

Soutenance du Projet 5

Segmentez des clients d'un site e-commerce

Par
Elisée TCHANA

Mentor
Cyril MONTI

Juillet 2021

Plan de Soutenance

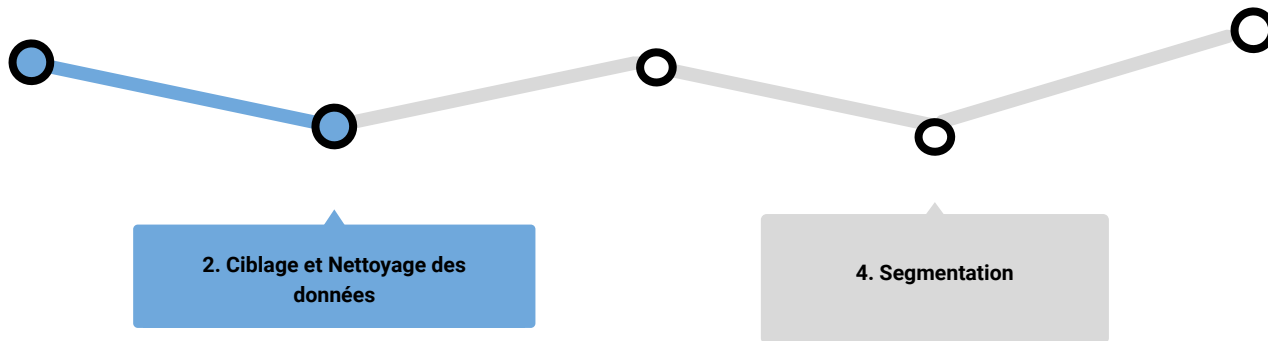
La présentation du projet

3. Analyse exploratoire

2. Ciblage et Nettoyage des
données

4. Segmentation

5. •Questions - Réponses



Description du Projet

Obtenir une segmentation marketing des clients

The logo for 'olist' is displayed in white lowercase letters on a solid blue square background.

Objectifs

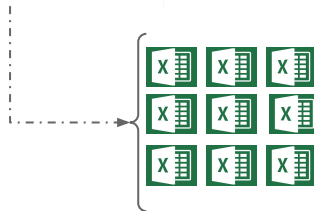
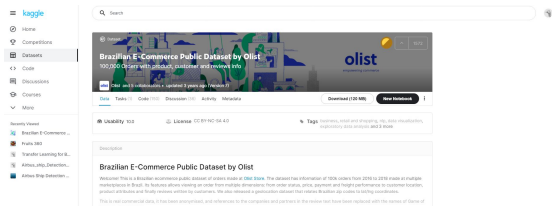
Comprendre les différents types d'utilisateurs

Fournir une description marketing actionnable de la segmentation

Évaluer la fréquence à laquelle la segmentation doit être mise à jour

Effectuer un devis de contrat de maintenance

Data Analysis Workflow



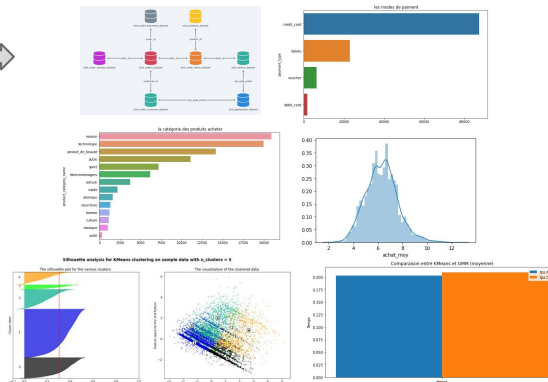
9 CSV



```
#Import packages
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import math
import pickle
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
from sklearn.preprocessing import StandardScaler
from scipy import spatial, stats

#Import data
data = pd.read_csv("openfoodfacts.csv", sep=";")
data.head()
```

```
customers = pd.read_csv('C:/Users/Elis e/Desktop/Projet (5)/archive/olist_customers_dataset.csv')
geolocalisation = pd.read_csv('C:/Users/Elis e/Desktop/Projet (5)/archive/olist_geolocation_dataset.csv')
order_items = pd.read_csv('C:/Users/Elis e/Desktop/Projet (5)/archive/olist_order_items_dataset.csv')
order_payments = pd.read_csv('C:/Users/Elis e/Desktop/Projet (5)/archive/olist_order_payments_dataset.csv')
order_reviews = pd.read_csv('C:/Users/Elis e/Desktop/Projet (5)/archive/olist_order_reviews_dataset.csv')
orders = pd.read_csv('C:/Users/Elis e/Desktop/Projet (5)/archive/olist_orders_dataset.csv')
products = pd.read_csv('C:/Users/Elis e/Desktop/Projet (5)/archive/olist_products_dataset.csv')
sellers = pd.read_csv('C:/Users/Elis e/Desktop/Projet (5)/archive/olist_sellers_dataset.csv')
translation = pd.read_csv('C:/Users/Elis e/Desktop/Projet (5)/archive/product_category_name_translation.csv')
```

