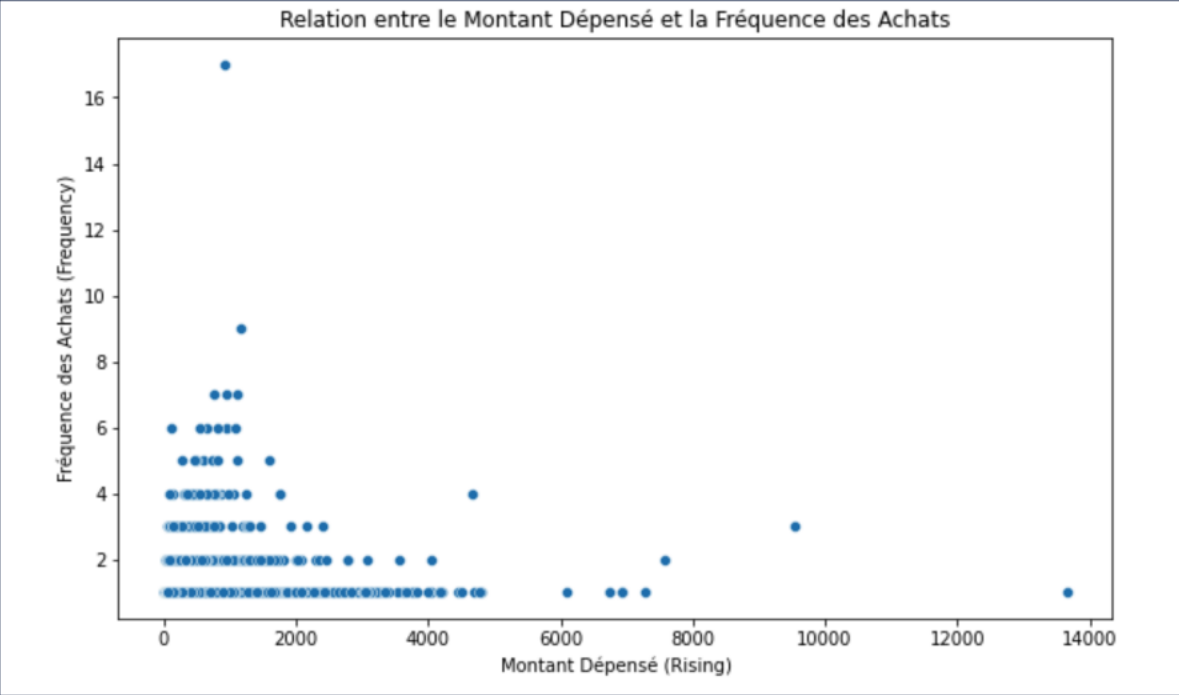
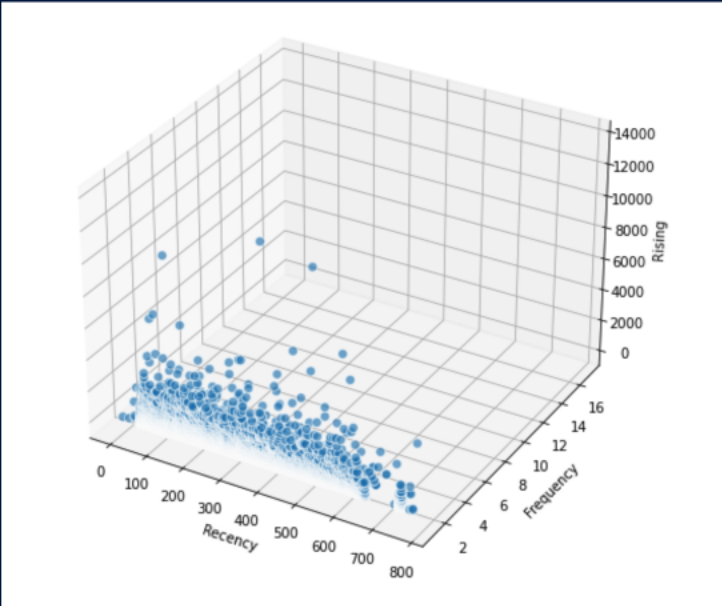
⚡	customer_unique_id ⚡	Importance ⚡	Recency ⚡	Frequency ⚡	Rising ⚡	Review_average ⚡
0	0000366f3b9a7992bf8c76cfd3221e2	False	2018-05-10 10:56:27	1	141.90	5.0
1	0000b849f77a49e4a4ce2b2a4ca5be3f	False	2018-05-07 11:11:27	1	27.19	4.0
2	0000f46a3911fa3c0805444483337064	False	2017-03-10 21:05:03	1	86.22	3.0
3	0000f6ccb0745a6a4b88665a16c9f078	False	2017-10-12 20:29:41	1	43.62	4.0
4	0004aac84e0df4da2b147fca70cf8255	False	2017-11-14 19:45:42	1	196.89	5.0