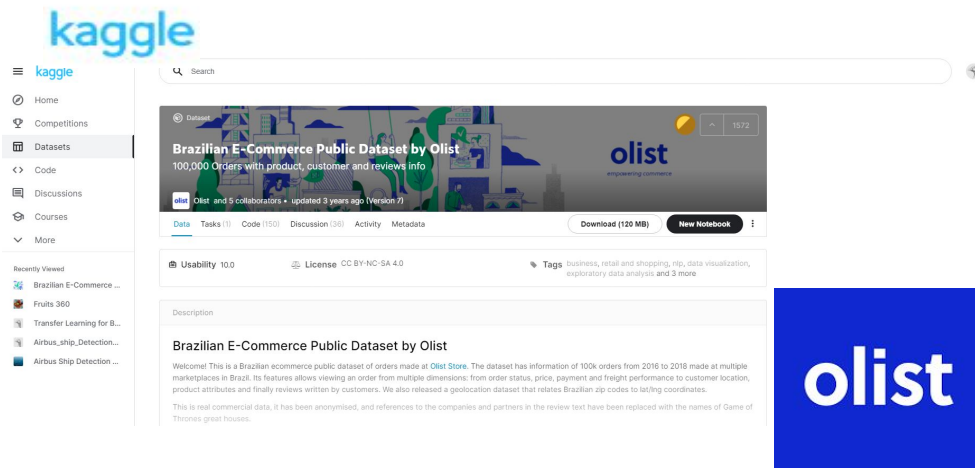


Data Cleaning & Analysis



Data Sharing

Pr sentation du jeu de donn es



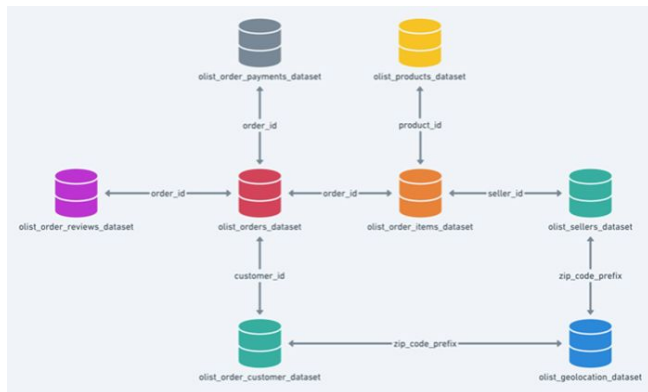
FICHIER	NB DE LIGNES	NB DE COLONNES	DESCRIPTION
Customers	99441	5	Donn�es clients
Geolocation	1000163	5	Donn�es ville Seattle
Order items	112650	7	Articles command�s
Order payments	103886	5	Paiement commandes
Order reviews	100000	7	Avis commandes
Orders	99441	8	Commandes
Products	32951	9	Produits
Sellers	3095	4	Vendeurs
Product Category Name Translation	71	2	Traduction des noms de cat�gorie

Donn es

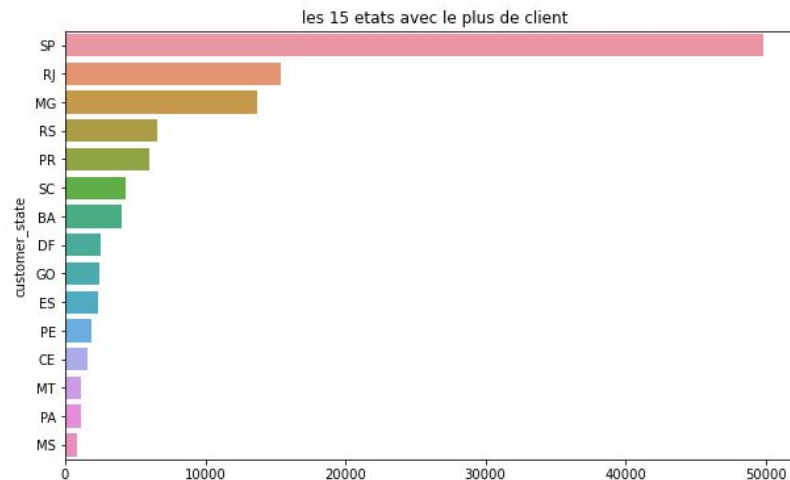
Base de donn es anonymis e ann es 2016   2018

1. 9 fichiers
2. 52+ indicateurs
3. couvrant 3 types d'informations:
4. Historique de commandes
5. Localisation clients et vendeurs
6. Satisfaction client
7. Conditions de paiement

Feature engineering pour segmenter les clients.

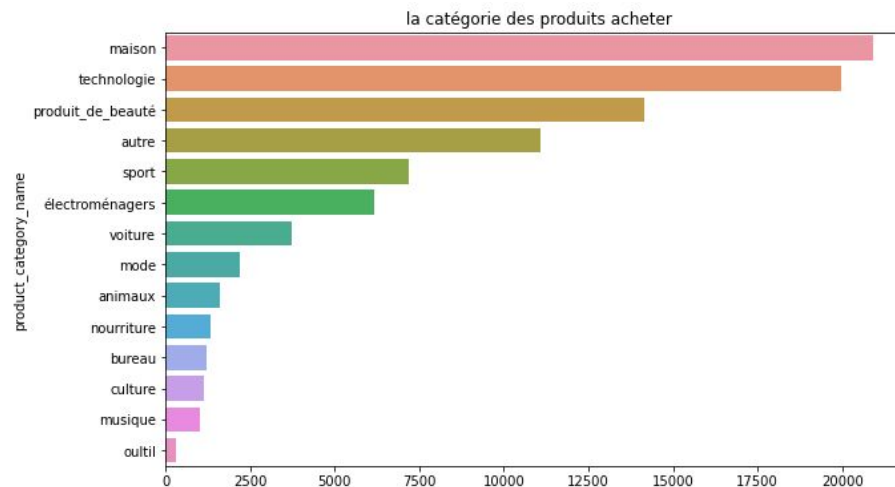
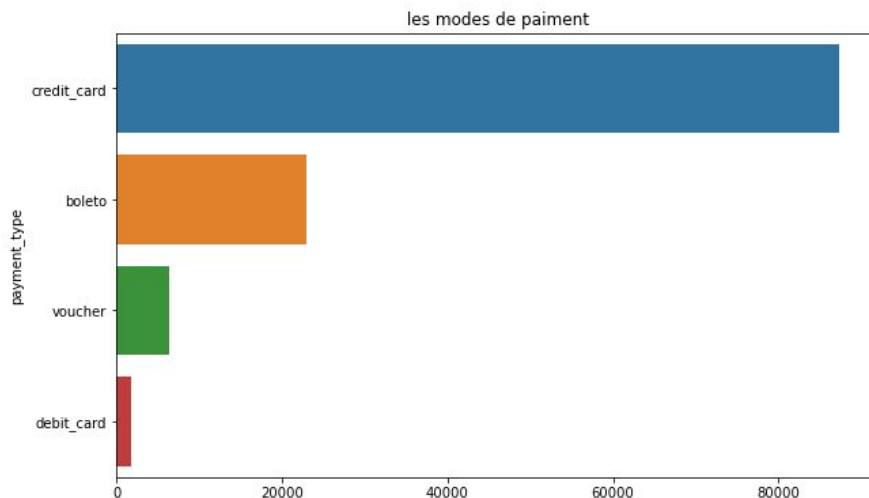


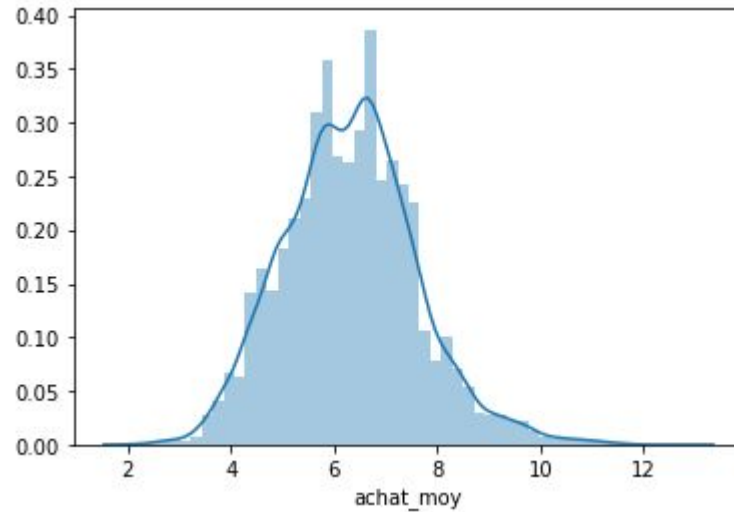
- Les achats moyens par client (en prix)
- Le nombre d'achats que le client a r alis  dans toutes ses commandes
- La note moyenne que le client a donn 
- Le type de paiement le plus utilis 
- La moyenne du nombre de paiement pour une commande (si la commande co te 30 euros, le client peut d cider de payer en 2 fois 15 euros)
- La ville d'achat
- L' tat de l'achat
- Le d lai de la commande
- Le jour de la semaine
- L'heure
- La taille du commentaire
- Le produit le plus achet 



Visualisation des features

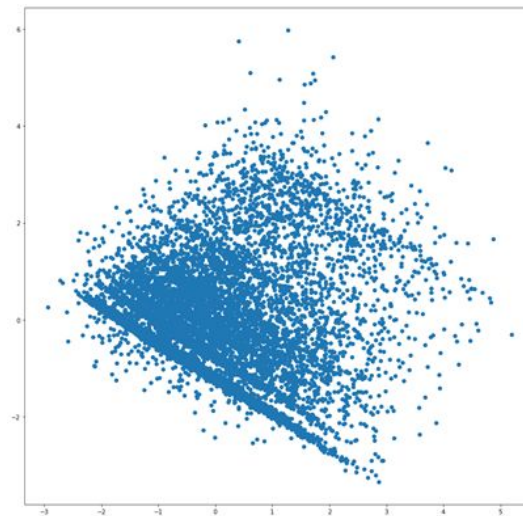
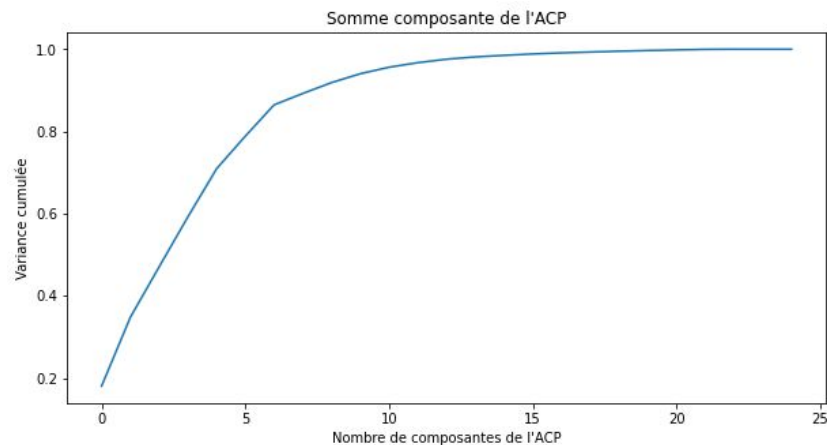
Visualisation des features