## Workflow



	Frequency	Rising	Review_average
count	96096.000000	96095.000000	95380.000000
mean	1.034809	166.594226	4.084989
std	0.214384	231.428912	1.341571
min	1.000000	0.000000	1.000000
25%	1.000000	63.120000	4.000000
50%	1.000000	108.000000	5.000000
75%	1.000000	183.530000	5.000000
max	17.000000	13664.080000	5.000000

	Variable	Shapiro-Wilk	Kolmogorov-Smirnov	Anderson-Darling	D'Agostino-Pearson
0	Review_average	False	False	False	False
1	Rising	False	False	False	False
2	Frequency	False	False	False	False
3	Recency	False	False	False	False





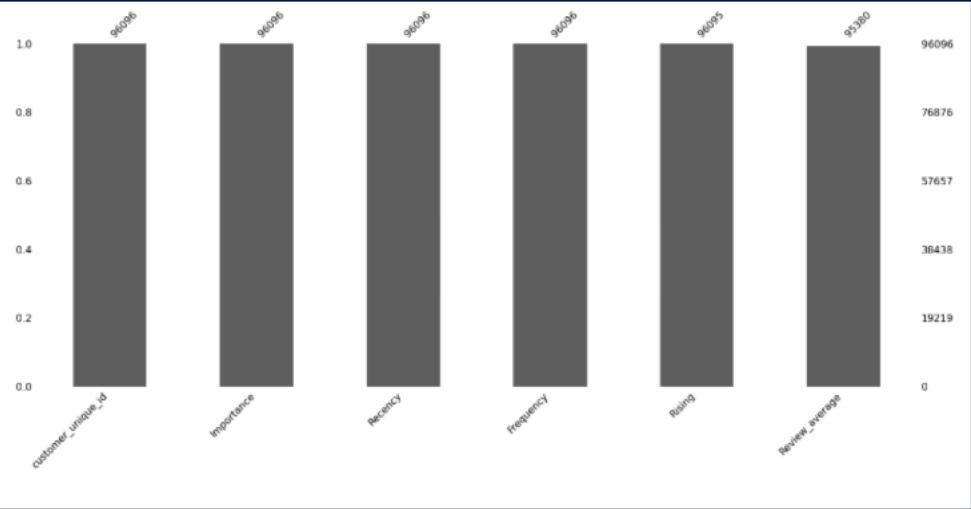
# Segmentez des clients d'un site e-commerce



# Nettoyage du jeu de données

## Workflow





Missing Values	Percentage
customer_unique_id	00.000000
Importance	00.000000
Recency	00.000000
Frequency	00.000000
Rising	10.001041
Review_average	7160.745088

Variable	Nom du test	Statistique de test	p-valeur	Normalité
0 Review_average	Shapiro-Wilk	0.691650	0.000	Non
1 Review_average	Kolmogorov-Smirnov	0.863798	0.000	Non
2 Review_average	Anderson-Darling	12118.263484	NaN	Non
3 Review_average	D'Agostino-Pearson	18164.113444	0.000	Non
4 Review_average	Lilliefors	0.324883	0.001	Non

Valeurs manquantes après nettoyage :

```
customer_unique_id 0
Importance          0
Recency             0
Frequency           0
Rising              0
Review_average      0
dtype: int64
```

Valeurs aberrantes détectées par variable :

	customer_unique_id	Importance	Recency	Frequency	Rising	Review_average
0	00053a61a98854899e70ed204dd4baf	False	231	1	419.18	0.000000
1	0005ef4cd20d2893f0d9fbd94d3c0d97	False	219	1	129.76	0.000000
2	000de6019bb59f34c099a907c151d855	False	425	1	257.44	0.250000
3	0010a452c6d13139e50b57f19f52e04e	False	463	1	325.93	0.000000
4	0012929d977a8d7280bb277c1e5f589d	False	365	1	155.65	0.000000
...	...	...	...	...	...	...
24516	f922bdd6bafcdf99cb90d7f39cea5b3	True	398	3	139.60	0.833333
24517	f9e254cc039740e17dd15a5305035928	True	563	2	80.12	0.500000
24518	009b0127b727ab0ba422f6d9604487c7	False	764	1	40.95	0.000000
24519	4854e9b3feff728c13ee5fc7d1547e92	False	772	1	75.06	0.000000
24520	b7d76e111c89f7ebf14761390f07d17	False	772	1	136.23	0.000000

```
duplicates = df_data_clean[df_data_clean.duplicated()]

if len(duplicates) > 0:
    print("Notre jeu de données contient", len(duplicates), "doublon(s).")
else:
    print("Notre jeu de données ne contient pas de doublons.")
```

Notre jeu de données ne contient pas de doublons.