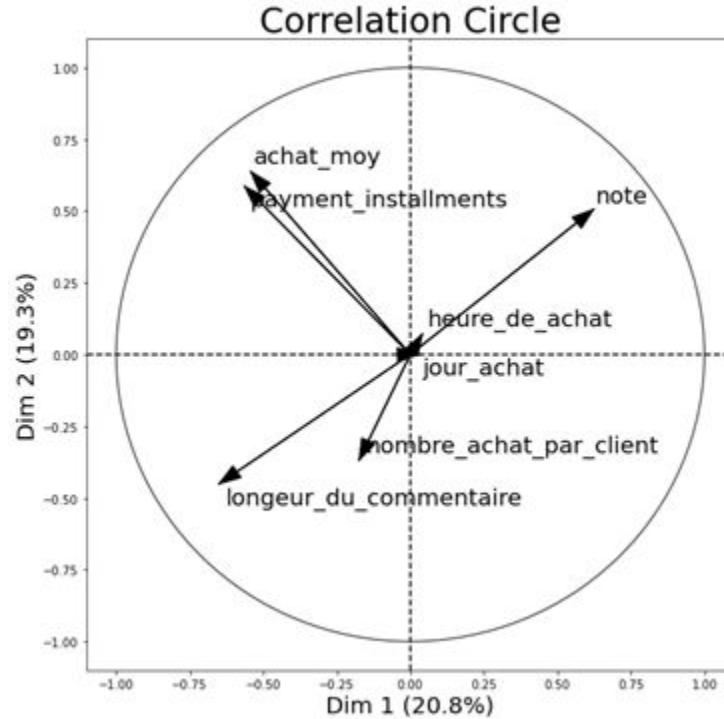


Passer les « features » avec un $\log+1$ réduire le df à 10 000

Modélisation ACP & ACP (2D)

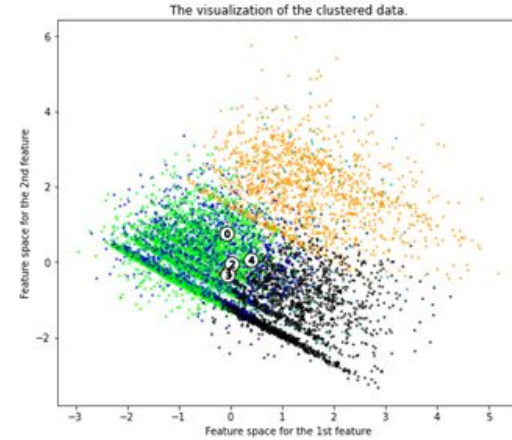
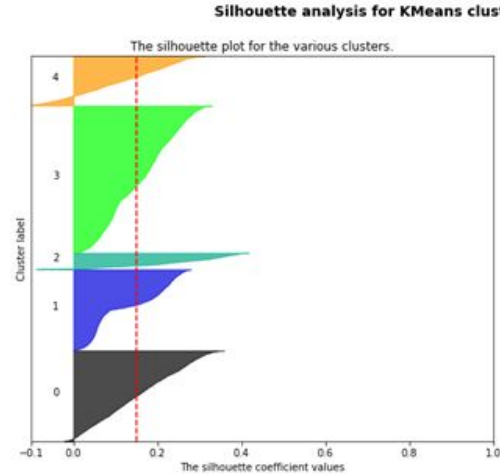
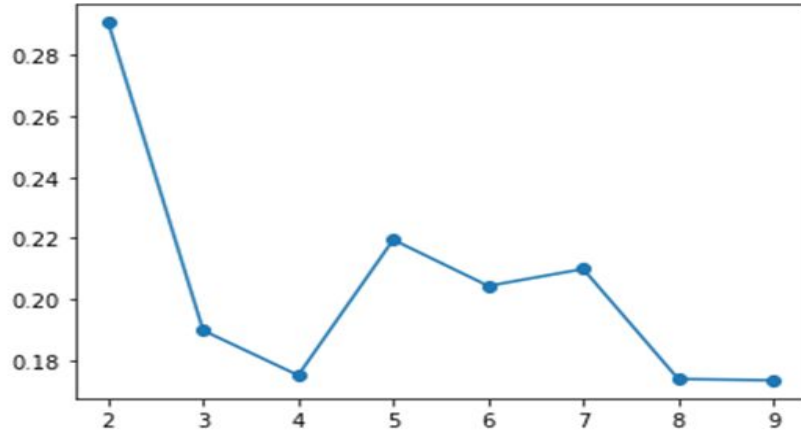


Projection ACP (corrélation)

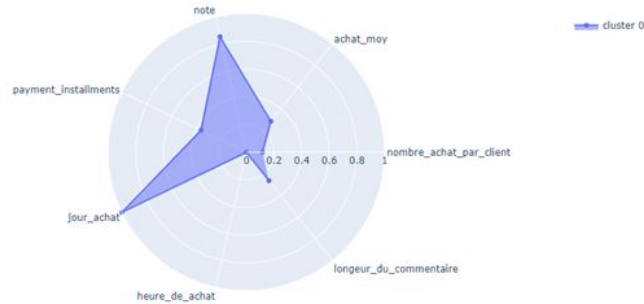


KMeans

KMeans est un modèle qui permet le clustering de données non étiquetées.

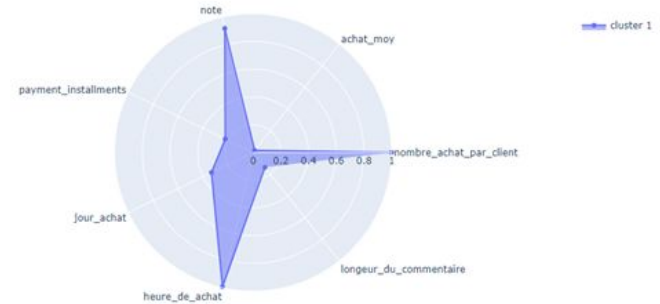


KMeans (Visualisation des clusters)



1er cluster que nous pouvons appeler
les clients du dimanche :

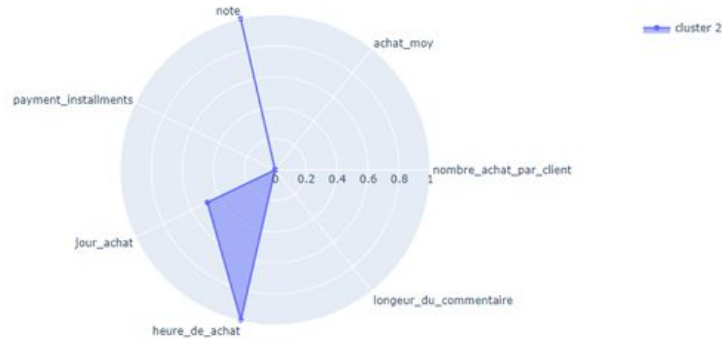
1. Bonnes notes
2. Peu de commentaires
3. Très peu d'achats
4. Achats pas très chers



2ème cluster que nous pouvons
appeler les clients réguliers :

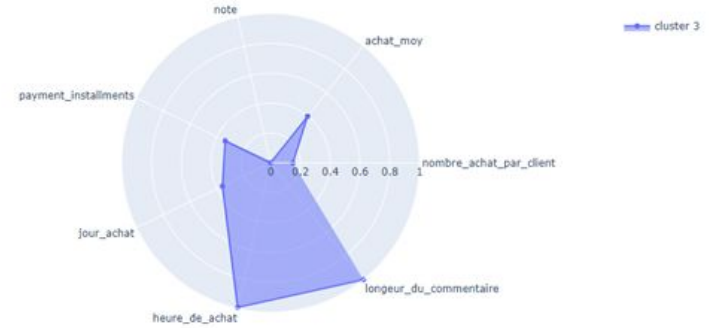
1. Bonnes notes
2. Peu de commentaires
3. Beaucoup d'achats
4. Achats pas très chers
5. Achats le soir

KMeans (Visualisation des clusters)



3  me cluster que nous pouvons appeler les clients satisfaits :

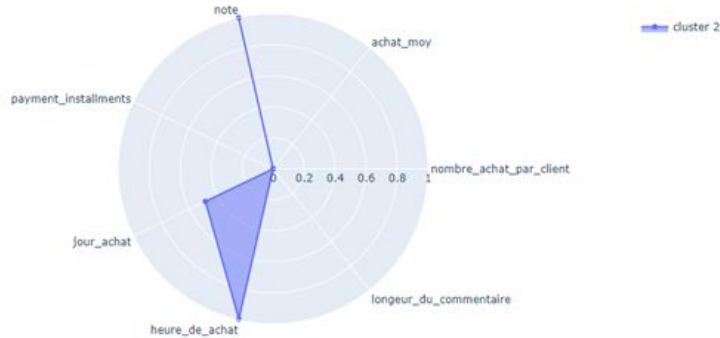
1. Excellente note
2. Pas de commentaire
3. Un seul achat
4. Achat pas tr s cher
5. Achat le soir



4  me cluster que nous pouvons appeler les clients m contents :

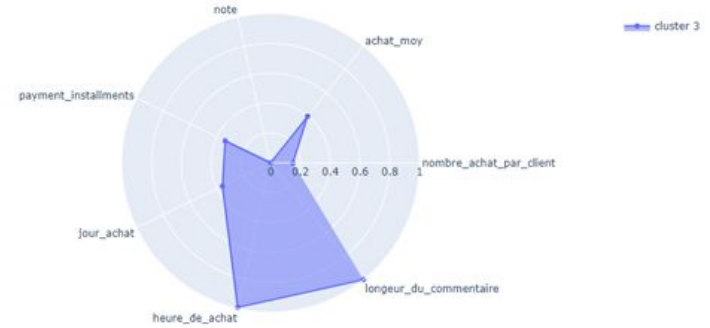
1. Mauvaises notes
2. Commentaires longs
3. Peu d'achats
4. Achats le soir