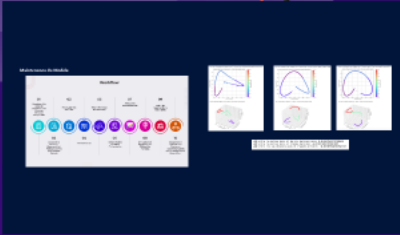
# Segmentez des clients d'un site e-commerce



## Introduction

Face à un marché en constante évolution, Olist souhaite transformer sa stratégie Data pour mieux comprendre et servir sa clientèle.

Pour cela nous devons développer une segmentation client dynamique, intuitive et évolutive.



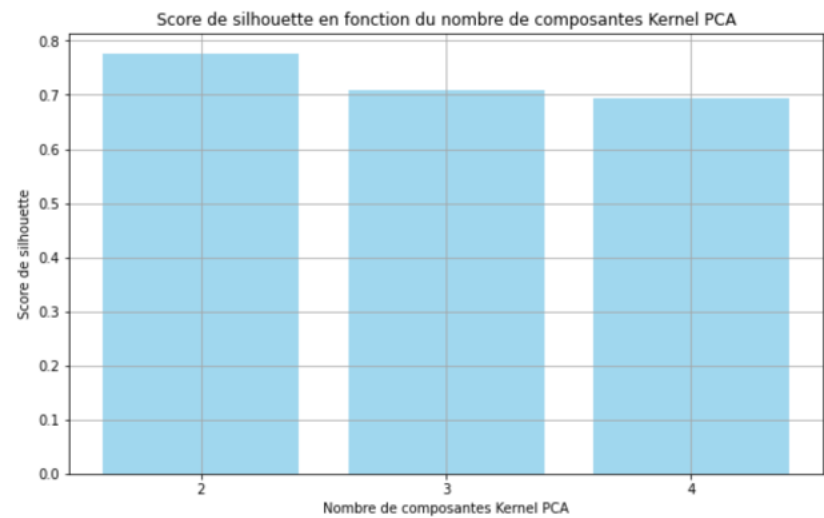
## Workflow



⚡	Normalisation ⚡	Silhouette Score ⚡	Davies-Bouldin Index ⚡	Temps de Calcul ⚡
0	Min-Max	0.459039	0.772232	194.893888
1	Standard	0.412250	0.964251	204.038569
2	Robust	0.502429	0.779281	203.990261

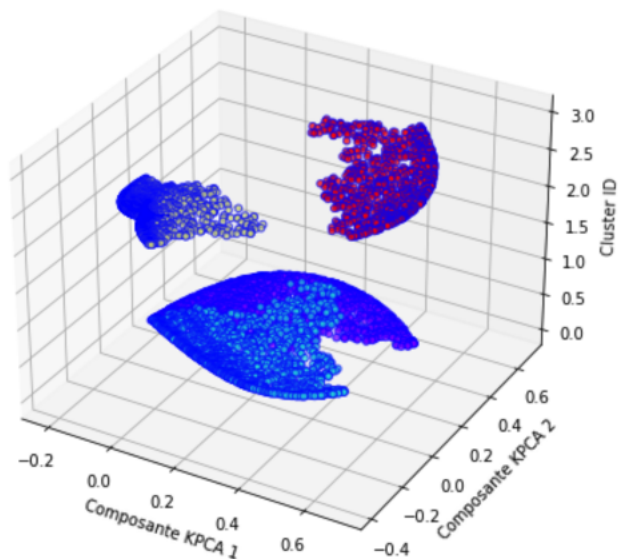
PCA Silhouette Score: 0.5622015328624331  
 Kernel PCA Silhouette Score: 0.7577758512305083  
 Isomap Silhouette Score: 0.6683052201065294  
 LLE Silhouette Score: 0.7407969080223582  
 t-SNE Silhouette Score: 0.3652634024620056

La méthode recommandée est : Kernel PCA avec un score de silhouette de 0.76



Meilleur score de silhouette: 0.78  
 Nombre de composantes Kernel PCA optimal: 2