KMeans (Visualisation des clusters)



3 me cluster que nous pouvons appeler les clients satisfaits :

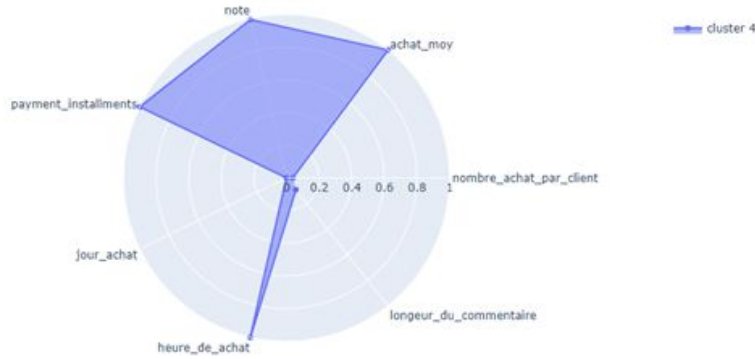
1. Excellente note
2. Pas de commentaire
3. Un seul achat
4. Achat pas tr s cher
5. Achat le soir



4 me cluster que nous pouvons appeler les clients m contents :

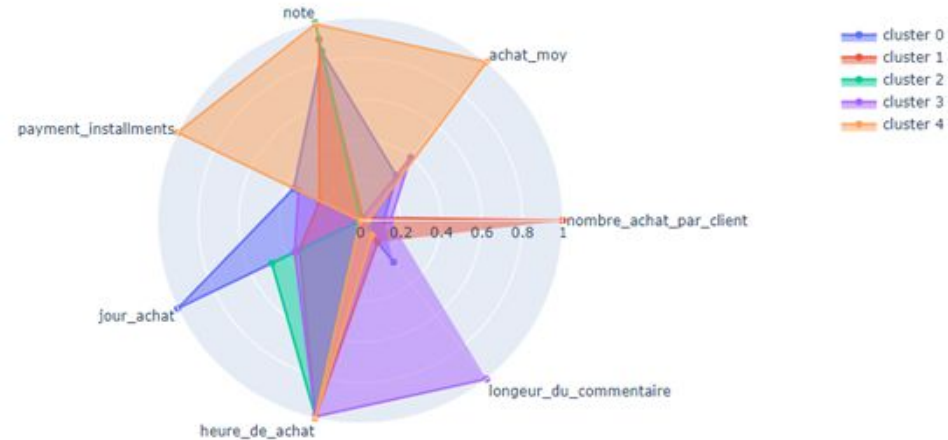
1. Mauvaises notes
2. Commentaires longs
3. Peu d'achats
4. Achats le soir

KMeans (Visualisation des clusters)



5ème cluster que nous pouvons appeler les client avec gros budget :

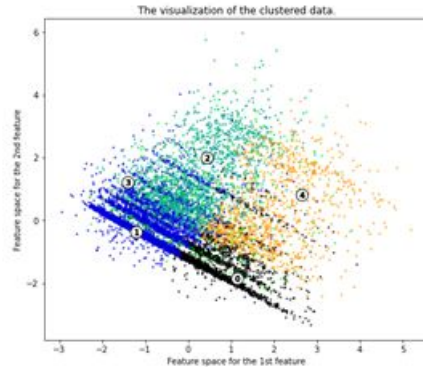
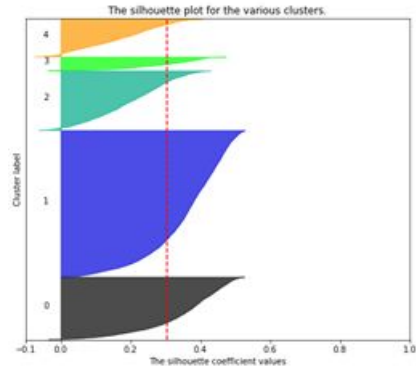
1. Excellentes notes
2. Commentaires très courts
3. Achats très chers
4. Paiement en plusieurs fois



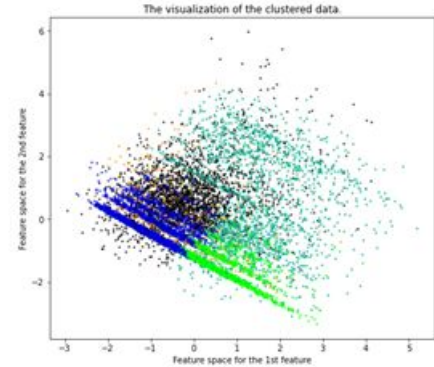
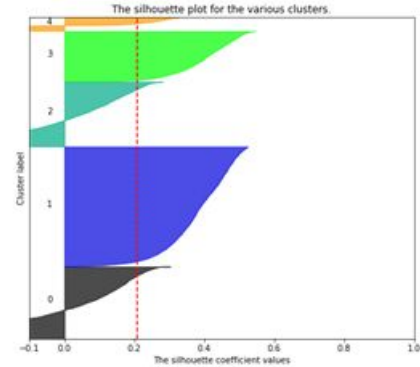
Modèles GMM

Comparaison entre KMeans et GMM

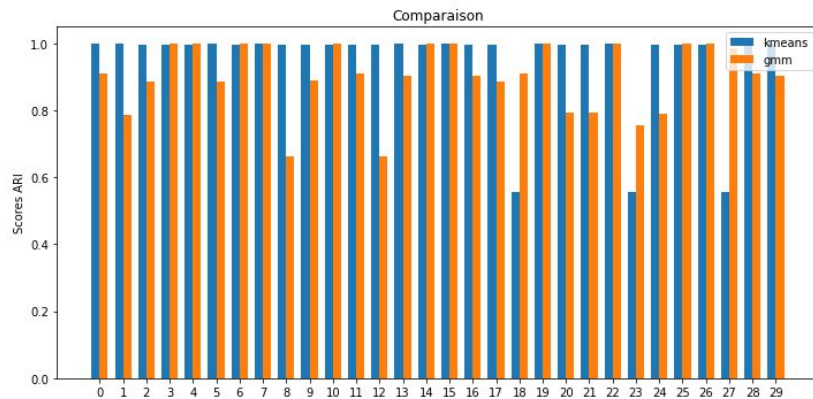
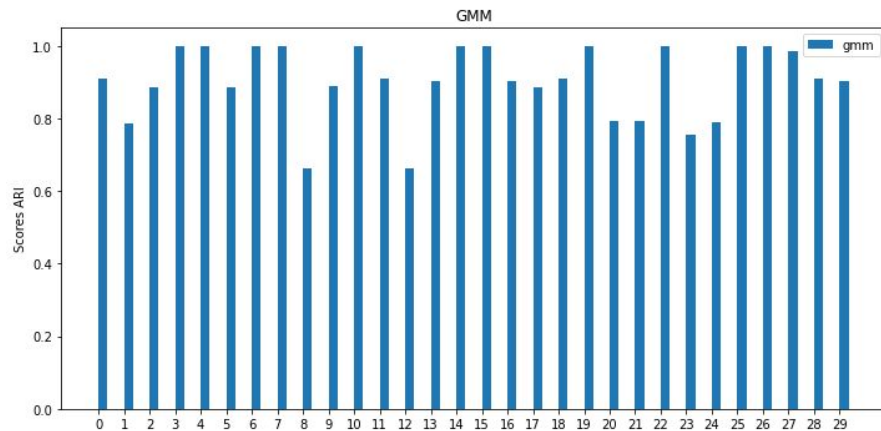
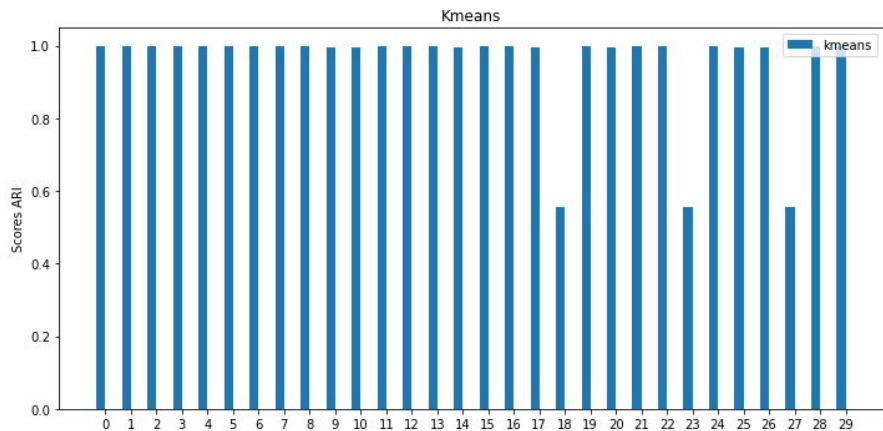
Silhouette analysis for KMeans clustering on sample data with $n_clusters = 5$



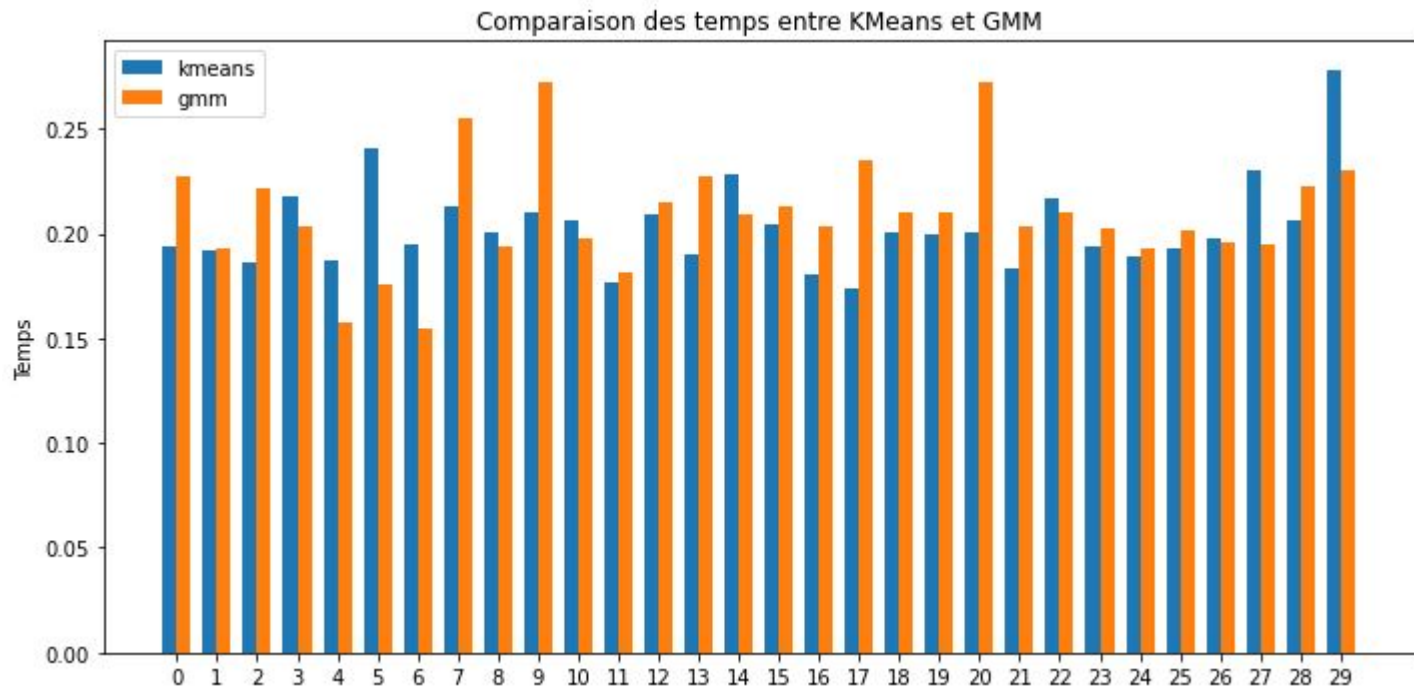
Silhouette analysis for KMeans clustering on sample data with $n_clusters = 5$