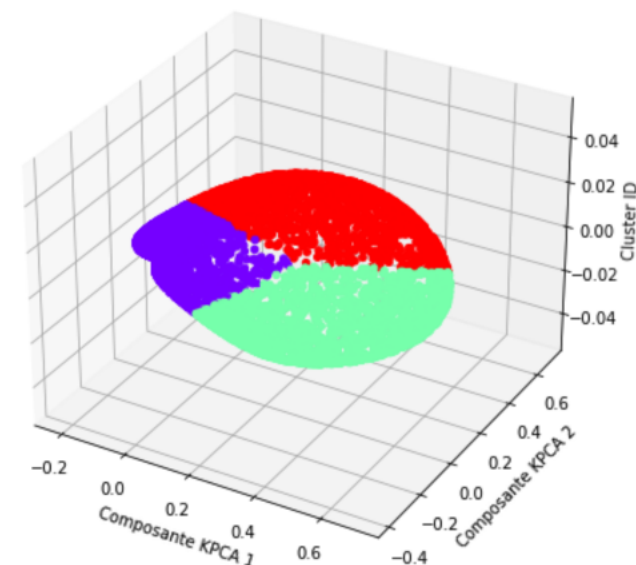
Vue "3D" des Clusters Hiérarchiques avec 4 clusters



Score de silhouette pour les clusters hiérarchiques : 0.76

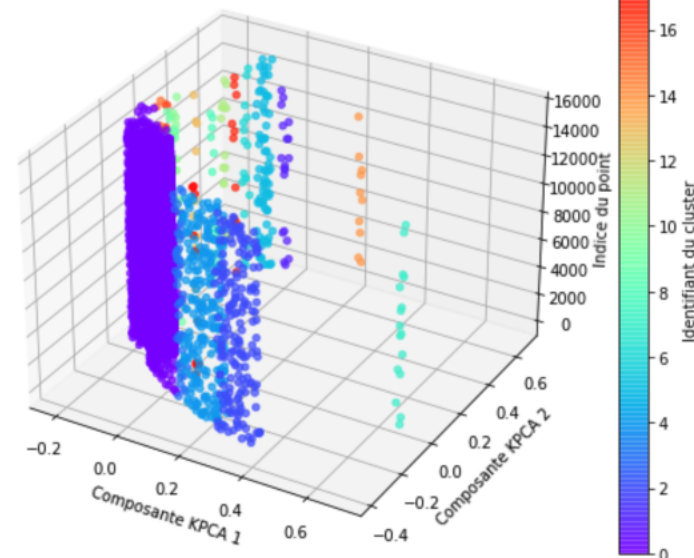


Vue "3D" des Clusters K-Means avec 3 clusters



Score de silhouette pour K-Means avec 3 clusters: 0.78

Visualisation 3D des Clusters DBSCAN



Score de silhouette pour DBSCAN avec eps=0.00900: 0.68



# Segmentez des clients d'un site e-commerce



## Introduction

Face à un marché en constante évolution, Olist souhaite transformer sa stratégie Data pour mieux comprendre et servir sa clientèle.

Pour cela nous devons développer une segmentation client dynamique, intuitive et évolutive.

## Table des matières

- Contexte organisationnel
- Exploration des Données
- Segmentation des Données
- Modélisation
- Recommandation de Produits

Contenu de la Table des Matières

1	Contexte organisationnel
2	Exploration des Données
3	Segmentation des Données
4	Modélisation
5	Recommandation de Produits

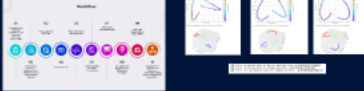
## Recommandation de Produits



## Recommandation



## Recommandation de Produits



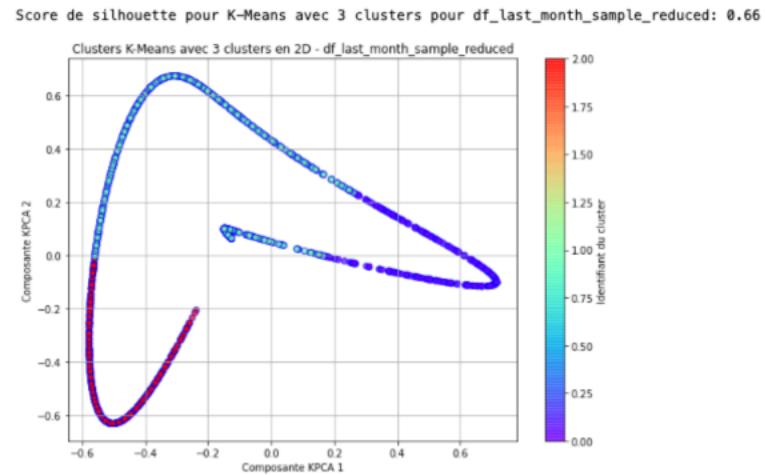


# Maintenance du Modèle

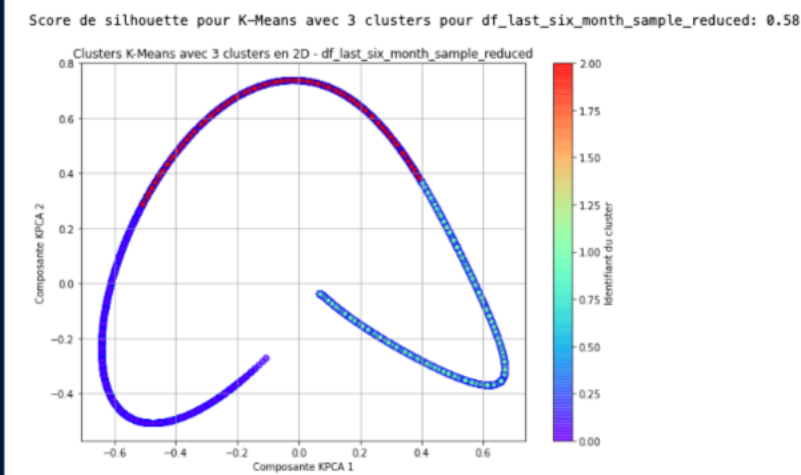
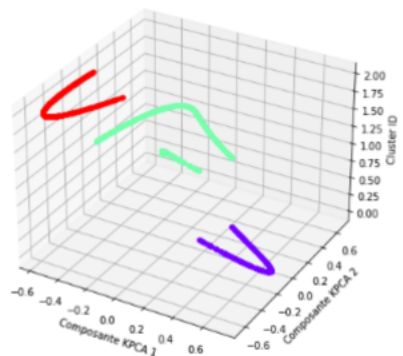
## Workflow



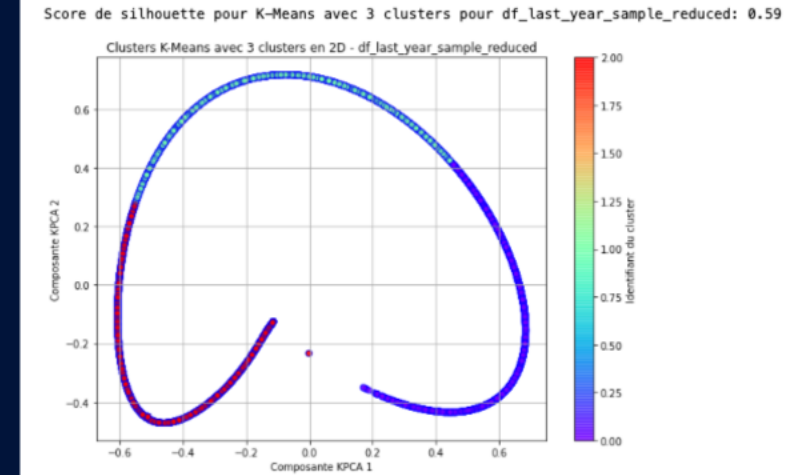
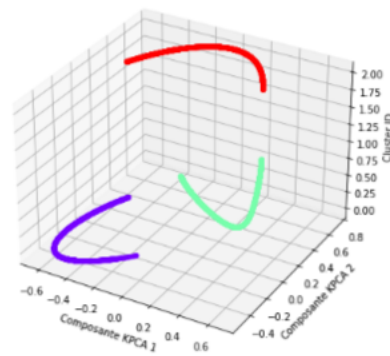




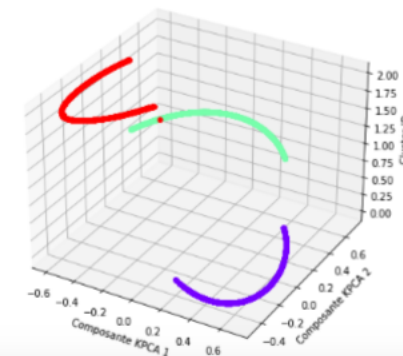
Vue "3D" des Clusters K-Means avec 3 clusters - df\_last\_month\_sample\_reduced



Vue "3D" des Clusters K-Means avec 3 clusters - df\_last\_six\_month\_sample\_reduced



Vue "3D" des Clusters K-Means avec 3 clusters - df\_last\_year\_sample\_reduced



ARI entre le dernier mois et les six derniers mois: 0.012827942779790618  
ARI entre le dernier mois et l'année dernière: -0.019316097283613005  
ARI entre les six derniers mois et l'année dernière: -0.001969560502945367

olist