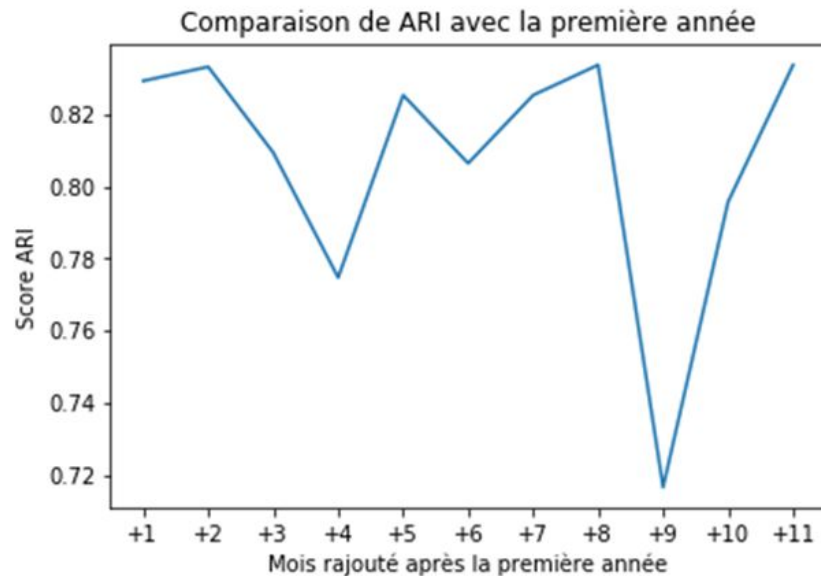
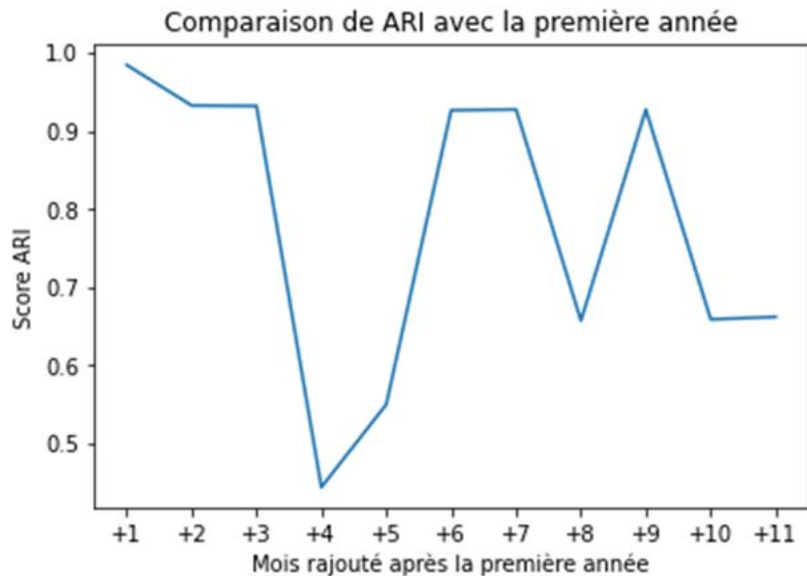
Stabilit     l'initialisation (KMeans - GMM /ARI)



Comparaison entre KMeans et GMM



Contrat de Maintenance Stabilisation (KMeans) - (GMM)



Conclusion

5 segmentations de clients :

1. Les clients qui achètent le dimanche
2. Les clients mécontents
3. Les clients avec un gros budget
4. Les clients réguliers
5. Les clients d'appoint

Pour le contrat de maintenance, après analyse de l'ARI, sur les deux ans de données que nous avons, nous pouvons voir qu'il doit avoir une mise à jour tous les 4 mois environ.

Merci pour votre attention !!